InfraNotes2017

This page was created in April of 2017 to help modernize our system records.

SysAdmins

Goals

KAM: Apache SpamAssassin is a framework for writing rules. I deliver rules to prove the code works but I don't view that the project has to provide rules. I use this as a guidance in where I spend my focus. Beyond that, my goal with belonging to the SysAdmin group is to ensure the project is supported with modern, secure hardware and software with a bus factor greater than one.

DAJ: Apache SpamAssassin should be an effective spam detection tool in a user's/admin's mail filtering toolbox. It cannot be perfectly tuned to each mail filtering environment due language, geographical, and other differences but users should be able to install SpamAssassin and follow some guidelines on the wiki to quickly and easily get their mail filtered accurately.

Acronymns

*ASF = Apache Software Foundation
*BZ = Bugzilla
*SA = Apache SpamAssassin
*PMC = Project Management Committee
*SVN = SubVersioN
*A.O = Apache.org
*S.A.O = SpamAssassin.Apache.org

Current Members

As of May 2017:
*Dave Jones - davej@apache.org
*Kevin A. McGrail - 703-798-0171 - kmcgrail@apache.org
*Bryan Vest - bvest@apache.org

Who's in Charge?

The PMC. There is no leadership hierarchy in the SpamAssassin SysAdmins.

NOTE: As with any ASF role, if you follow The Apache Way, you should feel empowered to Just Do It (TM Nike)

For a SysAdmin, your solution works (Merit), it's well documented (Open) and supports the project (Community), you're good to go though as a SysAdmin you need to realize we have control over private data. All SASA members have been asked to follow the LISA Code of Ethics.

Tenets we Follow

*The Apache Way Shane Curcuru's post has many good points.
*LISA/Sage Code of Ethics

Onboarding

Workflow

1. A PMC Member nominates a new SASA member as a committer since we store items in SVN for configs.
   NOTE: If they later produce code, they should request that permission from the PMC.
2. If the vote is successful, they then follow all the normal committer guidelines to get them an Apache ID including an appropriate committer license: New Committers Guide.
3. After the Apache ID is setup and given to the new team member,the new member will complete their setup of the **Important Resources** below.
   a. http://id.apache.org - setup strong password, SSH/PGP keys, mail forwarding, etc.
   b. Subscribe to the mailing lists as their apache.org email address
   c. Setup SVN repos using the RW https URL
   d. Make sure they can login to Bugzilla
   e. Create a Jira account with a secure password (not integrated)
   f. Create a wiki login using their full name. Spaces are allowed like "John Smith".
   g. Setup OPIE
   h. Setup their Apache home page with their PGP public key
4. Someone with Karma needs to:
   *Approve request to sysadmins mailing list
   *Add them to the wiki:
   *Contributor only
   *Contributor and Admin
   *Open a JIRA ticket at issues.apache.org similar to INFRA-14045 to get them access to SA servers

Important Resources

*New Committers Guide

Apache ID

Once your Apache ID is created and you receive email notification, setup a strong password, SSH keys, PGP keys, and mail forwarding at https://id.apache.org.

Mailing Lists

See Mailling Lists and subscribe with your **user@apache.org** address to:
sysadmins@spamassassin.apache.org - send email to sysadmins-subscribe-*user=apache.org@spamassassin.apache.org
ruleqa@spamassassin.apache.org - send email to ruleqa-subscribe-*user=apache.org@spamassassin.apache.org

SVN

Read-Only: http://svn.apache.org/repos/asf/spamassassin

Read-Write: https://svn.apache.org/repos/asf/spamassassin

Repo	Contents	Notes
sysadmins	Server and application configs	Encrypt passwords and sensitive information – NEED TO SPECIFY HOW WE WANT TO DO THIS
dns	Configs and records related to spamasssassin.org	Hosted by PowerDNS on sa-vm1.apache.org as hidden master
site	http://spamassassin.apache.org site contents	

Bugzilla

SpamAssassin Bugzilla

Jira

Sign up at Jira with your apache.org email address since it it doesn't use your password setup at https://id.apache.org.

