Measuring the Role of Greylisting and Nolisting in Fighting Spam

F. Pagani\textsuperscript{1} M. De Astis\textsuperscript{2} M. Graziano\textsuperscript{1} A. Lanzi\textsuperscript{2} D. Balzarotti\textsuperscript{1}

\textsuperscript{1}Eurecom
Sophia Antipolis, France

\textsuperscript{2}Università degli Studi di Milano
Milano, Italy

International Conference on Dependable Systems and Networks, 2016
A lot of research has been done on spam filtering techniques:

- Sender-based: blacklists, IP reputation, server auth...
- Content-based: bayesian filters, email prioritization...

Greylisting and Nolisting are two relatively-unknown sender-based approaches, not well studied.
A lot of research has been done on spam filtering techniques:

- Sender-based: blacklists, IP reputation, server auth...
- Content-based: bayesian filters, email prioritization...

**Greylisting** and **Nolisting** are two relatively-unknown sender-based approaches, **not** well studied
Very simple technique

Primary mail server non-existent

RFC-2821 compliant:

“To provide reliable mail transmission, the SMTP client MUST be able to try (and retry) each of the relevant addresses in this list in order, until a delivery attempt succeeds… In any case, the SMTP client SHOULD try at least two addresses.”
Intro
Nolisting

MTA

DNS

Primary MailServer
(foo.smtp.net)

Secondary MailServer
(foo1.smtp.net)

MX QUERY for foo.net

MX 0 smtp.foo.net
MX 15 smtp1.foo.net

A QUERY for smtp.foo.net
ANSWER: 1.2.3.4
HELO local.name

A QUERY for smtp1.foo.net
ANSWER: 5.6.7.8
HELO local.name

250 Hello local.name, I am glad to meet you
MTA

MX QUERY for foo.net

MX 0 smtp.foo.net
MX 15 smtp1.foo.net

A QUERY for smtp.foo.net
ANSWER: 1.2.3.4

HELO local.name

A QUERY for smtp1.foo.net
ANSWER: 5.6.7.8

HELO local.name

Primary MailServer
(foo.smtp.net)

Secondary MailServer
(foo1.smtp.net)

250 Hello local.name, I am glad to meet you

MTA

DNS

Primary MailServer
(foo.smtp.net)

Secondary MailServer
(foo1.smtp.net)

MX QUERY for foo.net

MX 0 smtp.foo.net
MX 15 smtp1.foo.net

A QUERY for smtp.foo.net
ANSWER: 1.2.3.4

HELO local.name

A QUERY for smtp1.foo.net
ANSWER: 5.6.7.8

HELO local.name

250 Hello local.name, I am glad to meet you
MTA

MX QUERY for foo.net
MX 0 smtp.foo.net
MX 15 smtp1.foo.net

A QUERY for smtp.foo.net
ANSWER: 1.2.3.4
HELO local.name

A QUERY for smtp1.foo.net
ANSWER: 5.6.7.8

DNS

Primary MailServer
(foo.smtp.net)

Secondary MailServer
(foo1.smtp.net)

HELO local.name

250 Hello local.name, I am glad to meet you
Intro
Nolisting

MTA  DNS  Primary MailServer (foo.smtp.net)  Secondary MailServer (foo1.smtp.net)

MX QUERY for foo.net
MX 0 smtp.foo.net
MX 15 smtp1.foo.net

A QUERY for smtp.foo.net
ANSWER: 1.2.3.4
HELO local.name

A QUERY for smtp1.foo.net
ANSWER: 5.6.7.8
HELO local.name

250 Hello local.name, I am glad to meet you
Message rejected for a certain amount of time (greylisting threshold)

The MTA keeps trying until the message is accepted

Further messages accepted without delay:

\(<\text{sender\_address}, \text{sender\_ip}, \text{recipient\_address}>\)

RFC-2821 compliant:

“The sender MUST delay retrying a particular destination after one attempt has failed...Retries continue until the message is transmitted or the sender gives up; the give-up time generally needs to be at least 4-5 days.”
Greylisting

MTA

Primary MailServer

(foo.smtp.net)

