Measuring the Role of Greylisting and Nolisting in Fighting Spam

F. Pagani[1] M. De Astis[2] M. Graziano[1]
A. Lanzi[2] D. Balzarotti[1]

[1]Eurecom
Sophia Antipolis, France

[2]Università degli Studi di Milano
Milano, Italy

International Conference on Dependable Systems and Networks, 2016

Spam Detection

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

Spam Detection

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

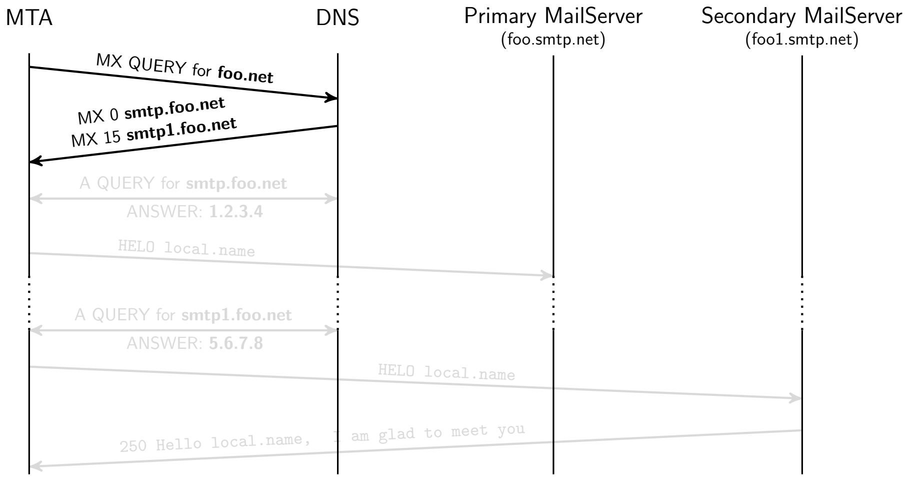
Intro

Nolisting

1. Very simple technique
2. Primary mail server non-existent
3. RFC-2821 compliant:

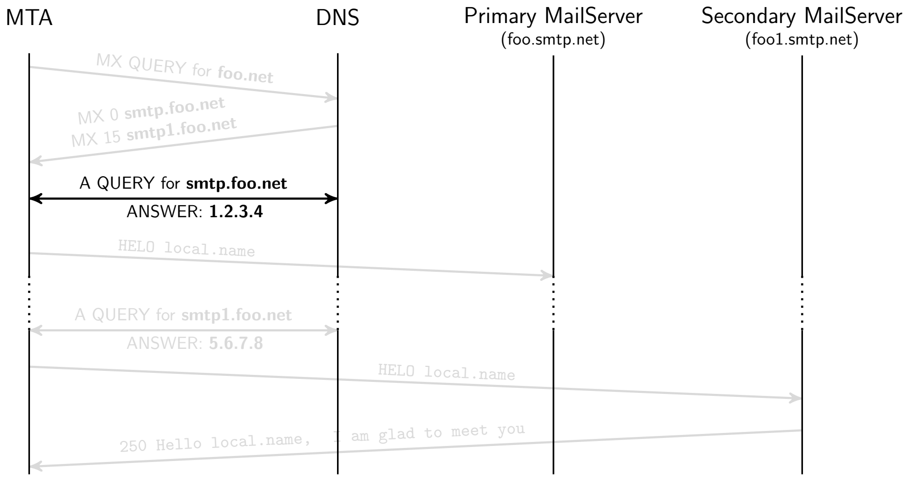
> *"To provide reliable mail transmission, the SMTP client MUST be able to try (and retry)* **each** *of the relevant addresses in this list* **in order***, until a delivery attempt succeeds... In any case, the SMTP client SHOULD try at least* **two** *addresses."*