ASF Infrastructure (Infra) Jira

NOTE: If you open a ticket that has comments and questions, use the WaitingForInfra button at the top of your ticket to alert Infra.

Wiki

1. Create an account at https://wiki.apache.org/spamassassin using your full name (i.e. Jane Doe).
2. Email sysadmins@spamassassin.apache.org to request access to the wiki:
   *Contributor only
   *Contributor and Admin

NOTE: Write access to the wiki is to anyone who has created a login name on the wiki whose name has been added to the page https://wiki.apache.org/spamassassin/ContributorsGroup

Write access to that page is to anyone whose wiki login name has been added to https://wiki.apache.org/spamassassin/AdminGroup

OPIE

OPIE is required to sudo to root. The basic idea is to setup an OPIE passphrase which is never entered into the ASF server but used to create a response to copy/paste at the sudo prompt.
*Apache reference: https://reference.apache.org/committer/opie
*Javascript client: https://reference.apache.org/committer/otp-md5

PGP Key

1. Add your PGP key in https://id.apache.org.
2. Setup your PGP key page at http://people.apache.org/~user:
   *sftp://user@home.apache.org
   *Create ~/public_html directory which is the DocumentRoot for /~user.
   *Setup index.html with PGP key and link to PGP asc file. TIP: wget http://people.apache.org/~kmcgrail as a starting point

Infrastructure

DNS Hosting

PowerDNS web interface for easy management of spamassassin.org DNS records:

1. Open an SSH tunnel: ssh -f sa-vm.apache.org -L 8090:localhost:8090 -N
2. Open web interface: http://localhost:8090
3. Login with admin. (Password is encrypted in sysadmins/accounts.)

Zone	Server	Contact	Notes
spamassassin.org	ns2.pccc.com	Kevin McGrail kevin.mcgrail@mcgrail.com, kmcgrail@apache.org	Instant updates via NOTIFY
	ns2.ena.com	Dave Jones djones@ena.com, davej@apache.org	Instant updates via NOTIFY
	dns-master.sonic.net	Grant Keller grant.keller@sonic.com	Hidden slave, 5 to 10 min delay of public slaves after NOTIFY
	ns.hyperreal.org	Brian Behlendorf	Currently not used since DJBDNS doesn't support NOTIFY or EDNS over TCP

Standards

*Apache Infrastructure standard is Ubuntu 16.04 LTS
*Cron entries should be in new standard locations /etc/cron.d, /etc/cron.daily, etc. and avoid using user's crontab
*Custom scripts should reside in /usr/local/bin if they are not direcly related to SpamAssassin processing that should be in /usr/local/spamassassin
*Symlink scripts from /usr/local/bin to /etc/cron.d, /etc/cron.daily, or /etc/cron.weekly. This provides easy discovery and future management by others on the sysadmins team.
*Scripts and cron entries should mail output to the sysadmins mailing list

Credentials

There are legacy shared credentials encrypted in SVN for elevated access on older machines. The project is slowly moving away from these concepts.

Legacy Servers

*minotaur.apache.org - handled various build and devel related tasks
*hyperion.apache.org - likely a Solaris box that had backup data of next server
*spamassassin.zones.apache.org - DIED - was replaced with spamassassin-vm
*spamassassin.zones2.apache.org - deprecated by Infra, replaced by sa-vm1.apache.org
*spamassassin-vm.apache.org - deprecated by Infra, replaced by sa-vm1.apache.org
*buildbot, ruleqa, etc. are aliases of above deprecated servers

