Open Relay 自動測試通報機制

金志誠
財團法人資訊工業策進會
elmo@iii.org.tw

摘要

本單位的經驗顯示，Greylisting 對於減少垃圾信的成效良好，且易於維護，但是卻擋不住來自 open relay 的垃圾信。為解決此問題，我們透過自動化機制，儘早識別並通報可能的 open relay；經過半年的運作，證明此機制每月可發現 1200 個以上的 open relay。

關鍵詞：垃圾信、open relay、ORDB、Greylisting。

1. 背景

去年七月間，為配合本單位推廣自由軟體的政策，而重新建置郵件閘道器時，我自網路上接觸到 Greylisting[2]的減少垃圾信機制。據估計，八成以上的垃圾信，都是透過 zombie PC 所寄出[4]。由於在遭遇暫時性傳輸失敗時，真正的郵件伺服器會依循 RFC 821[5]，於稍後重送郵件，而 zombie PC 則多缺少這種重送機制，故 Greylisting 便利用這種行為差異，以避免收到垃圾信。

Greylisting 把 SMTP 連線的發信 IP、寄件者、及收件者之組合，稱為 triplet。針對沒出現過的 triplet，Greylisting 會傳回 4xx 的 reply code，讓發信端以為遭遇暫時性傳輸失敗。此時發信端必須在指定的時間區段(我們設定為 triplet 首次連線後的 1 到 8 小時)內重送郵件，才能通過 Greylisting。

真正的郵件伺服器，是將待寄郵件存放在 mail queue。由於重送時郵件的 triplet 並未改變，便可在指定時間區段順利傳送郵件；而 Greylisting 也會保存對應的 triplet，未來屬於相同 triplet 的郵件就可直接通過。

反觀 zombie PC，也許基於資源考量，大多缺少儲存待寄郵件的 mail queue，而是在寄信時再以程式產生郵件。加上垃圾信業者為隱藏發信來源，經常變換發信 IP，或以亂數產生寄件者；這兩種方式都會產生新的 triplet。故 zombie PC 寄信時，便會一再遭遇暫時性傳輸失敗，而不能順利寄信。

經實際測試後發現，Greylisting 的確能減少許多垃圾信。進入 Greylisting 機制內的 triplet，其中 85%都未於指定時間區段內重送，推斷應是垃圾信；而且販賣色情光碟、藥物、及短期致富的垃圾信從此幾乎絕跡。由於完全不改變使用者的習慣，系統管理者也毋須時常調整設定，可算是成效良好又易於維護的減少垃圾信機制。

不過，Greylisting 也有擋不住的垃圾信；比方說，來自 open relay 的垃圾信。open relay 意指容許任何網路使用者，透過其寄送郵件的 SMTP 伺服器。透過 open relay 寄信，可增加追蹤發信來源的困難，又節省發信所需的頻寬，故常遭垃圾信業者濫用。由於 open relay 是真正的郵件伺服器，透過其原有的重送機制，自然就能順利通過 Greylisting。

雖然 Greylisting 無法阻擋 open relay 寄送的垃圾信，卻可為其他阻擋方式爭取時間。由於我們早已參考 Open Relay DataBase (ORDB)[7]的清單，以阻擋來自 open relay 的垃圾信；所以若在等待通過 Greylisting 的過程中，該 triplet 的發信 IP 已被 ORDB 識別為 open relay，我們便能阻擋它所寄的垃圾信。

因此，我們決定針對等待通過 Greylisting 的 triplet 進行測試；篩選出可能的 open relay 後，再通報 ORDB 做進一步確認，以減少來自 open relay 的垃圾信。經過半年的運作，證明此機制每月可發現 1200 個以上的 open relay。

2. 運作說明

此自動測試通報機制在現有的郵件閘道器上執行；其硬體規格為 AMD Athlon 900MHz 單 CPU 主機，搭配 256MB 記憶體，及一台 20GB 硬碟。閘道器完全使用自由軟體建置，包括：

(1) Linux：作業系統。
(2) Exim[1]：SMTP 伺服器。
(3) ClamAV：電腦病毒掃描程式。
(4) greylistd[3]：搭配 exim 的 Greylisting 伺服器。

除擔任本單位內部的 SMTP 伺服器，及對外的 MX 主機外，閘道器主要負責阻擋病毒信及垃圾信等問題郵件。為避免退信的額外負擔，此閘道器盡量在 SMTP 傳輸過程中，就直接拒絕問題郵件；因此每天雖有超過 85000 次的 SMTP 連線，在阻擋問題郵件後，實際收下的郵件則不到 25000 封。

整個機制分為測試通報及內部阻擋兩部份，由兩個安裝於 crontab 下，每 30 分鐘執行一次的 shell script 所構成；以下分別說明其處理步驟。至於下文所顯示的 mainlog 紀錄，已針對郵件帳號及 IP 位址進行處理，以免造成他人不必要的困擾。

