

行政院國家科學委員會專題研究計畫 成果報告

總計畫(3/3)

計畫類別：整合型計畫

計畫編號：NSC91-2219-E-009-033-

執行期間：91年08月01日至92年07月31日

執行單位：國立交通大學電信工程學系

計畫主持人：張仲儒

共同主持人：楊啟瑞，張仲儒，陳伯寧，陳耀宗

計畫參與人員：林立峰、黃鏗銘、黃慶喜、丁崇光、陳柏翰、鄭永宏、吳育葵、顏寧佑、翁昭源、潘家蘄、殷偉盛、花凱龍、林建光、王聖賢、姚建、詹益禎、郭國承、張均璋、何凱元、吳忠興、陳志偉、田伯隆、施汝霖

報告類型：完整報告

報告附件：國際合作計畫研究心得報告

處理方式：本計畫可公開查詢

中 華 民 國 93 年 2 月 16 日

計畫中文摘要

由於網際網路的蓬勃發展，使得網際網路上的訊務量急遽增加，且跨網域的比重也驟然提升甚多，除了最初且最基本的數據通訊之外，也加入了語音或影像等即時性類型的服務。如此，訊務流量(traffic flow) 的增加、訊務特性的多樣化與服務品質(QoS) 需求的大幅差異，使得原本只適合於非即時性數據(Data)傳輸的網際網路之各項應用的服務品質不易獲得保障。為了確保目前和未來可能的各項應用與服務能夠在網際網路上運作順暢甚或具有一定的品質，並維持網際網路的永續發展，故其需要對新型態的訊務特性有所瞭解，並透過頻寬的提升、與有效的 QoS 運作架構與訊務控制機制來確保使用者的服務品質可以獲得保障。

本整合型計畫擬以三年來研究網際網路上之訊務特性與提供服務品質保證的運作架構及訊務控制方式，並提出有效的解決方案，以完成網際網路端點對端點之服務品質保證。子計畫一由張仲儒教授主持，主要研究網際網路中高速路由選徑技術與提供服務品質保證之訊務控制法及其設計：直接針對路由器本身著手，增加路由器路由選徑的處理速度，或利用第二層網路的高速交換(Switch) 來取代第三層之路由選徑，以獲得等效於加快路由選徑速度之效果，並配合差別服務，提出精確且有效的允諾控制與使用參數控制法；子計畫二由陳伯寧教授主持，主要在探討網際網路中具有自我類似(Self-Similar) 性質的真實網路訊務，瞭解其成因與特性並提出訊務模型以作為其它子計畫方法設計之依據；子計畫三由陳耀宗教授主持，主要研究網際網路差異化服務的路由器中，多欄位分類器與負責提供差別服務的排程器之設計，驗證並比較出最適合者，並利用各種軟硬體方式以實現所完成之設計。

於計畫執行第二年度，新加入由楊啟瑞教授所主持的第四子計畫。子計畫四主要著眼於個人行動通訊時代的來臨，通訊網路已經由有形的網路延伸至無線的環境了，在最接近用戶的終端可以藉著可攜式多媒體通訊設備或電腦透過無線行動通訊科技進行通訊，甚或進入網際網路中擷取資料或取得服務，然而多變的無線環境也相對增加了各項訊務控制機制的複雜度。因而擬透過子計畫四探討如何以無線設備進行具備 QoS 保證的寬頻網際網路的存取，設計一個持續穩定且有高效能並支援 QoS 的無線媒體多重接取(Multiple Access) 競入技術，能夠使得競入訊號能以極高效率進入系統，同時保持區域網路的最大效能，並進一步於其上設計出一個動態頻寬分配機制。

關鍵詞：路由選徑、多協定標記交換、差異化服務 / 差別服務、訊務控制、呼叫允諾控制、使用參數控制

Abstract

With the blooming of Internet application services, the Internet traffic flow increases dramatically and the traffic flow from inter-network transmission also increases rapidly. Not only the basic data transmission but also some multimedia services (such as: voice, real-time video services) are also carried on the Internet. Henceforth, to provide the QoS-provisioning services, the larger bandwidth capacity, effective QoS-provisioning service framework, and traffic control mechanisms are the primary requirements in the design of the Internet.

The three-year integrated project is intended to propose an effective solution to provide end-to-end QoS guaranteed Internet services and the corresponding research interestings are: the traffic characteristics of the Internet services, a QoS-provisioning service framework and the traffic control mechanisms. The subproject one is to study the high speed Internet routing techniques, the QoS-provisioning traffic control mechanisms and the network configuration. To speed up the Internet routing, 2 possible techniques will be proposed to improve the performance of the router: faster the routing path searching techniques and layer-2 high speed switching techniques for replacing the layer-3 routing techniques. A precise and effective control admission control (CAC) and usage parameter control (UPC) will be also proposed for providing differentiated services. The subproject two is to study the self-similarity of traffic characteristics of the real Internet services and propose an appropriate traffic model that will be the reference model for other subprojects. The subproject three is to study and design the multiple field classifier and scheduler for the routers that provide differentiated services. The subproject four is to propose a persistently stable, efficient multiple access scheme for wireless networks, capable of providing signaling traffic high performance while retaining maximal throughput for local wireless networks. Many possible configurations will be verified and the most appropriate one will be proposed and implemented by hardware and software.

Keywords : Routing, Multi-Protocol Label Switch (MPLS), DiffServ, Edge Router, Core Router, Multi-field Classifier, Scheduler, Traffic Control, Call Admission Control, Usage Parameter Control, Self-similar

目錄

一、 計畫緣由及目的.....	1
二、 研究方法、成果與討論.....	3
1. 高速的路由選徑機制.....	3
1-1. 高速單一路由選徑(Unicast Routing)機制 - 階層式分群解析架構.....	3
1-2. 高速單一路由選徑(Unicast Routing)機制 - TCAM-based 架構.....	10
1-3. 高速群播路由選徑(Multicast Routing)機制.....	15
2. MPLS 網路之 VC-Merge 機制的效能分析.....	17
3. MPLS 網路之路徑保護及快速回復(Path Recovery) 機制.....	22
4. DiffServ 網路中精確的訊務監控調節(Traffic Contioner)機制....	29
5. 高速 IP 封包分類(Packet Classification) 演算法.....	37
6. 提供等比例封包延遲變異(jitter)之排程器(Scheduler).....	40
7. 具有自我類似性質之訊務特性研究與訊務源模型設計.....	44
7-1. 馬可夫過程(Markovian Process) 的自我類似性質探討.....	44
7-2. 自我類似特性之網路封包的產生.....	47
7-3. 自我類似特性之訊務流在資訊理論上的特徵.....	52
8. 支援 QoS 的無線媒體多重接取(Multiple Access)競入技術.....	56
三、 參考文獻.....	60
四、 計畫成果自評.....	67
五、 附錄.....	70
1. 附件一：國際合作研究計畫國外研究報告書.....	70

一、計畫緣由及目的

從 1969 年發展至今的網際網路(Internet)，已經由最初實驗性的研究成果，在歷經幾次的變革之後普及至教育和商業環境；更由於其跨網路、跨地域（無遠弗屆）極富有彈性(Flexibility) 與可能性(Possibility) 的特點，吸引了愈來愈多人投入此一新的網路世界，然而它的成功卻也加速顯現其發展的瓶頸。擁有超過兩千萬個節點及上億使用者的 Internet，必須進行大幅度的改造，才能進一步像電話一樣普及，事實上這些改造的研發工作也從未間斷。

