SPAM filtern mit Greylisting

Übersicht

Greylisting ist ein ziemlich wirkungsvolles Verfahren, um Spam-Emails abzuweisen. Es hat zwar auch einige Nachteile, nach einigen Wochen Praxiseinsatz haben wir uns jedoch dafür entschieden. Dieser Artikel enthält Infos sowie Konfigurationstipps.

Immer mehr SPAM

Die Spam-Flut auf meinem Server steigt immer weiter an: im Juni 2006 waren es noch etwa 500 - 1000 Spam-Emails pro Tag, inzwischen sind es schon 2000 bis 3000. Und das, obwohl dieser Server nur fünf Homepages hostet (die allerdings schon ziemlich lange im Netz sind). Es sind zwar einige Gegenmaßnahmen wie DNS-Blacklists, SpamAssassin und Clamav installiert, aber immer mehr Spam-Emails werden an Benutzer weitergeleitet.

Das größte Problem beim Filtern der Emails ist die Gefahr, dass legitime Emails als Spam eingestuft und dann später übersehen werden. Man ist also gezwungen, sich von allen Spam-Emails wenigstens den Betreff anzusehen, bevor man sie löscht. Hinzu kommt, dass die Spammer ihre Emails immer mehr verfeinern, um sie durch SpamAssassin hindurchzuschleusen. Das führt dazu, dass ein immer höherer Anteil an Spam-Emails nicht mehr als Spam markiert wird, oder - wenn man den SpamAssassin Bewertungsfaktor herabsetzt - die Wahrscheinlichkeit ansteigt, dass auch legitime Emails fälschlicherweise als Spam klassifiziert werden. Beide Alternativen sind gleich unangenehm für den Email-Nutzer.

Seit kurzem nutze ich zusätzlich zu den angegebenen Verfahren noch zusätzlich das Greylisting. Der Erfolg ist umwerfend: von ca. 3000 Spam-Emails pro Tag werden durch die Kombination von Blacklisting und Greylisting praktisch alle bis auf ein paar geblockt. Und der kümmerliche Rest wird von SpamAssassin markiert und in den „Junk" Ordner einsortiert, den man sich einmal am Tag kurz durchsieht und dann manuell löscht. Früher waren das immerhin einige zehn Spam-Emails pro Tag, bei denen man den Betreff lesen musste.

Wie Greylisting funktioniert

Das Greylisting Verfahren macht sich zunutze, dass die meisten Spam-Versende-Programme jede Email-Adresse nur jeweils ein einziges Mal verwenden. Kann die Spam-Email nicht beim ersten Mal ausgeliefert werden, dann macht das Programm einfach mit der nächsten Adresse weiter. Ein normaler Mailserver wird dagegen den Zustellversuch mehrfach wiederholen, notfalls mehrere Tage lang.

Beim ersten Zustellversuch wird die Annahme verweigert und die IP-Adresse des versendenden Mailservers sowie die Email-Adresse des Absenders und des Empfängers in eine Datenbank eingetragen. Versucht es der Mailserver (nach Ablauf einer konfigurierbaren Mindestwartezeit) ein zweites Mal, wird die Email dagegen entgegengenommen. In fast allen Fällen handelt es sich dabei um eine legitime Email, die über einen regulären Mailserver versendet wird.

Wenn der Absender ein zweites Mal eine Email an den gleichen Empfänger versendet, dann wird sie sofort übermittelt, da Absender, Empfänger und Mailserver ja bereits bekannt sind.

Nachteile

In der Theorie sollte es also nur geringfügige Nachteile geben. Nur die erste Email jedes Absenders wird ein wenig verzögert, ansonsten fällt der Einsatz des Greylistings keinem Email-Nutzer auf. Leider sieht die Praxis etwas anders aus.

Manche Emails werden erheblich verzögert, und zwar aus folgenden Gründen:

- Die meisten Mailserver verwenden ein Retry-Intervall von 30 Minuten, einige aber auch ein Intervall von 12 oder noch mehr Stunden.
- Große Unternehmen oder Webhoster verwenden meistens mehrere Mailserver. Es passiert häufig, dass erneute Zustellversuche dann von einem anderen Mailserver vorgenommen werden. Die Greylisting-Software sieht eine neue IP-Adresse, verweigert erneut die Annahme und startet erneut die Mindestwartezeit. Dies wiederholt sich solange, bis die Email von einem bekannten Mailserver versendet wird, der dann inzwischen freigeschaltet wurde.
  Da Unternehmen wie eBay oder Google dutzende von Mailservern einsetzen, kann sich das über viele Stunden hinziehen.

