特徵權重與數量對網路入侵偵測系統影響之研究

蘇民揚，張凱棊，魏華甫，林俊淵，莊淵全，謝瑞峰
銘傳大學資訊工程學系
minysu@mcu.edu.tw

摘要

異常行為偵測型(abnormal-behavior detection)的網路入侵偵測系統，其成功與否之重要關鍵在於所選取以供判斷的特徵是否恰當。另一方面，在強調即時處理的入侵偵測系統上，時間因素也至為重要；愈多的特徵也意謂處理時間愈長。如何選擇適量且有效的特徵實為網路入侵偵測系統設計中最重要的一環。本論文使用基因演算法結合 KNN (K-Nearest Neighbor)做為特徵選取的策略，並以大量阻斷服務攻擊(DoS attacks)做為實驗的對象。我們一開始使用 26 個特徵，訓練階段為每個特徵訓練出合適的權重值，再選取少量、重要的特徵進行測試。實驗顯示，對已知攻擊，14 個特徵表現最好，整體正確率(overall accuracy)達 99.83%；對未知攻擊，24 個特徵表現最好，整體正確率達 88.11%。

關鍵詞：網路入侵偵測、網路安全、阻斷服務攻擊(DoS)、基因演算法、特徵選取、KNN

1. 簡介

隨著資訊科技日新月異，帶動著電子商務的興起，網路安全也逐漸受到人們的重視。當人們都在享受著網路所提供便利以及功能時，網路攻擊也慢慢成為電子商務的隱憂；於是網路安全的重要性與日俱增，網路入侵偵測系統(Network Intrusion Detection System, NIDS)也因應而生。然而在強調可以即時偵測的 NIDS 裡，如何選取有效而又少量的特徵以降低處理時間同時提高偵測率，將關乎 NIDS 之成敗，也是本論文所要解決的問題。由於阻斷服務攻擊(DoS/DDoS)製作技術門檻最低，以致網路上隨處可以取得，為現今網路攻擊事件的主流之一，故本論文所提之方法也將以 DoS/DDoS 攻擊作為評估的實驗對象。

入侵偵測系統設計主要有兩大方向：誤用偵測(misuse detection)與異常行為偵測(anomaly detection)。誤用偵測是藉由系統管理者去定義出會造成電腦系統毀損或者是危害網路環境的型樣(patterns)，然後藉由型樣比對(pattern matching)的方式去偵測出可能對系統或網路造成危害的攻擊事件。異常行為偵測則是利用統計上的特性，將使用者在電腦上的操作習慣或者是將整個內部網路正常的資料流動趨勢作一個輪廓描述(profile)並記錄起來，若線上行為或網路流量資訊與此 profile 相差過大，則視為異常現象,也就是攻擊事件。誤用偵測與異常行為偵測各有優缺點，前者對已知攻擊幾可達 100%偵測率與 0%誤報率，但完全無法偵測未知或新型攻擊；後者在已知攻擊的偵測效果上略遜於前者，但對未知或新型攻擊則可以達到某種程度的偵測效果。

對於異常行為偵測系統而言，最困難的是如何設計所謂的正常行為輪廓描述(normal profile)，而這部份通常與所選取的特徵有密切關係。文獻上，屬於異常行為型的網路入侵偵測系統之研究有很多，不過大多集中在系統架構與偵測方法的設計，針對特徵選取探討的論文相對少很多;目前我們知道的有[1, 2, 3, 4, 5, 6, 7, 8]。

[7]是利用支向機(Support Vector Machine)的技術排序 KDD CUP99 [21]中 41 個特徵的重要性，[3]則大致與[7]相同，除了 SVM 外，也使用神經網路(Neural Network)。[4]討論基因演算法分別結合 Relief tree 與 Naïve Bayesian network 的特徵選取效果，也是使用 KDD CUP99 做為實驗的對象。[5, 6]均使用決策樹作為入侵偵測系統的偵測引擎，所採用之特徵選取方式：[6]是來自基因演算法，而[5]僅提及是結合 information gain, gain ratio 與 Gini index；至於實驗的部分，[5]是自行收集，而[6]採用 KDD CUP99 。[2]是用基因演算法結合 RBF (radial basis function)網路進行特徵選取，使用 KDD CUP98 中的 7 個攻擊做為實驗的對象。最後，[1, 8]是採用基因演算法結合 KNN 做特徵選取；[1]是用 KDD CUP99 TCPDUMP 做實驗，[8]則用 1988 DARPA BSM audit data 做實驗。不過，[4]並未提及 KNN 與 GA 內容細節，[8]則是把 audit data 視為文件，用文件分類的觀念:詞頻-反向文件頻率(term frequency – inverse document frequency; tf-idf)來權重特徵。

