異常電子郵件分析系統
A spam analysis system

張誌倫
銘傳大學資傳所
96166150@ss24.mcu.edu.tw

葉姵伶
銘傳大學資傳所
97166060@ss24.mcu.edu.tw

曾煜為
銘傳大學資傳所
97166033@ss24.mcu.edu.tw

江清泉
銘傳大學資傳所
ccchiang@mail.mcu.edu.tw

摘要

隨著近年來電子郵件通訊的盛行，垃圾郵件問題也漸成為眾人關切的焦點。由於目前垃圾郵件內容日漸多樣，憑單一分析方法可能無法做到全方位的防護；並且一般防護機制只服務單一郵件主機下的用戶，不提供非用戶的郵件服務，在資訊收集方面也僅限於單一主機，缺乏分散式的數據。因此往往遇到該主機未發生過的訊息或特徵時，false positive 或 false negative 的機率便會特別高。另一方面，部分的 anti-spam 技術並無與郵件作業平台結合，缺乏實際運作的測試數據。本研究提出一個整合式電子郵件分析服務系統，集合了多種分析方法，並將各分析結果做統整，同時針對貝氏分類法分析測試各種參數，提升整體的垃圾郵件防護效能。本系統不僅能服務本地伺服器下的用戶，也提供一個平台，讓外界伺服器的管理者和用戶透過日誌上傳以及信件轉遞的方式，由我們的系統代為做日誌與內容分析，再將結果以網頁和自動回信的方式回報。一方面透過分散式的資料上傳，能夠收集更多的垃圾郵件資訊，有助於日誌以及內容分析之研究，提高分析系統的效能，另一方面能夠服務各郵件伺服器的管理者及用戶，同時也可藉可系統作為未來 anti-spam 技術的測試平台。

關鍵字: 垃圾郵件, 整合式分析, 貝氏分類法, 分散式日誌分析。

ABSTRACT

With the popularization of the Internet, e-mail services become one of the tools of mass communication. In the same time, spam also annoys with people. Although there are many methods of analysis for spam, most of them only serve local users, and thus cannot provide more information about spam. Besides, single method can't detect all types of spam. This paper provides an integrated e-mail security service system which combines the distributed log analysis, Bayesian classifier, url check, etc. Mechanism integrated from these methods and parameters adjusted via experiments make this system be able to detect more types of spam. By using forward and auto-reply functions, this system provides an e-mail analyzing service to public users, thus allowing collecting and processing various e-mail messages to improve the precision of spam detection.

Keywords：Spam、Integrated、Bayesian Classification、distributed log analysis.

1. 前言

隨著近年來網際網路的普及化，電子郵件目前已成為眾人主要的通訊方式之一，同時也被廣泛應用在商業用途上(如網路購物存證，線上遊戲註冊驗證，網路股票交易訊息等)。而電子郵件通訊的低成本和高普及率也被許多業者以及黑客(Hacker)看中，利用這項工具散播大量的廣告和病毒，達成他們的商業或入侵目的，造成現今垃圾郵件氾濫的狀況。因此，目前如何有效阻擋垃圾郵件(spam)，是資訊安全研究中一個重要的議題。

垃圾郵件(spam) [5][6] ，一般泛指用戶信箱中任何未經許可就強行送達的信件，目前還沒有一個非常明確的定義。大量寄發垃圾郵件的行為會造成頻寬、儲存設備、時間、生產力等等的資源濫用，導致嚴重的經濟耗損。根據賽門鐵克 2008 年 10 月份垃圾郵件報告顯示[15]，在 2007 年的電子郵件中，垃圾郵件佔有率已從 70%逐漸調升為 80%。另外，國家通訊傳播委員會(NCC)[16]已於 2008/11/24 討論通過濫發商業電子郵件管理條例草案，未來明定民眾可向濫發商業信件者求償 500~2000 元的損害賠償，可見垃圾郵件泛濫的問題同樣引起政府關注。

由於目前垃圾郵件內容多樣，單憑任一分析方法可能無法做到全方位的防護，因此本研究整合了日誌分析、內容過濾、規則檢查等技術，希望提升垃圾郵件整體的分析效能。另外，雖然目前垃圾郵件的驗證和分析機制種類繁多，但大多郵件伺服器的分析服務只提供給

本機用戶，不提供非用戶的郵件服務，而所收集到的垃圾郵件資訊也僅限於本地伺服器，顯得不夠全面。有鑑於此，本研究提供一個平台，讓外界郵件伺服器的管理者或用戶透過上傳郵件日誌或轉寄信件的方式，由我們的系統代為做日誌及信件分析，並且透過網頁呈現以及自動回信(auto reply)的方式，將分析結果回報。不僅可達到全面的資訊收集，亦可服務任何郵件伺服器的用戶。

本論文第二部分說明與此系統相關的技術以及文獻資料，第三部分則敘述本系統的研究方法、架構以及運作流程，第四章將對現有的功能做效能評估以及測試，最後在第五章對目前所做的研究做一個總結。

2. 相關技術與研究

一般垃圾郵件的辨識機制主要分為兩類，身份驗證和內容分析。身分驗證，顧名思義是拿已知的各種資訊來驗證信件來源是否符合系統自訂的條件，符合則通過，反之則阻擋，如黑名單、白名單、灰名單、SPF[17]等。內容分析，是透過各種特徵以及規則，分析郵件本文後再評估目標為垃圾郵件的可能性，如Spamassassin(簡稱 SA)[18]。