- Manchmal enthält die Absender-Email-Adresse einen Time Stamp oder irgendeinen Code. Dies ist oft bei automatisch generierten Emails wie Versandbestätigungen, Webanfragen oder dergleichen der Fall. Aus Sicht des Greylisting-Systems sind das natürlich jedesmal neue Absender, die dann wieder verzögert werden.

Man kann also nicht eben mal am Telefon seinen Gesprächspartner bitten, eine Datei zuzusenden und hat sie dann zwei Minuten später vorliegen. Auch die Bestätigungsmails bei Registrierung in Diskussionsforen, Shops usw. werden mehr oder weniger lang verzögert.

Es kommt sogar vor, dass manche legitimen Emails nicht zugestellt werden. Das passiert immer dann, wenn der versendende Mailserver den Response Code des empfangenden Mailservers falsch interpretiert und als endgültige Fehlermeldung auffasst und keinen weiteren Sendeversuch unternimmt. Unter anderem ist mir dies bei manchen (nicht allen) Emails von eBay passiert.

Wenn der versendende Mailserver korrekt konfiguriert ist, erhält der Absender der Email wenigstens eine Fehlermeldung.

Installation und Konfiguration

Auf einem Debian System ist lediglich das Paket "greylistd" zu installieren.

Danach müssen neue ACL Regeln in exim (vexim-acl-check-rcpt.conf) eingetragen werden, damit das alles funktioniert. Die Regel, die auf der greylistd Installationsanleitung empfohlen wird, enthält zusätzlich zur Abfrage des greylistd - Status noch eine Abfrage der Whitelist. Es ist jedoch sinnvoller, die Abfrage der Whitelist separat am Anfang der ACL Regeln durchzuführen, so wird auch die Abfrage der DNS Blacklists vermieden, falls ein Mailserver in der Whitelist aufgeführt ist. Sonst kann es passieren, dass zwar ein Mailserver in der Whitelist drinsteht, aber die Email trotzdem abgelehnt wird, weil der Mailserver auch in einer der Blacklists steht. Das passiert gelegentlich bei kostenlosen Webmail-Diensten.

```
# accept all email from authenticated hosts

  accept    authenticated = *

# accept all email from whitelist

  accept
    hosts          = : +relay_from_hosts : \
                     ${if exists {/etc/greylistd/whitelist-hosts}\
                                 {/etc/greylistd/whitelist-hosts}{}} : \
                     ${if exists {/var/lib/greylistd/whitelist-hosts}\
                                 {/var/lib/greylistd/whitelist-hosts}{}}

# deny emails from any host in these blacklists
# note: check the policy of each blacklist provider to see wether it
matches
# your requirements. uceprotect.net and bl.spamcop.net may be too
aggressive

  deny
    message        = DNSBL listed at $dnslist_domain\n$dnslist_text
    dnslists       = sbl-xbl.spamhaus.org:combined.njabl.org:list.dsbl.org:
dnsbl-2.uceprotect.net:dnsbl-1.uceprotect.net:bl.spamcop.net

# handle greylisting

  defer
    message        = $sender_host_address is not yet authorized to deliver
\
                     mail from <$sender_address> to <$local_part@$domain>.
\
                     Please try later.
    log_message    = greylisted.
    !senders       = :
    domains        = +local_domains : +relay_to_domains
    condition      = ${readsocket{/var/run/greylistd/socket}\
                                 {--grey \
                                  $sender_host_address \
                                  $sender_address \
                                  $local_part@$domain}\
                                 {5s}{}{false}}
```

