

# 使用 MRU、AR-PST、SMA 與 WMA 建構網路節點服務需求之樣式組成預測模型與數量預測模型

湯政仁, 李宜亭, 戴妙如, 蔡宗育  
大同大學通訊工程所

ctang@ttu.edu.tw, d9810001, d9610002, g9710002@ms.ttu.edu.tw

林政廷, 宋鴻鈞  
台灣工業技術研究院

Tim\_lin@itri.org.tw, DavidSung@itri.org.tw

**摘要?** 網路總流量為週期性時間序列的重複樣式, 因為網路服務的使用習慣與行為。使用者的多寡將決定網路服務的需求量, 進而對網路上的資料流量造成影響, 而網路資料流量關係著電腦機房的能源消耗。為了維持穩定的網路服務品質, 電腦機房採用高可用性系統。在此叢集系統架構環境下會提高機房所產生的熱能, 亦造成冷卻系統的耗能問題。因此若能預先得知下個單位時間裡必需提供多少網路服務, 則可減少機房中機器的工作數量, 進而降低機房熱能的生成及冷卻系統消耗之能源。為找到真實需要的機器數量, 重點在於預測下個時間週期所可能發生的封包類型序列。本文除傳統的數量預測模型外, 提出了網路服務需求樣式組成預測模型, 其預測準確率能到達 90%。  
**關鍵詞:** 網路流量樣式、高可用性系統、網路服務叢集、冷卻系統、需求樣式預測

集系統機房內之冷卻以及相隨之產生之能源消耗問題亦成為近年備受矚目的焦點; 而叢集系統機房的溫度調控系統設計成為機房基礎建設中節能是否成功的重要環節。

網路服務業者為了保證客戶需求任何時刻皆能得到滿足, 因此經驗上會以服務所需的最大需求容量作為基本的運轉數量, 亦即同時啟動所有計算單元進入『服務』狀態, 而依據該計算單元的平均持續工作時間以及平均修復時間, 額外的備援計算單元也會啟動在『待機』狀態。然而, 網路需求並不是隨時都在滿載的情況; 實際的需求可能是最大需求量的極低比例; 亦即網路服務業者的能源使用消耗, 以及因此造成的用電費用都有極大比例是無端浪費的。

## 一、緒論

### (一) 研究動機

高可用性系統的設計透過對冗餘的硬體和軟體進行組合, 目前常見實現的方式是電腦叢集, 通過一組分散集成的電腦軟體或硬體連接起來高度緊密地協調運作完成計算工作。它們可被視為一台電腦。叢集系統中的單台電腦通常稱為節點, 通常透過區域網路連接。叢集電腦通常用來改進單台電腦的計算速度和可靠性。一般情況下叢集電腦之性能價格比較單台電腦, 如工作站或超級電腦要高得多。此類系統將大量的計算裝置整合於室內, 由於計算裝置的溫度會提高設備的失效比率, 冷卻系統在高可用性系統下成為必需設備, 因此在節能減碳等環保概念與訴求逐漸受到重視的今日, 叢

### (二) 研究目的

對於網路服務提供者而言, 網路需求負載將會受到客戶之群體類別與其進行網路活動時間的影響。因此推測, 網路服務提供者在訂定契約容量時, 若將網路需求樣式加入考慮, 進行需量預測; 並依據能夠被客戶接受的服務品質, 在網路需求尖峰以及需求離峰時, 做適當的機器啟動或停止之調整, 則將可以預期此技術對於能源的節省、電費負擔的降低, 以及對於因為以滿載情況為前提, 而啟動具備大量冗餘備援機組的叢集系統, 衍生冷卻系統運轉需求過量所造成的能源浪費情況將有所改善。本文因此提出一個新的技術, 進行網路節點服務需求預測模型之建構。首先採用傳統的 Simple Moving Average, SMA 與 Weighted Moving

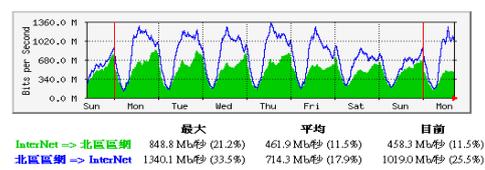
Average, WMA 建構網路節點服務需求量預測模型，提供未來下一個時間周期的封包數量預測，其結果能夠提供連線頻寬需求的預測。然而，兩者僅能針對某個單一性質數量或總量進行預估，甚少考慮封包樣式組成順序以及數量的連動關係；而僅由數量的變化產生應對策略，在網路需求的預測上，是不足夠的。有鑒於此，本文將提出使用 Most Recently Used, MRU 以及 Autoregressive Prediction Suffix Tree, AR-PST 演算法進行網路服務需求樣式組成預測模型的建構，設法對封包樣式的組成順序以及連線總數進行預測，以達到網路節點服務需求預測的目的。PST 之運算複雜度為  $O(n)$ ，能夠有效的提高計算效率。而根據對網路服務需求樣式序列的觀察，發現其中重複片段占了序列之極大比例，因此在進型網路服務需求樣式預測時，首先將運用簡單的 MRU 演算法進行樣式預測。

### (三) 本文貢獻

對於網路服務提供者而言，客戶的來源群體以及他們進行網路活動的時間是網路需求負載影響因素。若能將對網路需求之樣式加入考慮，進行需求預測，而得到最佳契約容量，並在網路需求尖峰以及需求離峰根據可被客戶接受的服務品質做適當的機器啟動或停止之調整，則電費的負擔將有可觀降低；而對於現今電腦機房內，因為以最大網路需求為基礎，而啟動具備大量冗餘備援機組的叢集，衍生出過量冷卻系統運轉需求所造成的能源浪費情況將會有極大的改善。

然而在了解網路需求樣式之前，透過對區域網路的流量觀察[1]，如圖一，可以發現網路總流量曲線為週期性時間序列的重複樣式，網路節點的需求量為總流量中的子集合，而目前並無單獨針對網路節點服務需求進行預測的模型；因此，本論文將在評估網路需求之後建構預測模型。