2.1 RBL

為了能夠應付現今多變的 spammer IP 或 Domain，網路上已有專門收集惡意來源，建立完整的黑名單資料庫，稱為「及時性黑名單」(Real-Time Blackhole Lists，RBL) [4] ，郵件伺服器可透過及時查詢 RBL 來判斷是否阻擋信件，可省去傳統黑名單查詢造成的負擔，亦可獲得較全面的資訊。但也由於此服務的及時性，資料庫的定期更新及維護也決定了過濾惡意來源的效能，若選擇過久未更新的資料庫可能會導致嚴重的 false positive 或 false negative，因此選擇有信譽的 RBL 作為黑名單查詢是非常重要的一環，如「ORBD」[19]是目前外界公認較準確且免費的 RBL。

2.2 階段式垃圾郵件行為分析技術

階段式垃圾郵件行為分析技術[7]主要以貝氏分類法為基礎，利用 SMTP 連接流程的四個階段對垃圾郵件的行為特徵進行分析，分別為 Helo、From、Rcpt to 與 Data。每階段分析的屬性各有不同，若在前面三階段中被識別為垃圾郵件則直接拒絕該郵件，有助於提升整體分析速度。但此方法採直接拒絕郵件的方式，難免會有遭誤判的合法郵件被刪除。並且在各階段皆為獨立分析，分析精準度可能沒有整合四階段特徵後共同分析來的高，因此本論文提出的系統，主要使用整合多樣分析結果的方式，將不同技術的分析結果經過統整，最後歸納出一個較全面的垃圾郵件評估。

2.3 分散式網路安全分析及偵測系統

考慮到內容分析機制會接觸郵件本文以及運算複雜度較高等因素，部分學者近期也著手研究以郵件日誌為基礎的分析技術[1]，希望藉由日誌檔上提供的郵件收送資訊以及系統訊息，達到近似內容分析機制的辨識效能，一方面可減低分析郵件的時間成本，另一方面可避免接觸他人隱私。

分散式網路安全分析及偵測系統[1]主要想法是建構一個網頁平台，提供外部使用者分析日誌，使用者可經由網頁介面將欲分析的日誌檔上傳，系統透過異常特徵比對、異常行為分析以及異常使用者追蹤對日誌檔做安全分析，再將回報結果呈現在網頁，供上傳者參考。但目前現有的日誌特徵還不成熟，無法辨識大部分的垃圾郵件，因此本研究也藉由分散式日誌的收集，對垃圾郵件的日誌特徵與寄件行為做研究，找出更多的相關資訊。

2.4 垃圾郵件陷阱

有些垃圾郵件研究者會為了能收集更多的研究數據，進而想辦法吸引更多的垃圾郵件發布者(spammer)寄信到自己的郵件伺服器。Spam Trap [6][8] ，便是其中一項技術。主要想法

為，建立一群專門接收垃圾郵件的帳號，再將這些帳號發佈到特定網頁上，並且想辦法讓人眼無法辨別(如白底白字)，如此則只有郵件位址捕獲程式能發覺我們所發佈的郵件位址，理所當然，這些帳號未來收到的郵件都將會是垃圾信。本研究也利用此觀念配合 postfix 的虛擬帳號功能達到收集垃圾郵件之目的。

3. 研究方法與系統架構

3.1 研究方法

為了能夠更全面的分析垃圾郵件，本研究整合了多項技術，包含分散式日誌分析、內容特徵比對、異常 url 檢測、郵件位址驗證以及貝氏分類。

3.1.1 分散式日誌分析

本研究使用的分散式日誌分析技術主要分為異常特徵比對、異常行為分析以及異常使用者追蹤[1] ，分析來源為透過網頁上傳的日誌檔以及從外部郵件伺服器遠端導入的日誌。異常特徵比對，是藉由比對已知的垃圾郵件特徵判斷是否為垃圾郵件的方法(圖 1)，如寄件者郵件位址異常(無@，空白，字串過長)、郵件 size 過大、收件人數過多等，都可作為垃圾郵件日誌特徵。當日誌分析程序開始進行時，會將資料庫中特徵資料表內的特徵值抓出與日誌進行比對，若發現符合特徵的日誌，則將該筆日誌資訊回報。此外，本研究使用正規表示法(Regular Expression)儲存特徵，能夠詮釋一般字串表示法無法呈現的組合(ex:「[^@]」代表字串不包括@)，提升整體分析效能。

異常行為分析則是根據多筆日誌的關連性來判斷寄件者過去是否發生異常行為的分析技術。異常行為在本研究指的是垃圾郵件寄件者(spammer)寄發垃圾郵件的跡象，而部分異常行為可經由分析多筆日誌後，從不同日誌間的關連性來辨別，如短時間內單一用戶收到多封來自相同寄件者的信件，或者不同帳號收到相同寄件者的信件、寄件者持續寄送郵件到不存在的位址等，都可解讀成發送垃圾郵件的行為。檢查流程為:將每筆日誌與 Log_behavior_pattern 資料表比對。若有符合資料表上任何異常行為的起始特徵 (如發現兩筆寄件者相同的日誌且時間點相近) ，則根據該異常行為的判斷規則(如一分鐘內接收多少相同寄件者的郵件)開始監控該寄件者的日誌。根據追蹤的結果，分析該寄件者是否符合某異常行為的條件(例如一分鐘內連續收超過七封視為異常行為)。

異常使用者追蹤，是檢查不同伺服器的日誌中有無相同的異常特徵或異常行為來源，並將檢查結果回報給伺服器管理者，提高管理者的警覺性。本研究的做法為，系統把分析到的異常日誌資訊(特徵或行為)匯入回報資料表前，先檢查該異常日誌的來源(郵件位址) ，過去是否同樣在其他郵件日誌中被偵測為異常，若有則在回報分析結果或做異常記錄查詢時將此資訊一併納入。

id	pid	ps	name	fname	pattern	infor
☐	2	1	From a drug	from	/\\b(?:cialis|levitra|phentermine|valium|viagra|vicodin|xanax)\\b/i	
☐	3	2	From who do ya say?	from	/*Hoodia/i	
☐	4	4	From invalid address at PayPal	from	(?:admin|services|support|update|verification)\\@paypal.com/i	