架構出網際網路的網際網路協定(IP Protocol) 原本只是適用於數據 (Data) 通訊的第三層網際網路協定，並藉著路由器連接不同的網路 / 網域形成一非連結導向(Connectionless) 的廣域網路。然而在網際網路蓬勃發展之際，也使得許多非單純數據傳輸的應用，例如語音或影像等具備即時 (Real-time) 傳輸要求的服務也採用 IP Protocol 進入網際網路中，再加上多媒體技術的突飛猛進與其服務的普及，使得對頻寬與網路服務品質保證的需求也相對應地提升。所以有許多的改革動作也應運而生，其中最主要的即是在高頻寬傳輸速度與服務品質保證上 (QoS guarantee) 的研究。

為了解決寬頻的問題，提升路由器的處理速度成為必然的趨勢。目前在這方面的研究可以分成兩大類：一是直接針對路由器本身著手，增加路由器路由選徑 (Routing) 的處理速度，即是要發展出高速的路由選徑機制；另一方面則是藉著改變架構，利用連結導向(Connection-oriented) 的第二層（電信）網路的高速交換(Switch) 動作來取代第三層非連結導向(Connectionless) 的 IP 協定網路的路由選徑動作，使可以獲得相當於在第三層上加快路由選徑動作速度的效果，我們也可以將其視為一種虛擬的路由選徑(Virtual Routing)方式，如 IETF 所提出的 MPLS(Multi-Protocol Label Switch) 技術。

提供服務品質保證方面，目前 IETF 已經針對這方面（服務品質保證）的需求成立了相關的 Working Group。其中的 DiffServ 並不針對 per-flow 的訊務提供服務品質保證，而是將訊務分為幾種不同的等級種類(Class)，再對每一種類提供不同的服務品質。不同種類的訊務在路由器上會被封包分類器(Classifier) 分開來，將相同種類的訊務再進行匯流整合(Aggregate) 視為一體的訊務流(flow)，然後只對此匯流整合的 per-class 訊務進行 QoS 的處理（例如排程與佇列管理），而不再對其中單一的 per-flow 訊務進行處理，如此便可以降低系統的複雜度與負擔，增加網路規模的擴充性。分類的依據可以是多樣的，根據 IP 封包中不同的欄位（例如：Source IP、Destination IP、Transport-layer Port、ToS/DSCP 等）與其上記載的訊息，或單一欄位、或多欄位組合的條件方式進行訊務的分類。DiffServ 在實際的運作方式上也必須引入連線導向（電信）網路中訊務控制的概念，例如以呼叫允諾控制對新進的連線作系統資源的確認，再決定是否接受此連

線的訊務，以更進一步確保能提供各種類的訊務所要求的 QoS。

此外，未來寬頻網際網路技術之走向，應是以解決目前與未來可能發生的網路問題為主軸。而在此研究上，如網路節點的排程控制，或是網路規約整體設計，多以傳統的波以松(Poisson)式的訊務為模擬基礎；而近年來的網路流量特性之量測研究顯示，真實網路封包序列(Packet-Train) 流量與波以松式的訊務有著極大的差距，甚至具有自我類似(Self-Similar) 的特性，因此要真正的了解乃至於提出真實網路問題的解決方案，勢必要對真實網路訊務有深入的了解。

而著眼於個人行動通訊的時代的來臨，通訊網路已經由有形的網路延伸至無線的環境了，在最接近用戶的終端可以藉著可攜式多媒體通訊設備或電腦都透過無線行動通訊科技進行通訊，甚或進入網際網路中，擷取資料或取得服務。如此大大增加了通訊的自由度與彈性，但也因為範圍涵蓋了較多變的無線環境，故相對地增加了複雜度，其中主要的研究課題是探討如何以無線設備存取寬頻網際網路。基本上，無線 ATM 網路已被視為下一代無線區域網路中最有潛力的一個可行方法，因為它具備了整合性多媒體高速服務，並能滿足各式不同 QoS 的需求，如 CBR、VBR 與 ABR。對 VBR 和 ABR 來說，它所要求的 QoS 分別為有限的傳送延遲與保證擁有最小的傳輸速率。在無線 ATM 網路上所面臨的主要挑戰有二：首先，是如何來設計一個有效率且支援 QoS 的無線媒體存取控制協定；其次，是如何於前述之協定上設計出一個動態頻寬分配機制。過去我們已經提出了一個無線 ATM 網路上的智慧型多重存取控制系統 (IMACS)，它是一個功能強大並可動態調整頻寬分配的設計。IMACS 包含了三個部份：媒體存取控制器 (MACER) 訊務推估器 (TEP)，與智慧型頻寬分配器 (IBA)。MACER 利用 TDMA 混合模型的架構，它的競入存取控制是基於動態樹狀分割 (DTS) 的碰撞解決演算法，以最佳化的分割深度 (SD) 為其參數，而無線網路的存取與訊務控制都可建立於此範例上。然而在此仍有三個相當富有挑戰性的問題等待我們去研究：第一個問題是 DTS 利用最佳化的 SD 可以不經任何額外的控制而達到 utilization 為 0.368。我們的**第一個目標**為設計出一套盡可能使用最少 feedback 的無線通訊方法，可以將這 utilization 升高，使其接近理論上限 0.58。第二個問題是在 IMACS 中，需服務品質保證的即時性訊務 (如 CBR 與 VBR) 和不需即時服務的訊務並不在一起考慮；我們**第二個目標**為設計出一個能同時考慮即時和非即時訊務的機制，且能滿足各自的 QoS 要求，並能達到最高的 utilization；最後，我們**第三個目標**為研讀長程(Long Range Dependent) 相關或自我相似性的論文，並設計以 Autocorrelation 函數為依據的一個智慧型訊務大小推估技術，以提供未來無線網路訊務工程的開發能力。

本整合型計畫之總計畫 (張仲儒主持) 旨在完成網際網路端點對端點之服務品質保證，研究網際網路上之訊務特性與提供服務品質保證的運作架構及訊務控制方式，並提出有效的解決方案。子計畫一 (張仲儒主持) 主要研究網際網路中高速路由選徑技術與提供服務品質保證之訊務控制法及其設計；子計畫二 (陳

伯寧主持)主要在探討網際網路中具有自我類似(Self-Similar)特性的真實網路訊務，瞭解其特性並可作為其它子計畫架構設計之依據；子計畫三(陳耀宗主持)主要研究網際網路差異化服務的路由器中，多欄位分類器與負責提供差別服務的排程器之設計。子計畫四(楊啟瑞主持)主要是針對區域無線網路而言，擬提出了一個具有持續穩定且有高效能的多重接取(Multiple Access)競入技術，能夠使得競入訊號能以極高效率進入系統，並能同時保持區域網路的最大效能。

接下來我們將分別就本計畫在「高速頻寬傳輸」與「服務品質保證運作機制」上的各項具體研究成果，進行方法說明與成果討論，

二、研究方法、成果與討論

1. 高速的路由選徑機制

這部分主要是發展適合硬體實現(Hardware-oriented)的高速路由選徑方法，期以硬體的運作方式加速路由選徑的運算速度，以滿足 Gigabit 超高速網路環境下以及未來更寬頻的網際網路應用的需求。

1-1. 高速單一路由選徑(Unicast Routing) 機制 - 階層分群解析架構

經過歸納與研究分析的結果，適合硬體實現的路由選徑方法應具備有下列的特性與概念：「固定資料長度」的資料運算動作，以及「規則化」的、「反覆運作」的處理程序(Process)，而「階層式分群解析」(或稱「多層次群組解析」)的方法即具備有上述的末兩項特點，再配合上以階層式的多元完全展開樹(Trie)來做為其將整個 IP 位址進行多層次的 IP 位址區段分群的參考架構後，便具備有「固定資料長度」的資料運算動作的條件，因而極適合於做為採用實際的硬體邏輯 & 運算電路來實現的路由選徑機制。「階層式分群解析」的方式是：將整個 IP 位址進行多個層次的 IP 位址區段分群 - 先進行第一層次的較粗略分群，再依據實際路由表格中路由字首(Route Prefix)的資訊，針對有需要做進一步細部分群解析的群組(即包含一個以上，對應至更小範圍 IP 位址區段的路由字首)進行下一層次更精細的分群展開，如此反覆運作至每一分群中沒有對應至更小範圍 IP 位址區段的路由字首為止。接下來將每一層次的每一 IP 位址區段分群與該 IP 位址區段路由選徑的結果進行對應並以表格紀錄(「分群-路由結果」對應表格)：若為毋須再進行下一階段細部分群解析的群組，將必然可以對應至一個該 IP 位址區段的(共同)路由選徑結果(即封包的輸出埠(output port))；若為需要再進一步細部分群解析的群組，則可以對應至一個「必須進行下一層次更細部分群解析」