每週圖表 (30 分鐘 平均)



每月圖表 (2 小時 平均)

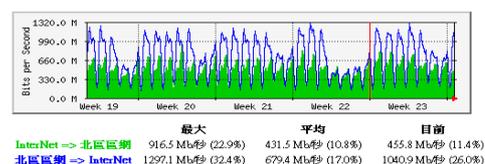


圖 1 2009/6/8 網路總流量曲線

首先將解析出網路上可被服務的需求；接著選用現有序列數值預測演算法，MRU、AR-PST、SMA 與 WMA 進行網路節點服務需求的預測。使用 MRU、AR-PST 進行網路服務需求樣式組成的預測模型的建構，並使用 SMA 以及 WMA 建構網路服務需求量預測模型。

## 二、背景知識

### (一) 網路流量樣式

決定叢集系統的合適運轉元件數量的關鍵在於網路需求。網路服務需求會依據使用者群集來源及其從事的網路活動、使用的網路服務而變化，網路流量的變化曲線能夠反應出使用者的使用時間、使用習慣、使用需求等網路服務使用特性，圖 1 為北區區網週流量以及月流量統計圖，由於北區區網下轄許多機關學校，工時固定，因此可顯示與電力需求曲線類似的效果，於工作日的流量樣式都十分接近，圖中顯示訊息量在每日上班上課時段與下班下課時段有極大的差異，另外在工作日與週末的訊息量差異亦相當顯著。由圖 1 可以觀察得到網路總流量曲線為週期性時間序列(Periodic Time Series)的重複樣式，由於對於網路服務提供業者而言，客戶的來源群體以及他們進行網路活動的時間是網路需求負載的重要影響因素。經由對網路上可被服務需求的解析，對網路需求樣

式進行統計與分析，並選用合適的序列數值預測演算法以建立網路服務需求之預測模型，進而能夠完成網路服務之需量預測。

## (二) TCP 連線數量與封包類型

為了能夠預知下一個時間單位所需的機器數量，首先必須能夠預測下一個時間單位將發生的連線數量，而傳統的網路服務需求數量預測結果僅能夠提供連線頻寬需求的預測，而對下一個時間單位之連線數量的預測無法提供協助，原因在於，傳統的網路服務需求數量預測著重於被提出的資料傳收服務需求，而當資料傳收服務需求可以被提出，表示一定有機器可以承接，也表示網路的連線已經完成了。為設法在連線發生以前，預估所需機器的數量。而透過 TCP 封包的類型分辨，可以針對特定封包類型的預測進而進行連線發生的預估。因此如能預測下一個時間週期所可能發生的網路服務需求類型序列，進而進行連線數量的預估，則能決定真正所需要使用的機器數量。舉例來說，假設有兩種情況：情況 A 為預估下一分鐘的網路服務需求總數量為 8,000,000 個，其中有 400,000 個連線需求網路服務需求；情況 B 為預估下一分鐘的網路服務需求總數量為 4,000,000 個網路服務需求，其中有 800,000 個連線需求網路服務需求。情況 A 網路服務需求總數量雖然較多，但是實際上所需的機器數量是情況 B 為多。除此之外，本文也考慮到斷線的需求會使原本的網路連線終止，而能空出新的連線空間，無須新增機器進行處理。因此，傳統連線數量預測此種不考慮狀態(Stateless)的預測方法，並不適用於有狀態相依特性的網路連線之上(如 TCP 連線)。

TCP 協定服務過程的動作大致可以分成三種類型，連線建立、資料傳收以及連線釋出。而實際上關係到連線數量的服務類型為連線建立與連線釋出，原因是當可以進行資料傳送這個動作時，表示連線已建立完成，因此在作連

線數量的預測時，這個動作所對應到的封包類型，資料傳送與資料回覆，對連線數量的預測不具助益，因此在進行封包樣式組成預測時，本文除直接考慮所有封包類型進行預測的狀況外，另將進行在去除資料傳送與資料回覆兩種封包的情況下再作預測，並比較兩種情況下之預測結果。

## (三) 預測演算法

### 1 Most Recently Used, MRU

MRU 原為快取記憶體所設計之演算法，原理為依據資料最近被存取的頻率高低來決定在快取記憶體中的存放順序。

由於觀察到所擷取的封包序列出現大量的重複片段[2]，故使用「Most Recently Used」建構網路服務需求樣式組成預測模型，預測下一個網路服務需求類型，預測方式為假設下一個需求會與目前最近收到的需求相同。

### 2 Autoregressive Prediction Suffix Tree, AR-PST)

自我回歸(Autoregressive, AR)模型是在統計學以及訊號處理上，一種常用的隨機過程類型，常用以模擬或預測各種不同的自然現象。其原理為利用時間序列裡每一筆資料與緊接其前的資料進行回歸計算，找出其相關關係。

本文中所建立的網際網路服務需求樣式組成預測模型，運用自我回歸模型之原理，將預測結果經由遞迴方式處理，不斷更新歷史資料，使得在預測下一個網路服務需求類型時，能夠經由歷史資料內統計得到之最新資訊，預測各種網路服務需求類型出現後，其下一個出現機率最大的網路服務需求類型。

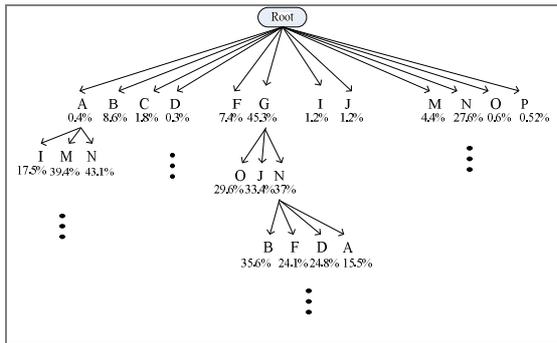


圖 2.1 統計後建立的 Prediction Suffix Tree

Suffix Tree，是以 Tree 做為資料結構，運算複雜度是目前最低的僅需  $O(n)$  就能完成資料結構的建立。以運算複雜度做評估，Suffix Tree 是在線性時間內完成資料結構的建立，本文的 Suffix Tree 應可稱為 Prediction Suffix Tree，在每個 edge 上有下一節點的發生機率，每一個分支實際上為一個 Markov Chain 如圖 2.1 所示。