圖 1.　異常日誌特徵

3.1.2 內容特徵比對

主要分為表頭特徵比對以及本文特徵比對。表頭特徵比對是針對信件表頭各欄位出現的異常資訊做比對的動作，包括了來源郵件位址異常、message-id 異常以及信件標題異常等。本文特徵則比對垃圾郵件常出現的詞彙(如賭場) ，以及當下發生的時事資訊。例如，在 2004 年勞力士仿冒品的垃圾郵件[20]，內容極力鼓吹大眾購買該公司的仿造品，因此當時許多與 Rolex 相關的單字組合皆被列為本文特徵。

其中，由於表頭特徵的部分欄位與日誌特徵相同，因此該部分將與日誌分析合作，透過查詢過去的日誌回報記錄先確認該郵件來源位址過去是否發生異常特徵，若有則可省去再次分析的時間，並且日誌上某些資訊不會顯示在

信件表頭(ex: size、寄件數) ，透過此合作方式可偵測到更多異常，增加分析效能。

3.1.3 異常url檢測

有鑑於現今許多釣魚網站(phishing website)皆透過釣魚郵件(phishing mail)[9]來達到散播目的，釣魚郵件主要是利用聳動或者不實的消息(如您的某會員密碼已快到期，若不更新將會封鎖帳號。) ，欺騙收件者進入釣魚網站輸入私人資訊，如帳號密碼、手機、工作場所、住家地址等，而釣魚信件的內容可能是文字、圖片、超連結、甚至就是一個網頁。對於釣魚網站防護，本論文主要針對信件中的 url 做分析。分析包含偽造 url 位址檢測與 url 特徵比對，其中 url 特徵比對與上述特徵比對技術相同，比較重要的是可以利用 Regular Expression 來定義 Cross-Site Scripting（XSS）的 url，作為偵測以 Email 為途徑的 XSS。偽造 url 檢測是指一般網頁連結都分為兩部分，分別為實際連接的位址以及在瀏覽器上顯示的名稱，而許多釣魚信件所附的 url 上通常顯示網址與實際連結網址會不一致(如顯示 yahoo 實際卻連向 google) 。本技術主要在比對有 url 的信件中，實際位址與顯示位址是否一致，若發現不一致的 url 則在顯示部分做出警告(圖 2)讓收件者提高警覺。

3.1.4 郵件位址驗證

本研究將各分析流程所擷取到的郵件位址存入特定 mail_address 資料表並利用 php 語法配合自動機的觀念對資料表內的郵件位址做驗證(圖 3) 。驗證每筆位址後再將結果匯入對應的 status 欄位。 每筆位址檢查的流程如下：

1. 檢查@是否存在，不存在則更新該位址的 status 欄位為 invlid。
2. 檢查該位址的 mx server 是否存在，不存在則更新該位址的 status 欄位為 invlid。
3. 嘗試連接該 mx server 的 25port，連不上則更新該位址的 status 欄位為 unknow。
4. 做 helo 指令測試，失敗則更新該位址的 status 欄位為 invlid。

To never aga1n : warning: URL名稱不一致

nice meeting you

Mitchel Atkinson
Copyist
Rephartox B.V., Lelystad, Netherlands
Phone: 228-618-7572
Mobile: 992-124-5177
Email: ayqhyaxjdsnxmq@world4fun.ch

please do not reply to this message

This product is a 96 second trial freeware

NOTES:

The contents of this paper is for your exclusive use and should not be span bulletin

alum crowberry pirogue

Time: Thu, 13 Jan 2005 21:34:26 -0800 ------_001_5710_95K30SC6.87HV1190-- 1

圖 2 假造 url 檢測回報

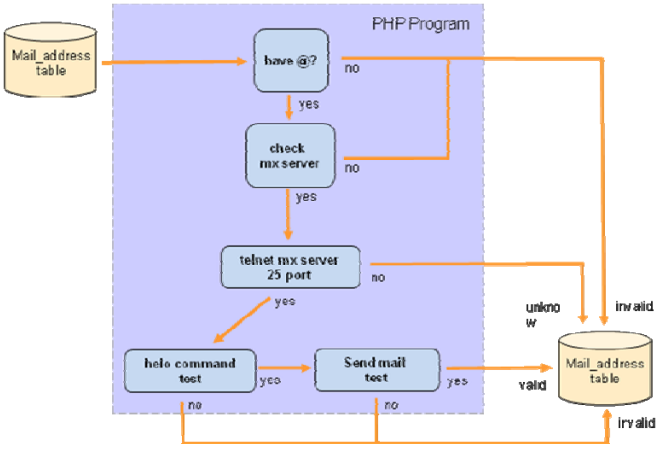

圖 3. 郵件位址檢測

5. 做寄件測試，成功則更新該位址的 status 欄位為 vlid。否則為 invlid。

其中流程 3 連不上設為 unknow 是因為某些合法郵件伺服器不開放外部 telnet 服務，因此無法判斷。驗證後的郵件位址可提供分散式日誌分析系統與郵件內文分析系統做查詢。