Servers

Hostname	Function	Software	Configs/Location	Resource/URL	SVN Location
apachesf.sonic.net	Donated by Sonic	CentOS 7		apachesf.spamassassin.org (64.142.56.146)	
colo.sonic.net	Retired			76.191.162.2	
trap-proc.spamassassin.org	Retired			a.k.a spam-trap.spamassassin.org (192.87.106.247)	
sa-vm1.apache.org	DNS Hidden Master	PowerDNS	/etc/powerdns/pdns.d/pdns.local.conf	spamassassin.org	dns (webserver API key redacted)
	Rsync Mirrors	rsyncd	/etc/rsyncd.conf	rsync.spamassassin.org	trunk/build/automc/etc-rsyncd.conf
	Web Server	apache2	/etc/apache2/sites-available/automc.conf	updates.spamassassin.org	trunk/build/automc/automc-apache2.conf
		apache2	/etc/apache2/sites-available/automc.conf	ruleqa.spamassassin.org	trunk/build/automc/ruleqa.cgi
<ac:structured-macro ac:name="unmigrated-wiki-markup" ac:schema-version="1" ac:macro-id="1d53d356-108f-4add-b325-d37f2551dcd3"><ac:plain-text-body><![CDATA[		SaUpdateMirrorSetup	svn	rsyncd [updates] for mirrors	spamassassin.apache.org/updates	[site/updates	http://svn.apache.org/repos/asf/spamassassin/site/updates/]/MIRRORED.BY	]]></ac:plain-text-body></ac:structured-macro>
	Nightly Masscheck	cron/scripts	/usr/local/spamassassin/automc/rsync/tagged_builds	ruleqa.spamassassin.org	trunk/backend/nitemc/README
	RuleQA web UI	cron/scripts	/usr/local/spamassassin/automc/html	ruleqa.spamassassin.org	RuleQaApp

Backups

An old backup exists in sa-vm1.apache.org:/usr/local/spamassassin/backups/spamassassin-vm. It's a large bzip'd tar file so make sure you don't extract it and fill up the filesystem.

We need to setup offsite backups that at least two of the SA sysadmins members can access. Crashplan is an option or we can setup BackupPC somewhere that can do backups via Rsync. BackupPC is a very simple backup tool with deduplication.

Specifically, what backups does KAM have as of 2017/05/08:
*hyperion.apache.org - N/A
*incoming.apache.org aka colo - Backup on KAM's Crashplan
*minotaur.apache.org (NOTE: Aka People) - N/A
*sa-vm1.apache.org - Backup on KAM's Crashplan
*Spamassassin-vm.apache.org - sa-vm1.apache.org:/usr/local/spamassassin/backups/spamassassin-vm Backup on KAM's Crashplan - Mar 15, 2017
*spamassassin2.zones.apache.org - sa-vm1.apache.org:/usr/local/spamassassin/backups/spamassassin-zones2 Backup on KAM's Crashplan from Approximately Jun 2015 last backup. Also have an Rsync copy from June 3, 2015 on PCCC TalonJR machine

Builds

The sa-vm1 server TZ is UTC so cron entries will be in UTC.

mkupdates

This section of scripts publishes new ruleset updates to the mirrors. There are currently (June 2017) two different rule daily updates. Both do lint tests against the latest version of SpamAssassin but the first one updates the 72_scores.cf based on the masscheck contributions while the second one is a "blind" rule promotion and tagged build of SVN rules for the masscheck area setup later.

