

CRM114 als SpamAssassin-Plugin

Martin Schütte



Inhalt

CRM114

Installation

Benutzung

SpamAssassin-Plugin

Einstellungen

TODOs

Amavis

Und nochwas

CRM114



- Filtersprache zur Text-Klassifizierung
- verschiedene Klassifizierungsalgorithmen
- wichtigste Anwendung: Spamfilter `mailreaver.crm`
- sehr gute Erkennungsraten
- Aufruf per Kommandozeile (`stdin/stdout`)

CRM114 installieren

- nötige Pakete: „TRE Regex Library“ und CRM114
- ggf. in neues Verzeichnis kopieren,
chown amavis:amavis o.ä.
- .css-Dateien (CRM114 Sparse Spectra files) anlegen:

```
cssutil -b -r spam.css
```

```
cssutil -b -r nonspam.css
```

CRM114 installieren

```
$EDITOR priolist.mfp – White- und Blacklist
```

```
+X-List-Administrivia: yes
```

```
+To:.*\<postmaster@
```

```
-To:.*\<spamfalle@
```

CRM114 installieren

`$EDITOR rewrites.mfp` – Ersetzungen

```
AgentSmith@the.matrix.net->MyEmailAddress  
SysOp@the.matrix.net->MyEmailAddress  
mail.matrix.net->MyLocalMailRouter  
192.168.10.5->MyLocalMailRouterIP  
[[ :space: ]] Agent Smith--> MyEmailName
```

CRM114 installieren

\$EDITOR mailfilter.cf – Konfiguration, darin wichtig:

```
:log_to_allmail.txt: /no/    # (default: yes)

:mime_decoder: /\usr\bin\b64decode -pr/
:add_headers: /yes/
:rewrites_enabled: /yes/
:text_cache: /reaver_cache/
```

Cache

- speichert jede Mail mit *Cache-ID* auf Platte
- ⇒ Policy-Frage

Cache

- speichert jede Mail mit *Cache-ID* auf Platte
- ⇒ Policy-Frage
- Training mit *Cache-ID* sucht Original aus Cache
- ⇒ Training per Mail-Weiterleitung
- ⇒ Mit *Message-Id* und Hilfs-Skript sogar Outlook-kompatibel (?)

Training per E-Mail

1. Passwort setzen
(:spw:/<password>/ in mailfilter.cf)
2. Sende E-Mail mit erster Zeile:
command <password> {good,spam}
3. Entweder
 - findet Cache-ID und lernt aus Cache
 - lernt Rest der Mail

Kommandozeile

zum Einbinden in procmail, mutt, etc

```
crm mailreaver.crm          < mail_in > mail_out
```

```
crm mailreaver.crm --spam < mail_in > mail_out
```

```
crm mailreaver.crm --good < mail_in > mail_out
```

Scoring/Klassifikation

- Mail bekommt Score/Punkte
- Klassifikation in
 - SPAM: $x \leq -10$
 - UNSURE: $-10 < x < 10$
 - GOOD: $x \geq 10$
- Meist $-50 < x < 50$
- Lern-Strategie: Train-on-Error (TOE/TONE)

SpamAssassin

- E-Mail-Spamfilter
- statische Regeln, Netzwerktests, Bayes-Klassifizierung
- modular, mit Plugins erweiterbar
- Aufruf per Kommandozeile (stdin/stdout), Server/Client (spamd/spamc) oder als Perl-Modul



SpamAssassin-Plugin

- Wrapper um die Kommandozeile
- konfigurierbar
- CRM114-Klassifikation gibt SA-Punkte
- Training per SA

Konfiguration: crm114.cf

Plugin einbinden:

- loadplugin
 Mail::SpamAssassin::Plugin::CRM114 crm114.pm
- full CRM114_CHECK eval:check_crm()
- crm114_command crm -u ~/.crm114 mailreaver.crm

Konfiguration: crm114.cf

SpamAssassin-Score:

- statisch

```
crm114_staticscore_good      -3.5
crm114_staticscore_prob_good -0.5
crm114_staticscore_unsure    0.0
crm114_staticscore_prob_spam 0.5
crm114_staticscore_spam      3.5
```

- dynamisch (nicht empfohlen)

```
crm114_dynscore              0
crm114_dynscore_factor      -0.05
```


Konfiguration: crm114.cf

Lernen:

```
crm114_learn      1
crm114_autolearn  0
```

exportierte Tags:

```
add_header all CRM114-Status  _CRM114STATUS_ \
                                ( _CRM114SCORE_ )
add_header all CRM114-CacheID  _CRM114CACHEID_
add_header spam CRM114-Version  _CRM114VERSION_
```

Bayes-Plugin

- seit Januar: Bayes-Klassifikation als Plugin-System
⇒ CRM114-Plugin dafür anpassen?
- Vorteile:
 - sa-learn zum Training nutzbar.
 - ???