### 3 移動平均(Moving Average, MA)

在統計學的領域裡，移動平均，是一種分析資料集合的方式，經由計算原始資料中不同子集合的平均，可產生一系列的平均值。移動平均在運算時，會固定子集合的長度以計算平均，每次運算完之後，將會將剔除這次子集合當中，最舊的一筆資料，並加入這次子集合之後的下一筆資料，作為下一次的平均運算。由於在資料序列上，觀察每次取平均的子集合範圍，如同在移動一般如圖 2.2 所示，故此稱之為移動平均(Moving Average, MA)。

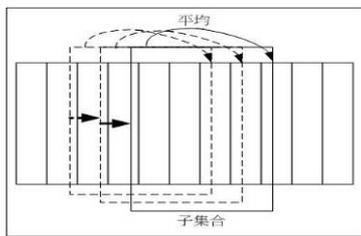


圖 2.2 移動平均的運算方式

簡單移動平均(Simple Moving Average, SMA)，是取預測值之前  $n$  個數值之算術平均。設前  $n$  每日之網路服務需求數量為  $P_1$  至  $P_n$ ，則

方程式為：

$$SMA = \frac{P_1 + P_2 + \dots + P_n}{n} \quad (2.1)$$

當計算連續的數值，一個新的數值加入的同時需將一個舊數值剔除，所以無需每次都重新逐個數值加起來，如下式所示：

$$SMA_{t_1} = SMA_{t_0} - \frac{P_1}{n} + \frac{P_{n+1}}{n} \quad (2.2)$$

加權移動平均 (Weighted moving average, WMA) 是指計算平均時，將個別數據乘以不同數值。在本文中，若取  $n$  個數值的 WMA，其權重的定義為，最近期一個數值乘以  $n$ 、次近期的乘以  $n-1$ ，如此類推，直到 1：

$$WMA = \frac{nP_M + (n-1)P_{M-1} + \dots + 2P_{M-n+2} + P_{M-n+1}}{n + (n-1) + \dots + 2 + 1} \quad (2.3)$$

## 三、網路節點服務需求預測模型之建構

### (一) 實驗環境

對於網路的服務需求研究必須先擷取並分類解析網路設備可能得到的封包，本文使用校園的電腦教室之網路環境來進行研究，此教室的使用狀況週一到週五為依照固定課表的安排提供學生使用。

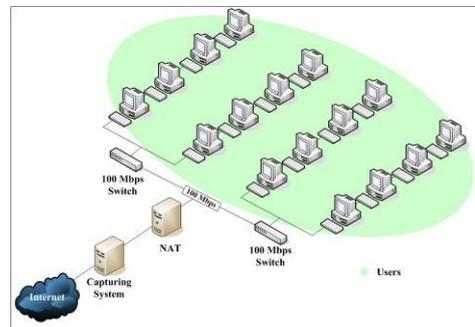


圖 3.1 封包擷取架構圖

### (二) 網路上可被服務需求之解析

為搜集網際網路上的封包入站出站情形，架

設一個網路環境，架構圖如圖 3.1 所示。在這個網路環境內，有一台伺服器被標名為 Capturing System，作為截取網路封包的伺服器，接於網際網路之後，於 NAT 伺服器之前，而接於 NAT 之後者為內部的使用者。在這架構下，Capturing System 可截取到網際網路上傳送進入內部或由內部傳送至網際網路的所有封包。Capturing System 使用 Linux 為系統，截取網路封包的方式是直接使用 Linux 所提供的命令 Tcpdump。在此擷取網路封包所使用的指令如下

```
tcpdump -i br0 -w packets$(date +%Y%m%d).pcap
```

將搜集下來的封包存成 pcap 檔案，目的是將所有封包搜集完整；若非使用參數-w 來將檔案存成 pcap 檔時，會造成封包不完整，例如：封包重傳。參數-w 的功用是將封包直接寫入檔案中，而不進行分析與輸出，因此能夠完整的將網路上入站及出站封包擷取下來。

在擷取下來的封包中，包含了各種協定，其中本文關注的是 TCP 協定封包。對於 TCP 進行下列的分析以得到網路需求與網路回覆的流量之辨識：入站與出站連線需求、入站與出站連線回覆、入站與出站連線確認、入站與出站資料需求、入站與出站資料回覆、入站與出站斷線需求、入站與出站斷線回覆、入站與出站連線重建。

進行分類之前必需先瞭解封包格式，封包格式如圖 3.2 所示。由圖中間那一列可知封包的最外層是 Ethernet Header 接著 IP Header 然後是 TCP Header 最後是 APP Header 之後才是資料部份。本計劃進行 TCP 封包的分类，因此主要是辨識 TCP Header 裡的各欄位資料。

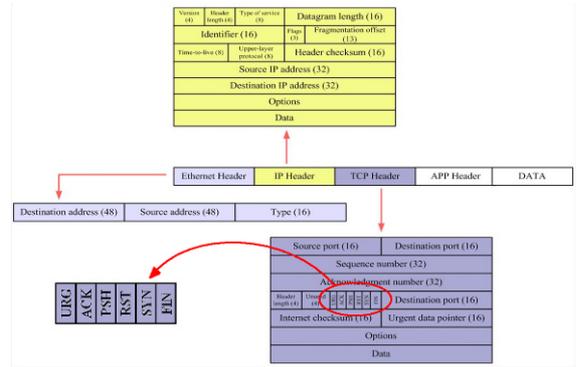


圖 3.2 以 EtherNet 作為實體層協定的封包 TCP Header 格式裡，其中一個欄位為 Flag，有 8 個位元，可標示八種型態，其中較常看到的是 ACK、SYN、RST 及 FIN。因此，可根據 TCP Header 中，Flag 這 8 位元的值辨識該封包屬於哪一類。另外是 Sequence number 的判別，TCP 封包的傳遞，無論是連線的建立或是資料傳遞，都會進行 Synchronize sequence number (SYN)，也就是 SYN 封包，TCP 主要就是根據這個數值來檢查封包是否漏掉或遺失，若發生封包遺漏，就會發出重新傳送的要求，因此 TCP 是一個擁有可靠性的網路協定。Sequence number 的動作在此以 TCP 建立連線所採用的 Three Way Handshaking 說明，連線建立的順序如下圖 4。

