以 maillog 為回饋機制的 spam mail 阻擋系統

李育旻
國立高雄師範大學資訊教育研究所
E-mail teach235@bsd.tscvs.ttct.edu.tw

摘要

郵件主機上的郵件記錄檔案 maillog 記載了詳細的連線記錄，看起來雜亂，但透過程式進行文字內容分析，抓取其中的 spam 來源 ip，阻擋其下一次的連線，可減輕系統過濾程式的負擔。為避免誤判的情形，有必要設計使用者反應機制，系統定時將擋信清單寄給使用者，讓使用者檢查，如有誤判的情形，可立即回報給系統管理者，由管理者針對誤判的 ip 予以移除管制連線。

關鍵詞：垃圾郵件、郵件過濾、maillog。

Abstract

The maillog in Mail server describes the messages about mail's log, it looks like bewilderment. But we can use program to filter the content, get the spam ip , and deny the ip at next connect. In this way, it can reduce the loading of the filter program. To avoid the filter program's mistake, this system will mail the deny mail's statistic to user. If user find the mistake, he can reply the mail to notify the system administrator. Then the administrator can determine to remove the rule.

Keywords: spam filter maillog

1. 前言

網路的興起，電子郵件已成為日常生活中一項重要的工具。但由於郵件傳送協定 smtp 設計不夠嚴謹，為了行銷，有人便利用電子郵件夾雜著商業廣告信件大量發送，發送給不特定的對象，稱之為垃圾郵件，帶給收件者很大的困擾。如何在一堆充斥廣告信的信件中，找出重要的信件，成為每次上網收信者的惡夢。因此管理良善的郵件系統現在幾乎都另外加裝過濾程式來幫助使用者找出垃圾郵件。由於過濾程式對垃圾郵件的判斷正確率並非百分之百，有些灰色地帶的信件內容也因人而異，不一定視為拒收的垃圾信件。所以如果過濾程式找出的疑似垃圾信件直接刪除危險性過高，一般的做法只在信件主旨加以變更，如加上{spam?}關鍵字，或將該信件搬移至特定的信匣存放，由使用者自己決定對信件的後續處理。站在郵件系統管理者的立場來看，垃圾信件的不良影響為：使用者抱怨垃圾信太多、垃圾信來源 ip 佔用過多連線頻寬與資源、垃圾信過多導至過濾程式佔用 cpu 使用率過高，更導致其它服務效能受到影響。郵件系統管理者有必要採取相對應的策略予以處理。當某 ip 有傳送 spam 之記錄，郵件系統管理者應針對此 ip 有一管制措施，而非任其一再傳送垃圾郵件，除非此 ip 是有特殊原因非開放不可的。郵件主機上的郵件紀錄檔是一個很重要的檔案。郵件記錄檔 maillog 中記錄了郵件傳送的過程、連線的 ip、寄件人、收件人、寄信時間等，管理者可分析此記錄檔，萃取其中有用的資訊，對散布廣告信的來源 ip 做一管制連線的措施。

2. 系統簡介

本系統以 perl 程式語言分析主機上信件記錄檔 maillog，定時抓取該檔內容中 SpamAssassin 程式所分析產生的 spam 記錄，並將所得 spam 來源 ip 寫入 mysql 資料庫中，同時系統定時將 mysql 中的 spam 來源 ip，新增至系統郵件服務 postfix 的設定檔 access 中來予以拒絕其連線。為預防 SpamAssassin 誤判 spam 而導致使用者重要信件被拒收，本系統另外將每天記錄檔中的擋信清單依個別使用者分類並寄回給使用者，如果有誤擋情形可立即回報給管理者，由管理者予以移除。為方便管理，本系統開發網頁管理界面，提供成效統計與被擋 spam ip 一覽，供使用者查看。長久下來，對於主機上郵件過濾系統負荷大為減輕，可確保郵件主機上的其他服務品質，如個人網頁(http)、檔案傳輸(ftp)，不會因為系統忙於過濾信件而發生延遲情形。

3. 系統特色

3.1 建置成本低

本系統採用 Fedora Core3 ，搭配 MailScanner +Postfix+SpamAssassin+Perl+PHP+Mysql+Apache ，使用硬體為雙 cpu，pentium III 800 Mhz，512MB 主記憶體之機器，伺服器系統與資料庫的選擇皆採用開放原始碼的免費軟體，不需額外的花費。

