Fighting Spam With Greylisting

Barbara Dijker
barb@netrack.net
Nov 11, 2004

Intro

- Greylisting is a new approach to blocking spam and virus email
- It is useful in conjunction with other mechanisms
- The concept is deceptively simple yet surprisingly effective
- There are no false positives
- However, there are issues with noncompliant SMTP servers

Context: Blacklisting

- Blacklisting rejects messages
  - Before accepting them in SMTP session
  - Based on only *one of*:
    - Recipient
    - Sender
    - Relay/Sending IP
- Lists are effectively static and manually maintained, even if obtained via DNS
- Can reject significant legitimate mail
- Few effective lists for virus plagued IPs

Context: Content Filtering

- Spamassassin and friends
  - Accept the message
  - Use server resources to analyze content
- Filter rules and lookups are effectively static or context dependent (learned)
- False positives are a concern
- Reducing false positives reduces effectiveness
- Easy workarounds, e.g. image body
- Separate tools needed for viruses
- Resource intensive

Context: Challenges

- ASK and friends:
  - Accept the message
  - Reply with a challenge for human response
  - Deliver the message to inbox after challenge response
- Extremely effective
- Clogs outgoing mail queue with failed attempts to bogus spam return addresses
- Doubles the volume of spam and then some by auto-response to every message

Greylisting Approach

- Dynamic SMTP session blocking based on "triple" of:
  - Envelope sender address
  - Envelope recipient address
  - Peer IP address
- Utilizes an existing feature of SMTP
  - 4yz Transient Negative Completion
  - The RFC says MTA "is encouraged to try again"
- The first occurrence of a triple results in the 451 temporary failure reply
- Subsequent tries are accepted

Why it works

- Well-behaved MTAs requeue and retry when a 451 is received
- Requeue and retry is expense:
  - requires cycles and disk space
  - takes time - enough to get caught / blacklisted
  - requires more programming, not simple
- Spam mail blasters and virus agents are simplistic and don’t requeue/retry

When it doesn't work

- Time sensitive email, e.g., ebay.
- Large corporate mail blasters
  - legitimate mail
  - intentionally don't retry
  - E.g., united.com, yahoogroups, daily news
- These are easily excepted / whitelisted through an existing list
- There a minor few broken MTAs out there

Parameters

- Delay
  - Length of time the server continues to refuse a new triple
  - Default 1 hour
- Lifetime of a new triple
  - Length of time the server remembers a new triple, in which the message must be retried
  - Default 4 hours
- Lifetime of an updated triple
  - Length of time the server remembers a retried triple, in which mail will be accepted
  - Default 36 days

Example

- Message attempted:
  - From billp@loneagle.com
  - To steve@netrack.net
  - IP 194.227.21.77
- 451 Please try again later (TEMP FAIL)
- Sending MTA requeues
- Message attempted on next queue run
- Message accepted and delivered
- Subsequent messages from Bill to Steve are delivered without delay

Example

- Greylisting acts on the full unique triple
- Messages forged from Bill to Steve yet sent from any other server are rejected the first time
- Messages from Bill to other users @netrack.net are rejected the first time
- The triple has to match exactly

Pros

- Over 95% effective “out of the box”
- Effective on spam and virus mail
- No false positives
- Not subjective
- Requires almost no maintenance
- Completely transparent to users
- Available for most open source MTAs
- Significantly reduces server load
  - server accepts up to 80% less mail
  - content filters and challenges only process real mail

Cons

- Delays mail the first time for each triple
  - At least as long as initial delay configured
  - Actually as long as *sending* MTA retry interval
- A few broken MTAs may require whitelisting
- Has no affect on mail relayed through a real MTA that retries, e.g., open relays
- No effect on forwarded email
- Spam gets through if they resend to the same list from the same server before they are on a blacklist/dcc

Won’t spammers adapt?