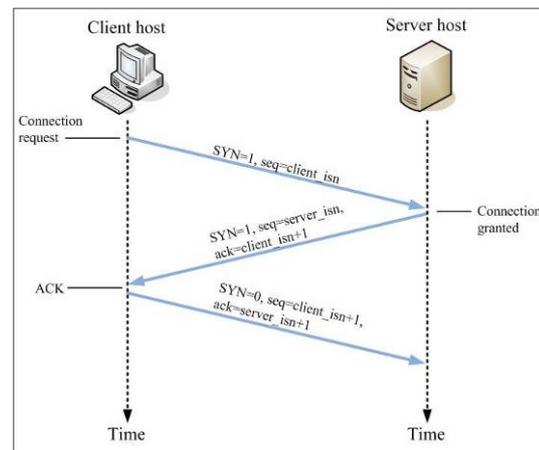


圖 3.3 TCP 的連線建立

由圖 3.3 可以發現，伺服器端回應客戶端所用的 ACK 的 Sequence number 會將收到的 SYN 的值做加 1 的動作，而且當客戶端收到 ACK 後會以

同一個 Sequence number 做回應。因此，根據 Sequence number 的值，可以判別哪些封包是屬於同一組(session)傳遞的資料。根據辨識方式，在所建立的網路環境下進行分類統計，圖 3.4 為以一週的時間為期的分類統計結果：

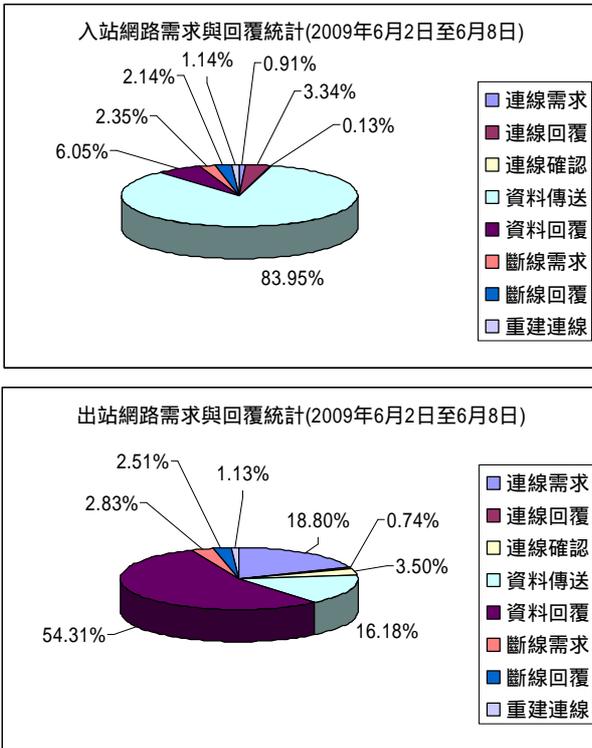


圖 3.4 網路需求與回覆數量統計

(三) 以 MRU 建立網路節點服務需求之樣式組成預測模型

MRU 網路服務需求樣式組成預測模型，建構原理為假設下一個需求會與目前最近收到的需求相同。預測模型將分為兩種結構：(A)以固定的封包數量為預測單位進行預測，使用固定的封包數量為預測單位，每次依此設定的單位進行下一個單位的網路服務需求樣式組成預測。如圖 3.5 為一次預測兩個時間單位的情形。

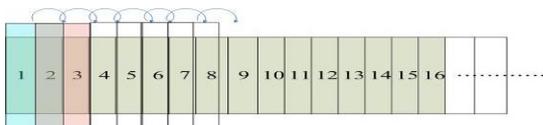


圖 3.5 以封包數量為預測單位

(B)以固定的時間間隔為預測單位進行預測，使用固定的時間間隔為預測單位，每次依此設

定的預測單位進行下一個單位的網路服務需求樣式組成預測。如圖 3.6 所示。

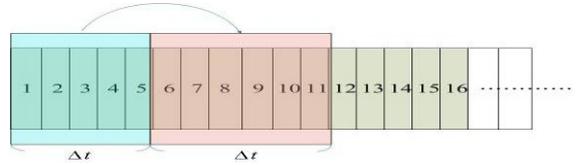


圖 3.6 以時間間隔為預測單位

(四) 以 AR-PST 建立網路節點服務需求之樣式組成預測模型

使用 AR-PST 建立網路節點服務需求樣式預測模型，是以固定的封包數量為預測單位進行預測，每次依此設定的單位依照預測當時 PST 對歷史資料的統計，得到每種服務需求樣式組成類型出現之後，下一個出現機率最大的服務需求樣式組成，以進行下一個單位的網路服務需求樣式組成預測。建構原理如圖 3.7 所示。

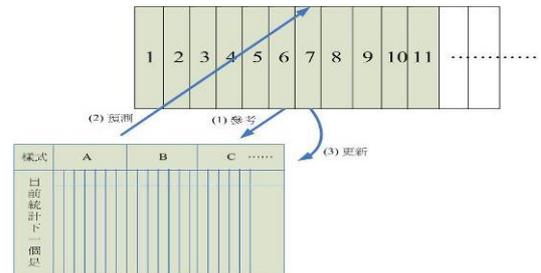


圖 3.7 AR-PST 預測模型建構原理

(五) 以 SMA 建立網路節點服務需求之數量預測模型

首先，在進行預測之前，必須先將所擷取的網路服務需求序列依據定義的最小時間周期為單位作切割。本文在建立此模型時，是取每十五分鐘為一個單位。SMA 網路節點服務需求預測模型的預測單位為時間單位的數量，亦即多少個十五分鐘，可自行定義，在定義預測單位裡所包含的時間單位個數之後，可進行所定義範圍內的每個時間單位之網路服務需求量的算術平均，以求得所定義範圍之後的下一個時間單位預測值，如圖 3.8。本文將使用不同長度的預測單位，針對十六種網路服務需求類型分