3.2 程式碼易於開發維護

利用 Perl 程式語言在字彙分析的強大能力來分析系統記錄，相對其他語言來說，更易於開發與維護，並結合 DBI 將資料寫入 Mysql 中，方便以網頁 PHP 程式整合開發管理界面。

3.3 擋信 ip 列表有自主控制權

Real-time Blackhole List（簡稱 RBL），是一個 ip 位址列表，稱做"即時黑名單"，此列表的 ip 為公眾認為的垃圾郵件發信主機。最常見的 RBL 為"relays.ordb.org"。郵件系統可設定加入 RBL 檢查，當發信來源連上郵件主機時，郵件主機會即時連上提供 RBL 資料庫的主機，將發信 ip 位址與 RBL 資料庫做比較，如果 ip 位址列於 RBL 資料庫中，信件就會被退回，否則就進入郵件主機中進行其他的處理。但由於 RBL 的資料是由用戶投訴及採樣累積建立的，有些誤報或失去時效，難以保證其正確性與時效性，發生誤判又不便刪除此筆記錄。因此本系統放棄採用 RBL 機制，而決定自行建立擋信 ip 列表，如有誤判情形，可有充份控制權放行該 ip 主機再次寄信進來。

3.4 有效率的 Smtp 端防制

廣告信過濾須要高效能的機器，本系統利用 smtp 端防制，透過分析郵件紀錄檔來蒐集廣告信來源 ip，可阻擋散發廣告信的 ip 於過濾系統之前，在 smtp 連線時就予以直接拒絕，節省因垃圾信增加過濾的硬體成本，可提供較佳的運作效率與防制功能。

3.5 自動化的 spam ip 來源阻擋

本系統利用 crotab 定時更新阻擋 ip 列表，自動化阻擋有散布垃圾郵件不良記錄的 ip 來源。

4. 架構及原理

本校郵件系統以 Postfix 為 MTA(Mail Transfer Agent)，版本為 postfix-2.1.5-5，這個版本支援 cidr table，可以直接設定可連線與不可連線之網段，如圖 1 所示。

```
69.17.117.3/32          REJECT sorry, beacuase
61.48.24.113/32         REJECT sorry, beacuase
221.221.17.91/32        REJECT sorry, beacuase
221.221.31.198/32       REJECT sorry, beacuase
72.10.198.251/32        REJECT sorry, beacuase
221.221.20.1/32         REJECT sorry, beacuase
70.167.19.250/32        REJECT sorry, beacuase
221.221.17.55/32        REJECT sorry, beacuase
61.48.24.235/32         REJECT sorry, beacuase
206.82.202.235/32       REJECT sorry, beacuase
```

圖 1 postfix 擋信格式

```
#限制能連port 25的client ip
smtpd_client_restrictions =
        permit_sasl_authenticated,
        cidr:/etc/postfix/ipaccess,
        hash:/etc/postfix/access,
        permit
```

圖 2 /etc/postfix/main.cf 設定

要支援限制能連 port 25 的 client ip 網段，要在/etc/postfix/main.cf 中設定 smtpd_client_restrictions = cidr:/etc/postfix/ipaccess，如圖 2 所示，該設定表示允許用戶端以 sasl 認證連線，拒絕 ipaccess 檔中 cidr 格式的 ip 連線，及拒絕 access 檔中的機器列表連線，其餘則許可。所以當 client 以 smtp 協定連接郵件主機 port 25 時，郵件主機會先檢查該 ip 是否允許寄信連線的動作，存在 ipaccess 檔案中的 ip 列表，皆無法對郵件主機成功連線。在圖 1 中，重要關鍵字為 REJECT，意謂將符合的連線寄信 ip 直接予以拒絕，且會在系統記錄檔 maillog 中留下記錄文字 Client host rejected，如圖 3 所示。此設定方式會給予 client 一拒絕連線訊息。對於大量連線之 spam 可設定關鍵字 DISCARD，直接予以刪除，對 client 不回覆任何訊息，減輕系統回應的負擔。

```
postfix/qmgr[4917]: 0894E1C63AF: removed
postfix/smtpd[4935]: connect from unknown[220.163.64.196]
postfix/smtpd[4935]: NOQUEUE: reject: RCPT from unknown[220.
known[220.163.64.196]>: Client host rejected: sorry, beacuas
, we reject you ip; from=<jamie50815@gtvs.info> to=<teach90@
w> proto=SMTP helo=<gtvs.info>
postfix/smtpd[4935]: lost connection after RCPT from unknown
```