2.1 測試通報

這部分負責識別出 Greylisting 機制內的疑似 open relay，並通報給 ORDB；詳細步驟如下：

(1) 整理發信 IP

每個收到暫時性傳輸失敗，而必須等待通過 Greylisting 的 triplet，都會在 exim 的 mainlog 留下類似這樣的紀錄：

2006-03-29 00:00:54 H=(xxx-axx5d8d248c) [5x.4x.12x.x] F=<xxxx@xx.xxnet.net> temporarily rejected RCPT <xxx@iii.org.tw>: Greylisted defer: 1143561654

在"F="之前，中括號內的字串，就是發信 IP。我們會找出最近 30 分鐘內的 Greylisting 紀錄，並取出對應的發信 IP，進行 relay 測試。此外，因為同一 IP 位址可能連續傳送數封信，而有多筆紀錄，故也須去除重複的資料。

(2) relay 測試

在此我們以 rlytest[8]，一個 open relay 測試程式，試著透過步驟(1)整理出的 IP 位址，寄一封測試信給自己；如果對方收下測試信，則有可能是 open relay。但由於部分 SMTP 伺服器，是先接受所有信件，再進行郵件過濾，所以單單收下測試信，並不代表對方真的會傳遞測試信件；我們必須等收到測試信時，才能確定該受測 IP 是 open relay。

我們也會保留每封測試信的 Message-Id，及其對應受測 IP，以便在收到測試信時，驗證其真實性。

(3) 通報 ORDB

從寄出測試信開始，至收到測試信為止，快的話只要幾秒鐘，慢的話卻可能要二十幾天的時間。為避免時間延誤，我們直接將所有收下測試信的疑似 open relay，都通報給 ORDB，以便進一步測試及確認。

ORDB 除了接受網頁通報外，也可利用電子郵件進行通報。其規定如下：

A. 接受通報的郵件帳號為 relays@ordb.org。
B. 須以"Relay: xxx.xxx.xxx.xxx"的格式，將通報的 IP 位址填寫在信件內文。
C. 每次不得通報超過 100 個 IP 位址；每天不得通報超過 200 個 IP 位址。

我們會將疑似 open relay 的 IP 位址，依規定格式寄到 ORDB；如此就完成通報。由於 ORDB 具備完整的測試、查詢、及移除 open relay 紀錄的機制，後續的處理便完全不必我們操心。

2.2 內部阻擋

一般說來，我們不常收到在測試通報的步驟(2)中所寄出的測試信。由於 open relay 的 IP 位址，往往已先列於 ORDB 的清單內，因此測試信傳回時，便會直接被閘道器阻擋。

所以當我們收到測試信時，代表：這個 IP 位址的確是 open relay，但是尚未列在 ORDB 上。這時我們會設置內部的 open relay 阻擋清單，暫時拒絕其來信。詳細步驟如下：

(1) 整理測試信，並驗證信件真實性

每一封寄回來的測試信，都會在 exim 的 mainlog 留下類似這樣的紀錄：

2006-03-29 17:35:17 1FOX5B-0005wY-AW <= nobody@xxx.com H=(xxx.com) [1x.23x.9x.19x] P=esmtp S=1507 id=rlytest-1143624907-22826@mta2.iii.org.tw for xxxx@iii.org.tw

緊接著"id="，帶有"rlytest"的字串，就是測試信的 Message-Id；我們會過濾出最近 30 分鐘內的測試信紀錄。由於測試信的 Message-Id，是由寄送時間與 process id 所構成；所以若來信的 Message-Id，與當初保留的一致，則可確認其真實性。

(2) 併入並清理內部 open relay 阻擋清單

接下來藉由測試信的 Message-Id，取出對應的受測 IP，以併入內部 open relay 阻擋清單；此清單會依照 SMTP 伺服器所接受的格式產生，以做為阻擋垃圾信的參考資料。此外，我們並未提供外界自行由內部清單移除紀錄的功能；為避免曾經是 open relay，但已經解決問題的主機，因以往的紀錄，無法順利寄信到本單位，所以凡併入清單內超過四天的 IP 位址，都將自動刪除。

3. 效益分析

由於向 ORDB 通報後，ORDB 會以電子郵件回報後續的狀況，例如 IP 位址是否已被通報過，是否確認為 open relay 等訊息；故整理相關郵件，便可了解此機制的運作情形。

先看經 ORDB 確認為新的 open relay 的統計。

表 1 ORDB 回報統計

回報時間	新發現的 open relay 數	平均回報時間 (秒)	通報後 30 分鐘內回報個數
2005 年 11 月	1450	26276	1155
2005 年 12 月	1513	69653	1031
2006 年 1 月	1341	43130	619
2006 年 2 月	1639	41155	389
2006 年 3 月	1937	34495	285
2006 年 4 月	1389	73985	14