Problem: fork()

- CRM114 nicht in Perl
- ⇒ nicht mit im Speicher
- Jeder Aufruf startet neuen CRM-Interpreter

Problem: fork()

- CRM114 nicht in Perl
- ⇒ nicht mit im Speicher
- Jeder Aufruf startet neuen CRM-Interpreter
 - Auswirkung unklar
 - auf reinem Mailserver: alles im Cache?
 - CRM114-Daemon ist nicht geplant
 - Benchmark nötig!

Amavisd-new

A Mail Virus Scanner

- Daemon zur Integration von Virenscannern und SpamAssassin
- Einbinden per SMTP/LMTP
- Meist auf MTA in der Mail-Queue

Amavisd-new

SpamAssassin-Aufruf

- amavisd-new ruft auf:
Mail::SpamAssassin->check(\$mail_obj)
- empfängt von SA:
(`$spam_level`, `$sa_tests`,
`$spam_report`, `$spam_summary`)
- seit Version 2.5 auch `%supplementary_info`
- alles andere wird verworfen!

Amavisd-new

Probleme:

- BDB-Cache nur für Level, Status, Report und Summary
- Header-Erzeugung (Cache-ID) nicht frei konfigurierbar

Bisher per Patch behoben; „Richtige Lösung“ fehlt noch.

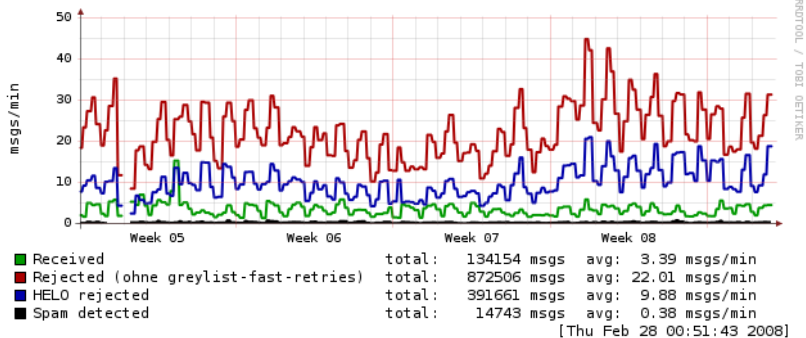
Wirksamkeit

oder auch: SMTP- gegen Inhaltsfilter

Lohnt sich der Aufwand?

Wirksamkeit

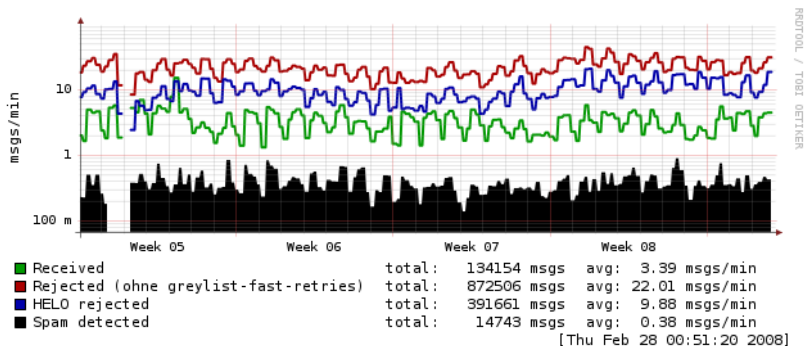
oder auch: SMTP- gegen Inhaltsfilter



(linear)

Wirksamkeit

oder auch: SMTP- gegen Inhaltsfilter



(logarithmisch)

Wirksamkeit

oder auch: SMTP- gegen Inhaltsfilter

- SMTP-Restriktionen bringen viel
- ⇒ erstmal SMTP gut konfigurieren
- Inhaltsfilter nur letztes Mittel von 90 % Ablehnen auf 99,x % Ablehnen/Erkennen

Quellen/Links

- CRM114 – the Controllable Regex Mutilator:
<http://crm114.sourceforge.net/>
- Apache SpamAssassin: <http://spamassassin.apache.org/>
- amavisd-new: <http://www.ijs.si/software/amavisd/>
- SpamAssassin CustomPlugins:
<http://wiki.apache.org/spamassassin/CustomPlugins>
- CRM114 Spamassassin-Plugin: <http://mschuette.name/wp/crm114-spamassassin-plugin/>