新軟週報
http://www.nusoft.com.tw
Internet Security Fighter
2007/9/17- 2007/9/23 第43期

多功能 UTM / MS 系列報導

技術淺談與應用 - 為何初架設的新軟UTM，其貝氏過濾資料庫中並無資料

現今網路垃圾郵件氾濫成災，是所有電子郵件使用者的夢魘。數年前，垃圾郵件頂多只是讓收件者感到麻煩，現在卻成為耗損企業生產力的嚴重問題；許多垃圾郵件內容不只不堪入目且無實際效益，還會嚴重佔用企業網路頻寬，同時浪費公司郵件伺服器主機的儲存空間與運算效能。也就是因為垃圾郵件的氾濫已經開始拖累經濟發展（虛耗企業成本、佔用網路頻寬...），並且延伸出許多社會問題（網路釣魚、情色廣告...）。因此各國政府開始針對垃圾郵件訂定各種法案來遏止其氾濫，但至今仍無太大作用。在還未有強效解決方案之前，垃圾郵件與反垃圾郵件的戰爭只會越演越烈。

為了協助企業防堵垃圾郵件，許多資訊安全公司推出各種反垃圾郵件機制。但道高一尺，魔高一丈，在進步的不只是反垃圾郵件機制，垃圾蟲（Spammer，指濫發垃圾郵件者）也在研擬更新的垃圾郵件發送方法，使得部份較早推出的反垃圾郵件機制逐漸失去舞台（例如：垃圾郵件黑名單－RBL）。但是，也有種反垃圾郵件機制在這場「垃圾郵件大戰」中仍能歷久彌新，並廣泛被應用於各種反垃圾郵件產品中－「貝氏過濾法」（Bayesian Filtering）。

「貝氏過濾法」是利用「貝氏定理」為基礎而推出的機制。「貝氏定理」是在西元1763 年貝士（Thomas Bayes）的遺著中所發現。這個歷史長達兩世紀的古老定理是透過“事前機率”與“條件機率”，來計算出“事後機率”，常常運用在統計學當中。這定理雖為古老，但它所延伸出來的「貝氏過濾法」卻對付垃圾郵件卻出奇的好用。

「貝氏過濾法」的運作方式是將整封信件（含信件 Header、信件內文...，附加檔案除外）分割成一個個單一詞句（Token），再利用特定演算法分析每個詞句的出現機率給予信件評分。最後，以信件分數的多寡來評斷該信件是否為垃圾信件。

上述「貝氏過濾法」之運作，所依賴評分的標準就是擁有大量詞句的「貝氏過濾資料庫」。「貝氏過濾資料庫」分為兩個－“垃圾郵件資料庫”與“非垃圾郵件資料庫”。“垃圾郵件資料庫”專門存放各種垃圾郵件用詞，而“非垃圾郵件資料庫”則保存了企業往來信件中的常用詞。「貝氏過濾法」就是在這兩個資料庫中比對每個詞句之出現機率，來判斷該信件是否為垃圾信件。

有一點要注意的是，新軟系統產品－多功能 UTM 在出廠時，「貝氏過濾資料庫」上並無任何資料，因此「貝氏過濾法」在一開始時並無法正常使用。唯企業以“辨識學習”方式，提供「貝氏過濾資料庫」正常信件、垃圾信件各 200 封後，方可正常運作。

為甚麼新軟多功能 UTM 在出廠時就不事先提供預設的「貝氏過濾資料庫」呢？還要客戶提供正常信件、垃圾信件學習不是很麻煩嗎？其實主要原因有以下兩點：

1. 信件是否屬於垃圾郵件是依收件者的主觀認知來判定；就同一信件而言，有可能每位使用者、企業的判斷都不同，更何況是處於模糊地帶的“電子報”。依目前情況，絕大部分“電子報”皆屬於推銷產品的垃圾郵件，但是假如收件者剛好有此種產品之購買需求，他也就不會把這封“電子報”視為垃圾郵件。

2. 多功能 UTM 在出廠時並沒有企業往來郵件的正確資料。倘若「貝氏過濾法」使用預設的「貝氏過濾資料庫」來過濾信件時，將有可能導致正常信件被誤判為垃圾信件的情況發生。