以上大部份論文的實驗對象都是 KDD CUP99

中的 TCPDUMP 資料集，這存在一個主要問題，它們都是屬於離線(off-line)型態的分析，並無法滿足偵測系統即時處理的需要。KDD CUP99 中所公佈的 41 個特徵大都是取自連線(connection)而非個別封包，也就是要先將封包收集、重組成連線後才能取出這 41 個特徵。而且這 41 個特徵的選法相當複雜且多樣性[1,10]：特徵 0 到 9 是取自封包標頭，特徵 10 到 22 是取自連線的負載(payload)並與網域資訊相關(domain knowledge)，特徵 23 到 31 是考慮流量(traffic-based)並以兩秒為單位所統計的數值，最後特徵 32 到 40 也是考慮流量但改以 100 個連線(connections)為單位所統計的數值。如此複雜的特徵取得方式，使得上述論文的方法，一但要實作成可即時處理的網路入侵偵測系統(NIDS)變的不可行。另外，KDD CUP98/99 的資料集以現今來看也略顯過時，例如 DoS 類攻擊 CUP98 僅有 12 隻，而 CUP99 多了 5 隻，達 17 隻。

本篇論文所採用之特徵全來自封包(packet)，而非連線(connection)，主要以封包標頭欄位為主。如此方能在很短的單位時間內，統計出這些特徵的量並做出即時判斷。本文採用基因演算法結合 KNN，在 KNN 距離計算上直接將特徵個別權重帶入計算，如此當演進結束時，我們就可獲得所有特徵的權重值。本文後面章節內容陳述如下，第二節簡短介紹基因演算法與 KNN 演算法，第三節提出我們的研究架構與實驗方法，第四節呈現並分析實驗數據，結論與未來展望則放在第五節。

2. 背景知識

基因演算法(Genetic Algorithm; GA)藉由染色體中基因的交換與突變，不斷演進，最後找出最佳的染色體。染色體(chromosome)由許多基因(genes)組成，一個基因就代表一個特徵。基因演算法中染色體的適應函數(fitness function)計算，往往擁有著決定性的影響，我們使用 KNN 演算法來設計適應函數。

2.1 基因演算法

Holland 認為[11]，生物的演化主要發生在染色體的基因中，然而每種生物其特徵主要源於該生物親代的基因序列，演化指的是每一代基因所發生的變化情形。所謂適者生存是指這一代的基因排列優於母代的基因排列，而產生比上一代更能適應環境生存的世代。

GA 是強調基因型的轉變，將欲求解問題的參數經過編碼成為基因格式，利用遺傳運算進行演化來找到問題的最佳解。這些遺傳運算是模擬自然界的演化程序，包括有複製(reproduction)、交配(crossover) 與突變(mutation) 等等，基因演算法處理流程可用圖一來描述。群體(population)是由固定數目之染色體(或稱個體 individual)組成，每條染色體由固定長度之基因構成，第 0 代群體之染色體是由亂數產生。

通常一個基因代表一個特徵，一條染色體(or 個體)包含解決一個問題所需的全部特徵。所以一條染色體也代表一個可能解(feasible solution)，GA 就是要藉由演化找出最佳解(optimal solution)。如圖 1 所示。

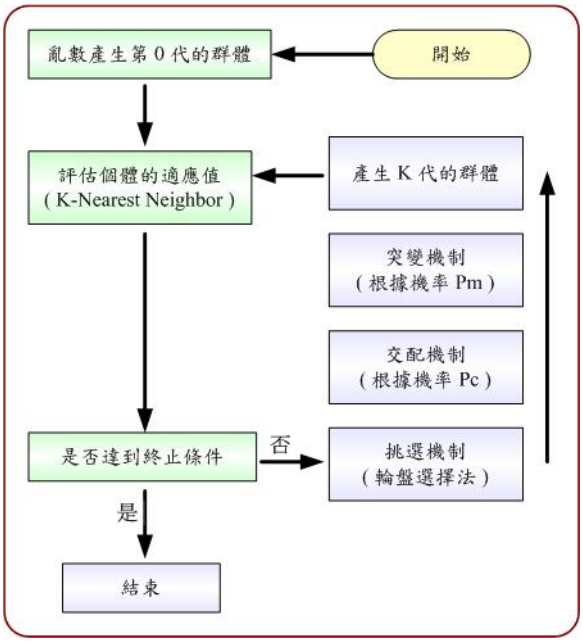

圖 1 基因演算法流程圖

每代之染色體皆透過適應函數計算出適應值，若符合終止條件，演化結束。終止條件包括達到預設的演化代數，或最佳適應值大於一個預設的門檻植。若不符合終止條件，則依照染色體其適應值高低，依照比例挑選染色體，進行交配及突變，如此反覆進行直到符合終止條件。交配及突變各有預設的機率，交配程序可分單點、兩點、多點交配...等。圖 2 所示以單點交配為例，交配點位於染色體中間，交配點後方的基因互換形成兩個子代。突變是指在染色體上，決定一個、兩個或多個基因，利用亂數的機制進行突變，進而達到改變適應值收斂的現象。