Intro
Nolisting



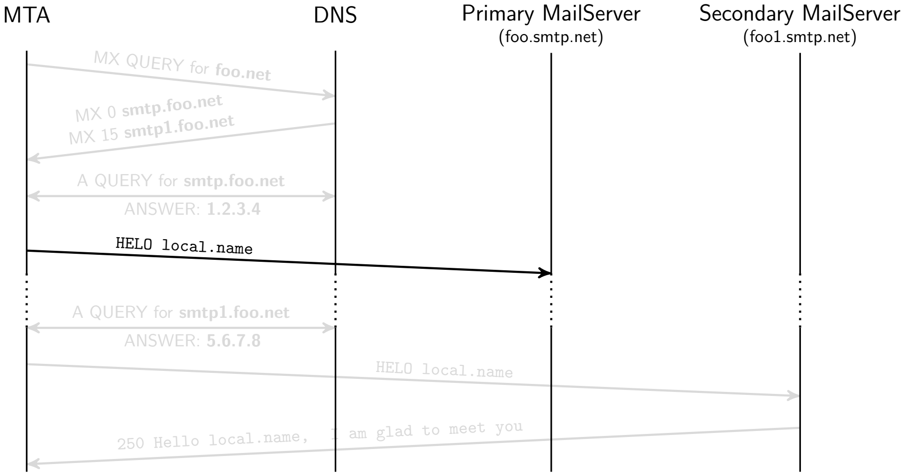
Intro

Nolisting



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**

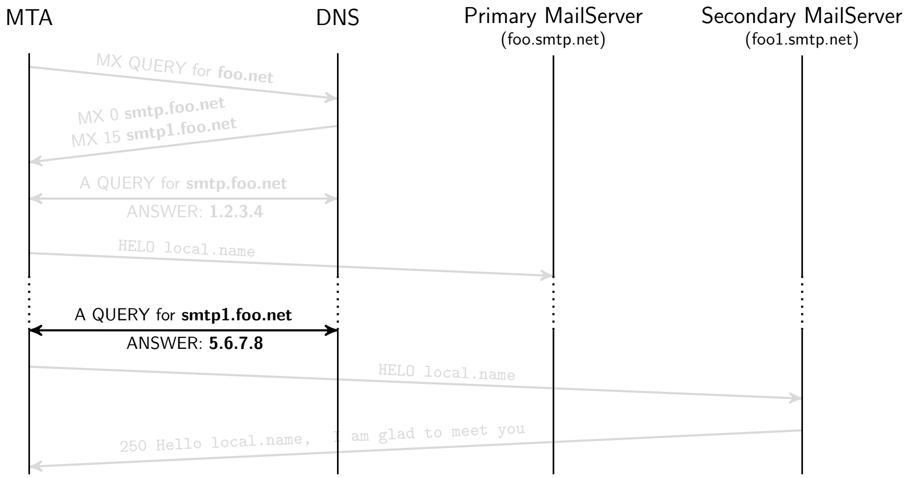
ANSWER: **5.6.7.8**

`HELO local.name`

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

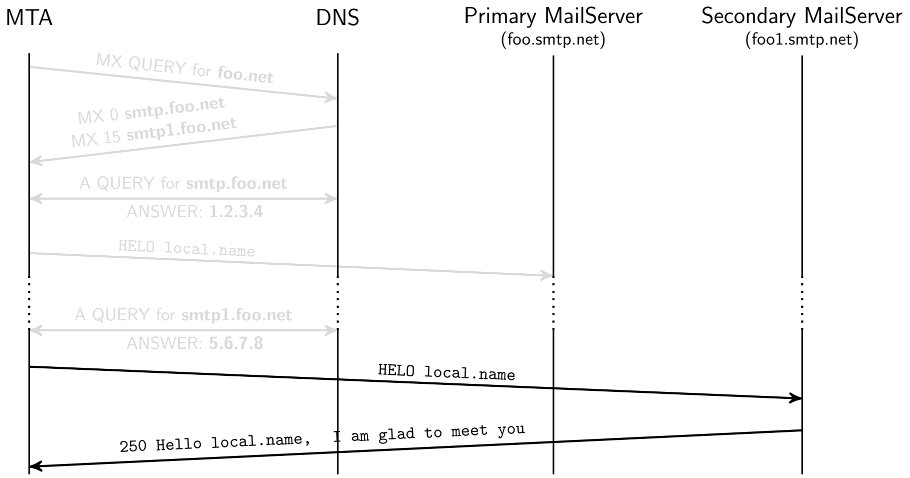
Intro

Nolisting



Intro

Nolisting



Intro

Greylisting

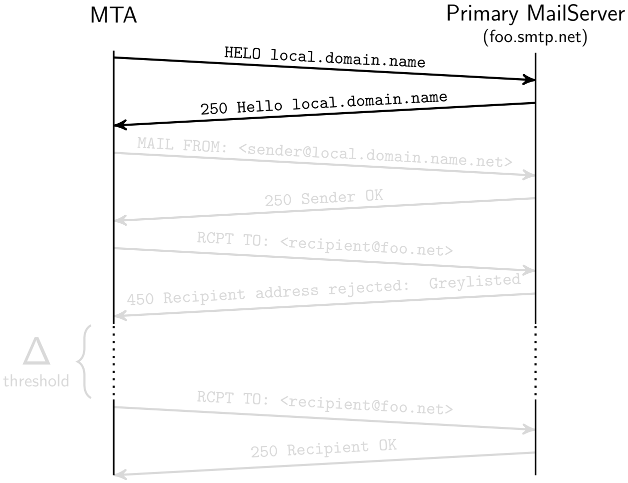
- Message rejected for a certain amount of time (**greylisting threshold**)