的指示，並且指向連結至該進一層次的「分群-路由結果」對應表格。進行實際路由選徑的應用時，只要將所欲查詢的目的 IP 位址(Destination IP Address) 與各層次的 IP 位址區段分群進行比對，尋找其所屬的最細的分群，待確定目的 IP 屬於何層次的某一分群後，即可以由該層次的「分群-路由結果」對應表格直接查表得知其路由結果輸出埠。而此「比對」動作是規則地、次第從第一層分群開始，再視需要逐步往分群更精細的層次檢視、比對。由於是將路由選徑之搜尋演算動作化為「規則」化的、「反覆運作」的多層次比對與查表動作，因此已具備適合硬體實現的初步條件了。

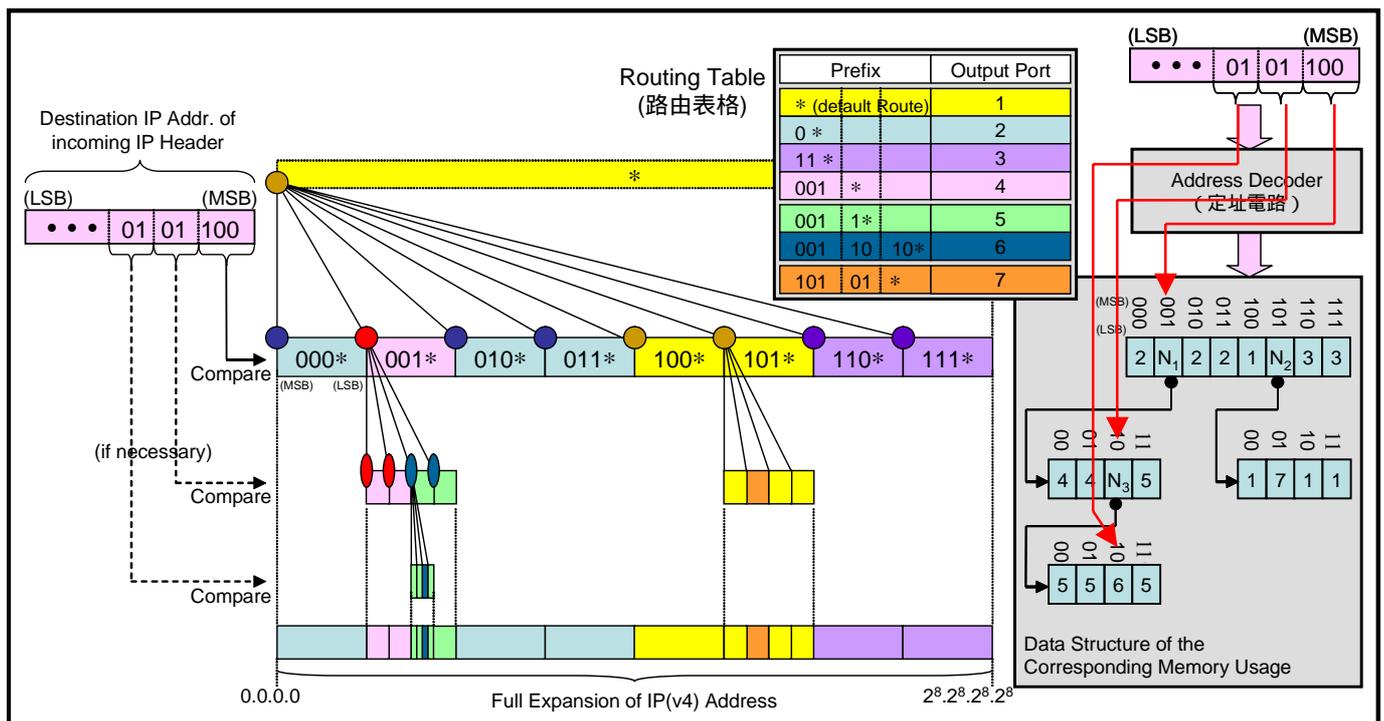


圖 0: Trie-based 階層式分群解析路由方法

因此我們將以「階層式分群解析」的方式為主，來發展適合硬體實現的高速路由選徑方法。而我們所提出的作法，是以階層式的多元完全展開樹 (Trie) 來做為「階層式分群解析」方法中，將整個 IP 位址進行多層次的 IP 位址區段分群的參考架構——也就是使每層次的 IP 位址區段分群恰好為一 N -bit ($1 \leq N < 16$) 的完全展開樹(Trie)，如此一分群的 IP 位址區段範圍皆可以用一個 IP 位址字首 (Prefix) 的形式來表示 (例如：140.113/16) ——再依據實際路由表格中路由字首 (Route Prefix) 的資訊，進行實際階層式的多元完全展開樹的建置工作，在每一層次中，僅針對有需要做進一步細部分群解析的群組 (即包含一個以上對應至更長路由字首的次 IP 位址區段) 進行下一層次更精細的 N' -bits Trie ($1 \leq N' < 16$) 的分

群展開（以上的敘述可以參見圖 0 所示）。每一層次皆對應至一組表格，紀錄此層次中各分群 IP 位址區段(Segment) 及其所對應的路由選徑結果，或是必須進行下一層次更細部分群解析的指示。而針對有需要才進一步細部分群解析的方式也比起純粹的（階層式）多元完全展開樹方法大大減少所需的記憶體容量。實際應用於進行路由選徑查詢動作時，即是將所欲查詢封包的目的 IP 位址(Destination IP Address) 與各層次的分群進行比對：由第一層次分群開始，直至其所屬的分群不再有進一步的分群解析為止，此時只要查詢該層次的「分群-路由結果」對應表格即可得到該封包的路由輸出埠。而由於我們是以階層式的多元完全展開樹（Trie），來做為其將整個 IP 位址區段進行多層次分群的參考架構，每層次的 IP 位址區段分群恰好為一 N-bit ($1 \leq N < 16$) 完全展開樹(Trie)，因此可以如同圖 0 右半部份所示，進一步將目的 IP 位址在各層次的「分群比對」動作，轉換為 N-bit 固定長度的「定址」動作，而所定址到的記憶體內容即儲存該階層中該 IP 位址群組(Segment) 的路由選徑結果(output port)，或是指向儲存著下一層次分群解析的路由選徑結果的 Pointer，如此也相當於將原本分群比對之後的「分群-路由結果」表格的查詢動作也一併整合進來了。也就是說，只要透過反覆的（固定資料長度的）定址動作，即可以得到路由選徑的結果。至此，我們已將傳統路由選徑之搜尋演算動作，轉化為一套系統化的階層式路由資料結構，與一「規則」化的、「反覆運作」的多層次「定址」動作，而這些都可以採用實際的硬體邏輯 & 運算電路來加以實現並獲得加速的效果，例如其中的定址動作便能夠採用定址電路來達成。此時 Routing 的速度則完全取決於「記憶體定址」的「存取次數」與「定址電路運作速度」。