3.1.5 貝氏分類

貝氏分類法[2][3][7][10][11][12][13][14][21]是利用「貝氏定理」發明的文件分類法。「貝氏

定理」是將事前機率與條件機率結合，最後導出事後機率的過程。此技術主要分為資料庫訓練與信件分類兩部分。資料庫訓練主要是透過個別匯入大量單一種類的信件(ex:1000 封 spam) ，將每封信切割成斷字(Token)後存入該類別的斷字資料表，做為日後信件分類的依據。信件分類部分同樣將欲分類的信件分割成斷字，再透過與各信件類別的斷字資料表做比對，計算每個斷字在各信件類別出現的機率(本研究之類別為 spam 與 nonspam)，進而推算出為垃圾信的可能性。假設將一封信 M 切成斷字 $w_1 \ldots w_n$，則其分類公式如(1)所示，p 表示該信為 spam 類別的機率，$p_1 \ldots p_n$ 分別是 $w_1 \ldots w_n$ 為 spam 類別的機率，$(1\text{-}p_1) \ldots (1\text{-}p_n)$則是 $w_1 \ldots w_n$ 為 nonspam 類別的機率。(2)為 p_m 算式(m=1…n) ，其中 $P(w_m|S)$為 w_m 在 spam 類別出現的機率，$P(w_m|H)$為 w_m 在 nonspam 類別出現的機率，P(S)與 P(H)在此分別代表 spam 與 nonspam 類別的評分加權值，相加為 1。透過此公式，可計算出各類別生產出該郵件的機率值，最後將該郵件歸類至機率值最大的類別。

$$p = \frac{p_1 \cdot p_2 \cdot \ldots \cdot p_n}{p_1 \cdot p_2 \cdot \ldots \cdot p_n + (1-p_1) \cdot (1-p_2) \cdot \ldots \cdot (1-p_n)} \quad (1)$$

$$p_m = P(S|w_m) = \frac{P(w_m|S) \cdot P(S)}{P(w_m|S) \cdot P(S) + P(w_m|H) \cdot P(H)} \quad (2)$$

3.2 系統架構

本系統實驗環境以 Linux (Fedora 10)系統為基礎，搭配 Postfix Server，Web Server、C 語言、PHP、PostgreSQL 作為伺服器端平台，主要研究項目為：

1. 整合內容過濾、日誌分析、位址認證以及 url 檢測等技術，提升垃圾郵件的分析效能。
2. 提供一個能服務不同郵件伺服器用戶及管理者的分析平台。

因此，根據上述幾點，本系統架構分為三部分(圖 4)：分散式日誌分析系統、內文分析系統以及網頁監管系統。分散式日誌分析系統負責接收郵件日誌並針對日誌上的特定行為及特徵進行分析與回報。內文分析系統則是對所有進入郵件傳遞代理單元(MDA)的郵件進行內容過濾、特徵比對、病毒掃描等程序後再依據結果做及時的文件評估。

上述兩部分都會將寄件端的郵件位址存入資料表中，再透過郵件位址檢測程式進行驗證。網頁監管系統負責將資料庫內所有數據以網頁方式呈現，並且監控所有系統相關服務的運作狀態。資料庫系統主要存放接收到的數據、分析結果以及本系統會用到的特徵與行為資訊。接下來就各部份的架構及流程做詳細說明。

3.2.1 分散式日誌分析系統

此部分目的為透過大量的日誌分析，製做出一份黑名單供郵件伺服器使用，由於此系統是做離線型分析，不會影響收寄件流程，因此可提升郵件伺服器整體的分析效能與速度。

為了能夠收集全面的郵件日誌資訊，系統提供各郵件伺服器的管理者透過網頁將日誌上傳到本系統，再將分析結果回報給管理者並同時存入資料庫系統。另外也支援外界郵件伺服器將日誌同步導入至系統中。主要流程分為上傳日誌與導入日誌兩部分(圖 5) 。

上傳日誌部分，使用者一開始會先透過系統的網頁介面將欲分析的日誌檔上傳，網頁系統會將收到的日誌檔存入分析系統。在分析系統中，先將日誌檔透過用 php 撰寫的資料處理單元做處理後，把該日誌檔的重要資訊，包含上傳者 IP、上傳時間、寄件者位址、收件者位址、日誌記錄時間等，分別匯入個人日誌暫存表以及常駐的上傳日誌資料表，並且同時將所有郵件位址匯入郵件位址資料表，並且同時向郵件位址資料表查詢日誌中是否有無效郵件位址存在，若發現則視為異常。接下來個人日誌暫存表中的內容將透過異常特徵比對、異常行為分析以及異常寄件者追蹤機制來判斷是否有異常日誌存在，再將分析結果匯入個人回報暫存表以及常駐回報資料表，最後把個人回報資料表以網頁方式呈現給使用者後刪除所有暫存

資料表。

導入日誌與上傳日誌流程的差別僅在於導入日誌流程中，不用向使用者回報分析結果，因此不需建立臨時資料表。而伺服器日誌導入的方式為遠端伺服器將日誌透過 syslogd 設定後同步傳送至本系統的 port 514，由我們撰寫的 C 語言接收程式收下後做分析與記錄。