圖 3 maillog 中拒絕連線訊息

Mailscanner 是一個電子郵件通訊閘的程式，可以外加掃毒引擎及廣告信判斷引擎來增加其過濾功能。Postfix 收到電子郵件後，會將郵件交由 Mailscanner 處理，暫存在佇列中，由 Mailscanner 自己或呼叫掃毒引擎及廣告信判斷引擎來進行過濾。本系統是由 Mailscanner 搭配 SpamAssassin 這套程式來進行廣告信的過濾、學習與評分。過濾結束後，Mailscanner 再將控制權還給 postfix 程式繼續完成信件的傳送。SpamAssassin 程式使用大量的預設規則來檢查垃圾信，這些規則包括檢查郵件的標頭、內文、送信者，並給予評分，由於垃圾郵件往往具有某種特徵，所以若某郵件評分後的總分高於某項標準，可以判定其為垃圾郵件。其在系統記錄 maillog 中留下的記錄如圖 4 所示。SpamAssassin 更可以透過不斷的廣告信學習強化其正確率，可高達 95%以上。當某郵件被 SpamAssassin 評分，評斷為廣告信後，Mailscanner 此時可設定修改其郵件主旨加入{spam?}關鍵字，方便使用者分類信件。

```
23 bsd MailScanner[31493]: Message 5385C1C63BE.4BD21
@flexipay.cc) to bsd.tscvs.ttct.edu.tw is 廣告郵件,
分數=20.651, 被需要 6, autolearn=spam, BAYES_99 3.50
 DOMAIN_RATIO 3.18, HTML_90_100 0.02, HTML_FONT_FACE
0 0.45, HTML_IMAGE_RATIO_02 0.02, HTML_MESSAGE 0.00,
FROM_MTA_ID 1.72, URIBL_JP_SURBL 2.46, URIBL_SC_SURB
IBRARY 2.75)
23 bsd MailScanner[31493]: Spam Checks: Found 1 spam
23 bsd MailScanner[31493]: Spam Actions: message 538
eliver
```

圖 4 廣告信在 maillog 中的記錄

由於 maillog 中記錄了被 SpamAssassin 判斷為廣告信件的來源 ip，該 ip 可能由於管理不當，允許轉寄廣告信或已被入侵成為轉信跳板而不自知，所以我們可以撰寫程式分析 maillog 抓取此類 ip，並將此 ip 列入 postfix 擋信格式中，拒絕其 smtp 連線，這樣做的優點是寄信來源會收到被拒絕的訊息，可提醒該 ip 管理者，另外一項優點是有些廣告信具有同 ip 重覆寄送不同主旨的特徵，這類廣告信在本系統只會成功 1 次，留下紀錄，接下來就

會被加入 postfix 擋信列表，拒絕其 smtp 連線，不再對此來源 ip 轉寄的信進行過濾，減少系統的負擔。為避免 SpamAssassin 誤判或該 ip 只是一時被轉寄廣告信，後來已修正，另設計一反應機制，同樣分析 maillog，將擋信記錄中的收件者，寄件者，寄信時間予以記錄並存至 Mysql 資料庫中，每天將使用者被擋的來源 mail 郵件帳號彙整寄給使用者本人，由使用者自己檢查清單上是否有誤擋的情形發生。如果有可以直接在信件上做回覆，系統管理者每天收信後即可調整擋信機制。整個系統架構流程如圖 5 所示。

