

# SpamAssassin!

**Matt Sergeant**



# SpamAssassin

- SpamAssassin Identifies Spam
  - ◆ It does not delete spam
  - ◆ It does not bounce spam
- Here I'll tell you all about:
  - ◆ Heuristic Anti Spam Software
  - ◆ Client, Server or filter operation
  - ◆ Rules
  - ◆ Score Genetics
  - ◆ Plugins



# SpamAssassin Kicks Ass

- Over 90% accuracy
- Highly customisable
- Gives very detailed reporting:

```
SPAM: ----- Start SpamAssassin results -----  
SPAM: This mail is probably spam.  The original message has been altered  
SPAM: so you can recognise or block similar unwanted mail in future.  
SPAM: See http://spamassassin.org/tag/ for more details.  
SPAM:  
SPAM: Content analysis details:   (6.9 hits, 5 required)  
SPAM: PLING_PLING                (0.8 points)  Subject has lots of exclamation marks  
SPAM: PLING                      (0.1 points)  Subject has an exclamation mark  
SPAM: FRONTPAGE                  (4.4 points)  BODY: Frontpage used to create the message  
SPAM: SUBJ_ALL_CAPS              (-0.1 points) Subject is all capitals  
SPAM: CTYPE_JUST_HTML            (1.7 points)  HTML-only mail, with no text version  
SPAM:  
SPAM: ----- End of SpamAssassin results -----
```

- Millions of users worldwide



# Spam Heuristics

- What makes an email spam?
- Unsolicited, commercial, bulk, junk
- Heuristics examine the email to determine whether the email is one of the above



# Heuristics Examples

- HTML with forms
- mailto: with "remove" or "unsubscribe" in subject
- Discussing "opt-in" (yeah right!)
- "This is a one time mailing"
- Empty "To" address
- Possibly forged "From" address



# Heuristics Scoring

- Each rule is assigned a score from -5 to +5
- All matching rules are added up
- If total  $\geq$  threshold score, its probably spam
- This avoids blocking based on 1 or 2 matching rules

```
From: me  
To: my_wife  
Subject: Wooohooo!
```

```
I had a great time last night with that viagra, honey!
```



# SpamAssassin Operation

- SpamAssassin works as a pipe
- Email goes in one end, and pops out the other augmented with information
- Default actions:
  - ◆ Change the Subject to "\*\*\*\*\*SPAM\*\*\*\*\* [original]"
  - ◆ Add a spam report to the top of the email
- All of this is configurable
- One nice option is spam level stars
  - ◆ Changes subject to "\*\*\*\*\* [original]" for possible spam
  - ◆ Changes subject to "\*\*\*\*\* [original]" for definitely spam



# Client Operation

- DeerSoft Inc sell an Outlook Plugin
- and one is also available for free
- KMail can be made to run SpamAssassin
- Many other clients also support SA via the "piped" operation



# Server Operation

- Plugs directly into Sendmail via MIMEDefang
- Plugs into QMail via either qmail-scanner or a QMAILQUEUE binary (supplied)
- Postfix/Cyris - spamchec.py in contrib/ directory
- Exim - needs to use procmail
- No support for MS Exchange (yet)



# SpamAssassin via Procmail

```
:0fw  
| spamassassin -P  
  
:0:  
* ^X-Spam-Status: Yes  
caughtspam
```



# Spamd - the spamassassin daemon

- There's a lot of perl code to parse on each hit
- So we can remove that overhead using spamd
- Spamd listens on a TCP/IP socket for connections via the spamc protocol
- Spamc is a small C client stub that sends the email over the socket



# Rules

- Rules all defined in config files
- Examples:

```
header INVALID_MSGID      Message-Id !~ /^<(?:\".+\"|[\^\\s]+)\@(?:\".[.+\"]|[\^\\s]+)
describe INVALID_MSGID    Message-Id is not valid, according to RFC-2822
```

```
body UNIVERSITY_DIPLOMAS /\b(?:college|university)\s+diplomas/i
describe UNIVERSITY_DIPLOMAS    University Diplomas
```

```
uri HTTP_CTRL_CHARS_HOST  /^https?:\/\/\.[^\s\/]*[\x00-\x08\x0b\x0c\x0e-\x1f]/
describe HTTP_CTRL_CHARS_HOST    Uses control sequences inside a URL's hostname
```

- Four types of rules:
  - ◆ header - a test for a particular header
  - ◆ body - a test for the *text* in the body of the message
  - ◆ rawbody - a test for the raw body text
  - ◆ uri - a test on the URIs in the document



# Auto Whitelist

- Detects who has sent you clean mail
- Bumps "whiteness" up for good citizens
- Also works as a blacklist for bad citizens
- Unfortunately only works on received mail - working on sent mail would be better



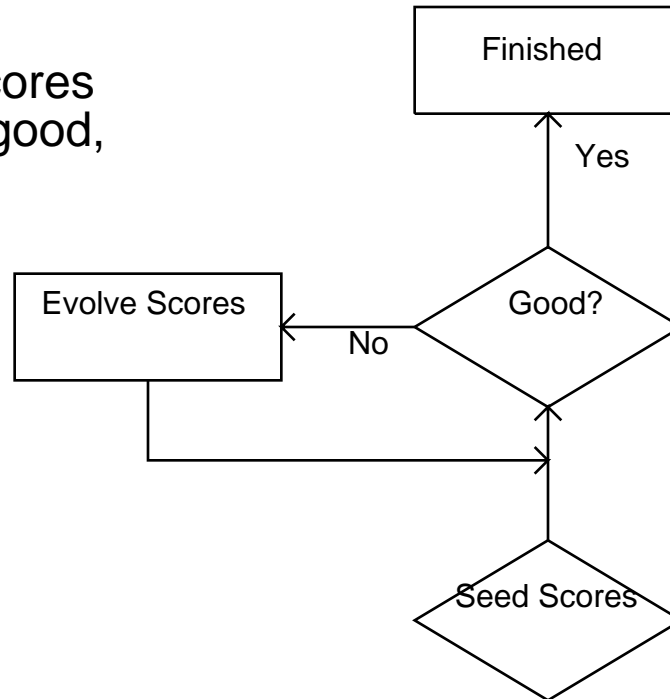
# Configuration Options

- Change the report template
- Turn off the report template
- Place to get user config from (SQL, filesystem)
- Score threshold
- ok\_locales, ok\_languages
- Spam stars



# Genetic Algorithms

- 500+ rules
- Too many to assign scores manually - is this rule good, bad, etc
- So we use a GA



# Plugins

- Can plugin different modules
- Example: Razor
- Example: DCC
- Example: Alternate mail parsers
- Example: Machine Learning



# Future Research

- More machine learning
- Performance Issues
- Better mail parsing
- Phrase Matching improvements

