

Mail-Server: SpamAssassin: sa-learn anlernen

Problem:

SpamAssassin ist konfiguriert und läuft. Erkennt aber nur einen Bruchteil der eingehenden Spammails.

Erklärung:

SpamAssassin arbeitet zu dem Zeitpunkt nur mit ein paar Mustererkennungen und Regeln. Ein Großteil der aktuellen Spam-Wörter ist SpamAssassin noch nicht bekannt. Durch Einrichtung des s.g. Bayes-Filter erfolgt dann eine bessere Erkennungsrate.

Lösung:

SpamAssassin kann aktuelle Spam's lernen. Zum Anlernen gibt es zwei Strategien:

a) Automatisch

b) Manuell

Gerade in der Anfangszeit lohnt es Zweigleissig zu fahren. Denn der Bayesfilter braucht 200 gelernte Spam's bevor er richtig zu arbeiten anfängt.

Später kann man den Autolearn abschalten, wenn es z.B. Performance-Probleme gibt. Ansonsten hilft die Einstellung grundsätzlich uptodate zu bleiben.

Für a) stellt man einfach den `auto_learn` in der `local.cf` ein:

```
# use Bayes-Filter
use_bayes 1
bayes_path /var/spool/spamassassin/bayes
bayes_file_mode 777
# auto-learn einschalten
# fuer Version 2.x:
auto_learn 1
# ab Version 3.0:
bayes_auto_learn 1
```

Bei b) nutzt man das beiliegende Programm `sa-learn` und füttert das mit gesammelter Spammail:

```
# mbox-Format:
sa-learn --spam --mbox /var/mail/spam
```

Zurecht kommt jetzt die Frage: Woher bekomme ich die Spammails?

Mail-Server: SpamAssassin: sa-learn anlernen

Es gibt überwiegend 2 Antworten:

- Man hat bereits selbst gesammelt (z.B. in Outlook oder Thunderbird).
Der Thunderbird kann direkt einen ganzen Ordner im mbox-Format speichern.
Für Outlook braucht man externen Software wie z.B. den [MailNavigator](#).
- Die schnellere Methode ist allerdings sich aktuelle Spams im mbox-Format von [spamarchive.org](#) runter zuladen.
Update (da spamarchive.org offline gegangen ist):
Nicht aktuelle Spam-mboxen finden sich auf der [inoffiziellen deutsche TheBat! Seite](#).
Weitere [Spam-Mailboxen](#)

Eindeutige ID: #1176
huschi
2007-09-11 09:21