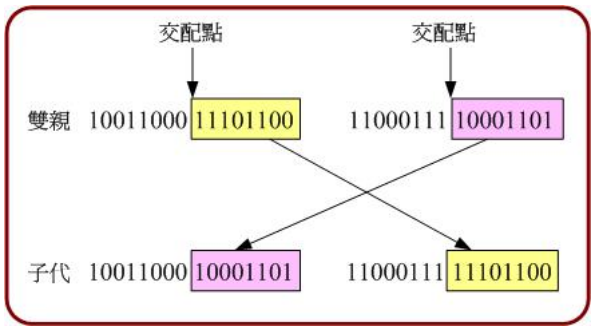

圖 2　單點交配示意圖

適應性函數的設計，對基因演算法擁有決定性的影響，因為有效的適應性函數可以降低因繁衍的代數，進而提升系統效能。本論文利用 KNN 來設計適應性函數，所有的特徵主要來自封包的標頭欄位。

2.2 KNN (K-Nearest Neighbor)

K-Nearest Neighbor [12] 是屬於案例學習(Instance-based learning)中的一種方法，KNN 將每個案例(instance)表示為 n 維空間上的一個點。假設空間中存在若干已分類好的樣本案例(sampling instances)，面對一個未知類別的案例『x』時，KNN 演算法會找出樣本案例中與『x』最接近的 k 個案例，其中多數者之類別即判斷為『x』之類別。圖 3 以 k = 3 為例，其中未知類別之案例將被判斷為 A 類，因其最接近的 3 個點中有 2 個點(多數)是屬於 A 類。

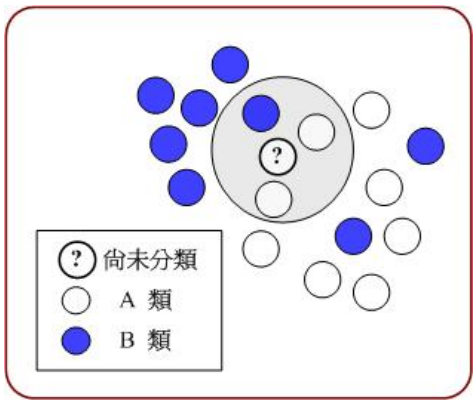

圖 3　KNN 示意圖，k=3

所謂最接近的 k 個點，一般指的是歐幾里得距離最小的 k 個點。兩點 $X=(x_1, x_2, ..., x_n)$ 與 $Y=(y_1, y_2, ..., y_n)$ 的歐幾里得距離 *dist(X, Y)*算法如下：

$$dist(X,Y)=\sqrt{(x_1-y_1)^2+(x_2-y_2)^2...+(x_n-y_n)^2} \quad (1)$$

3. 本文所提之方法

我們依基因演算法架構，分三部份介紹本文所提的方法。首先說明所選之特徵，接著說明本文適應函數之設計，最後說明交配與突變的方式。

3.1 特徵

本文針對 DoS/DDoS 攻擊所產生之異常流量提出一個快速偵測的方法，故特徵主要源自封包標頭欄位，所考慮的標頭包括 IP, TCP, UDP, ICMP, ARP 與 IGMP，附表一詳列出所有被考慮的特徵，共 26 個。表中有不少特徵是需要同時檢查多個欄位，例如『S.IP + ACK Flag + ACK number』與『Port (20)和 length(>1400)』。前者主要用於檢查，在相同的來源端 IP 位置下，標記 ACK Flag = 1，但標頭內沒有 ACK number；或擁有 ACK number 卻尚未標記 ACK Flag = 1 的封包個數。由於大多數的阻斷服務攻擊程式，只有標記 ACK Flag=1 或只出現 ACK number，藉由這種方式可以統計出這些異常封包的數量。至於『Port (20) + length(>1400)』這特徵用來區分出正常的 FTP 下載與 DoS 攻擊，在我們觀察眾多的攻擊程式之中，發現其實一般 DoS/DDoS 攻擊的封包數量很多，可是負載不大，然而正常的 FTP 下載，傳送資料是使用 TCP Port 20 同時會填滿整個負載。