為了決定實際運作時，所使用的實際階層分群（參數）的設定，包含階層數目與每一階層中的分群大小（因為每一階層是為一個 N-bit 的多元展開樹，所以這裡也相當於是在決定每一階層的 N 值大小，亦即多元展開樹的大小），以及所需要的記憶體容量大小，因此我們進一步探討 Routing Table 的特性，並再深入瞭解所設計方法的特點，以及兩者之間的關係。對於每一種路由選徑方法而言，不同的 Routing Table 資料皆會因為其中路由字首數據的不同，諸如路由字首數量多寡、分佈的型態 / 趨勢、分佈的量值大小差異，造成其不同的運作記憶體容量需求。在觀察我們所設計的路由選徑方法之後發現，對於一筆 Routing Table 資料而言，不同的階層分群方式會影響儲存路由結果的資料結構所需要的記憶體容量，這也表示我們可以藉由適當地設定階層分群方式，來獲得最小記憶體容量。透過初步的檢驗程序發現，增加分群階層數可以使每一階層的完全展開樹規模較小，相對上每一分群範圍較廣，因此可以較有效率地針對有需要再進一步分群解析的 Subnet 才進行下一階層的展開，所以可以相當有效地減少記憶體需求，然而卻有平均記憶體存取次數隨之遞增的缺點，因而我們必須在分群階層數與記憶體存取次數之間權衡一最佳點。此外觀察也發現，記憶體容量需求也並非隨著階層數的增加而永遠呈等速率地減少，而是會趨向一飽和值。因此在綜合考量階層數變化對於記憶體容量以及存取次數的效應之後，我們決定採用 5 層次

的分群解析架構。

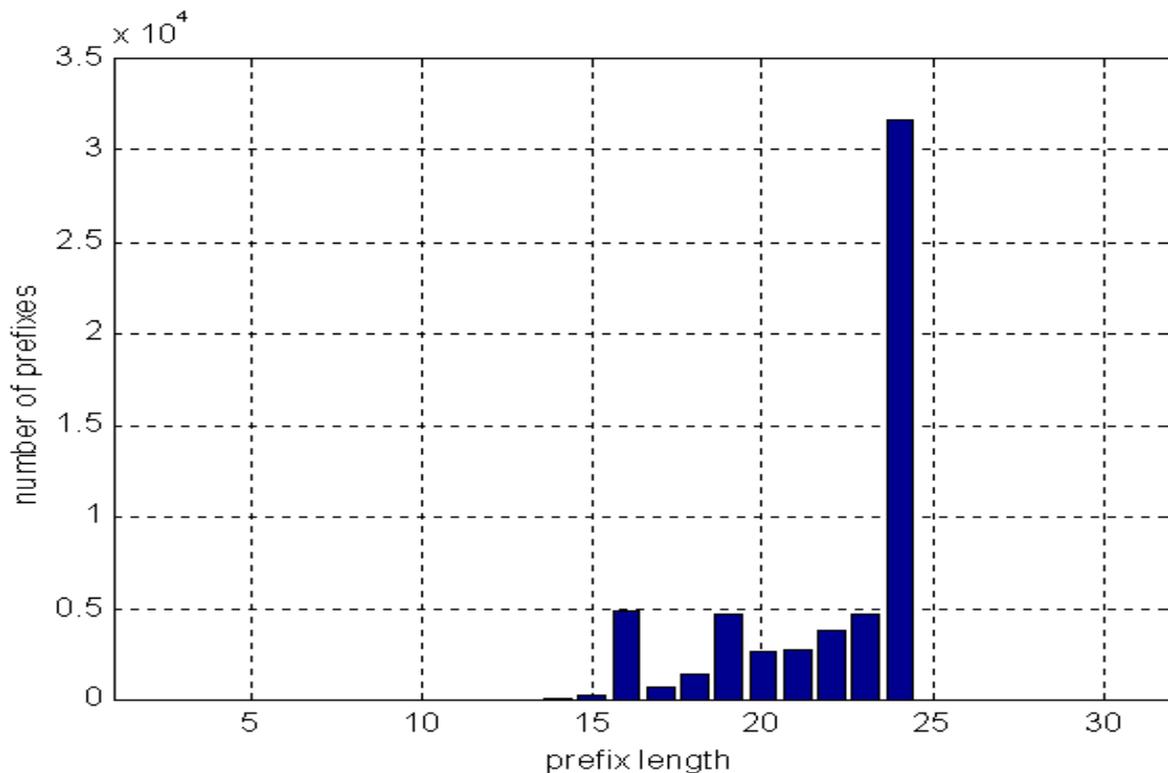


圖 0: Routing Table 中路由字首的分佈狀態圖
(本圖數據參考自 MaeEast 2000/07/18 的統計資料[4])

最後我們更進一步地分析實際 Routing Table 中路由字首的分佈狀況，作為階層式多元完全展開樹各層次中分群的依據，期望能以對應路由字首分佈特性來最佳化的分群設計，獲得更有效率的記憶體使用。圖 0 為一般 Routing Table 中普遍的路由字首分佈狀態統計圖，從圖上可見大多數的路由字首長度集中在 16 bits 至 24 bits 之間，因此若是在此長度之間有一個階層分界展開點，便可以在此階層的展開中完成大多數的路由選徑查詢動作（如果實際的 IP 封包標頭的目的 IP 位址不特別集中在長度超過 24 bits 的路由字首的話）而根據之前分群階層數考量與決策過程的經驗來看，如果在 1 至 15 bit 的字首長度之間定有分層展開點，則應該可以有較小的記憶體容量需求，然而卻也會因而增加平均記憶體存取次數。在同樣權衡記憶體大小與平均存取次數後之後，我們擬定以 16 bits 的路由字首長度做為第一分群階層的展開，期望以此展開長度在第一次的分群展開中便能夠完成大多數的路由選徑查詢，並進一步如圖 0 所示檢驗多種分群組合方式，最後我們擬定用於 Unicast Routing 的 32-bit IPv4 位址而言的最佳 IP 位址階

層分群方式為「16-1-7-1-7」，因為此方式可以得到最佳的表現：平均記憶體存取次數較小，以及最小的記憶體需求。

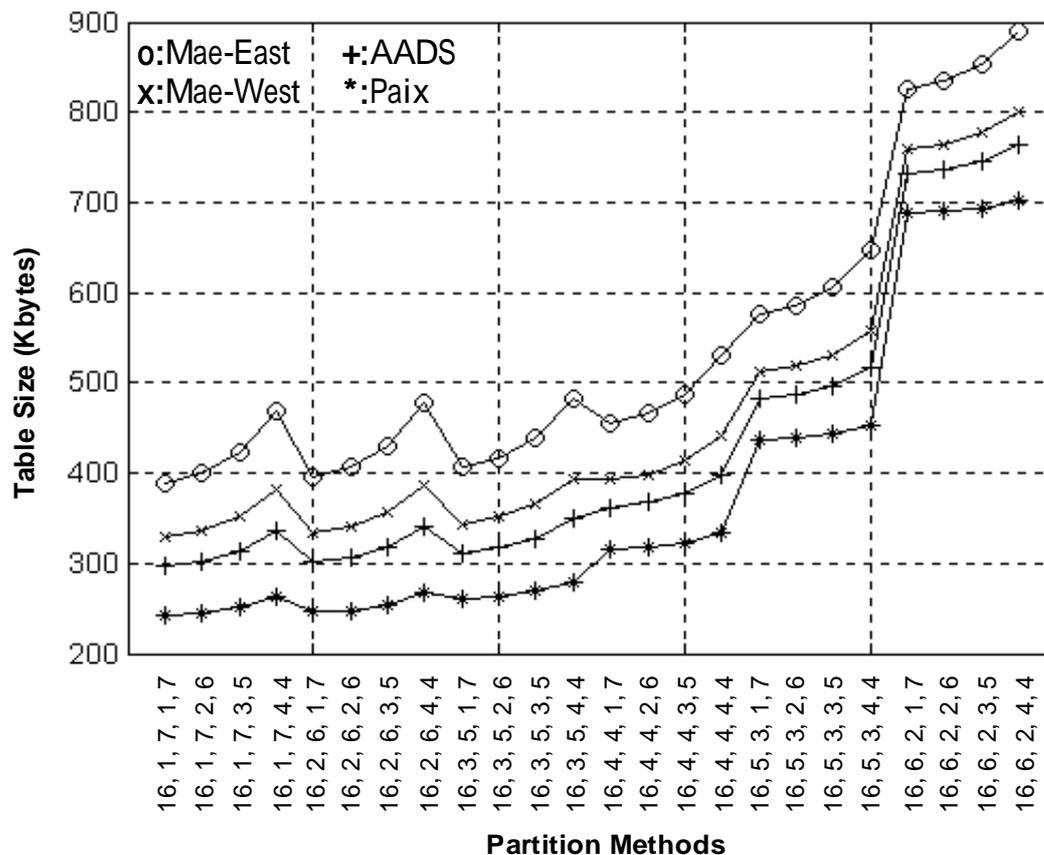


圖 0: 不同的 IP 位址 (32-bit IPv4) 階層分群方式及其所需之記憶體容量
(對不同 ISP/NSP 的 Routing Table 而言)