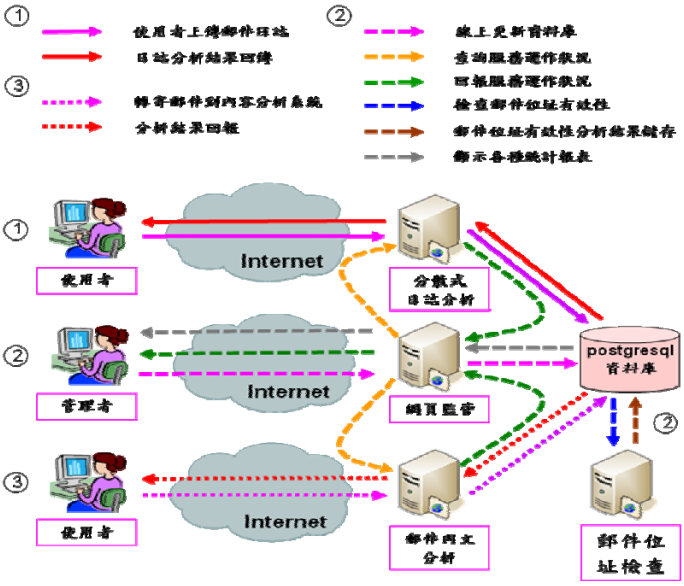

圖 5.　系統整體架構

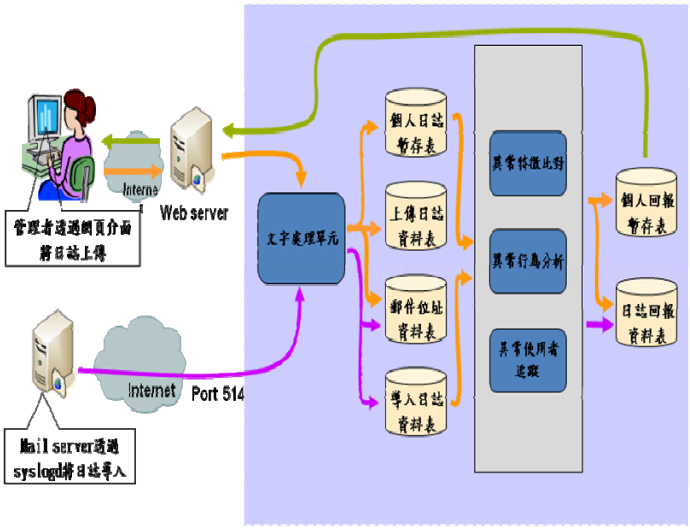

圖 5.　分散式日誌分析系統

3.2.2 內文分析系統

本系統主要是針對所有進入郵件傳遞代理單元(MDA)的郵件進行內容過濾、特徵比對、病毒掃描等程序後再依據分析結果做及時的文件評估。若評估後發現信件異常，則以修改標題及表頭的方式提醒收件者。除了上述提供日誌的分析服務外，我們也希望對外界共享內文分析系統的服務，因此本系統架構分為外界服務信件與一般信件兩部分(圖 6)。

首先，針對外部服務信件，本系統提供了一個固定帳號「checkspam」，用戶可將欲分析之信件透過轉寄到該帳號的方式，經由本系統做內容分析後，再將分析結果以自動回覆的方式寄返。

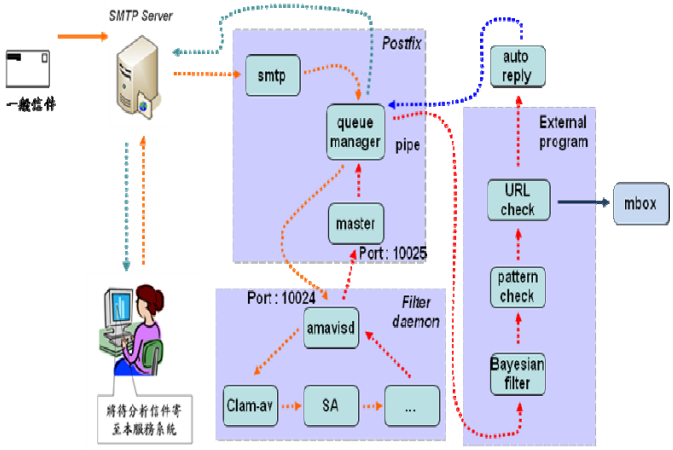

圖 6.　內文分析系統

內部流程如下:

1. 信件透過 smtp 協定從 25port 收下。
2. 交由 queue-manager 做郵件位址確認。
3. 將信件送到 10024port 由 amavisd 接收。
4. 透過 Clam-Av 與 SA 對郵件進行掃描&過濾。
5. Amavisd 將信件傳送到 10025port 由 postfix 重複進行寄件流程。
6. 再次交由 queue-manager 做郵件位址確認。
7. 透過 postfix 的 pipe 設定將信件導入我們設計的外部過濾程序。
8. 分別對信件做日誌紀錄查詢、貝氏過濾、內容特徵比對以及 url 檢查。
9. 透過 auto-reply 程式將信件寄回原寄件者。

其中 port 10025 是 postfix 為了接收 amavisd 回傳信件而另外由 master 元件設定的 smtp port，並且從這個 port 所接收的郵件不會再度啟動過濾機制，避免陷入無窮迴圈。若信件在上述流程中發現異常，系統會以修改標題的方式提醒收件者，並在表頭上註明詳細資訊。

一般信件流程與外界服務信件的差別在於一般信件做完外部程式分析後，會直接寫入 mailbox，不需經過 auto-reply 程式。

3.2.3 網頁監管系統

為了有效監控各系統的分析結果與接收數據以及方便管理自訂的分析規則，本論文提出一個網頁監管系統(圖 7)，將資料庫系統內各類分析結果與自定分析規則，以網頁方式呈現，達到監控與管理的需求。主要分為監控面與管理面兩部分，在監控方面，系統將至今的各子系統的分析結果製作成不同類型的報表後以網頁方式呈現，讓管理者便於分析當前垃圾郵件的趨勢以及各垃圾郵件與日誌間的關連性，並且透過指令回報的方式，可隨時在網頁上監控其餘兩個子系統的運作狀態，方便管理者隨時監控。而管理部份，透過本系統，能讓管理者透過自訂的網頁介面直接對分析規則做更新，隨時對分析系統做出適當的管理。