- The easiest way is by using open relays
- We already have a way to deal with this: open relay blacklists
- Spam sent via bounce may also go through
- Otherwise, spammers...
  - have to incur the cost of queuing and retries
  - may be on a blacklist or dcc by the time they retry
- This cost shifting is a good thing - it changes the value proposition of spam

Greylisting Server Load

- Doesn’t greylisting add load to the server?
- Every message has to be received twice?
- After about a week, over 95% of delayed messages are spam and viruses
- Regular correspondence is already auto-whitelisted
- It adds a few database lookups
- The load it saves by NOT accepting spam and viruses far outweighs the load it adds, by about a factor of 4

Noncompliant MTAs

- Mail server software that does not properly requeue and retry after receiving a 451 STMP response:
  - Novell Groupwise 6
  - InterMail 4.0
  - ISMail 1.7.8
- All of these are down rev
- Senders who receive 451 bounces should upgrade

Implementations

- SA-Exim (variant w/spamassassin)
- OpenBSD (built into spamd)
- FreeBSD (comes w/relaydelay)
- Exim (5 add-on varieties)
- Sendmail (relaydelay perl milter)
- DCC milter extension
- Qmail and Qmail spp
- Qpsmtpd
- Postfix (6 add-on varieties)
- SMTPwrap (via inetd)

Sendmail w/relaydelay

- Requires
  - Sendmail 8.12+ with milter compiled in
  - Mysql (which can be running externally)
  - Threaded Perl 5.8.0+
  - Sendmail::Milter
  - DBI and DBD::Mysql
- Relaydelay
  - Perl script
  - Current development version 0.05
- Everything is at greylisting.org

Installation: conf

- Modify relaydelay.conf

```
$database_type = 'mysql';
$database_name = 'relaydelay';
$database_host = 'localhost';
$database_port = 3306;
$database_user = 'xxxxx';
$database_pass = 'xxxxxx';
$delay_mail_secs = 60 * 60;  # 60 minutes
$auto_record_life_secs = 4 * 3600;  # 4 hours
$update_record_life_secs = 36 * 24 * 3600; # 36 days
```

- Several other parameters too
- Example file is well commented
- Install into /etc/mail
- Should not be world readable

Installation: db

- Create Mysql database, tables, users
  - Modify mysql.sql script with username and password you put in relaydelay.conf
  - Run mysql with mysql.sql input
- Create basic whitelist entries for known exceptions like ebay
  - Run mysql with whitelist.mysql input
- Create your own whitelist entries
  - Use whitelist.mysql as an example
  - Whitelist everything you allow to relay as relay entries in mail access will be ignored

Installation: sendmail

- Add milter hook to sendmail mc file

```
define(`MILTER',1)
INPUT_MAIL_FILTER(`relaydelay', `S=local:/var/run/relaydelay.sock, F=T,
T=S:1m;R:2m;E:3m')dnl
```

- The define is only needed once in your mc file
- F=T (optional)
  - If the milter is unavailable, all connections will be rejected with 451
- Rebuild your cf file and install

Installation: relaydelay

- Install relaydelay.pl where you want it
- Make sure it is executable
- Restart sendmail with new cf
- Run relaydelay.pl
- Make sure it looks fine
- Turn verbose off in relaydelay.conf
- Restart relaydelay.pl
- Modify your boot scripts to run it at boot time