最後，我們採用一簡單的位元圖(Bit Map) 壓縮法(Compression Bit Map, CBM) 將儲存路由結果的資料結構做壓縮，以進一步減少所需的儲存記憶體空間。而我們所設計的階層式分群解析法原本在運作架構特性上就相當適合模組化的運作，可以將每一個階層的分群解析路由查詢動作都視為一獨立而完整的運作模組，因此透過適當的電路規劃與安排，可以達成硬體上超管線式(Pipeline) 平行多工架構的運作方式及其優點，當有 IP 路由選徑查詢進入第二階層後，便可以馬上接受下一 IP 路由查詢的要求，維持高度的路由查詢 Throughput，使每個封包路由選徑動作所要存取記憶體的次數減少至極致，可達到相當於在一次的記憶體存取動作與時間，便可以完成一筆 IP 路由選徑查詢的動作，此時 Routing 的速度將幾乎完全取決於記憶體存取定址的硬體定址電路運作速度。

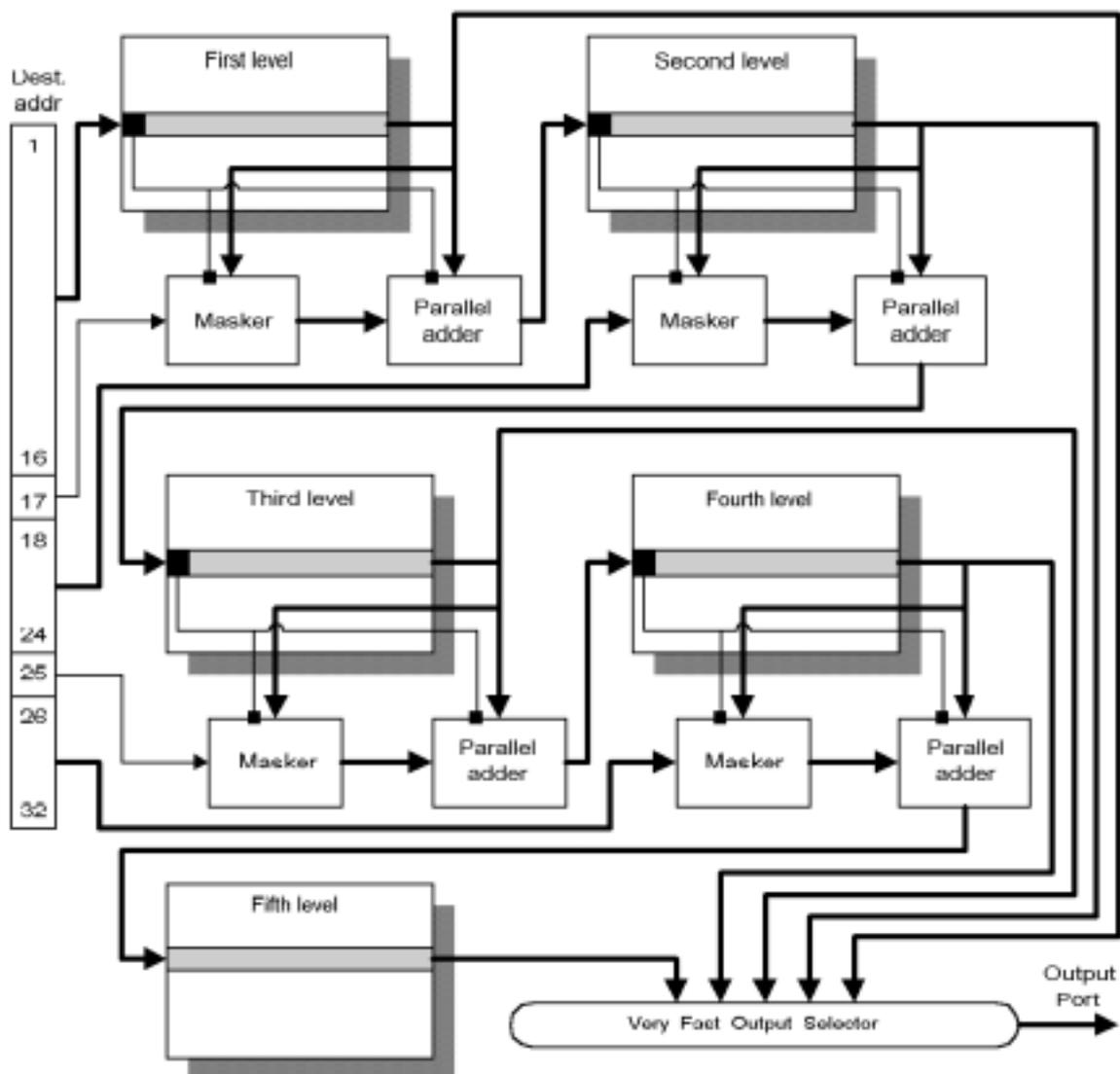


圖 0: 快速路由選徑機制的硬體邏輯架構

圖 0 所示即為我們所設計的快速路由選徑機制的硬體邏輯架構。從表 0 的模擬結果看來，目前我們所設計適合於 IPv4 Unicast 的高速路由選徑方法相較於其他類似概念的方法 [3]，僅需要不到 400 Kbytes 的記憶體空間，而且透過硬體上 Pipeline 平行多工的方式，可達到每個封包在平均一次記憶體定址存取的动作即可以獲得路由選徑結果。未來我們將考量實際商用化系統環境與規格（速度與記憶體容量需求）進行調整與最佳化設計，改進其路由表格資訊更新（Routing Table Update）時所必須對應的相關資料結構與內容的替換或更新，提出更有效率的方式。除此之外也擬配合次世代網際網路的潮流，進一步發展適用於 IPv6 環境的高效能路由選徑技術。

NAP (Number of Route Prefixes)	Enhanced Forwarding Table Size (Kbytes) with (16, 1, 7, 1, 7)	Forwarding Table Size (Kbytes) in [3]
<i>Mae-East</i> (57,701)	388	464
<i>Mae-west</i> (34,319)	350	438
<i>AADS</i> (30,705)	337	431
<i>Paix</i> (16,274)	270	357