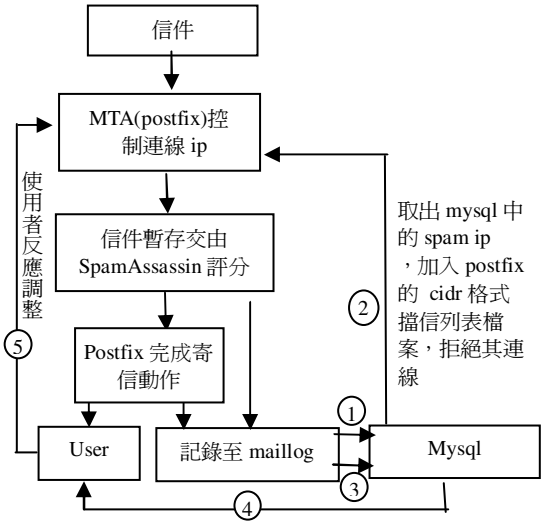

圖 5 系統架構流程

本系統分成 5 個部分，分別撰寫 perl 程式及 php 網頁管理界面，如表 1，其功能如下。

表 1 主要程式說明

系統名稱	功能說明
① 分析子系統 spamfilter.pl	分析 maillog，抓取被過濾程式 SpamAssassin 標示為廣告信的信件記錄，將其中判斷出的 spam ip 寫入 Mysql 資料庫
② 擋信子系統 spamadd.pl	由 Mysql 資料庫讀出 spam ip 欄位，寫入 postfix 的 cidr 格式擋信列表檔案，由 postfix 在 smtp 連線階段就拒絕其再度連線
③ 統計子系統 spamnotify.pl	分析 maillog，抓取拒絕連線記錄，將被擋的 spam ip 企圖要傳送的寄件人帳號，收件人帳號，寄信時間寫入 Mysql 資料庫
④ 反應子系統 spamstat.pl	由 Mysql 資料庫讀出每一收件人的擋信清單，並寄出彙整名單，含申訴超連結，可反應誤判情形
⑤ 網頁管理 子系統	提供被擋來源 ip 一覽、誤擋 ip 申訴、成效分析、管理者登入、申訴結果等網頁(圖 10)

5. 系統運作情形

在未使用 maillog 分析回饋機制阻擋系統時，對於廣告信件，本校只採取在信件主旨加入{spam?}關鍵字的方式，並設定規則，將其分類至廣告信匣中，幫助使用者可以快速過濾不必要的信件。系統每天要負擔對信件的病毒過濾與內容分析，同時學校教師的網頁瀏覽，檔案傳輸的使用也很頻繁，系統負荷有過重的現象，SapmAssassin 程式更有連結逾時的系統記錄發生。加入此擋信機制後，spam 來源 ip 只會成功 1 次，之後就被加入擋信機制中，拒絕其 smtp 的連線，阻擋成效如圖 6，每天的阻擋率將近一半以上，意謂阻擋了一半以上原本要過濾的垃圾郵件，進而系統要處理的信件數下降，SpamAssassin 也不再有連結逾時的情形。

圖 6 阻擋成效

為避免 SpamAssassin 程式的誤判，而造成正常信件被拒絕連線，本系統加入個別使用者通知擋信名單的功能，如圖 7 所示，可以清楚知道有那些寄件人在什麼時間企圖寄信給使用者卻被擋無法寄信成功，使用者可以檢查是否有誤判的情形，直接點選信件上的連結回報給管理者，如圖 8 所示。管理者可由網頁管理界面予以移除誤判 ip，如圖 9 所示。從擋信名單可以證明，廣告信會利用相同 ip 重覆發送，以及寄件人帳號欄位不實的情形。本系統確實發揮抵制廣告信來源 ip 再度對本機散發廣告信的功能。