- The MTA keeps trying until the message is accepted
- Further messages accepted *without* delay:

  <**sender_address, sender_ip, 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."*

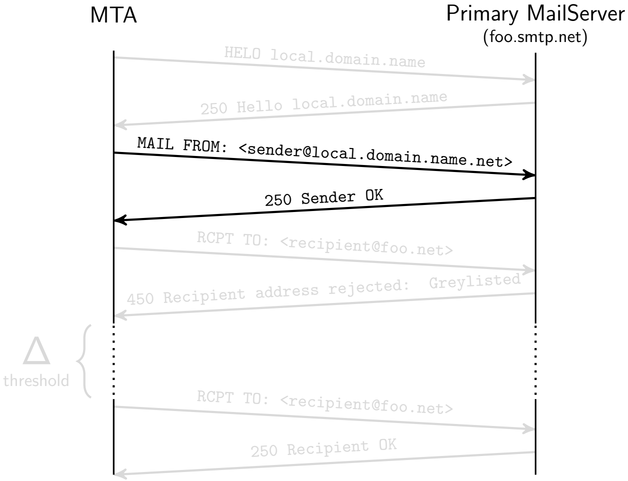
Intro

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
Δ
threshold
RCPT TO: <recipient@foo.net>
250 Recipient OK

Intro

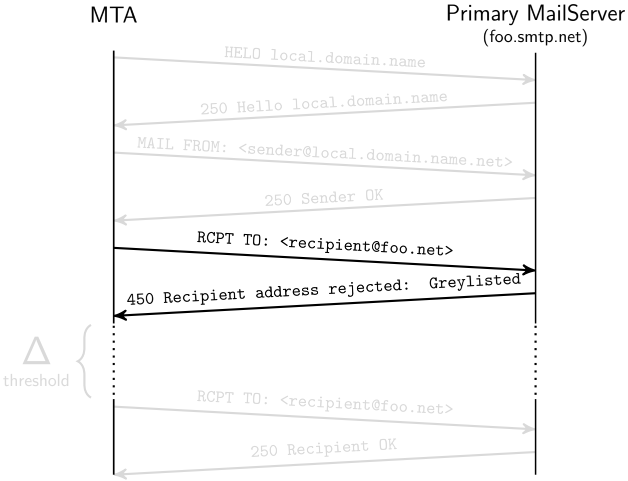
Greylisting



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
Δ
threshold
RCPT TO: <recipient@foo.net>
250 Recipient OK