別進行網路服務需求量預測與探討。

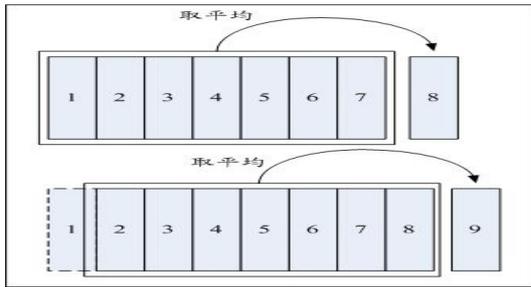


圖 3.8 SMA 預測模型之建構原理

(六) 以 WMA 建立網路節點服務需求之數量預測模型

本文在建立此模型時，一樣是取每十五分鐘為一個最小時間周期單位。WMA 網路節點服務需求量預測模型的預測單位為時間單位的數量，亦即多少個十五分鐘，可自行定義，在定義預測單位裡所包含的時間單位個數之後，可進行所定義範圍內的每個時間單位之網路服務需求量的加權平均，以求得所定義範圍之後的下一個時間單位預測值，如圖 3.9。本文將使用不同長度的預測單位，針對十六種網路服務需求類型分別進行網路服務需求量預測以及探討。

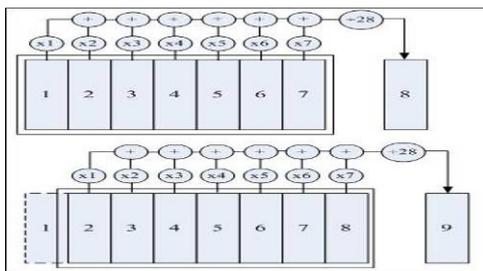


圖 3.9WMA 預測模型之建構原理

#### 四、預測結果與分析

(一) MRU 網路服務需求樣式組成之預測結果與分析

網路服務需求樣式組成之預測結果

MRU 網路服務需求樣式組成預測的結果分為以封包個數與以時間間隔為預測單位。以封包個數為預測單位之預測方式，其結果如圖 4.1

所示，藍色曲線為考慮所有封包類型去進行預測的結果，而紅色曲線為忽略資料傳送與資料回覆後的預測結果。橫軸表示所選用的預測單位，也就是封包個數的多寡，而縱軸表示預測準確率。

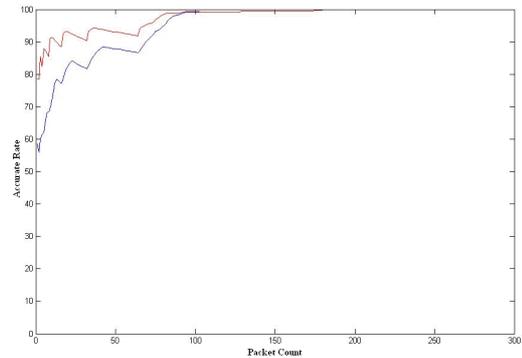


圖 4.1 以封包個數為預測單位之預測結果

以時間間隔為預測單位之預測方式，其結果如圖 4.2 所示，藍色曲線為考慮所有封包類型去進行預測的結果，而紅色曲線為忽略資料傳送與資料回覆後的預測結果。橫軸表示所選用的預測單位，也就是時間間隔的長短，而縱軸表示預測準確率。

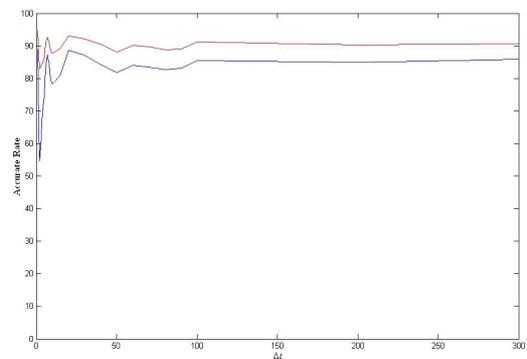


圖 4.2 以時間間隔為預測單位之預測結果  
網路服務需求連線數量之預測結果

接著為連線數量的預測結果，使用 MRU 網路服務需求樣式組成預測模型之樣式組成預測結果，計算得到預測的連線數量，與實際的連線數量進行比較。首先為以封包數量為預測單位的結果，圖 4.3 為以一個與十個封包為預測單位之預測結果，觀察預測單位的大小變化對於預測準確變化的影響，橫軸為代表實際時間，

縱軸代表預測與實際連線數量的各數差異。

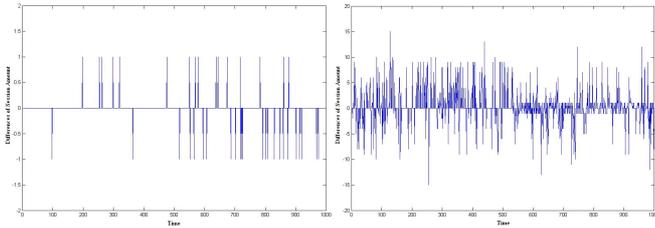


圖 4.3 以一個(左)與十個封包(右)為預測單位

圖 4.4 為以一秒與十秒為預測單位之預測結果，觀察預測單位的大小變化對於預測準確變化的影響，橫軸為代表實際時間，縱軸代表預測與實際連線數量的各數差異。

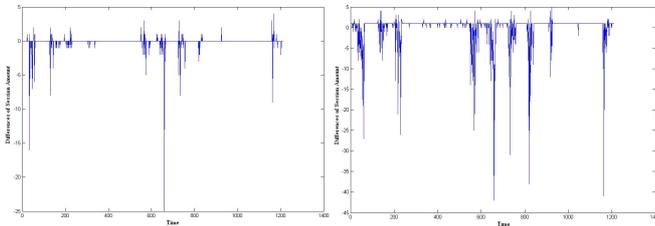


圖 4.4 以一秒(左)與十秒(右)為預測單位  
預測結果分析

針對 MRU 網路服務需求樣式組成預測模型之預測結果的觀察，歸結出兩項影響預測結果的因素進行探討，首先是預測所選用的預測單位，封包數量的多寡或是時間間隔的大小，對網路服務需求樣式組成預測的影響。第二項要探討的因素為建立模型時，歷史資料中是否包含資料傳送與資料回覆對兩種類型，對於網路服務需求樣式組成預測的影響。