Und hier meine /etc/greylistd/config Konfigurationsdatei mit einigen wichtigen Änderungen:

```
##############################################################################
### FILE:       /etc/greylistd/config
### PURPOSE:    Configuration settings for the "greylistd(8)" daemon
##############################################################################

[timeouts]
# Initial delay before previously unknown triplets are allowed to pass
# Default is 1 hour = 3600 seconds
retryMin     = 300

# Lifetime of triplets that have not been retried after initial delay
# Default is 8 hours = 28800 seconds
retryMax     = 259200

# Lifetime of auto-whitelisted triplets that have allowed mail to pass
# Default is 60 days = 5,184,000 seconds
expire       = 5184000


[socket]
# Path to the UNIX domain socket on which greylistd will listen.
# The parent directory must be writable by the user running 'greylistd'.
# Default path is "/var/run/greylistd/socket".
path         = /var/run/greylistd/socket

# UNIX filemode of that socket.  See "chmod(1)" for the meaning of this.
# Default mode is 0660.
mode         = 0660


[data]
# Update interval -- save data to the filesystem if if has been more
# than this many seconds (default 600) since the last save.
update       = 600

# Path to the file containing the current state of each data item
(triplet),
# along with some general statistics.
# Default is "/var/lib/greylistd/states".
statefile    = /var/lib/greylistd/states

# Path to the file that will contain the original, unhashed data for the
# "list" command.  Default is "/var/lib/greylistd/triplets".
tripletfile  = /var/lib/greylistd/triplets

# Whether or not to retain unhashed triplets, for the "list" command.
# Default is "true"
savetriplets = true
```

Anpassungen

retryMin ist die Zeitspanne, die mindestens verstreichen muss, damit eine Email akzeptiert wird. Eine Zeitspanne von 300 Sekunden scheint gut zu funktionieren. Die Idee hinter der längeren Zeitspanne von 1 Stunde ist, dass Spam-Emails nach dieser Zeitspanne evtl. schon in die bekannten DNS-Blacklists eingetragen sind. Nutzt der Server diese Blacklists, dann würde die Zustellung verhindert werden, auch wenn der Spammer die Zustellversuche mehrfach wiederholt.

retryMax ist die Zeitspanne, in der Einträge in der greylist gehalten werden. Es empfielt sich, den Defaultwert von 1 Tag deutlich zu vergrößern. Erstens gibt es Server, die ein sehr langes Retry-Intervall haben, zweitens sollte man sich gelegentlich die Einträge mal durchsehen, um problematische Emails (siehe Abschnitt Nachteile) ausfindig machen zu können und evtl. Anpassungen machen zu können.

expire ist die Zeitspanne, in der Einträge in der whitelist gehalten werden. Der Default ist 60 Tage, wesentlich längere Einträge schaden sicher auch nicht. Man will nur vermeiden, dass die whitelist ins Unendliche wächst.

Die Whitelist

Wie schon erwähnt, gibt es verschiedene Unternehmen, die entweder sehr viele oder falsch konfigurierte Mailserver haben. Die IP-Adressen dieser Server sollte man in die /etc/greylistd/whitelist-hosts Datei aufnehmen. Es gibt noch eine zweite Whitelist unter /var/lib/greylistd/whitelist-hosts, die mit einigen vordefinierten Einträgen versehen ist, daher kommen die eigenen Einträge in die /etc/greylistd/whitelist-hosts, damit diese bei einem eventuellen Update nicht überschrieben werden.

Diese whitelists werden nicht vom greylistd ausgewertet, sondern von der jeweiligen Mailsoftware, in diesem Falle von exim. Die Syntax richtet sich daher nach dem jeweiligen Programm.

Hier ein paar Beispiele, wie man an die Adressen der Mailserver kommt. Man durchkämmt gelegentlich die "greylist" nach problematischen Einträgen. Hier beispielsweise einige Emails von eBay, die nicht oft genug wiederholt wurden und folglich nie angekommen sind. Man sieht auch, dass immer andere IP-Adressen verwendet wurden.

```
$ greylist list --grey | grep arne@schirmacher.de | grep ebay
  2006-11-09 13:58:43      1  66.135.197.13 emailconfirm@ebay.com
arne@schirmacher.de
  2006-11-09 17:47:14      1  66.135.209.207 endofitem@ebay.de
arne@schirmacher.de
  2006-11-09 17:47:18      1  66.135.209.215 endofitem@ebay.de
arne@schirmacher.de
  2006-11-10 10:28:49      1  66.135.197.23 checkout@ebay.de
arne@schirmacher.de
  2006-11-10 10:41:21      1  66.135.197.28 bidconfirm@ebay.de
arne@schirmacher.de
```

Man sollte also die eBay-Mailserver in die Whitelist aufnehmen. Viele Unternehmen teilen die IP-Adressen oder Adressbereiche ihrer Mailserver in einem sogenannten SPF-Record mit. Diesen SPF-Record bekommt man mit folgendem Befehl:

```
$ dig ebay.com txt

...

;; ANSWER SECTION:
ebay.com.               3292    IN      TXT     "v=spf1 mx include:s._spf.
ebay.com include:m._spf.ebay.com
                                                 include:p._spf.ebay.com
include:c._spf.ebay.com ~all"
ebay.com.               3292    IN      TXT     "spf2.0/pra mx include:s.
_sid.ebay.com include:m._sid.ebay.com
                                                 include:p._sid.ebay.com
include:c._sid.ebay.com ~all"

...
```