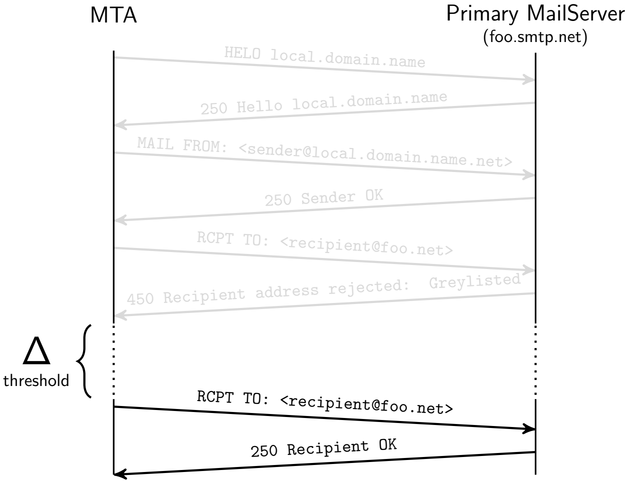
Intro

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
Δ
threshold
RCPT TO: <recipient@foo.net>
250 Recipient OK

Intro

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
Δ
threshold
RCPT TO: <recipient@foo.net>
250 Recipient OK

Greylisting & Nolisting

The main assumption of the two techniques is that spam-bot are **not** RFC-compliant (fire-and-forget).

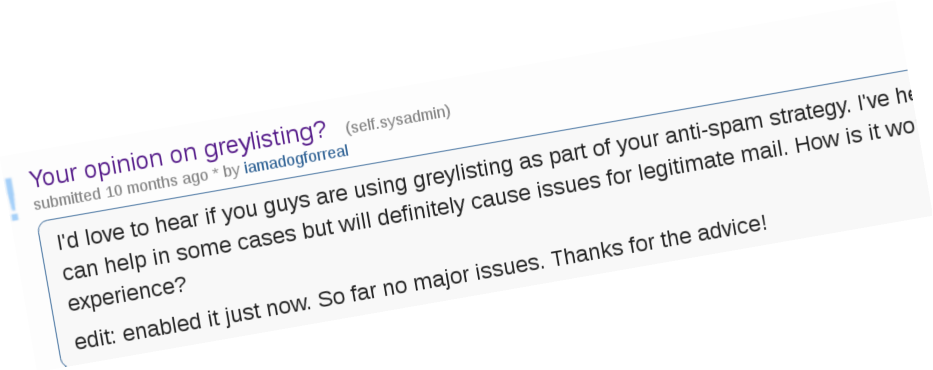
Pros

- Easy to implement
- RFC Compliant
- Do work

Cons

- Easy to evade
- Benign email lost/delayed
- Don't work

Motivation



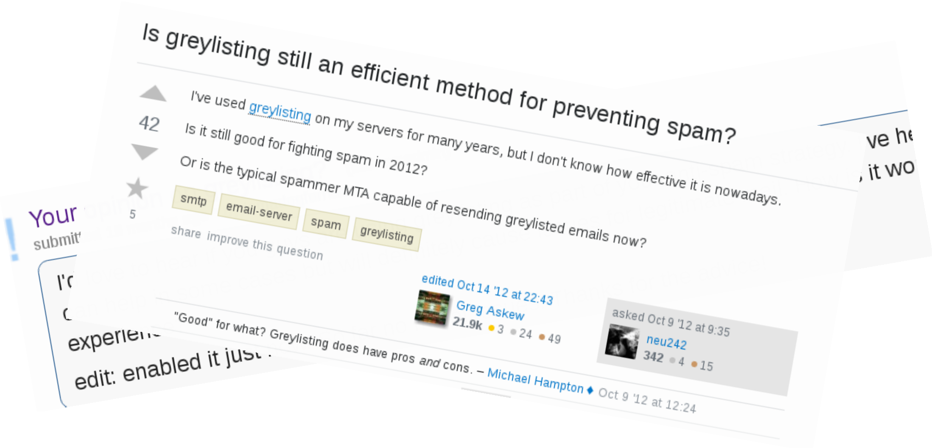
Your opinion on greylisting? (self.sysadmin)

submitted 10 months ago * by iamadogforreal

I'd love to hear if you guys are using greylisting as part of your anti-spam strategy. I've he
can help in some cases but will definitely cause issues for legitimate mail. How is it wo
experience?

edit: enabled it just now. So far no major issues. Thanks for the advice!