HELO local.domain.name

250 Hello local.domain.name

MAIL FROM: <sender@local.domain.name.net>

250 Sender OK

RCPT TO: <recipient@foo.net>

450 Recipient address rejected: Greylisted

RCPT TO: <recipient@foo.net>

250 Recipient OK

\[\text{threshold}\]
Intro

Greylisting

MTA

HELO local.domain.name

250 Hello local.domain.name

MAIL FROM: <sender@local.domain.name.net>

250 Sender OK

RCPT TO: <recipient@foo.net>

450 Recipient address rejected: Greylisted

RCPT TO: <recipient@foo.net>

250 Recipient OK

Primary MailServer

(foo.smtp.net)

\[\text{threshold}\]
Greylisting

MTA

Primary MailServer
(foo.smtp.net)

HELO local.domain.name

250 Hello local.domain.name

MAIL FROM: <sender@local.domain.name.net>

250 Sender OK

RCPT TO: <recipient@foo.net>

450 Recipient address rejected: Greylisted

RCPT TO: <recipient@foo.net>

250 Recipient OK
The main assumption of the two techniques is that spam-bots are not RFC-compliant (fire-and-forget).

**Pros**
- Easy to implement
- RFC Compliant
- Do work

**Cons**
- Easy to evade
- Benign email lost/delayed
- Don’t work
I'd love to hear if you guys are using greylisting as part of your anti-spam strategy. I've heard it can help in some cases but will definitely cause issues for legitimate mail. How is it working for you?

edit: enabled it just now. So far no major issues. Thanks for the advice!
Is greylisting still an efficient method for preventing spam?

I've used greylisting on my servers for many years, but I don't know how effective it is nowadays.

Is it still good for fighting spam in 2012?

Or is the typical spammer MTA capable of resending greylisted emails now?
Motivation

F. Pagani, M. De Astis, M. Graziano, A. Lanzi, D. Balzarotti
Measuring the Role of Greylisting and Nolisting in Fighting Spam

I didn't have a handle on what greylisting actually entails until the past couple of weeks when we first used it as a last-ditch effort to get a large (for us) customer's spam situation under control. It worked, and this is highly effective for stopping spam.

Greg Askew
Oct 12 at 22:43

Is greylisting still an efficient method? I've used greylisting on my server for years. It's still good for fighting spam.

Michael Hampton
Oct 12 at 12:47

42

greylisting is highly effective for stopping spam. It might be common knowledge, but I didn't have a handle on what greylisting actually entails until the past couple of weeks when we first used it as a last-ditch effort to get a large (for us) customer's spam situation under control. It worked.

Adams
Oct 9 at 12:35

42

greylisting is highly effective for stopping spam. It might be common knowledge, but I didn't have a handle on what greylisting actually entails until the past couple of weeks when we first used it as a last-ditch effort to get a large (for us) customer's spam situation under control. It worked.

Adams
Oct 9 at 12:35

42

greylisting is highly effective for stopping spam. It might be common knowledge, but I didn't have a handle on what greylisting actually entails until the past couple of weeks when we first used it as a last-ditch effort to get a large (for us) customer's spam situation under control. It worked.

Adams
Oct 9 at 12:35

42
Motivation

Greylisting: The worst thing to happen to email since spam

By Marco on April 5, 2007

Greylisting (or graylisting) is a new anti-spam measure implemented on email servers. It’s starting to be a common known spam. (sell.sysadmin) submitted 2 years ago by bluesoul 21.9k • 3 • 24 • 49

I didn’t have a handle on what grey does have pros and cons. – Michael Hampton • Oct 9 '12 at 12:24
Contributions

- Worldwide adoption of Nolisting
- Impact on spam delivery
- Greylisting and the Real World
Adoption of Nolisting

We used two dataset from scans.io (zmap):

1. DNS records (135M domains):
   - d.com mx 0 smtp.f.net
   - d.com mx 15 smtp1.f.net
   - smtp.f.net a 1.2.3.4

2. Full IPv4 SMTP:
   - 1.1.1.1
   - 1.2.3.10
   - 1.3.4.5

Steps

- D → MX₁, MX₂...
- MXᵢ → IPᵢ
- Nolisting:
  - IP₁ ∉ IPv4SMTP
  - IP₂ ⊂ IPv4SMTP
Adoption of Nolisting

We used two dataset from scans.io (zmap):

1. DNS records (135M domains):
   
   d.com   mx 0  smtp.f.net
   d.com   mx 15 smtp1.f.net
   smtp.f.net   a   1.2.3.4