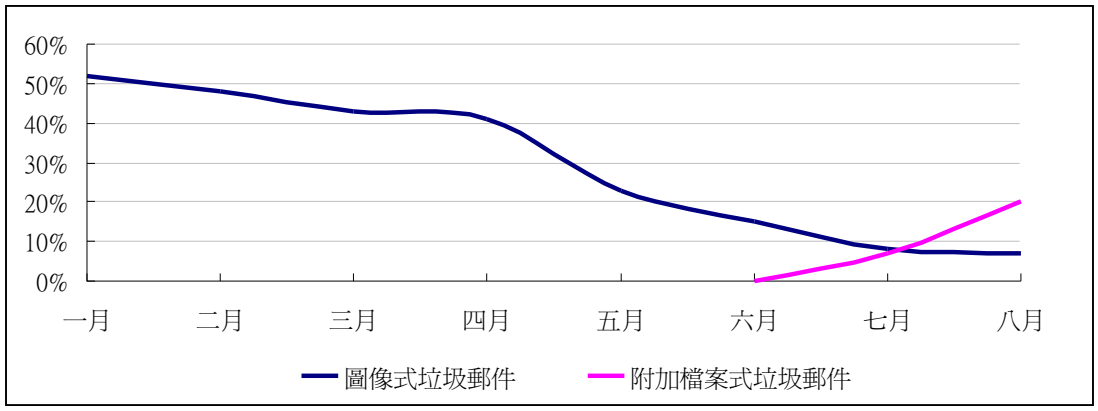
因此企業如須啟用「貝氏過濾法」來濾除垃圾信件，就必須先用“辨識學習”教會「貝氏過濾法」甚麼是垃圾信件，甚麼是非垃圾信件。往後如有誤判的情況發生，也可使用“辨識學習”矯正此錯誤。當然，企業提供的資料越多，「貝氏過濾法」也會越精確，甚至可高達九成九以上的辨識率！！（理想的“辨識學習”比例 － 垃圾郵件：正常信件 ＝ 2：1）

文 黃贊中 isaac@nusoft.com.tw

市場行銷報導 - 垃圾郵件新趨勢：附加檔案形式垃圾郵件

在最近一兩個月來，您是否會收到一些奇怪的信件？沒有郵件內容，卻夾帶了PDF、Excel、Zip...檔案，打開來看看，卻發現裡面都是廣告！！您猜的沒錯，這就是目前開始流行的新型態垃圾郵件－“附加檔案式垃圾郵件”。

在年初所流行以圖片方式（jpg、gif...）散佈垃圾郵件訊息的“圖像式垃圾郵件（Image Spam）”，由於垃圾蟲（Spammer，指濫發垃圾郵件者）使用了各種“防機器判讀技術”替圖片加料（文字變形、變色、雜點...）以躲避反垃圾郵件系統的查緝，使得“圖像式垃圾郵件”一度氾濫嚴重，而無法阻攔。這問題在各種新式反垃圾郵件系統陸續推出之後得到解決，使得“圖像式垃圾郵件”的比例持續減少，漸漸取而代之的就是“附加檔型式垃圾郵件”。



圖一 圖像式垃圾郵件 與 附加檔案式垃圾郵件 佔全體垃圾郵件比例

“附加檔案式垃圾郵件”通常夾帶著Word、Excel、PDF、FDF（Adobe Acrobat表單文檔文件）...這幾種常用文件類型的廣告，或將這些文件以Zip壓縮後再夾帶於信中。就因一般反垃圾郵件機制不會針對附加檔案查驗，所以這些新式的垃圾郵件能夠輕易的穿過各種反垃圾郵件機制。也因為“附加檔案式垃圾郵件”通常偽裝成正常郵件，所以相當容易促使收件者開啟，達到廣告宣傳的目的。

為了因應這些變化多端之垃圾郵件，新軟系統的垃圾郵件過濾機制採用多層架構過濾方式（又稱為雞尾酒式），以複合方法層層濾除垃圾郵件。其中，“垃圾郵件特徵”便是專門處理新式垃圾郵件之一大利器；可有效濾除各種新式垃圾郵件，當然也包含“附加檔案式垃圾郵件”。

每當有新型態垃圾郵件開始出現時，新軟系統的工程師們便會分析其各項特徵，匯整成「垃圾郵件特徵碼」，供新軟系統產品「多功能UTM」及「郵件伺服器」下載。藉此方式有效處理來襲的各式垃圾郵件，還給企業一個乾淨的電子郵件環境。

文 程智偉 rayearth@nusoft.com.tw