第一個因素，預測單位的大小，根據實驗結果的觀察，發現所選取的預測單位越大，其樣式組成預測準確率有上昇的趨勢，兩者之間呈現明顯正相關的關係，無論是以封包數量或是時間間隔為單位的預測模型，其準確率皆能超越 90%。但是在連線數量的預測上，卻是預測單位越小，越能與實際連線數量相符合，當預測單位增大，其預測誤差將會隨之增加。

第二個因素，建立模型時，是否將資料傳送與資料回覆兩種網路服務需求類型列入考慮。比較這兩種需求類型的考慮與否，發現忽略這

兩種需求類型之後，準確度更快的攀升到定值。實質上，在達成本文所關注的連線(session)總數預測的目標上，同一連線之間的資料傳送與回覆，僅可預測連線頻寬的需求，對於所需機器的數量預測上，幫助並不大。原因在於，當資料封包被建立後，表示一定有機器可以承接，也表示網路的連線已經完成了。而原始的封包序列裡，大量的資料傳送與資料回覆兩種網路服務需求類型，穿插在連線需求、斷線需求等本文所關切的需求類型當中，將會對預測模型的預測效率造成影響，一旦不考慮資料傳送與資料回覆兩種類型之後，對於預測的準確度提升有正面的影響。

## (二) AR-PST 網路服務需求樣式組之預測結果與分析

AR-PST 網路服務需求樣式組成預測的結果，首先為樣式組成預測的結果，圖 4.5 與圖 4.6 為以一個與三個封包個數為預測單位之預測方式之結果如圖。下方的曲線皆為考慮所有封包類型去進行預測的結果，而上方的曲線為忽略資料傳送與資料回覆後的預測結果。橫軸表示實際時間，而縱軸表示預測準確率。

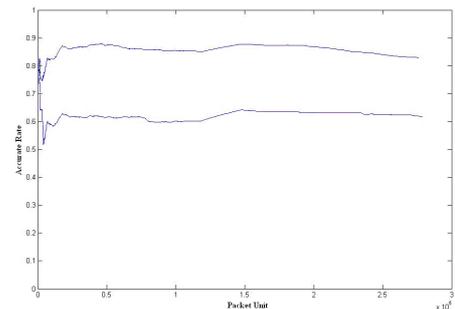


圖 4.5 以一個封包為預測單位

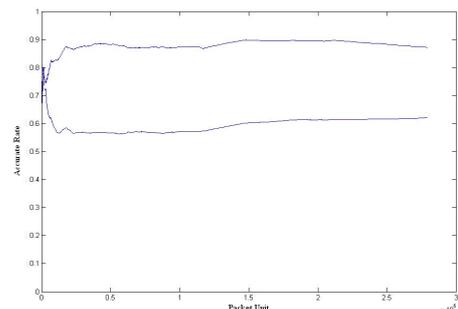


圖 4.6 以三個封包為預測單位

比較預測單位的不同對於預測準確率的影響，結果如圖 4.7 所示，藍色曲線、紅色曲線與綠色曲線，分別為以一個封包、兩個封包以及三個封包進行預測的結果，橫軸為實際時間，縱軸為預測準確率。

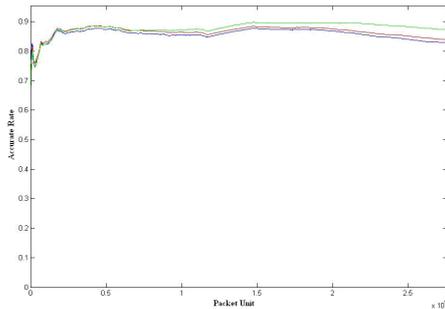


圖 4.7 預測單位與預測準確率之關係  
網路服務需求連線數量之預測結果

接著為連線數量的預測結果，使用 AR-PST 網路服務需求樣式組成預測模型之樣式組成預測結果，計算得到預測的連線數量，與實際的連線數量進行比較。圖 4.8 與圖 4.9 為以一個與三個封包為預測單位，之連線數量預測結果，橫軸為代表實際時間，縱軸代表連線數量，藍色曲線表示實際的連線數量、紅色曲線表示考慮所有封包類型進行預測的結果、綠色曲線表示忽略資料傳送與資料回覆後的預測結果。

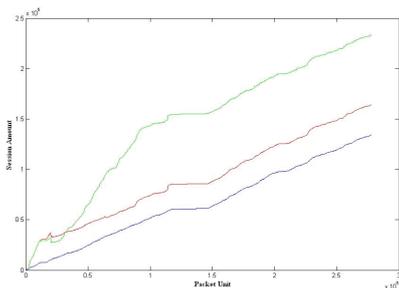


圖 4.8 以一個封包為預測單位

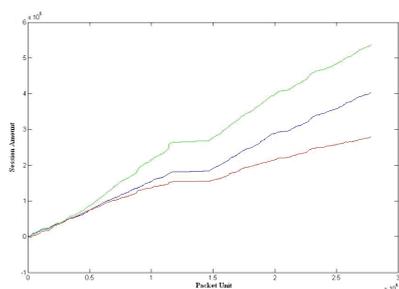


圖 4.9 以三個封包為預測單位

### 預測結果分析

針對 AR-PST 網路服務需求樣式組成預測模型之預測結果，亦可觀察到與 MRU 預測結果相同的兩項影響預測結果的因素，首先是預測所選用的預測單位，封包數量的多寡對網路服務需求樣式組成預測的影響。第二項要探討的因素為建立模型時，歷史資料中是否包含資料傳送與資料回覆對兩種類型，對於網路服務需求樣式組成預測的影響。