2. Full IPv4 SMTP:
   
   1.1.1.1
   1.2.3.10
   1.3.4.5

Steps

- D → MX<sub>i</sub>, MX<sub>j</sub>...
- MX<sub>i</sub> → IP<sub>i</sub>
- Nolisting:
  
  IP<sub>1</sub> ∉ IPv4SMTP
  IP<sub>2</sub> ⊂ IPv4SMTP
Adoption of Nolisting

We used two dataset from scans.io (zmap):

1. **DNS records (135M domains):**
   - d.com mx 0 smtp.f.net
   - d.com mx 15 smtp1.f.net
   - smtp.f.net a 1.2.3.4

2. **Full IPv4 SMTP:**
   - 1.1.1.1
   - 1.2.3.10
   - 1.3.4.5

Steps

- **D → MX₁, MX₂..**
- **MXᵢ → IPᵢ**
- **Nolisting:**
  - \( IP₁ \not\subset IPv4SMTP \)
  - \( IP₂ \subset IPv4SMTP \)
Adoption of Nolisting

- One entry: 47.73%
- DNS misconf.: 5.78%
- Nolisting: 0.52%
- Not using Nolisting: 45.97%
Adoption of Nolisting

Notes

- 0.52% represent more than 500k domains
- Five in Alexa top-1000:
  - 1 domain top 15
  - 2 domains top 500
  - 2 domains top 1000

Not very well known, but used by large organizations!
Adoption of Nolisting

Notes

- 0.52\% represent more than 500k domains
- Five in Alexa top-1000:
  - 1 domain top 15
  - 2 domains top 500
  - 2 domains top 1000

Not very well known, but used by large organizations!
Impact on Spam Delivery

Goals

Questions

- Are the techniques still working against modern malware?
- If not, how malware is able to bypass them?
- What is the “best” Greylisting threshold?
Impact on Spam Delivery

Setup

- Win7
- DNS Server (Nolisting)
- Postfix Server (Greylisting)

F. Pagani, M. De Astis, M. Graziano, A. Lanzi, D. Balzarotti Measuring the Role of Greylisting and Nolisting in Fighting Spam 14 / 27
Impact on Spam Delivery

Approach

- Spamming botnets from Symantec Internet Security Threat Report
- Samples collected from different sources (malwr.com, virustotal.com, virusshare.com)

<table>
<thead>
<tr>
<th>Malware Family</th>
<th>Percentage of Botnet Spam</th>
<th>Number of Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cutwail</td>
<td>46.90%</td>
<td>3</td>
</tr>
<tr>
<td>Kelihos</td>
<td>36.33%</td>
<td>6</td>
</tr>
<tr>
<td>Darkmailer</td>
<td>7.21%</td>
<td>1</td>
</tr>
<tr>
<td>Darkmailer(v3)</td>
<td>2.58%</td>
<td>1</td>
</tr>
<tr>
<td>Total Botnet Spam</td>
<td>93.02%</td>
<td>11</td>
</tr>
<tr>
<td>Total Global Spam</td>
<td>70.69%</td>
<td></td>
</tr>
</tbody>
</table>

- Each sample executed in **isolation**, collecting network traces and server logs
### Impact on Spam Delivery

Are the techniques still working against modern malware?

<table>
<thead>
<tr>
<th>SAMPLE</th>
<th>GREYLISTING</th>
<th>NOLISTING</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cutwail:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sample1</td>
<td>✓</td>
<td>x</td>
</tr>
<tr>
<td>sample2</td>
<td>✓</td>
<td>x</td>
</tr>
<tr>
<td>sample3</td>
<td>✓</td>
<td>x</td>
</tr>
<tr>
<td><strong>Kelihos:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sample1</td>
<td>x</td>
<td>✓</td>
</tr>
<tr>
<td>sample2</td>
<td>x</td>
<td>✓</td>
</tr>
<tr>
<td>sample3</td>
<td>x</td>
<td>✓</td>
</tr>
<tr>
<td>sample4</td>
<td>x</td>
<td>✓</td>
</tr>
<tr>
<td>sample5</td>
<td>x</td>
<td>✓</td>
</tr>
<tr>
<td>sample6</td>
<td>x</td>
<td>✓</td>
</tr>
<tr>
<td><strong>Darkmailer:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sample1</td>
<td>✓</td>
<td>x</td>
</tr>
<tr>
<td><strong>Darkmailer(v3):</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sample1</td>
<td>✓</td>
<td>x</td>
</tr>
</tbody>
</table>

A ✓ sign means the technique was **effective** to prevent spam
A X sign means the technique was **ineffective** against that malware
Nolisting Bypass

How the malware is able to bypass Nolisting?