表 0: 路由選徑方法所需記憶體容量之比較

綜合上述的內容來看，我們在高速路由選徑法之研究設計方面，藉由適當的查詢架構與路由表格結構的設計，快速而硬體架構導向的路由查詢演算法得以實現，配合上實際硬體化操作與超管線(Pipeline) 平行多工運作架構的設計，使得路由選徑查詢得以達到平均約一次記憶體存取動作即可完成的高通透率(Throughput)，可以向上支援至超高速乙太網路(Gigabit Ethernet) 甚或更高速網際網路頻寬的需求。根據階層式分群解析概念，配合壓縮化完全展開樹以及 Compression Bit Map 壓縮法而發展出來的路由選徑方法，確實能使路由選徑機制運作所需的記憶體容量降低許多，相對於平面展開式的壓縮方式，更小的記憶體容量亦更有利於硬體架構的設計與運作，甚至可以將記憶體與周邊相關的邏輯電路整合入同一單晶片中，成為一獨立的硬體路由選徑搜尋引擎，除了可以更大幅提升速度之外，也符合系統化晶片(System on Chip, SoC) 的發展潮流與趨勢。另外，我們也將所設計的方法針對 Routing Table 中，大量的路由字首長度大於 24 的此種特殊路由字首分佈情況做進一步的檢驗與分析，發現其所造成的記憶體容量需求成長的幅度相當低，並未隨之成比例地大量增加，從此結果也可見我們的方法還具有相當的空間擴展性的優勢。藉由對單一路由(Unicast Routing) 之選徑技術研究，我們可以很快進入多點群播路由(Multicast Routing) 及下一代網際網路 IPv6 之相關路由選徑技術的研究，甚至是封包分類器(Packet Classifier) 中複雜度更高、查詢參照資訊(欄位) 更多的查表搜尋演算機制的設計。

1-2. 高速單一路由選徑(Unicast Routing) 機制 - TCAM-based 架構

除了上述根據階層式分群解析概念，配合壓縮化完全展開樹而發展出來的硬體架構導向的路由選徑方法之外，另一類硬體化操作的路由選徑機制，即是著眼於因半導體製程技術的進步而開發出來的 TCAM (Ternary Content Addressable Memory) 記憶體特性，而採用其為基礎的路由選徑方法，可稱之為「直接 TCAM 路由比對方法 (Direct TCAM Match)」。因此在本計畫中，我們也提出一套採用 TCAM 記憶體特性的硬體化操作路由選徑機制。相較而言，如上一節中採取階層式分群解析概念，或以完全展開樹為基礎的這一類方法 [2, 3, 6]，相當於是將原本路由選徑所需要的查詢演算方式進行轉換，成為一些單純、規律的邏輯電路運作和記憶體存取動作的操作程序，之後只要將 IP 封包標頭的 Destination IP 位址輸入，透過此規則化的操作程序即可以獲得路由選徑的結果。由於並不是直接透過一般 IP 路由選徑搜尋演算方法的過程，而是間接地透過轉換後的規則化操作程序即可獲得等效於路由選徑搜尋演算法的結果，因此也被歸類稱為「間接路由查詢方法(Indirect Lookup)」。

隨著半導體技術的進步，記憶體的種類與功能也不斷推陳出新，除了運作速度與記憶容量之單位面積密度的提升與價格的下降之外，也從最初單純的資料儲存目的發展至以專屬應用為主的功能性記憶體，「內容定址記憶體 (Content Addressable Memory, CAM)」即是做為「資料搜尋」用途的專門記憶體。其應用的方式為：記憶體中的每一個儲存單位存入的為某一應用的一筆候選內容資料以及該內容的相關聯數據資料，待該應用需要進行內容搜尋的時候，僅需將該內容輸入記憶體做為定址用途，CAM 記憶體自動會將內容比對吻合的該儲存單位的內容相關聯數據資料輸出。省去在傳統記憶體架構的操作程序中，必須自行將記憶體中的候選內容資料一一按址取出並分別比對，再將內容吻合項目的相關聯數據資料另外按址讀出所必須花費的時間。而具三元資料比對能力的 TCAM (Ternary CAM) 記憶體的提出，更是讓 CAM 記憶體的資料搜尋能力因為具備更彈性的應用方式而進一步提升：儲存的內容可以包含 don't care 萬用字元(*)，因此可做到多對一的模糊化搜尋方式，也就是一筆候選內容可以包含多個可能性，輸入資料時不再需要完全吻合候選內容才會得到輸出，只要與候選內容近似，在其包含的可能性範圍內，就可以得到相對應的關聯數據輸出。這樣的資料搜尋應用方式與 IP 路由選徑有著相近似的運作方式 (Routing Table 中每一筆路由字首資料皆可視為是以萬用字元對應至一 IP 位址區段，也就是多個 IP 位址，而同一 IP 位址可能會被多個路由字首所對應的 IP 位址區段範圍所涵蓋。當欲查詢某一 IP 位址的路由結果時，根據 Longest Prefix Match 的原則，即是將包含到此待查 IP 位址的所有路由字首都搜尋出來並比較其字首長度，而以最長字首所對應的路由結果做為此待查 IP 位址的路由選徑查詢結果)，因而使得 TCAM 也開始被考慮應用在 IP 路由選徑的機制中。

以 TCAM 為基礎的 IP 路由選徑方法即是一個硬體化操作的路由選徑方式，具備有多方面的優點：由於屬於硬體的運作架構，所以運作的速度相當快；也因為其本身的操作特性即相當符合 IP 路由選徑搜尋演算模式，所以僅需要搭配相當簡單的周邊邏輯電路便能夠進行路由選徑查詢的應用，而運作所需要的資料結構的建立與更新速度也很簡單、迅速，可以直接使用 Routing Table 的路由字首資料而不需做任何轉換或處理的動作，所需要額外的記憶體容量也很小。然而其目前唯一、也屬重大的缺點是，價格仍然過高，使得其雖然具備執行路由選徑機制最佳且優秀的能力條件，但是真正商用化的路由器仍未見有採用其做為路由選徑機制的應用。在本研究中，我們提出了一個整合直接 TCAM 路由比對和間接路由查詢方法的路由選徑演算法，而設計的主要動機與概念便是：充分利用 TCAM 的特性並兼顧其價格缺點，整合間接路由查詢方法，並採用兩者平行處理、分工合作的概念，截長補短—以間接路由查詢方法彌補 TCAM 因價格高而數量不足，無法完全負擔路由選徑搜尋應用需求的缺點；利用少量的 TCAM 搭配 Priority 處理邏輯單元，來負擔部分（路由字首長度較長的）路由資訊的路由選徑應用，減少間接路由查詢方法所需負責的路由資訊數量，因而降低其記憶體需求。而此新的複合式路由選徑方法，最多只需要 2 次的記憶體查詢時間，便可以得到路由選徑的結果，而且也一併減少路由搜尋資料表的更新時間。其較細部的設計與運作程序如下面的內容所述。

令 l_i 和 h_i 表示路由器 (Router) 路由表格中第 p_i 筆路由字首的長度及其對應的路由器輸出埠。我們所設計採用 TCAM 記憶體特性的硬體化操作路由選徑方法的架構圖如圖 0 所示：上半部是為一個既有的間接路由查詢方法，處理 l_i 小於或等於 24 的路由字首 p_i ，下半部即是直接 TCAM 路由比對方法，處理 l_i 大於 24 的路由字首 p_i 。此兩部份在實際的操作中是為平行處理的運作方式，對於一筆輸入欲進行路由選徑查詢的 IP 位址，會同時被輸入至兩部份。若只有上半部分的 Indirect Lookup 有輸出，則 Selector 單元會將此結果直接做為自己的輸出，成為該待查 IP 路由選徑的最終結果；若上下兩部份都分別得到路由選徑的結果，則 Selector 單元將會因為 Longest Prefix Match 的路由選徑原則，而以下半部分 Direct TCAM Lookup 的結果做為自己的輸出，同時也表示是該待查 IP 路由選徑的最終結果。上半部的 Segment Table 是存放以 IP 位址的前 16 位元為第一分群解析階層展開後的每一分群 IP Segment 所對應的路由結果 Pointer；而邏輯處理單元 (Logic Process Unit) 則是根據 Segment Table 所輸出的路由結果 Pointer，進一步指向 ADH (Associated Default Hop) 取得對應的路由選徑結果（也就是路由輸出埠），或是指向儲存著下一層次分群解析的路由選徑結果的 NHA (Next Hop Array)。而下半部的 1st、2nd、3rd 和 4th TCAM 表示 4 群 TCAMs 硬體單元，分別儲存並處理 Routing Table 中字首長度為 25 至 26 bits、27 至 28 bits、29 至 30 bits 和 31 至 32 bits 的路由字首，並依序具有由小至大的輸出優先權；Priority Resolve Unit 單元則如同一個 Filter，依 1st、2nd、3rd 和 4th TCAM 的實際輸出情形，選擇當中具最高優先權的 TCAM 「有效」輸出做為自己的輸出，若四個