第一個因素，預測單位的大小，根據實驗結果的觀察，發現所選取的預測單位越大，其樣式組成預測準確率有上昇的趨勢，這樣的情況在忽略資料傳送與資料回覆的預測項目上更為明顯。但是在連線數量的預測方面，卻是在較小的預測單位，其預測誤差較小，造成連線數量預測誤差的原因為 AR-PST 在建構時，預測出其會經歷的學習階段，嘗試錯誤所造成的，而在預測單位越大時，學習所需要的時間越久，造成其誤差越大，另外，在連線數量的預測結果，發現以所有封包類型進行預測的結果將與實際連線數量較接近。

第二個因素，建立模型時，是否將資料傳送與資料回覆兩種網路服務需求類型列入考慮。比較以一個、兩個與三個封包進行樣式組成的預測結果，發現忽略這兩種需求類型之後，預測準確度皆有明顯的躍升。與 MRU 的預測結果相同指向相同的結論。

### (三) SMA 網路服務需求量預測結果與分析

#### 網路服務需求量之預測結果

使用 SMA 網路服務需求量預測模型，進行十六種類型封包的數量預測，由於不同類型其預測結果相當近，故取其中的入站資料回覆作報告。探討不同的預測單位之預測結果如圖 4.10 所示，藍色曲線為實際的封包、數量、紅色為以兩個時間單位也就是三十分鐘為預測單位的預測結果、綠色曲線為以五個時間單位為預測

單位的預測結果而紫色為以十個時間單位為預測單位進行預測的結果。橫軸表示實際時間，而縱軸表示封包數量。

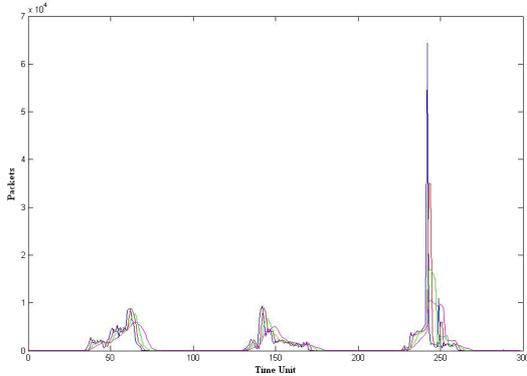


圖 4.10 不同的預測單位之預測結果

預測單位與預測準確率的關係如圖 4.11 所示。橫軸為預測單位的大小，縱軸表示預測準確率。

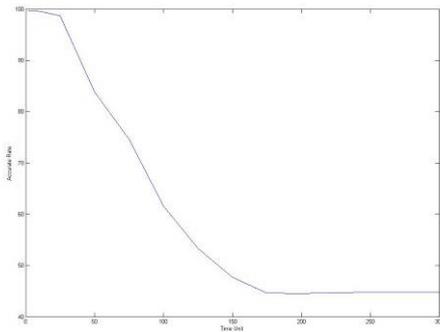


圖 4.11 預測單位與預測準確率的關係

#### 預測結果分析

針對 SMA 網路服務需求量預測模型的數量預測結果，歸結出兩個影響預測結果的因素進行探討，首先是預測所選用的預測單位的大小，亦及時間單位的數量多寡，對網路服務需求量預測的影響。第二項要探討的因素為實際網路服務需求量變化的劇烈程度，對於網路服務需求量預測的影響。第一項因素，所選取的預測單位不同對於預測準確率的影響，觀察圖 4.10 與圖 4.11，可以發現，預測單位越大時，預測準確率越低，預測數量與實際數量的差距越大。

第二項因素為實際網路服務需求量變化的劇烈程度，觀察圖 4.10，發現實際的網路服務

需求量變化越劇烈，將會造成 SMA 所預測得到的結果與其差異越大，且在預測單位越大的狀況下，這個情況將會更加明顯。

#### (四) WMA 網路服務需求量預測結果與分析

##### 網路服務需求量之預測結果

使用 WMA 網路服務需求量預測模型相同於 SMA，以入站資料回覆作為結果報告，以利後續進行兩模型的比較。探討不同的預測單位之預測結果如圖 4.12 所示，

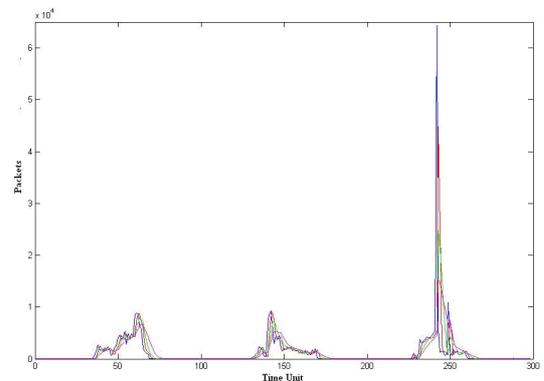


圖 4.12 不同的預測單位之預測結果

藍色曲線為實際的封包、數量、紅色為以兩個時間單位也就是三十分鐘為預測單位的預測結果、綠色曲線為以五個時間單位為預測單位的預測結果而紫色為以十個時間單位為預測單位進行預測的結果。橫軸表示實際時間，而縱軸表示封包數量。