所有實驗的案例(instances)都是統計單位時間(本文設兩秒)網路流量中這 26 個特徵的封包數做為其特徵值。接著我們對案例中的特徵值做正規化如下，如此可以避免單一特徵值過大所造成的影響。經過正規化處理後，所有案例中的特徵值都介於 0 至 1 之間。本文提出的正規化方式，如公式(2)。

$$\text{Normalize } f_i = \frac{1-e^{-kf_i}}{1+e^{-kf_i}} \quad (2)$$

其中 f_i 是指特徵 i，e 為歐拉常數(Euler number，e≒2.7182818284)，k 為介於 0 到 1 的常數，其值在訓練階段由專家決定，隨著特徵的不同而有所變動。上述這種正規化方式，更適用於即時性的資料。

3.2 基因、染色體與適應函數

雖然附表一列出偵測 DoS/DDoS 攻擊的可能特徵，但他們的權重應該是有所不同的。本文之目的是利用基因演算訓練這 26 特徵的權重，也就是藉 GA 演化，賦予每個特徵不同的權重，使反應不同的重要性。演化結束後，再移除權重較輕的特徵，使特徵數目降低以增進 NIDS 處理時間而又可以維持準確率。

令權重向量 $W = [w_1, w_2, \ldots, w_n]$（本文中 n = 26），一個權重向量代表一條染色體，向量中的值 w_i, $1 \le i \le n$, 代表基因，且 $0 \le w_i \le 1$. 假設群體(population)是由 k 條染色體(本文中 k = 30)構成，第 0 代之 k 條染色體內的基因值皆由亂數產生。

在適應性函數的計算上，我們先改變歐幾里得距離公式，使之可以反應不同特徵的個別重要性。假設空間有兩個案例 $X = (x_1, x_2, ..., x_n)$ 與 $Y = (y_1, y_2, ..., y_n)$，本文中它們的距離算法如下：

$$dist(X,Y) = \sqrt{w_1^2(x_1 - y_1)^2 + w_2^2(x_2 - y_2)^2 ... + w_n^2(x_n - y_n)^2} \quad (3)$$

我們選擇利用 w_i^2 來加權特徵 i 的重要性。接著透過 KNN，就可以利用訓練資料集中已標示類別的案例來計算每條染色體的適應函數值(fitness value)，我們的適應函數定義如下：

$$\text{Fitnesss} = \frac{Total - FP - FN}{Total} \quad (4)$$

其中 Total 為訓練案例的總數，FP 為將正常誤判為攻擊(假警報)的個數，FN 為將攻擊誤判為正常(漏報)的個數。圖 4 是本文實驗中的一個例子，圖中顯示第 0 代 3 個染色體的適應函數值及其基因的組成，第 0 代染色體內的基因值是由亂數產生。

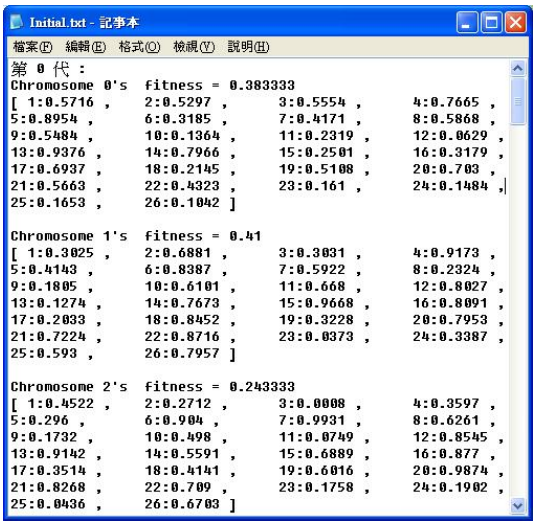

圖 4 第 0 代染色體及其適應函數值

3.3 挑選、交配與突變

在挑選的機制中，我們採輪盤式法則，也就是適應值越大之染色體，越有機會被挑選衍生下一代。如圖五所示，假設有 6 條染色體，其適應值分別為 fit_1，fit_2，....，fit_6，若某條染色體之適應值 fit_j, $1 \le j \le 6$，被挑選的機率是 $fit_j / \sum_{i=1} fit_i$