TCAM 皆無有效輸出，則 Priority Resolve Unit 便以一個預設輸出代替；Associated Memory 存放四個 TCAM 中所有（長度大於 24 bits 的）路由字首所對應的路由選徑結果（也就是路由輸出埠），並會根據 Priority Resolve Unit 的結果，輸出所對應的路由選徑結果；若 Priority Resolve Unit 的輸出是為預設輸出，則 Associated Memory 將不會有輸出。

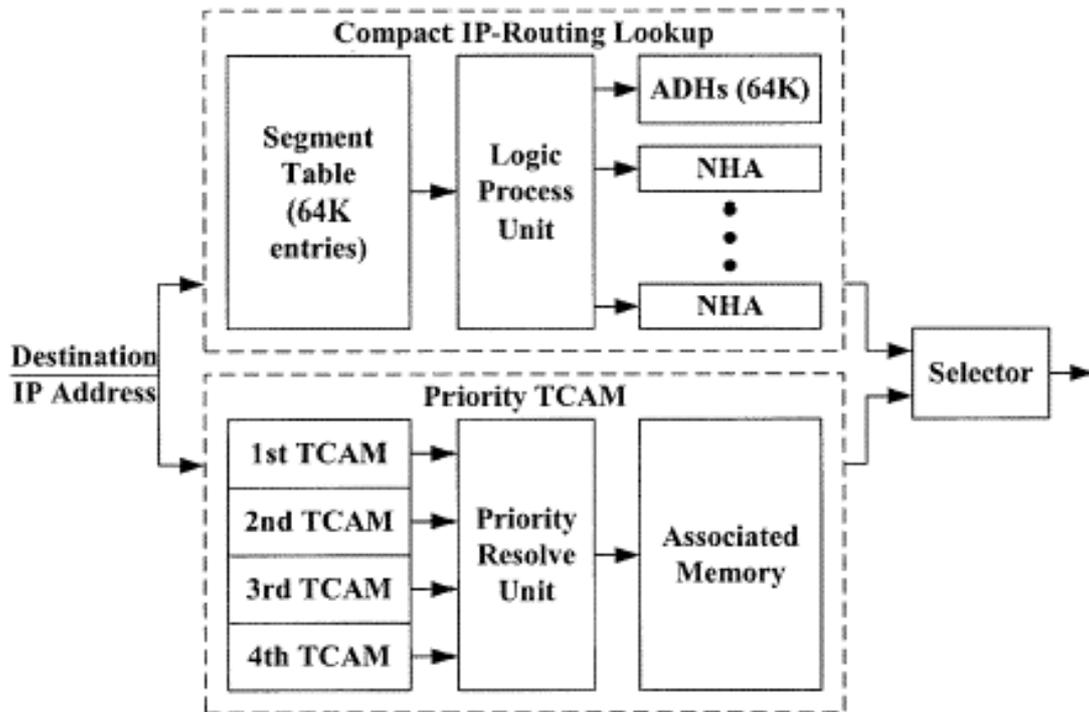


圖 0: Priority TCAM IP 路由選徑機制的功能架構圖

令 $p_i(x, y)$ 表示 p_i 的 x 位元至 y 位元。表 0 顯示 IP segment 192.168 的對應路由表。因為 192.168 IP segment 的最大長度 l_i 為 32 bits。因此若利用傳統上單純的間接路由查詢方式 Huang's scheme [3] 來查詢的話，此 192.168 IP segment 需要 $2^{(32-16)}$ 個對應的 NHA 路由器輸出埠 entry。利用我們提出的複合式路由選徑查詢方法（如表 0 的最後一個欄位所示）：我們把 l_i 大於 24 bits 的路由字首 p_i ，用直接 TCAM 路由比對方法來處理；把 l_i 小於或等於 24 bits 的路由字首 p_i ，才由間接路由查詢方法來處理。因此在間接路由查詢方法這部分只需要 $2^{(23-16)}$ 個（因為把 l_i 大於或等於 24 bits 的路由字首 p_i 排除，最大的 l_i 為 23 bits）對應的 NHA 路由器輸出埠 entry，大幅降低 NHA 路由器輸出埠 entry 的記憶體需求，同時也把最長的 IP 路由選徑查詢的時間降低在 2 次記憶體存取的時間內。

Entry i	Routing Entry $P_i / l_i / h_i$	$p_i(x, y)$ $x=17, y=24$	Compact IP-Routing Lookup or Priority TCAM
0	192.168 / 16 / 0	xxxx xxxxb	Compact IP-Routing Lookup
1	192.168.20 / 22 / 1	0001 01xxb	Compact IP-Routing Lookup
2	192.168.84 / 22 / 2	0101 01xxb	Compact IP-Routing Lookup
3	192.168.68 / 23 / 3	0100 010xb	Compact IP-Routing Lookup
4	192.168.68.16 / 28 / 4	0100 0100b	Priority TCAM
5	192.168.68.16 / 32 / 5	0100 0100b	Priority TCAM

表 0: Routing prefixes of the 192.168 segment

不僅如此，我們還進一步提出一個「同值位元整合壓縮法(Common Bit Integration)」，來改善間接路由查詢方法的 (NHA) 記憶體需求。以表 0 中 entry 1 至 entry 3 (l_i 小於或等 24 bits 的路由字首) 為例，其第 17、19、21 和 22 位元是相同的，因此我們只需紀錄第 18、20 和 23 位元的 3-bit 的 pattern 即可，因此對應的 NHA 路由器輸出埠 entry 可進一步由原來的 $2^{(23-16)}$ 減少至 $2^{(23-16-4)} = 2^3$ ，更進一步減低所需要的記憶體大小。

圖 0 表示 TCAM 硬體單元的架構圖。TCAM 硬體單元由路由字首 p_i 暫存器，對應的 p_i mask bit pattern、32 個 3 位元比較器，和 1 個 32 位元的 AND 邏輯運算單元所構成。由於 TCAM 硬體單元完全是由硬體構成，因此以其為基礎的路由選徑方法比起同為硬體操作架構的間接路由查詢方法而言，仍是具有較快的運作速度。

以網際網路上運作的實際 IP 路由表為例，表 0 列出三個大型 ISP 處的實際 Routing Table 的資料，以及兩種具代表性的間接路由查詢方法 (Huang's scheme 及 Chen's scheme) 與我們提出的複合式路由選徑查詢方法對記憶體需求的比較。由表中可見，增加少量的 TCAM 硬體單元，所需的記憶體可大幅降低，不僅使 IP 路由輸出埠的查詢時間降低在 2 次的記憶體讀取的時間內，亦可降低更新搜尋的資料表的時間。