表 1 顯示，我們通報的 IP 位址，有 1450 個由 ORDB 在 2005 年 11 月回報為新的 open relay；由 ORDB 收到通報，到回報確認結果為止，平均花費

26276 秒(超過 7 小時)的時間；在這 1450 個 open relay 中，有 1155 個(約八成)是在通報後 30 分鐘內便回報確認。其餘紀錄可類推。

其實回報所需時間的差異相當大；紀錄中，最短的回報時間是 127 秒，但是最久的卻要 9540366 秒，也就是超過三個月才回報！

此機制原本為儘早發現 Greylisting 內的 open relay 而設計。由於此機制每 30 分鐘執行一次，所以我們也統計，在到達 Greylisting 指定時間區段(60 分鐘)前，即通報後 30 分鐘內就回報的情況。一開始，近八成的 open relay 都能在 30 分鐘內回報；但經過六個月，能在 30 分鐘內回報的已不到 1%，已無法達成當初的目的。

不過，若以發現 open relay 的個數來看，則此機制的表現仍相當不錯；每月都能發現 1200 個以上的 open relay。

接下來是測試信的收回統計。

表 2 測試信收回統計

寄送時間	收回測試信數	平均傳遞時間(秒)	30 分鐘內傳回數
2005 年 11 月	400	485517	125
2005 年 12 月	207	188717	97
2006 年 1 月	199	77058	131
2006 年 2 月	92	243651	36
2006 年 3 月	129	43585	61
2006 年 4 月	141	8119	70

表 2 顯示，我們收到 400 封在 2005 年 11 月寄出的測試信；由寄出到收回測試信，平均需要 485517 秒(超過 5 天)的時間；在 400 封收到的測試信中，有 125 封在寄出後 30 分鐘內便傳回來。其餘紀錄可類推。

表 2 內的資訊與表 1 類似；但可發現，我們收回的測試信，都遠少於同時期 ORDB 所回報的數目。這是因為郵件閘道器參考 ORDB 的清單；若已被 ORDB 確認為 open relay，閘道器便會直接拒絕其郵件。

最後是此機制所通報的 open relay，佔同時期 ORDB 確認的 open relay 總數的比例。

表 3 此機制通報佔 ORDB 當月比例統計

回報時間	此機制發現的 open relay 數	ORDB 確認的 open relay 總數	此機制所佔比例
2005 年 11 月	1450	3041	48%
2005 年 12 月	1513	3513	43%
2006 年 1 月	1341	2348	57%
2006 年 2 月	1639	2538	65%
2006 年 3 月	1937	2857	68%
2006 年 4 月	1389	2030	68%

ORDB 公佈了每天新發現及修正的 open relay 個數[6]；表 3 的“ORDB 確認的 open relay 總數”一欄，就是加總其上資料後的結果。

表 3 顯示，在 2005 年 11 月，經此機制通報到 ORDB，且確認無誤的 open relay 有 1450 個；同時期所有經 ORDB 確認的 open relay 共 3041 個；所以此機制所通報的 open relay，佔總數的 48%。其餘紀錄可類推。

若由每月有接近甚至超過五成的 open relay，是由此機制率先通報來看，此機制對於儘早發現 open relay 的助益相當大。

4. 結論

本文所討論的 open relay 自動測試通報機制，其優點在於：

(1) 資源需求不高，可與郵件閘道器現有服務在同一主機上執行。
(2) 完全自動化運作，不需要人員介入。
(3) 依照客觀的測試方法來辨識 open relay，不牽涉人員的主觀判斷，不致發生誤判。
(4) 藉由通報 ORDB，進行資訊分享，協助全球其他郵件伺服器，儘早阻擋來自 open relay 的垃圾信。

若純粹考慮本單位的需求，此機制阻擋垃圾信的效果，是比不上 Greylisting 的；畢竟八成以上的垃圾信皆由 zombie PC 寄出，這是 Greylisting 才使得上力的部分。但若以儘早通報 open relay，從而協助其他郵件伺服器過濾問題郵件來看，受益於此機制而阻擋的垃圾郵件總數，也許遠超過本單位使用 Greylisting 所減少的垃圾郵件；這也算是本單位對網路社群的一些回饋吧。

參考文獻

[1] Exim, http://www.exim.org/

[2] Greylisting, http://projects.puremagic.com/greylisting/

[3] greylistd, http://packages.debian.org/unstable/mail/greylistd

[4] IronPort, "Spammers Continue Innovation: IronPort Study Shows Image-based Spam, Hit & Run, and Increased Volumes Latest Threat to Your Inbox", http://www.ironport.com/company/ironport_pr_2006-06-28.html

[5] Jonathan B. Postel, RFC 821, "Simple Mail Transfer Protocol", Aug 1982.

[6] New and fixed relays, http://www.ordb.org/statistics/changes/

[7] Open Relay DataBase, http://www.ordb.org/

[8] rlytest, http://www.unicom.com/sw/rlytest/