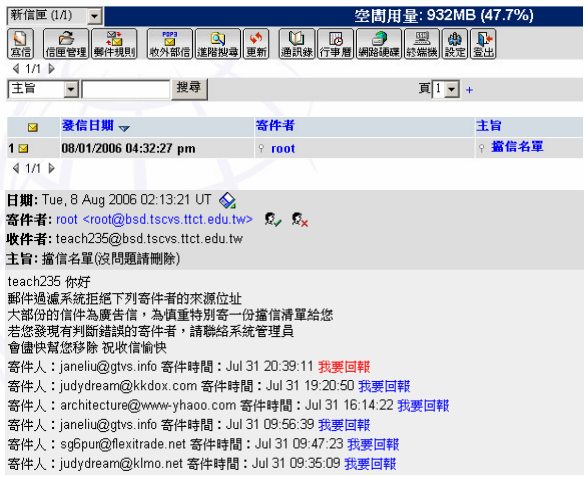

圖 7 個別使用者擋信名單

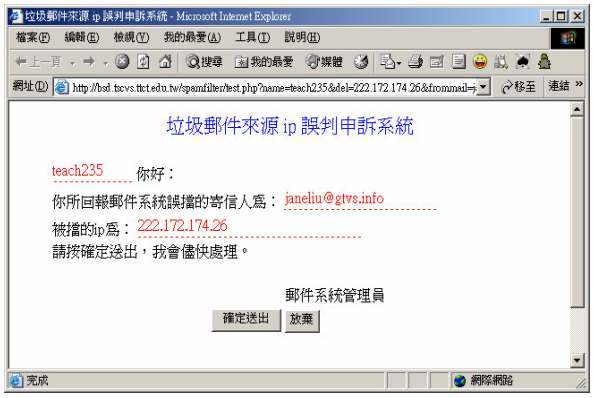

圖 8 使用者回報

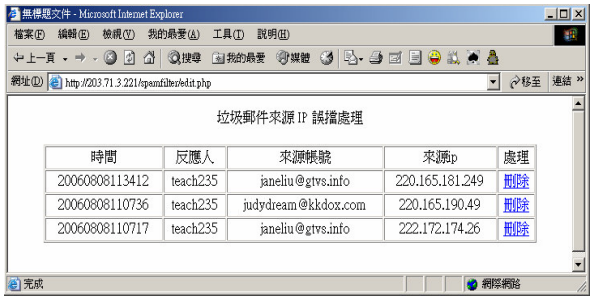

圖 9 管理者管理畫面

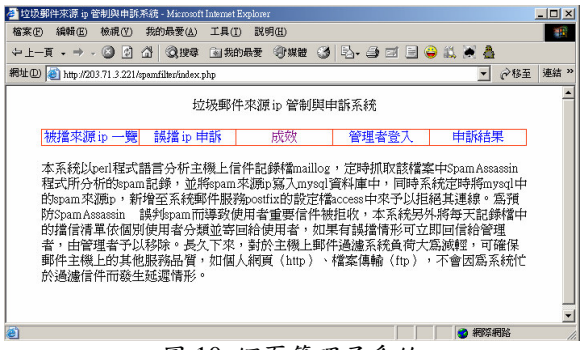

圖 10 網頁管理子系統

6. 結論及未來研究方向

Maillog 為郵件主機上重要的記錄檔，記錄了很多有意義的訊息，透過程式的分析，取出特定的資料加以應用，是減輕網路管理工作的重要方式。由於大部份廣告信程式會在網路上搜尋開放 relay 信件的郵件主機，藉由此主機散布廣告信，本系統的做法是只要曾透過此類 ip 寄送廣告信，就暫時予以阻擋，透過與使用者的互動來進一步決定是否為誤判。觀察記錄檔發現，被阻擋的廣告信來源 ip 會重覆嘗試要求連線而被 postfix 拒絕，擋信清單愈長表示本系統確實發揮抵制此類 ip 的作用。此系統也發現一個有趣的問題，對於某些使用者感興趣的信件如電子報，卻因其他使用者認為是廣告信件而反應給系統管理者，系統管理者若將此類信件加入 SpamAssassin 程式學習 spam 特徵，成為公用的過濾規則，此類信件就會進而被判斷為 spam，造成電子報來源 ip 被系統拒絕連線，經欲收電子報使用者反應後，因為本系統加入的擋信機制會影響整個郵件系統，管理者無法針對個別使用者開放或拒絕，衡量之下，只好將此 ip 開放，並請訂閱電子報的使用者利用收信界面的郵件規則，將此來源帳號移入所設定的電子報信匣中，避免被分類至廣告信匣。未來研究將加入流量管制，對於大量連線疑似 spam 來源的 ip 限制其 smtp 連線流量，確保系統其他服務有穩定的連線服務品質。

參考文獻

[1] 簡信昌，“Perl 學習手札”，上奇，2004。

[2] 賴守全、謝木政，”網路資訊使用管制技術之實現與挑戰”，2004 年台灣網際網路研討會論文集，pp.588-589，2004。

[3] 張傑生、唐瑤瑤、許凱平、陳啟煌、李秀惠、賴飛羆，”使用 Open Source 軟體進行 SPAM Mail 防制處理-以台灣大學電子郵件系統為例”，2005 年台灣網際網路研討會論文集，2005。

[4] http://spamassassin.apache.org/

[5] http://www.mailscanner.info/

[6] http://www.mysql.org/

[7] http://www.perl.org/

[8] http://www.php.net/

[9] http://www.postfix.org