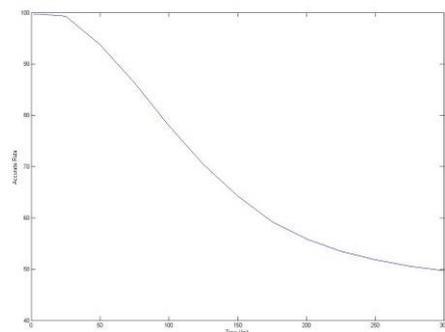


圖 4.13 預測單位與預測準確率的關係

預測單位與預測準確率的關係如圖 4.13 所示。橫軸為預測單位的大小，縱軸表示預測準確率。

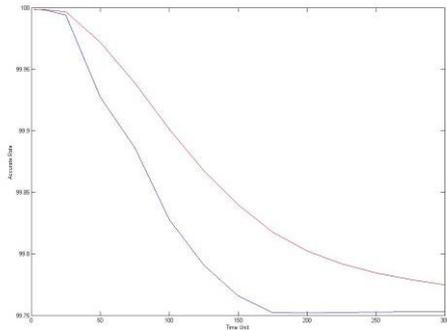


圖 4.14 預測單位與預測準確率的關係(SMA vs. WMA)

SMA 與 WMA 在預測入站資料回此類封包時，預測單位的變化對預測準確率的影響如圖所示，紅色曲線為 WMA 的預測結果，藍色曲線則為 SMA 的預測結果。橫軸為預測單位的大小，縱軸表示預測準確率。

#### 預測結果分析

針對 WMA 網路服務需求預測模型的數量預測結果，一樣可以歸結出：預測所選用的預測單位的大小，以及實際網路服務需求變化的劇烈程度，兩個影響預測結果的因素進行探討。第一項因素，所選取的預測單位不同對於預測準確率的影響，觀察圖 4.12 與圖 4.13，可以發現，預測單位越大時，預測準確率越低，預測數量與實際數量的差距越大。

第二項因素為實際網路服務需求變化的劇烈程度，觀察圖 4.12，發現實際的網路服務需求變化越劇烈，將會造成 WMA 所預測得到的結果與其差異越大，且在預測單位越大的狀況下，這個情況將會更加明顯。

根據預測結果，比較 SMA 與 WMA 網路服務需求預測模型在進行網路服務需求預測的結果，發現對於所選取的預測單位的不同，亦即預測時，選取的時間單位數量多寡所造成之準確率變化，WMA 的容忍程度較 SMA 高。圖 4.14 為使用 SMA 與 WMA 進行入站資料回覆此需求類型之數量預測，其預測單位與預測準確度的變化，可以觀察到 WMA 對於因為預測

單位的增大所造成的預測準確度降低的狀況較 SMA 平緩，原因是 WMA 預測模型在設計上，多了加權的動作，使得此模型對於數量變化的敏感度較高。

## 五、結論與未來展望

### (五) 結論

根據統計數據可了解，各種入站及出站的網路需求與回覆的網路服務需求數，並且可發現資料流量是週期性的變化。而若能經由預測得之未來一個週期時間的網路服務需求量、未來一個週期時間內網路機房機器需提供多少工作量，管理機房耗能情形以進行機房設備管理，在網路需求尖峰以及需求離峰時，做適當的機器啟動或停止之調整，則將可以預期此技術對於能源的節省、電費負擔的降低，以及對於因為以滿載情況為前提，而啟動具備大量冗餘備援機組的叢集系統，衍生冷卻系統運轉需求過量所造成的能源浪費情況也將有所改善。

因此本文提出一組對於網路節點服務需求的預測系統，此系統包括網路服務需求樣式組成的預測模型以及網路服務需求數量的預測模型。網路服務需求的樣式組成預測模型是使用 MRU 演算法與 AR-PST 模型所建構。而網路服務需求數量預測模型則是運用 SMA 以及 WMA 之預測方式所設計。

根據對預測結果之觀察，本文發現使用 SMA 與 WMA 建立的網路服務需求數量預測模型在預測效能上，具有優異的表現。然而，兩者僅對網路服務需求類型之數量或是整體數量等傳統預測方向作預估，甚少考慮不同類型網路服務需求的出現順序以及數量的連動關係；而僅由數量的變化產生應對策略，在網路需求的預測上，這點是不足夠的。本研究可以利用傳統的 SMA 或 WMA 可以達到 90% 左右甚至更高的預測準確度。但是，資料網路服務需求數量僅可預測連線頻寬的需求，對於所需機器的

數量預測上，幫助並不大。

原因在於，當資料網路服務需求被建立後，表示一定有機器可以承接，也表示網路的連線已經完成了。本文的重點在於設法在連線發生以前，預估所需機器的數量。為了找到真實需要的機器數量，重點在於預測下一個時間週期所可能發生的網路服務需求類型序列，方可決定所真正需要使用的機器數量。

本文首先利用了簡單的 MRU 模型，預測下一個網路服務需求類型，也就是假設下一個需求會與目前最近收到的需求相同，以此結果去進行預測，其預測準確率約在 90%。

此外本文也發展了 AR-PST 模型，用來預測下一個網路服務需求類型，目前最高可利用三種類型的組合之計算其比例進行預測，其準確率約為 62%，而在不考慮資料傳送與資料回覆兩種服務需求類型之狀況下，其準確率可達到 90%，未來也將針對此部份的預測模型進行參數的調整以提高其預測準確度。

#### (六) 未來展望

本文最終目標為預測未來一個週期時間內機房機器需提供多少工作量，管理機房耗能情形。經由得知網路服務需求量後，進行機房設

備管理。因此，如何提升對未來一個時間週期的網路服務需求量之預測準確度是本文未來研究的重點。後續將進行主機之耗電量測結合網路流量預測，取得網路流量大小與機房主機耗電之間的關係式，之後只需要預測出網路流量即可計算取得可能的主機耗電情況，接著可根據主機耗電量計算求出冷氣的需求量，達成能源效率評估模型。

#### 誌謝

本研究由台灣工業技術研究院『智慧型建築電腦機房內高可用性叢集與空調系統能源整合之最佳化控制研究』研究案補助研究經費，特此致謝。

#### 參考文獻

- [1] 台北區網中心, <http://www.tp1rc.edu.tw/>
- [2] C.-J. Tang, and M.R. Dai, "Identifying the Repeating Music Patterns: Performance and Correctness," 2008 Symposium on Digital Life Technologies, Human-Centric Smart Living Technology, June, 2008.