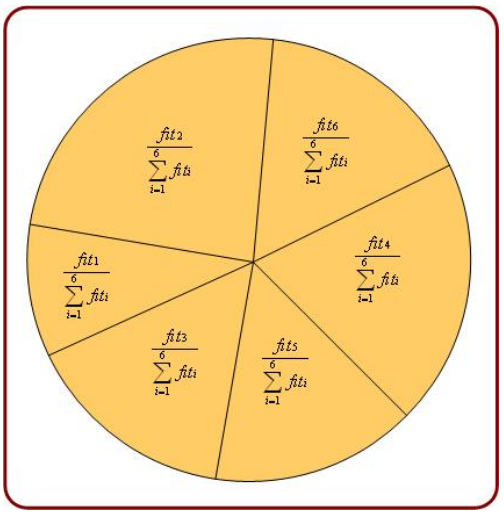

圖 5 輪盤式挑選法

在交配方面，我們使用雙點交配，如圖 6 所示。四個交配點分別位於親代兩個染色體『000111』與『111100』前後；兩個親代交配點中間的基因互換，即形成兩個新的子代。突變上，我們隨機挑選一個染色體，隨機更換 2~5 個基因值。若突變後的染色體適應函數值大於突變前的體適應函數值，則取代之，否則不做任何改變。

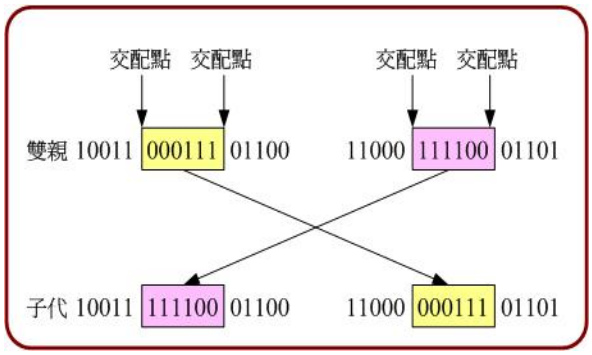

圖 6　雙點交配

4.　實驗數據與分析

實驗使用的平台作業系統為 Windows XP Professional，程式語言為 C++，開發工具為 Microsoft Visual Studio 2005，封包蒐集透過 WinpCap，而網路拓樸連接情形如圖六。壹台電腦負責執行攻擊程式，多台電腦負責製造正常封包流量，包括 FTP 下載、WEB 瀏覽與下載、Telnet、MSN、E-mail 等正常網路應用。所有的攻擊程式下載自 VX Heavens (http://vx.netlux.org/)，這個網站是由著名的防毒軟體公司 Kaspersky 所維護，共收錄 6 萬多個惡意程式且持續增加。屬於 DoS 類別且運作於 Windows 平台的攻擊程式共 207 隻，但因部分需配合特定 DLL 檔方可執行，直接下載就可在我們的電腦上執行的程式共 65 隻，本節實驗之 DoS 攻擊程式共 65 隻，實驗使用之攻擊程式及案例 (instances) 詳列於 http://163.25.149.115/NCS2007/。

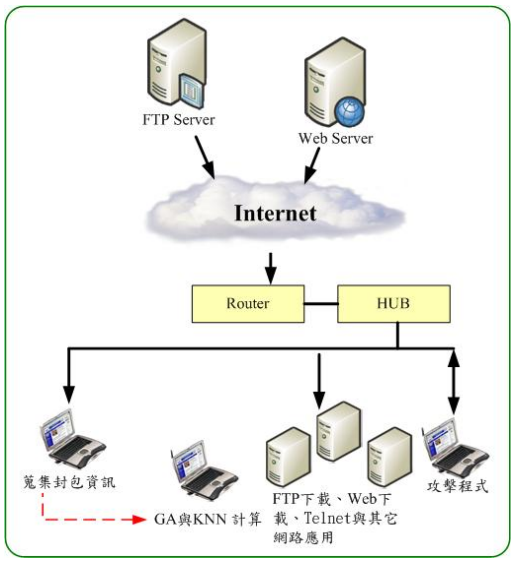

圖 7 網路拓樸

我們共有三類案例，分別為樣本案例(sampling instances),訓練案例(training instances)與測試案例(testing instances). 每個案例是兩秒封包流量中 26 個特徵(附表一)的統計值，並經 3.1 節公式(2)所述的正規化方式處理過。

樣本案例共有 200 筆，用來供 KNN 演算法分類使用，其中標示為攻擊 100 筆，正常 100 筆。訓練案例共 900 筆，用來供給基因演算法訓練特徵權重使用，其中標示為攻擊 300 筆，正常 600 筆。

樣本案例 (sampling instances) 與訓練案例 (training instances) 中的攻擊案例皆源自相同的 20 個攻擊程式，包括正常流量與 DoS 攻擊程式，基因演算法訓練過程主要目的是為附表一每個特徵標示出權重，一開始(i.e., 第 0 代)隨機產生 30 個染色體，每個染色體代表一個權重向量 $W = [w_1, w_2, \ldots, w_n]$, $n=26$ 且 $0 \le w_i \le 1$, $i = 1,2, \ldots, 26$。每個訓練案例會計算與所有樣本案例的距離用以做 KNN 歸類，而所用的距離算法如 3.2 節公式(3)。對每個染色體而言，所有的訓練案例跑一遍 KNN 就可以得到這個染色體的適應函數值 (fitness value)，計算方法如 3.2 節公式(4)。