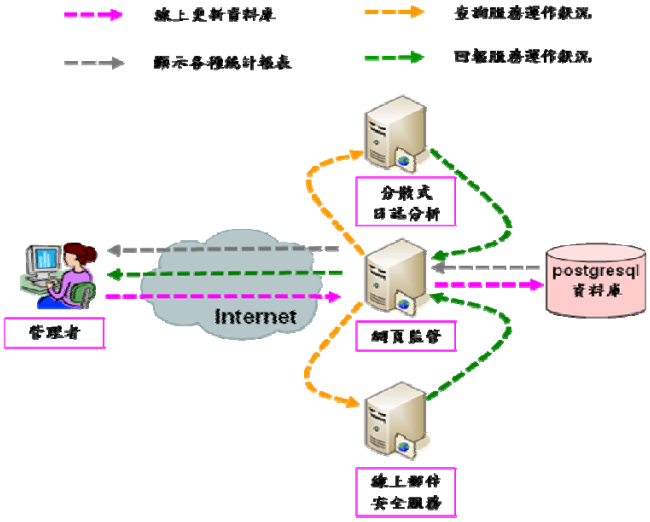

圖 7. 網頁監管系統

4. 研究結果與分析討論

本部分主要針對系統功能做效能分析及統計，包括了貝式過濾器，內容特徵分析以及 url 檢測，並設法將這些分析結果統整成一個明確的數值，讓使用者更容易理解系統的分析結果。

4.1 系統效能測試

目前測試的主要數據來源有三組，分別是 TREC[22] 、Spamassassin[23] 以及 Enrron[24]。其中 Enron 共有 52675 封信件，包括 19088 封正常信與 33587 封垃圾郵件；TREC 共有 92192 封封信，包括 39402 封正常郵件與 52790 封垃圾郵件；Spamassassin 共 9352 封信件，其中含 2401 垃圾郵件與 6951 正常郵件。貝氏資料庫訓練方面，為了讓分類器學習到較多樣的郵件內容， 我們從三組來源的 spam 及 nonspam 信件中各抽出 2000 封，做為訓練樣本。

測試效能的方法，是將各郵件來源(Enron、TREC、Spamassassin)分為 nonspam 與 spam 兩部分去測試個別功能(先測 spam 再測 nonspam)最後算出個別功能的 precision(3)、miss_rate(4)以及 false_alarm_rate(5) 。其中 SS 為是 Spam 且被判別為 Spam 的數量，NS 為非 Spam 卻被判別為 Spam 的數量，SN 為是 Spam 但被視為正常郵件的數量。

$$\text{precision} = \frac{SS}{SS + NS} \qquad (3)$$

$$miss\_rate = \frac{NS}{nonspam\text{總量}} \qquad (4)$$

$$false\_alarm\_rate = \frac{SN}{spam\text{總量}} \qquad (5)$$

首先，針對貝氏分類器的部分，本研究定義了三種不同的 P(S)及 P(H) 來測試這兩項數值對貝氏分類結果的影響。分別為重覆單字比例，實際類別比例以及同等權重。重複斷字比例是依據各類別目前經資料庫訓練得到的斷字數對總斷字數的佔有率來定義 P(S)及 P(H) ，包

含重覆的斷字，假設 spam 類別現有斷字數為 1000，nonspam 為 3000，則 P(S)=0.25，P(H)=0.75(在本研究 P(S)= 0.5566，P(H)= 0.4434)。實際類別比例是以第一部分提到的賽門鐵克的垃圾郵件報告做依據， P(S)=0.8，P(H)=0.2。同等權重是指不使用加權機制來測試分類結果，因此 P(S)及 P(H)皆為 0.5，分析結果如(表 1)所示。另外我們也針對不同的 P(H)與 P(S)比例做效能測試(0.1 / 0.9 ~ 0.9 / 0.1)，測試結果為(圖 8)(圖 9)(圖 10)。

表 1. 不同 P(S)&P(H)定義比較結果

P(S)/P(H)	precision	miss_rate	false_alarm_rate
重覆單字比例 P(S) = 0.5566 P(H) = 0.4434	0.9447	0.0566	0.0316
實際類別比例 P(S) = 0.8 P(H) = 0.2	0.7140	0.3963	0.0105
同等權重 P(S) = 0.5 P(H) = 0.5	0.9686	0.0308	0.04523

