

jgc's spam and anti-spam newsletter

Feature Article: Greylisting Explained December 31, 2004

jgc's spam and anti-spam newsletter is a biweekly email newsletter written by POPFile author John Graham-Cumming. Each newsletter contains a feature article on spam or anti-spam techniques.

To subscribe to the newsletter visit <http://www.jgc.org/>

This article is Copyright © 2004 John Graham-Cumming. If you want to use all or part of this article in your own publication or presentation please drop an email to antispam@jgc.org.

Greylisting

Greylisting is a technique designed to stop spam before it arrives in your inbox. It's implemented on the mail server that initially receives your mail, and works by delaying mail from senders that the mail server has not encountered before.

Delaying the mail works because the SMTP protocol¹ that is used to send all Internet email has a mechanism for handling mail servers that are overloaded or unable to process mail. Greylisting fakes the overload situation, when an email arrives from a previously unseen sender, causing the message to be delayed.

If the sender is legitimate then their mail server will try again (typically within 4 hours). Spammers and email viruses tend to be fly-by-night operations when sending mail and will not check to see if the mail got through and will not resend undelivered messages.

By implementing greylisting on a mail server most spam and virus email is automatically eliminated. Naturally, there are problems that make greylisting a less than perfect solution, but it's a viable way of eliminating spam and viruses if you control your personal or organization's incoming mail server.

How Greylisting Works

Greylisting was first proposed by Evan Harris in a whitepaper² in August 2003. In his original proposal he suggestion the following action be taken when an email arrives at an incoming mail server:

1. Make a note of three things: the IP address of the mail server sending the message, the sender's address and the recipient's address. This is called the triplet.
2. If that triplet has been seen before then accept the message and deliver it.
3. If that triplet is new then tell the sending server that the server is too busy right now. Make a note of the triplet so that if the sending server retries the message will be accepted.

The system quickly learns who usually sends mail to the server, and rejects email that is not resent. For example, when an email arrives from a sender that is known to the system, the following exchange of messages occurs between the sending and receiving machines:

<i>Sending Machine Says...</i>	<i>Receiving Server Replies...</i>
HELO sending-machine.org	
	250 Hello there, 192.168.123.1
MAIL FROM: <user@sending-machine.org>	
	250 Sender OK

1 <http://www.faqs.org/rfcs/rfc821.html>

2 <http://projects.puremagic.com/greylisting/whitepaper.html>

<i>Sending Machine Says...</i>	<i>Receiving Server Replies...</i>
RCPT TO: <you@receiving-machine.com>	
	250 Recipient Looks Good
<i>Sends the email message and disconnects</i>	

As soon as the sending machine has sent the MAIL FROM and RCPT TO commands the receiving machine knows everything it needs to implement greylisting. In the example above, the triplet consists of the IP address of the sending machine (which the receiving machine has found to be 192.168.123.1), the sender (user@sending-machine.org) and the recipient (you@receiving-machine.com). If that particular combination has been seen before the receiving server gives the go ahead to the sending machine to start sending the email message by replying 250 Recipient Looks Good. The code 250 is the SMTP way of saying "OK, I did what you asked".

If that particular triplet had not been seen before the conversation would go like this:

<i>Sending Machine Says...</i>	<i>Receiving Server Replies...</i>
HELO sending-machine.org	
	250 Hello there, 192.168.123.1
MAIL FROM: <user@sending-machine.org>	
	250 Sender OK
RCPT TO: <you@receiving-machine.com>	
	451 Server busy, try again later
QUIT	
	221 Goodbye, come back soon

The 451 reply means that the receiving machine is unable to carry out the request made by the sender, but that this is not a permanent error and so the machine should retry. At this point the receiving machine makes a note of the triplet (192.168.123.1, user@sending-machine.org, you@receiving-machine.com) so that if the sending machine is a real mail server, and not a spammer's machine, when it replies, the message will be accepted without further delay (just as in the first example conversation above).

Almost all real SMTP email servers will retry the conversation within 1 to 4 hours of the temporary rejection. Naturally, once a particular triplet has been accepted there is no further delay in receiving mail from that sender. The greylisting system essentially learns who the regular, real correspondents of a particular organization are and only delays mail for spammers and previously unseen senders.

If the mail is retried within one hour the receiving server will respond with the 451 message to prevent fast retries by a spammer. After one hour the message will be delivered.

Spammer Adaptation

Spammers could adapt to greylisting by keeping track of messages that were delayed and retrying them. This adds cost to the spammer's business because they must maintain the database of delayed messages and track which servers were used to send each message. Since spammers typically use compromised machines or open proxies to send their messages they must keep track of the server used on a per-message basis and retry through the same proxy or compromised machine.

In addition the spammer must not retry too quickly, because typical greylisting implementations have a one hour delay before they will accept a retried email. This means that the spammer is forced to reuse the IP address after one hour. If greylisting is used in combination with a blacklist³ which is maintaining a list of IP addresses that are being used for spamming then greylisting is even more effective because even if the spammer retries after one hour it is likely that their spam will be rejected based on the blacklist.

What's Good

1. Greylisting is conceptually simple and easy to implement. There are now many open source implementations of greylisting available for every popular SMTP email server⁴.
2. Greylisting adapts as spammers change their message types. Since greylisting only relies on three things, the IP address, sender's email address and recipient's email address, it does not matter how spammer's obfuscate or forge their emails.
3. Greylisting works better as spammers make use of zombie networks. Because spammers frequently change their IP address by sending their spam through networks of compromised computers (known as zombies) greylisting automatically removes their spam.
4. Greylisting also fights email viruses and worms. Because many include their own SMTP code that doesn't handle retries, greylisting automatically prevents viruses and worms from being delivered.
5. Greylisting rejects email before it is sent which saves the recipient from paying bandwidth costs associated with spam email. Only a small conversation consisting of a few bytes (the MAIL FROM and RCPT TO messages) is enough to reject a spam without seeing the body of the message.

What's Bad

1. Good mail from senders that have not been seen before by the system is delayed for up to 4 hours. Users may start worrying ,for example, why their airline flight reservation confirmation has not arrived.

³ e.g. <http://www.spamhaus.org/sbl/index.lasso>

⁴ <http://projects.puremagic.com/greylisting/links.html>

2. You must be the administrator of your email server to implement Greylisting. This isn't something an individual user can decide to install.
3. There are some legitimate mailers that may retry messages from a different IP address because they have a block of servers that handle their outgoing mail. For these situations it is necessary to configure the greylisting software to recognize the different IP addresses as being the same.
4. It may be unacceptable for certain types of email to be delayed (for example, email to a company's `support@` email alias might require a very fast answer). In which case the mail administrator must manage a whitelist of recipient email addresses that bypass the greylisting system.

Where to find out more

The best starting point for learning about greylisting is Evan Harris' own web site: <http://projects.puremagic.com/greylisting/>. There's a lot of information there about the idea, about actual implementations and there's a email list for discussion.