圖 8 為演化過程最佳染色體的適應函數值變化情形，本實驗群體之個體數(也就是染色體個數)為 30，每次進化 100 代，重複執行 50 次。

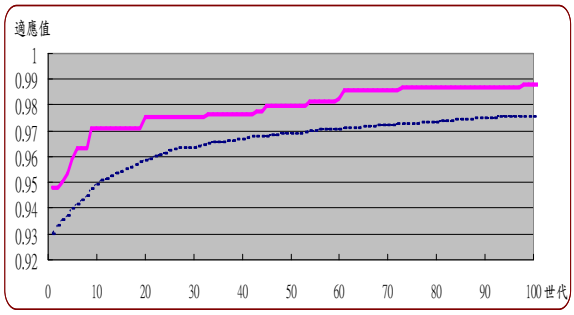

圖 8　適應值的進化改變

圖中實線代表全部 50 次重複實驗中每代所有群體出現之最大適應函數值(best of best)，而虛線代表 50 次重複實驗中每代群體中最佳染色體的平均適應函數值(average of best)。最大值之適應函數值為 0.987778，此時特徵向量如下：

[1:0.9222 , 2:0.9597 , 3:0.2003 , 4:0.2371 ,
5:0.1298 , 6:0.9199 , 7:0.9408 , 8:0.04 ,
9:0.7707 , 10:0.827 , 11:0.0125 , 12:0.001 ,
13:0.1123 , 14:0.1968 , 15:0.0293 , 16:0.0071 ,
17:0.0863 , 18:0.6683 , 19:0.6389 , 20:0.9958

21:0.8817 , 22:0.6473 , 23:0.6252 , 24:0.0756 , 25:0.0561 , 26:0.3785]

我們將這 26 個特徵，根據權重由大到小排列，如附表二。為了能更深入評估訓練出的權重是否真具代表性，同時也為驗證是否具偵測未知攻擊的能力，我們另產生兩種不同的型態的測試案例(testing instances),TestA 與 TestB，以供評估。TestA 是為測試已知攻擊，所以攻擊案例源自與樣本及訓練相同的 20 隻攻擊程式。至於 TestB 則設計為測試未知攻擊，所以攻擊案例源自與樣本及訓練完全不同的 45 隻攻擊程式。TestA 與 TestB 中的攻擊案例包含單一與多隻 DoS 程式混合攻擊。用上面 26 特徵之權重向量對 TestA 與 TestB 做 kNN 做分類，所得到之數據如表一。整體準確率(overall accuracy)的計算方式為(TP+TN)/(TP+FP+TN+FN)。

表一 26 個特徵的偵測效果

	TestA	TestB
TP	99.59%	55.17%
FP	0.23%	0.50%
TN	99.77%	99.50%
FN	0.41%	44.83%
Overall Accuracy	99.72%	84.72%

由表一得知在已知攻擊中，偵測率為 99.59%、誤報率為 0.41%、整體準確率為 99.72%。在未知攻擊中，偵測率為 55.17%、誤報率為 44.83%、整體準確率為 84.72%。

接著，我們根據全部 26 個特徵的權重，從權重值最輕的開始，逐一拿掉特徵，再從新訓練特徵向量並測試，對 TestA(已知攻擊)所得到的結果顯示於圖九，對 TestB(未知攻擊)所得到的結果顯示於圖十。

圖九中最好的整體準確率發生在特徵數為 14，其值達到 99.83%，而若採用全部 26 個特徵，整體準確率略小，為 99.72%，圖九的詳細數據列於附表三。圖十中最好的整體準確率發生在特徵數為 24，其值達到 88.11%，而若採用全部 26 個特徵，整體準確率略小，為 84.72%，圖十的詳細數據列於附表四。