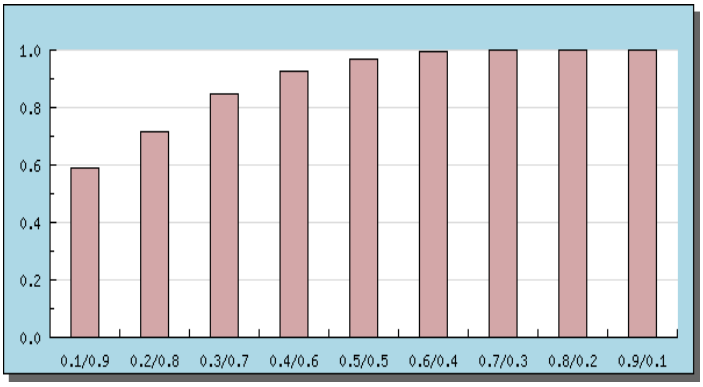

圖 8 各 P(H)／P(S)比例測試結果- precision

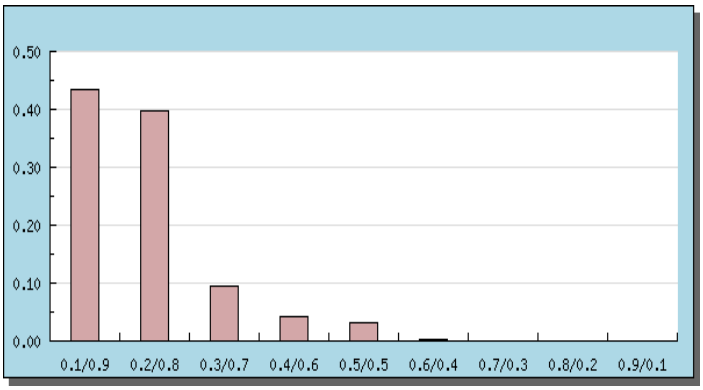

圖 9 各 P(H)／P(S)比例測試結果- miss_rate

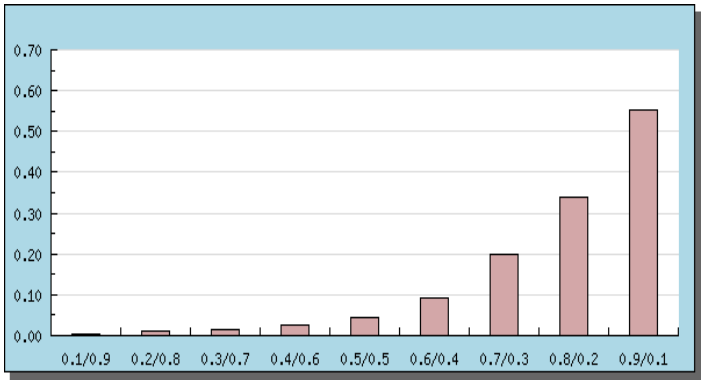

圖 10 各 P(H)／P(S)比例測試結果- false_alarm_rate

從上述兩項測試結果皆可發現一個規律，P(S)越高，則 precision 及 miss_rate 的數值被劣化，false_alarm_rate 被優化。因為 spam 的整體分數提高，表示信件被判別為 spam 的機率相對提升，不管是 spam 類或是 nonspam 類都一樣。若在 spam 類別判斷正確率（SS 出現的機率）已經很高的情況下再將 spam 評分權重提高，則 NS 的成長幅度，絕對會比 SS 來的大，因此這樣的結果是可預期的。相對的，P(H)越高代表全部類別判斷為 nonspam 的比例都會增加，因此 precision 及 miss_rate 數值會被優化，false_alarm_rate 數值則被劣化。由此測試可得知，P(S)及 P(H)在貝氏分類器中佔有舉足輕重的地位，如何定義一個合適的評分加權值，是郵件分析系統必須考量的參數；另外，系統可以藉由使用者的更正行為（將 nonspam 信件從 spam 中移出），來動態調整 P(S)及 P(H)。而在電子郵件的考量裡，大部分收件者會寧願多收幾封垃圾信也不希望合法信件被判斷成垃圾信（low miss_rate）。因此，在 false_alarm_rate 是合理範圍的情況下，本研究採用同等權重的方式定義評分加權值。

各功能評估結果如(表 2)所示，其中 Spamassassin 為 Postfix 郵件系統內建的過慮套件，主要透過各種規則配合不同權重的分數，將郵件所符合的規則分數加總後，再由管理者自訂的門檻分數判斷其為垃圾郵件的可能性。而 content_chk 與 url_chk 所用的 rule 則是參考 SARE 網站[26]以及 snort 官網[25]所發佈的部份

規則。從結果可看出，這三種方法的 miss_rate 都很低，在 url 檢測與內容特徵比對的部分，雖然 false_alarm_rate 很高，但是一旦偵測到符合特徵的郵件，其 precision 高達 95%，雖不能分析全方位的垃圾郵件，但在能偵測到的部分卻有很高的精準度，因此可拿來作為整合式系統的一部分，負責檢查有 url 異常特徵與異常本文特徵的郵件。而貝氏分類法的測試結果在各方面數值上都很優異，但還是有少部分無法正確判別的郵件，可能是該垃圾郵件有某些特徵是無法靠斷字統計做判別，而這些特徵剛好可以放入 url 異常特徵或異常本文特徵中，達到良好的互補效果。

4.2 整合式評估

我們依據得到的測試結果將各功能做整合，做法採用計分制，並且各方法在分析結束後使用測試結果的 precision 做為加權值，最後將各功能產生的分數做加總，依據總分判斷該郵件是否為 spam。因此，以上述測試得到的結果為例，我們設定一封信件如果經貝氏分類判別為 spam 算 1Hit，內容特徵與 url 特徵部分，符合一個特徵算 1hit，則最後總分計算如(6)所示，其中 TH 為 total_hit。

$$TH = SA\_Hit * 0.9682 + Bayesian\_Hit * 0.9686 + content\_Hit * 0.9164 + url\_Hit * 0.9541 \quad (6)$$

此外，本研究為了證明內容特徵比對與異常 url 檢測功能存在的必要性，先從三個郵件來源中取樣 30000 封 spam 對貝氏分類器做測試，測試結果共有 1347 封 spam 誤判為 nonspam，再將這些郵件分別做內容特徵比對與異常 url 檢測，測試結果如(表 3)所示，可看到兩種方法皆可偵測出貝氏分類器遺漏的垃圾郵件，證明了此兩種方法皆有存在的必要，也證明本系統確實能提高整體的郵件分析效能。

表 2. 系統效能測試結果

分析方法	precision	miss_rate	false_alarm_rate
Spamassassin	0.9682	0.0287	0.1507
url_chk	0.9541	0.0097	0.8145
content_chk	0.9164	0.0083	0.9321
Bayesian	0.9686	0.0308	0.0452