Motivation



Is greylisting still an efficient method for preventing spam?

42

5

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?

edited Oct 14 '12 at 22:43
Greg Askew
21.9k ● 3 ● 24 ● 49

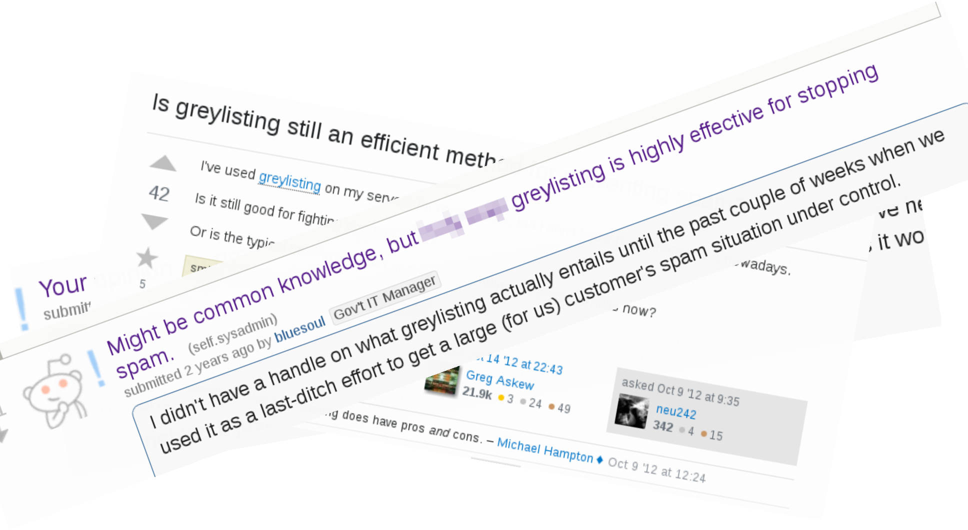
asked Oct 9 '12 at 9:35
neu242
342 ● 4 ● 15

"Good" for what? Greylisting does have pros and cons. – Michael Hampton ♦ Oct 9 '12 at 12:24

Your
submit
I'
experien
edit: enabled it just

ve he
it wo

Motivation



Is greylisting still an efficient meth

42

5

I've used greylisting on my serv

Is it still good for fightin

Or is the typic

wadays.

now?

14 '12 at 22:43

Greg Askew 21.9k ● 3 ● 24 ● 49

asked Oct 9 '12 at 9:35

neu242 342 ● 4 ● 15

g does have pros and cons. – Michael Hampton ♦ Oct 9 '12 at 12:24

Your opinion

submitt

Might be common knowledge, but [redacted] greylisting is highly effective for stopping spam. (self.sysadmin)