Keeping an eye on

- syslog, patch available from Phil Kizer <pckizer@nostrum.com>

  Nov 11 02:43:38 cybox greylist[1451]: iAB9haJY012306: **triplet never seen**, inserting #765734: [211.222.119.111] <xuzbla@dreamwiz.com> <dijker@labyrnith.com>
  Nov 11 02:43:39 cybox greylist[1451]: iAB9haJY012306: **triplet never seen**, inserting #765735: [211.222.119.111] <xuzbla@dreamwiz.com> <barb.dijker@labyrnith.com>
  Nov 11 02:44:29 cybox greylist[1451]: iAB9iHHA012368: **triplet known and block expired**, passing for #159121: [195.186.1.207] <stefania.xxxx@xxxx.ch> <xxxx@agentxxxxxx.com>
  Nov 11 02:44:40 cybox greylist[1451]: iAB9icmN012395: **triplet known and block expired**, passing for #89137: [199.239.138.162] <daily_headlines@ms1.lga2.nytimes.com> <lasvegas@bigboxxxxxx.com>
  Nov 11 02:47:48 cybox greylist[1451]: iAB9lmRO012788: **Whitelisted relay**: 206.168.112.154 entry #16922
  Nov 11 02:48:31 cybox greylist[1451]: iAB9mL2s012894: **triplet known but block still active**, TEMP FAIL for #765758: [218.27.234.24] <cathleen.pugh_gc@a-vinstouw.dk> <barb.dijker@labyrnith.com>

Multiple Servers

- Greylisting must be enabled on all MX hosts for your domain
  - Otherwise spam will just come through the non-greylisting MX servers
- Use a central greylist database
  - Otherwise messages will be rejected with 451 by each MX

Poprelayd

- If you use poprelayd or another dynamic relay whitelist, it needs to work with greylisting
- I have a poprelayd patch that writes to the relaydelay mysql database
- www.netrack.net/misc/poprelayd.diff
- Requires these in poprelay.conf

```
# Greylist database to update
$greylist = "relaydelay";
$greylist_conf = "/etc/mail/relaydelay.conf";
```

Issues: server pools

- Some large organizations have clusters of mail servers
- Each attempt to retry a message comes from a different IP in the cluster!
- Mitigate with
  - $do_relay_lookup_by_subnet = 1;
    - Matches if relay ip was from same subnet
  - $check_wildcard_relay_ip = 1;
    - Allows whitelist wildcards like 64.

Issues: no retry, etc

- Many well-known corporate servers don't retry intentionally because
  - Information is timely and useless if delayed, or
  - Volume of mail too large for their resources, or
  - They're just lazy
- Some servers use a unique envelope sender with every try, usually mailing list servers or marketing mail
- Whitelist these, e.g., ebay, united, etc.

Issues: retry too slowly

- Some servers, notably Comcast, retries once after 24 hours
  - Sigh
  - $auto_record_life_secs = 25 * 3600; # 25 hours
- This does not appreciably increase the amount of spam allowed (yet)
- They either retry within seconds or not

Case Example

- About 800 users: residential & business
- Before Greylisting
  - 300 connections per hour blacklisted
  - 400 messages per hour accepted
  - 125 m/hr tagged by spamassassin
- After Greylisting
  - 300 connections per hour blacklisted
  - 250 c/hr refused by greylisting and not retried
  - 150 m/hr accepted
  - 30 m/hr tagged by spamassassin
- 80% of connections are rejected!
- Content filter load reduced by 5x

Users love it

- ISP employees gained about 1 hr / day with a cleaner mailbox
- After an initial adjustment/education period, nothing but raves
- Set initial delay to 0 to “learn” for 1-4 weeks for a more gentle transition
- My personal mailboxes have 300+ fewer messages a day

Resources

- Mailing lists
  - projects.puremagic.com/greylisting
- Whitepaper
  - Also at puremagic

Other stuff: SPF

- Sender Policy Framework
- spf.pobox.com
- NOT an anti-spam tool
- SPF is a mechanism which makes it easier to identify spoofed envelope sender (return-path) addresses
- You should be interested if you don’t want your domain spoofed

Other stuff: bounce relay

- Mail servers are vulnerable if they accept a message in SMTP before actual validation of recipients
- This is a “feature” to minimize smarts in the smtp server to make it more secure
- Qmail can be vulnerable, e.g.
- Spammer sends to known invalid addresses with the intended recipient as the envelope sender
- The mail server accepts the message, fails delivery, and queues the bounce