Entscheidend ist die ANSWER SECTION. Hier stehen entweder die IP-Adressen oder Adressbereiche direkt drin, oder sind indirekt mit einer include-Direktive angegeben. In letzterem Fall muss man noch einmal dig für die angegebenen Domains aufrufen:

```
$ dig s._spf.ebay.com txt

...

;; ANSWER SECTION:
s._spf.ebay.com.        2790    IN      TXT     "v=spf1 ip4:66.135.209.192
/27 ip4:66.135.197.0/27
                                                 ip4:64.4.240.64/27 ip4:
64.4.244.64/27 ~all"

...
```

Diese IP-Adressen-Ranges kann man nun in die Whitelist eintragen (jedoch ohne den ip4: Prefix). Dies ist für alle Einträge zu wiederholen, evtl. auch rekursiv.

Hat ein Unternehmen keinen SPF-Eintrag, kann man zumindest die in den MX-Records gelisteten Mailserver in die Whitelist aufnehmen. Diese Server bekommt man mit "dig ebay.com mx" angezeigt.

Anbei meine Whitelist, die einige wichtige Server enthält (eBay, Google Mail, 1&1 Domains, GMX, web.de):

```
12.155.144.75 # ebay.com according to their SPF records
62.22.61.131
63.104.149.126
63.80.14.17
64.127.115.252
64.4.240.64/27
64.4.244.64/27
64.68.79.253
64.94.204.222
65.110.161.77
66.135.195.180
```

```
66.135.195.181
66.135.197.0/27
66.135.209.192/27
66.135.215.134
66.135.215.224/27
67.72.12.29
67.72.12.30
67.72.99.26
80.66.137.58
80.93.9.10
194.64.234.129/27
195.234.136.12
203.49.69.114
204.13.11.49
204.13.11.51
206.165.246.83
206.165.246.84
206.165.246.85
206.165.246.86
209.63.28.11
210.80.80.136
212.110.10.2
212.147.136.123
212.208.64.34
213.219.8.227
216.113.168.128
216.113.175.128
216.113.188.112
216.113.188.96
216.177.178.3
216.33.244.6
216.33.244.7
216.33.244.84
216.33.244.96/27
217.149.33.234
220.248.6.124

207.171.160.0/19 # amazon.com according to their SPF records
87.238.80.0/21
72.21.196.0/24
72.21.208.0/24

64.233.160.0/19 # gmail.com according to their SPF records
66.249.80.0/20
72.14.192.0/18
216.239.56.0/23

209.85.146.0/23 # googlemail.com

212.227.126.128/25 # moutng.kundenserver.de (1&1) according to their SPF
records

81.169.144.0/20 # Strato AG (this range is too large)
```

```
213.165.64.0/23 # gmx.net

217.72.192.221 # web.de (fmmailgate01.web.de - fmmailgate09.web.de)
217.72.192.227
217.72.192.234
217.72.192.242
217.72.192.243
217.72.192.247
217.72.192.248
217.72.192.249
217.72.192.184

194.25.134.80 # t-online.de
194.25.134.17
194.25.134.81
194.25.134.18
194.25.134.82
194.25.134.19
194.25.134.83
194.25.134.20
194.25.134.84
194.25.134.21
194.25.134.85 152.163.225.0/24 # aol.com according to their SPF records
205.188.139.0/24
205.188.144.0/24
205.188.156.0/23
205.188.159.0/24
64.12.136.0/23
64.12.138.0/24
205.188.158.121
64.12.137.184
64.12.137.249
205.188.156.185
205.188.157.25
64.12.137.168
64.12.138.185
205.188.155.89
205.188.157.217
205.188.159.57
64.12.138.57
64.12.138.120
64.12.138.152
205.188.156.249
205.188.159.217
64.12.138.89
```

Links

http://de.wikipedia.org/wiki/Greylisting

http://en.wikipedia.org/wiki/Greylisting

http://www.greylisting.org/

http://projects.puremagic.com/greylisting/whitepaper.html

Eure Erfahrungen?

Falls ihr Greylisting ebenfalls einsetzt, Fragen oder weitere Vorschläge für die Whitelist habt, hinterlasst bitte einen Kommentar oder schickt mir eine Email.