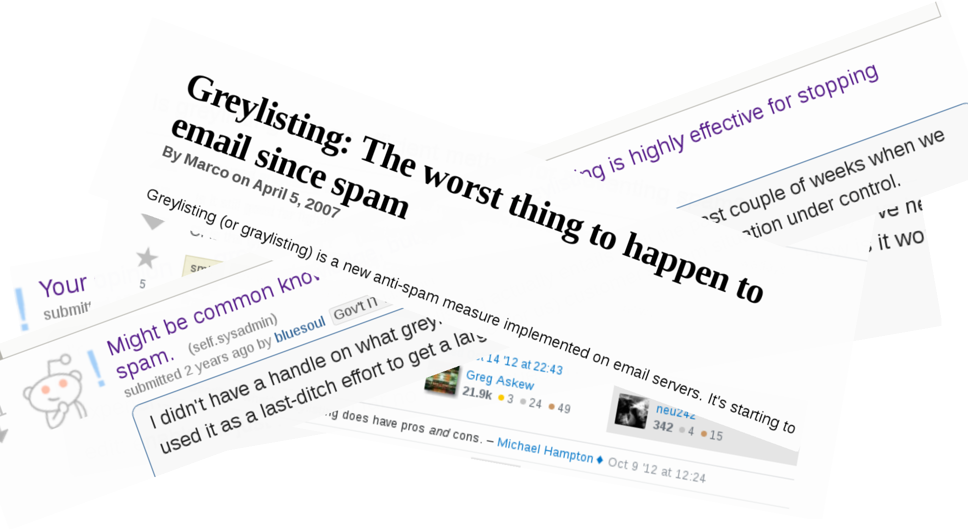
submitted 2 years ago by bluesoul Gov't IT Manager

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

ve ne

it wo

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

ng is highly effective for stopping

st couple of weeks when we
ation under control.

ve ne
it wo

Your
submit

Might be common kno
spam. (self.sysadmin)
submitted 2 years ago by bluesoul Gov't I

I didn't have a handle on what grey
used it as a last-ditch effort to get a lar

14 '12 at 22:43
Greg Askew
21.9k ● 3 ● 24 ● 49

neu242
342 ● 4 ● 15

g 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 → $MX_1, MX_2..$
- $MX_i \rightarrow IP_i$
- Nolisting:
  $IP_1 \not\subset$ IPv4SMTP
  $IP_2 \subset$ 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

- $D \rightarrow MX_1, MX_2..$
- $MX_i \rightarrow IP_i$
- Nolisting:
  $IP_1 \not\subset$ IPv4SMTP
  $IP_2 \subset$ 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
```

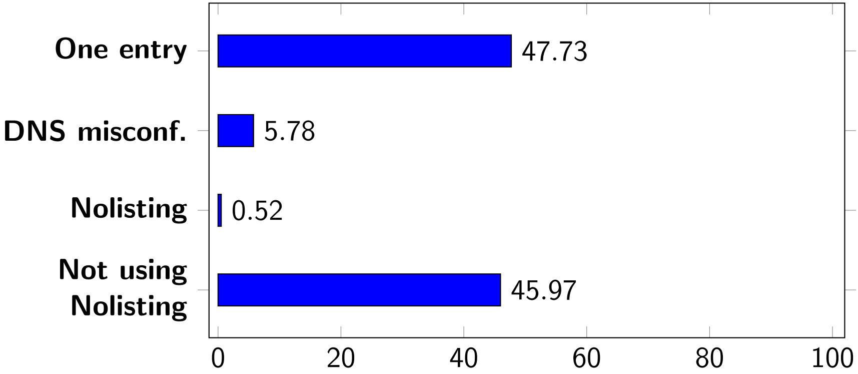
2. Full IPv4 SMTP:

```
1.1.1.1
1.2.3.10
1.3.4.5
```

Steps

- D → MX_1, MX_2..
- $MX_i \rightarrow IP_i$
- Nolisting:
  $IP_1 \not\subset$ IPv4SMTP
  $IP_2 \subset$ IPv4SMTP

Adoption of Nolisting



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

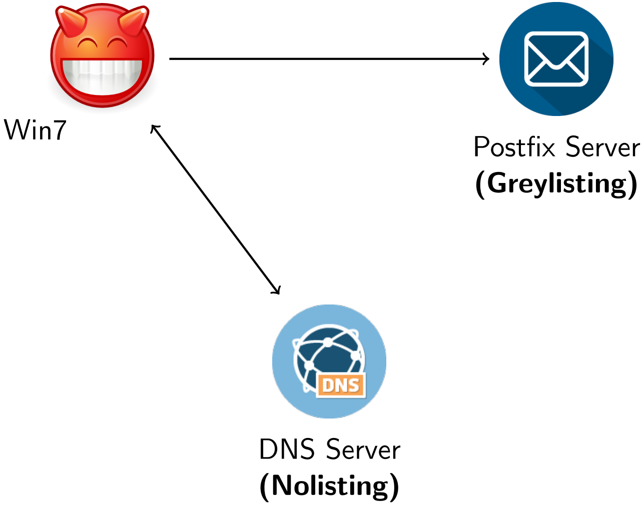
Goals

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



Impact on Spam Delivery

Approach

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

Malware Family	Percentage of Botnet Spam	Number of Samples
Cutwail	46.90%	3
Kelihos	36.33%	6
Darkmailer	7.21%	1
Darkmailer(v3)	2.58%	1
Total Botnet Spam	93.02%	11
Total Global Spam	70.69%	

- Each sample executed in **isolation**, collecting network traces and server logs

Impact on Spam Delivery

Are the techniques still working against modern malware?

SAMPLE	GREYLISTING	NOLISTING
Cutwail:		
sample1	✔	✗
sample2	✔	✗
sample3	✔	✗
Kelihos:		
sample1	✗	✔
sample2	✗	✔
sample3	✗	✔
sample4	✗	✔
sample5	✗	✔
sample6	✗	✔
Darkmailer:		
sample1	✔	✗
Darkmailer(v3):		
sample1	✔	✗

A ✔ sign means the technique was **effective** to prevent spam

A ✗ sign means the technique was **ineffective** against that malware

Nolisting Bypass

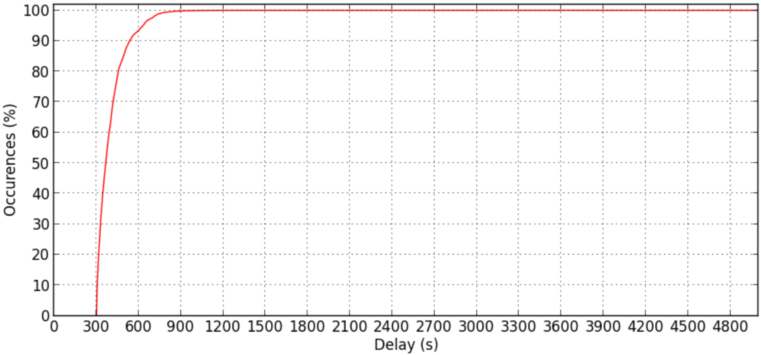
How the malware is able to bypass Nolisting?

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

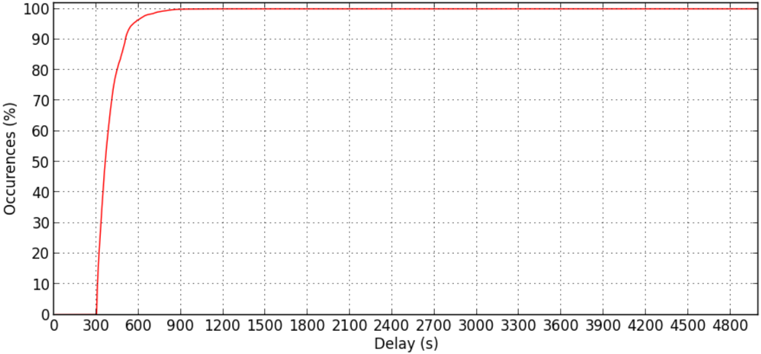
How does the threshold affect spam delivery?



CDF of the spam delivery delay with greylisting at **300** seconds

Greylisting Threshold

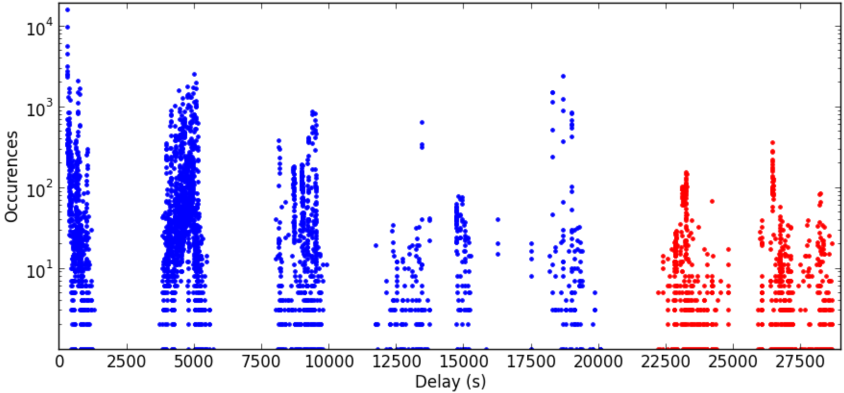
How does the threshold affect spam delivery?