表 3. 功能必要性測試結果

分析方法	Spam_hit數
url_chk	48
cmtent_chk	40

5.結論

本論文建構一個整合式垃圾郵件分析系統。主要分為分散式日誌分析、內文分析以及網頁監管三部分，其中分散氏日誌分析是透過開放外界用戶上傳日誌到本系統，透過分析他人的日誌讓系統接收到更多外部郵件資訊。內文分析部分透過 postfix 的 pipe 功能將信件導出至外部程式，為郵件做進一步的分析，其中包括了特徵比對，url 檢測以及貝氏分類，並且同時提供外部用戶透過轉寄到特定帳號的方式代為分析郵件，再透過 auto reply 的方是將信件連同分析結果寄回。網頁監控部分主要是提供一個網頁界面，讓管理者能夠隨時隨地透過網頁監控整個系統。效能分析上，我們使用 Enrron、Spamassassin、TREC 提供的郵件數據做個別功能的測試，再經由測試結果計算每個功能的權重，最後將不同功能的分析結果統整成簡單易懂的資訊。未來將繼續提升現有功能

的準確度與辨識範圍並且蒐尋更多能加入本系統的分析方法。

參考文獻

[1] 章博傑，“分散式網路安全分析及偵測系統之研究”,私立銘傳大學資訊傳播工程研究所碩士論文，(民 97)。

[2] 張僩鈞，“兩階層式垃圾郵件過濾機制之研究”,私立銘傳大學資訊傳播工程研究所碩士論文，(民 95)。

[3] 張年旺，“增進 Postfix 系統之垃圾郵件過濾效能”,國立高雄第一科技大學電腦與通訊系研究所碩士論文，(民 95)。

[4] 王文政，“垃圾郵件過濾系統之分析研究”,國立台灣科技大學資訊管理研究所，(民 94)。

[5] PJ Sandford, JM Sandford, DJ Parish,“Analysis of SMTP Connection Characteristics for Detecting Spam Relays” This paper appears in: Computing in the Global Information Technology, 2006. ICCGI '06. International Multi-Conference on, Publication Date: Aug. 2006, On page(s): 68 – 68.

[6] Wilfried N. Gansterer, Michael Ilger,“Analyzing UCE/UBE traffic”, This paper appears in:ACM International Conference Proceeding Series, Year of Publication: 2007, on Pages: 195 – 204.

[7] Liu Ming , Li Yunchun , Li Wei,” Spam Filtering by Stages” This paper appears in: : Convergence Information Technology, 2007. International Conference on, Publication Date: 21-23 Nov. 2007, On page(s): 2209 – 2213.

[8] Taverira,D.M., Duarte,O.C.M., “A monitor Tool for Anti-spam Mechanisams and Spammers Behavior” This paper appears in: Network Operations and Management Symposium Workshops, 2008. NOMS Workshops 2008. IEEE Publication Date: 7-11 April 2008 On page(s): 101 – 108。

[9] Aburrous,M.,Hossain,M.A., Thabatah,F., Dahal, K. “Intelligent Phishing Website Detection System using Fuzzy Techniques”, This paper appears in: Information and Communication Technologies: From Theory to Applications, 2008. ICTTA 2008. 3rd International, Publication Date: 7-11 April 2008 Conference on, On page(s): 1 - 6。

[10] Biju Issac, Wendy Japutra Jap, Jofry Hadi Sutanto, “Improved Bayesian Anti-Spam Filter – Implementation and Analysis onIndependent Spam Corpuses”, This paper appears in: Computer Engineering and Technology, 2009. ICCET '08. International Conference on, Publication Date: 22-24 Jan. 2009, On page(s): 326 - 330。

[11] Chun-Chao Yeh, Soun-Jan Chiang, “Revisit Bayesian Approaches for Spam Detection”, This paper appears in: Young Computer Scientists, 2008. ICYCS 2008. The 9th International Conference for, Publication Date: 18-21 Nov. 2008, On page(s): 659 - 664。

[12] Yun Wang, Zhiqiang Wu, Runxiu Wu, “Spam filtering system based on rough set and Bayesian classifier”, This paper appears in: Granular Computing, 2008. GrC 2008. IEEE International Conference on, Publication Date: 26-28 Aug. 2008, On page(s): 624 - 627。

[13] Abu-Nimeh, S., Nappa, D., Xinlei Wang, Nair, S., “Bayesian Additive Regression Trees-Based Spam Detection for Enhanced Email Privacy”, This paper appears in: Availability, Reliability and Security, 2008. ARES 08. Third International Conference on, Publication Date: 4-7 March 2008, On page(s): 1044 - 1051 。

[14] Chui-Yu Chiu, Yuan-Ting Huang, “Integration of Support Vector Machine with Naïve Bayesian Classifier for Spam Classification”, This paper appears in: Fuzzy Systems and Knowledge Discovery, 2007. FSKD 2007. Fourth International Conference on, Publication Date: 24-27 Aug. 2007, On page(s): 618 - 622。

[15] The State of Spam A Monthly Report – October 2008, Generated by Symantec Messaging and Web Security

[16] NCC-NEWS http://blog.nownews.com/alexandros/textview.php?file=0000193471

[17] Sender-Policy-Framework, http://www.openspf.org/

[18] SpamAssassin, http://spamassassin.org/

[19] ORBD,http://www.coolacid.net/the-news/99-orbdorg-marks-all-as-spam

[20] Rolex-NEWS, http://www.epochtimes.com/b5/4/10/29/n703447.htm

[21] BayesianFiltering, http://forum.icst.org.tw/phpbb/viewtopic.php?t=3641

[22] 2005 TREC Public Spam http://plg.uwaterloo.ca/~gvcormac/treccorpus/

[23] SpamKANN, http://www.trudgian.net/spamkann/

[24] Enron-Email-Dataset,http://www.cs.cmu.edu/~enron/

[25] Snort, http://www.snort.org/snort-rules/

[26] SARE http://www.rulesemporium.com/rules.htm