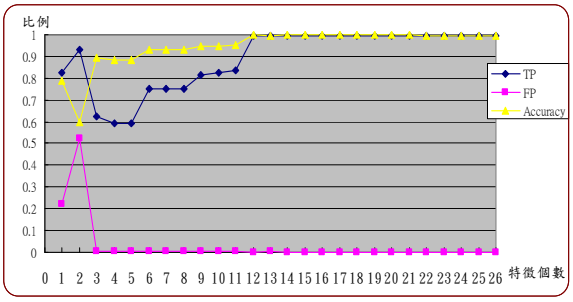

圖 9 Test A 的特徵個數與偵測效果

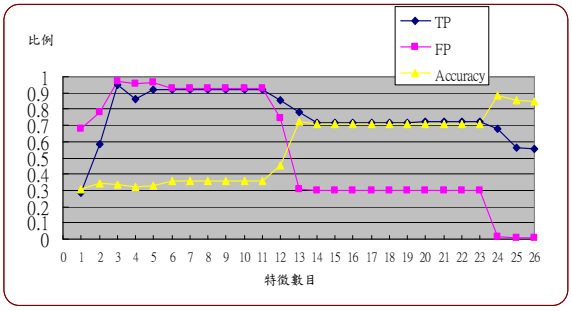

圖 10 Test B 的特徵個數與偵測效果

5. 結論

本文提出一個有效率的方式用來權重特徵，並分析特徵數目與偵測效果之間的關係。這個方法並且可以很容易的實做成一個網路入侵偵測系統，用來即時分析網路流量是否正遭受 DoS 攻擊。我們的結論是選取適量且有效的特徵，不僅可以減少 NIDS 處理時間，並且可以增進偵測效果。

誌謝:感謝行政院國科會專題研究計畫之補助(NSC 95-2221-E-130-003)，使本論文得以順利完成。

參考文獻

[1] Melanie J. Middlemiss and Grant Dick, “Weighted Feature Extraction using a Genetic Algorithm for Intrusion Detection”, Evolutionary Computation, 2003. CEC '03,8-12 Dec. 2003 pp. 1669- 1675, Vol.3.

[2] Alexander hofman, Timo Horeis, and Bemhard Sick, ”Feature Selection for Intrusion Detection: An Evolutionary Wrapper Approach,” IEEE Proc., pp. 1563-1568, 2004.

[3] A H. Sung, S. Mukkamala.” Identifying Important Features for Intrusion Detection Using Support Vector Machines and Neural Networks.” In Proceedings of the 2003 International Symposium on Applications and

the Internet Technology, IEEE Computer Society Press, pp.209-216.

[4] Chi Hoon Lee, Jin Wook Chung, Sung Woo Shin: “Network Intrusion Detection Through Genetic Feature Selection,” Proc. Of the Seventh ACIS International Conference on Software Engineering, Artifical Intelligence, Networking, and Parallel/Distributed Computind (SNPD), pp. 109-114, 2006.

[5] T. Abbes, A. Bouhoula, and M. Rusinowitch, ”Protocol analysis in intrusion detection using decision tree” Proceeding of the International Conference on Information Technology: Coding and Computing (ITCC), pp. 404-409, 2004.

[6] Gary Stein, Bing Chen, Annie S. Wu, and Kien A. Hua, “Decision tree classifier for network intrusion detection with GA-based feature selection,” ACM Southeast Regional Conference (2) 2005: 136-141

[7] Srinivas Mukkamala and Andrew H. Sung, ”Feature ranking and Selection for Intrusion Detection Using Support Vector Machines,” Proceedings of International Conference on Information and Knowledge Engineering, pp. 503-509, 2002.

[8] Yihua Liao and V. Rao Vemuri, “Use of K-Nearest Neighbor Classifier for Intrusion detection,” Journal of Computer & Security, Vol. 21, No. 5, pp. 439-448, 2002.

[9] DARPA 1999 Intrusion Detection Evaluation, http://www.ll.mit.edu/IST/ideval/data/data_index.html

[10] The UCI KDD Archive,

http://kdd.ics.uci.edu/databases/kddcup99/kddcup.names

[11] John H. Holland, Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence, The MIT Press, 1992.

[12] Fabrizio S., “Machine Learning in Automated Text Categorization”, ACM Computing Surveys, 2002.

附表一 特徵表列

No	protocol	feature	no	protocol	feature
1	IP	Source Address	15	TCP	S.IP + ACK Flag + ACK number
2	IP	Destination Address	16	TCP	Length
3	IP	Header length	17	TCP	Port (20) + length(>1400)
4	IP	Total length>1400 || <40 &&TTL = = 64	18	TCP	S.port number
5	ICMP	Length	19	TCP	D.port number
6	ICMP	Type	20	TCP	SYN number
7	IGMP	Length	21	TCP	URG Flag + URG data
8	IGMP	Type	22	TCP	ACK Flag + ACK number
9	ARP	Length	23	UDP	S.port number
10	ARP	S.IP + ARP count	24	UDP	D.port number
11	TCP	S.IP + S.port number	25	UDP	Length
12	TCP	S.IP + D.port number	26		封包總數
13	TCP	S.IP + SYN count			
14	TCP	S.IP + URG Flag + URG data			