25 2 * * * automc *~/svn/trunk/build/mkupdates/do-stable-update-with-scores
***~/svn/masses/rule-update-score-gen/do-nightly-rescore-example.sh**
***~/svn/masses/rule-update-score-gen/generate-new-scores.sh**
*uses ~/tmp/generate-new-scores for SVN work area
*sorts out the usable corpus from the latest 'SVN revision' at the top of the submitter's log file which should match the latest tagged build of SVN rules
*${REVISION} LINE 123 NEEDS IMPROVEMENT!!! THIS SVN REVISION NEEDS TO BE CLOSELY TIED TO THE REVISION THAT WAS STAGED IN THE MASSCHECK RSYNC DIR.
*checks the sorted corpus for a minimum number of valid contributors and ham/spam
***~/svn/trunk/build/mkupdates/mkupdate-with-scores**
*uses ~/tmp/sa-mkupdate for SVN working area
gets latest SVN ${REVISION} from rulesrc/scores/score-set
*masses -> perl Makefile.PL && make (complete build of SA and test)
*perl hit-frequencies
*garescorer - compiles and runs it, requires build/pga
*sends email if not enough masscheck submitters or usuable ham/spam for the latest SVN revision
*creates ${REVISION}.tar.gz ${REVISION}.tar.gz.sha1 and ${REVISION}.tar.gz.asc in /var/www/automc.spamassassin.org/updates for mirrors to pull
*updates DNS TXT entries [0-3].3.3.updates.spamassassin.org and 0.4.3.updates.spamassassin.org – versions >= 3.4.1 have a CNAME to 3.3.3.updates.spamassassin.org
*Script rewrite notes:
*Make each primary step modular since these steps are commmon in other scripts
*Should check for minimum contributors of ham/spam up front and not waste resources if requirements not met
*These 3 scripts above all share the same temp working dir. This should be determined from config file or relative path of user's home dir for flexibility.
*Should be able to run the ham/spam processing in parallel and merge the results together to cut this time in half
*Temp working dir for the corpus should be persistent so the rsync copy will be faster.
*Usuable corpus symlink setup could be improved. Invalid stale corpus should be removed into an archive/excluded dir.

30 8 * * * automc *~/svn/trunk/build/mkupdates/run_nightly > /var/www/automc.spamassassin.org/mkupdates/mkupdates.txt
*Currently ${SA_VERSION} = "3.4.2"
*${REVISION} = latest SVN revision THIS NEEDS TO BE ADDRESSED!!! NEED TO PREVENT REVISION FROM MESSING UP THE MASSCHECK PROCESSING.
*creates new rules/active.list
*commits new rules/active.list
*runs spamassassin lint against the updated rules and checks in a tagged version of 'sa-update_${SA_VERSION}_${TSTAMP}'
*commits "promotions validated" and emails dev@spamassassin.apache.org
*creates ${REVISION}.tar.gz ${REVISION}.tar.gz.sha1 and ${REVISION}.tar.gz.asc in /var/www/automc.spamassassin.org/updates for mirrors to pull
*updates DNS TXT entries [0-3].3.3.updates.spamassassin.org and 0.4.3.updates.spamassassin.org – versions >= 3.4.1 have a CNAME to 3.3.3.updates.spamassassin.org
*Script rewrite notes:
*Uses many of the same primary steps previous section so reuse the code and not have to maintain multiple versions
*Should be turned into generic script that can be run on demand via SVN trigger/polling

nitemc

These run shortly after the build/mkupdates/run_nightly to setup the masscheck download area based on the latest tagged build of SVN rules.
34 8 * * 0-5 automc *~/svn/nitemc/corpora_runs >> ~/rsync/corpus/nightly-versions.txt
36 8 * * 0-5 automc *~/svn/nitemc/extract_to_rsync_dir nightly ~/rsync/corpus/nightly-versions.txt
34 8 * * 6 automc *~/svn/nitemc/corpora_runs >> ~/rsync/corpus/weekly-versions.txt
36 8 * * 6 automc *~/svn/nitemc/extract_to_rsync_dir weekly ~/rsync/corpus/weekly-versions.txt

ruleqa

This updates the web interface for http://ruleqa.spamassassin.org.
5 2-20 * * * automc . /etc/profile; */usr/local/bin/runRuleQArefresh.sh
*$HOME/svn/masses/rule-qa/corpus-hourly --dir=$HOME/rsync/corpus
*$HOME/svn/masses/rule-qa/automc/gen_info_xml
*$HOME/svn/masses/rule-qa/automc/ruleqa.cgi -refresh