CDF of the spam delivery delay with greylisting at **5** seconds

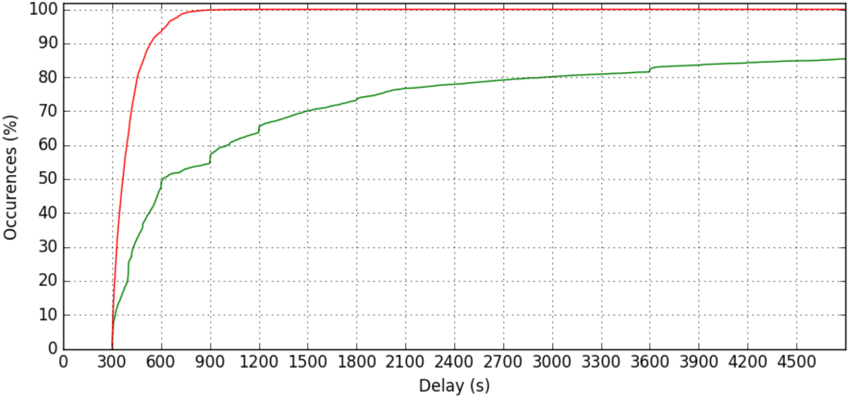
Greylisting Threshold

How does the threshold affect spam delivery?



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).

Greylisting and the Real World



CDF of spam delivery delay with threshold at 300 seconds:

real-world mailbox

vs.

malware samples

Greylisting and the Real World

PROVIDER	SAME IP	ATTEMPTS	DELIVER	DELAYS (min:sec)
gmail.com	✗ (7)	9	✔	6:02, 29:02, 56:36, 98:44, 162:03, 229:44 309:05, 434:46
yahoo.co.uk	✔	9	✔	2:07, 5:39, 12:58, 27:16, 55:13, 109:35 216:47, 430:36
hotmail.com	✔	94	✔	1:01, 2:03, 3:04, 5:06, 8:07, 12:08, 16:10 . . . every 4 minutes . . . , 362:11
qq.com	✗ (2)	12	✗	5:05, 5:11, 5:17, 6:19, 8:22, 12:25, 20:29, 52:31, 84:35, 144:42, 204:56
mail.ru	✗ (7)	13	✔	1:18, 19:15, 49:14, 79:49, 113:20, 154:18, 187:53, 235:20, 271:03, 305:50, 340:38, 373:45
yandex.com	✔	28	✔	1:05, 2:58, 6:53, 14:55, 30:28, 45:41, 61:01 ...every 15:30 minutes..., 369:21
mail.com	✗ (2)	10	✔	5:02, 12:37, 23:59, 41:03, 66:38, 105:01, 162:35, 248:56, 378:28
gmx.com	✗ (3)	10	✔	5:01, 12:33, 23:50, 40:46, 66:09, 104:14, 161:22, 247:04, 375:36
aol.com	✔	5	✗	5:32, 11:32, 21:32, 31:32
india.com	✔	10	✔	6:21, 16:21, 36:21, 76:21, 146:22, 216:21, 286:21, 356:21, 426:21

Table: Webmail delivery attempts with a 360-minute (6h) greylisting threshold.

Takeaways

Nolisting blocks ~27% of spam

Takeaways

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

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

Conclusion

- 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?