Inspecting the DNS logs revealed that:

- **Kelihos (✓):** Only target the primary mail server
- **Cutwail (✗):** Targets the lowest priority mail server
- **Darkmailer (✗):** RFC compliant - from highest to lowest
- **Darkmailer v3 (✗):** RFC compliant - from highest to lowest
Greylisting Threshold

How does the threshold affect spam delivery?

CDF of the spam delivery delay with greylisting at 300 seconds
Greylisting Threshold
How does the threshold affect spam delivery?

CDF of the spam delivery delay with greylisting at 5 seconds
Retransmission delays of Kelihos with a greylisting threshold of 21600 seconds. In blue the failed attempts (below the threshold) and in red the delay of delivered emails (above the threshold).
CDF of spam delivery delay with threshold at 300 seconds:
real-world mailbox
vs.
malware samples
<table>
<thead>
<tr>
<th>PROVIDER</th>
<th>SAME IP</th>
<th>ATTEMPTS</th>
<th>DELIVER</th>
<th>DELAYS (min:sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>gmail.com</td>
<td>✗ (7)</td>
<td>9</td>
<td>✓</td>
<td>6:02, 29:02, 56:36, 98:44, 162:03, 229:44 309:05, 434:46</td>
</tr>
<tr>
<td>hotmail.com</td>
<td>✓</td>
<td>94</td>
<td>✓</td>
<td>1:01, 2:03, 3:04, 5:06, 8:07, 12:08, 16:10 362:11</td>
</tr>
<tr>
<td>qq.com</td>
<td>✗ (2)</td>
<td>12</td>
<td>✗</td>
<td>5:05, 5:11, 5:17, 6:19, 8:22, 12:25, 20:29, 52:31, 84:35, 144:42, 204:56</td>
</tr>
<tr>
<td>mail.com</td>
<td>✗ (2)</td>
<td>10</td>
<td>✓</td>
<td>5:02, 12:37, 23:59, 41:03, 66:38, 105:01, 162:35, 248:56, 378:28</td>
</tr>
</tbody>
</table>

Table: Webmail delivery attempts with a 360-minute (6h) greylisting threshold.
Nolisting blocks ~27% of spam
Greylisting blocks ~43% of spam, and delays the remaining for 300s...

...but it also introduces a considerable delay in some legitimate emails
Spamhaus response time

From greylisting.org website:

“...there is a large chance that the mass mailer/spammer has been **identified** by the more conventional anti-spam software. Thus, when he retries it, is likely that we will know him for what he really is!”

Over 170 days:

- 99561 passed greylisting / whitelisted
- 28556 never retried (**stopped** by greylisting)
- 31 not blacklisted the first time but were when the mail was accepted
Spamhaus response time

From greylisting.org website:

“...there is a large chance that the mass mailer/spammer has been identified by the more conventional anti-spam software. Thus, when he retries it, is likely that we will know him for what he really is!”

Over 170 days:

- 99561 passed greylisting / whitelisted
- 28556 never retried (stopped by greylisting)
- 31 not blacklisted the first time but were when the mail was accepted
From greylisting.org website:

“...there is a large chance that the mass mailer/spammer has been identified by the more conventionally anti-spam software. Thus, when he retries it, is likely that we will know him for what he really is!”

Over 170 days:

- 99561 passed greylisting / whitelisted
- 28556 never retried (stopped by greylisting)
- 31 not blacklisted the first time but were when the mail was accepted
Greylisting and Nolisting (could) play an important role in fighting spam (~70%), but might be outdated easily.

Nolisting is not very well deployed but 5 domains in Alexa Top-1000.

Malware is not able to exploit a short Greylisting delay.

A high threshold is useless and delay too much benign email.

Webmail providers need to be whitelisted.
That’s all folks!

Thank you for your attention!
Any Question?