附表二 特徵權重與排序

rank	feature	weight	rank	feature	weight
1	UDP length	0.9958	15	S.IP_Dport	0.2003
2	S.IP_Sport	0.9597	16	TCP URG_err	0.1968
3	S.IP_ARP	0.9408	17	S.IP_URG_err	0.1298
4	IP S_Address	0.9222	18	TCP SYN	0.1123
5	S.IP_ACK_err	0.9199	19	TCP len_port 20	0.0863
6	ICMP type_err	0.8817	20	IGMP big_len	0.0756
7	IP len_TTL 64	0.827	21	ARP size err	0.0561
8	IP Hdr_len_err	0.7707	22	IP D_Address	0.04
9	UDP Sport	0.6683	23	TCP ACK_err	0.0293
10	ICMP big_len	0.6473	24	TCP Sport	0.0125
11	UDP Dport	0.6389	25	TCP length	0.0071
12	IGMP type_err	0.6252	26	TCP Dport	0.001
13	封包總數	0.3785			
14	S.IP_SYN	0.2371			

附表三 Test A 的特徵個數與偵測效果

特徵數目	1	2	3	4	5	6	7	8	9	10	11	12	13
TP	0.8269	0.9328	0.6253	0.5906	0.5906	0.7536	0.7536	0.7536	0.8167	0.8228	0.8371	0.9939	0.9939
FP	0.2231	0.5248	0.0031	0.0038	0.0038	0.0038	0.0038	0.0038	0.0038	0.0038	0.0038	0.0008	0.0046
TN	0.7769	0.4752	0.9969	0.9962	0.9962	0.9962	0.9962	0.9962	0.9962	0.9962	0.9962	0.9992	0.9954
FN	0.1731	0.0672	0.3747	0.4094	0.4094	0.2464	0.2464	0.2464	0.1833	0.1772	0.1629	0.0061	0.0061
Acc.	0.7906	0.6000	0.8956	0.8856	0.8856	0.9300	0.9300	0.9300	0.9472	0.9489	0.9528	0.9978	0.9950

附表三(續)

特徵數目	14	15	16	17	18	19	20	21	22	23	24	25	26
TP	0.9959	0.9959	0.9959	0.9959	0.9959	0.9959	0.9959	0.9959	0.9959	0.9959	0.9959	0.9959	0.9959
FP	0.0008	0.0008	0.0008	0.0008	0.0015	0.0015	0.0015	0.0015	0.0023	0.0023	0.0023	0.0023	0.0023
TN	0.9992	0.9992	0.9992	0.9992	0.9985	0.9985	0.9985	0.9985	0.9977	0.9977	0.9977	0.9977	0.9977
FN	0.0041	0.0041	0.0041	0.0041	0.0041	0.0041	0.0041	0.0041	0.0041	0.0041	0.0041	0.0041	0.0041
Acc	0.9983	0.9983	0.9983	0.9983	0.9978	0.9978	0.9978	0.9978	0.9972	0.9972	0.9972	0.9972	0.9972

附表四 Test B 的特徵個數與偵測效果

特徵數目	1	2	3	4	5	6	7	8	9	10	11	12	13
TP	0.2816	0.5850	0.9466	0.8616	0.9200	0.9166	0.9166	0.9166	0.9166	0.9166	0.9166	0.8550	0.7783
FP	0.6816	0.7833	0.9708	0.9541	0.9641	0.9266	0.9266	0.9266	0.9266	0.9266	0.9266	0.7458	0.3033
TN	0.3183	0.2166	0.0291	0.0458	0.0358	0.0733	0.0733	0.0733	0.0733	0.0733	0.0733	0.2541	0.6966
FN	0.7183	0.4150	0.0533	0.1383	0.0800	0.0833	0.0833	0.0833	0.0833	0.0833	0.0833	0.1450	0.2216
Acc.	0.3061	0.3394	0.3350	0.3178	0.3306	0.3544	0.3544	0.3544	0.3544	0.3544	0.3544	0.4544	0.7239

附表四(續)

特徵數目	14	15	16	17	18	19	20	21	22	23	24	25	26
TP	0.7150	0.7150	0.7150	0.7150	0.7183	0.7183	0.7217	0.7217	0.7217	0.7200	0.6783	0.5617	0.5517
FP	0.2967	0.2967	0.2967	0.2967	0.2967	0.2967	0.2967	0.2967	0.2967	0.2967	0.0175	0.0050	0.0050
TN	0.7033	0.7033	0.7033	0.7033	0.7033	0.7033	0.7033	0.7033	0.7033	0.7033	0.9825	0.9950	0.9950
FN	0.2850	0.2850	0.2850	0.2850	0.2817	0.2817	0.2783	0.2783	0.2783	0.2800	0.3217	0.4383	0.4483
Acc.	0.7072	0.7072	0.7072	0.7072	0.7083	0.7083	0.7094	0.7094	0.7094	0.7089	0.8811	0.8506	0.8472