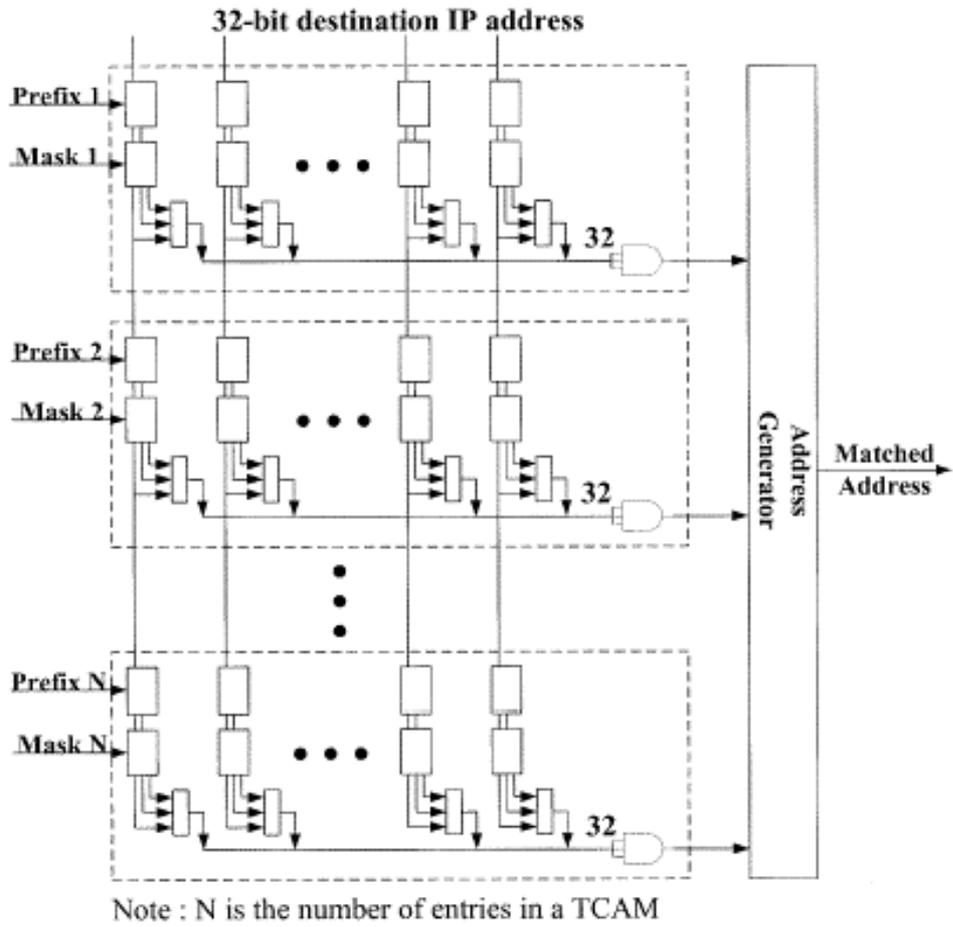


圖 0: TCAM 硬體單元的功能架構圖

	AADS	Mae-West	PAIX
Prefixes	33,931	37,523	18,569
Segments	5,813	6,126	3,571
Length > 24	431	433	443
Huang's [3]	599K	610K	507K
Chen's [6]	659K	781K	543K
Priority TCAM IP-Routing Lookup	423K +431 TCAM	463K +433 TCAM	377K +443 TCAM

表 0: 各種路由選徑方法所需記憶體容量之比較

1-3. 高速群播路由選徑(Multicast Routing) 機制

群播路由選徑(Multicast Routing) 與一般單一路由選徑(Unicast Routing) 的不同點在於：一個 Multicast IP 即代表著一個唯一的 Multicast Group，其包含(對應到)多個分佈在不同區域的主機(Host)，並非如其他(一般的) Unicast IP 多半是以一個連續的區段(Segment) 為單位，來指派、對應到某一區域的單一主機。所以，①在路由表格中，Multicast 的 Routing entry 是以(單一)完整長度的路由字首來呈現，與一般 Unicast 的 Routing entry 是以不完全長度的字首來代表一個連續 IP 位址區段的形式有所不同；②同時，其路由選徑結果也不再只是對應到單一的輸出埠，而可能是一「組」的多個輸出埠；③另外一個與 Unicast Routing 形式上較大的不同是，為了讓 Multicast 的運作更有彈性，可以自由地根據不同的資料發送端來設定不同的資料群播遞送的方式，Multicast Routing 採取多(雙)欄位資料搜尋比對的路由選徑動作，除了 Multicast 的 Destination IP 之外，也必須同時根據 Source IP 位址來決定路由結果輸出埠為何。因此，Multicast Routing 的關鍵技術是為一個高速的多欄位資料完全相符(Exactly Match) 的比對搜尋機制，並且能同時迅速地解析出所需要進行封包轉送的多個路由輸出埠。從高速單一路由選徑方法的設計經驗得知，內容定址記憶體 CAM 和 TCAM 本身即可視為一個簡單而易於實現的硬體架構的資料比對搜尋裝置，而 CAM 的運作方式即相當於一個資料完全相符(Exactly Match) 的比對搜尋機制，因此相當適合應用在

高速硬體化操作的 Multicast Routing 路由選徑機制的設計中；此外，有鑑於 Compression Bit Map (CBM) 的概念對於單一路由選徑機制運作所需的資料結構記憶體空間上的壓縮有很大的幫助，因此我們同樣將此概念應用在群播路由 (Multicast Routing) 的選徑機制設計上，以有效降低、控制路由表格的大小。綜合上述的考量，我們提出一套以 CAM 為基礎，並採用 CBM 技術為輔的高速硬體化操作的 Multicast Routing 路由選徑方法。此方法的群播路由查詢速度可以相當快，一次群播路由查詢只需要三次的記憶體存取次數(Memory Accesses)，再加上管線式(Pipeline) 運作架構的可行性，也同樣能使得平均查詢次數降至一次的記憶體存取次數。其細部的設計與運作如下面的內容所述。

整個群播路由表格的組成基本上包含來源位址(Source Address)、目的位址(Destination Address) 以及由此兩個位址所決定出唯一的一組輸出埠號碼(Output Port Numbers)。根據群播路由查詢法則，當路由器收到一個封包，發現其目的位址為群播網際網路位址(Multicast IP Address) 後，接著必需檢查其來源位址，確定這是此群播組(Multicast Group) 裡的成員，以及是哪一位成員所發出的封包，由這兩個位址便可決定出一組輸出埠號碼；每一個號碼表示路由器必須要複製一份此封包往這個輸出埠送。基本上，我們以底下所示的方式表示一個群播組和其所包括的成員：

Group G1: Sources H1, H4, H6, H7, H8.

Group G2: Sources H3.

Group G3: Sources H2, H5, H8.

G_x 表示目的位址， H_x 表示來源位址。根據這個法則，我們把路由器上經由路由協定(Routing Protocol) 所獲知的來源位址和目的位址(這裡特別是針對對應至 Multicast Group 的 Multicast IP 位址) 作成一個二維空間的棋盤狀對應表，此表格的縱軸表示來源位址，橫軸表示目的地位址，而兩軸的交叉點基本上有一個圓圈，黑色圓圈表示這是一個有效的群播組和組員的關係，(群播組(Multicast Group), 組員(Membership Source))，白色圓圈則表示是一個無效的群播組和組員的關係，如圖 0(a) 所示。將這些圖形資訊做一個二進位數字編碼，黑色圓圈為 1，白色圓圈為 0，便可形成如圖 0(b) 的一串二進位數字串列。接著我們將此種 Bit Map 的資訊編碼方式也同樣運用在路由輸出埠組上，我們將每一個有效的 (Multicast Group, Membership Source) 關係所對應的輸出埠號碼組作以 Next Hop Bit Map Stream 的形式來表示，每一筆 Next Hop Bit Map Stream 的每個位元依序對應到一個輸出埠，而以其位元值表示是否要複製一份封包往這個輸出埠送，位元 1 表示需要，位元 0 表示不需要，如圖 0(c) 表格中每一個 Entry 所示。最後再將所有有效的 (Multicast Group, Membership Source) 所對應到的 Next Hop Bit Map Stream 以一個如圖 0(c) 所示的 Next Hop Bit Map Stream Array 陣列來依序置放，做為查詢之用。

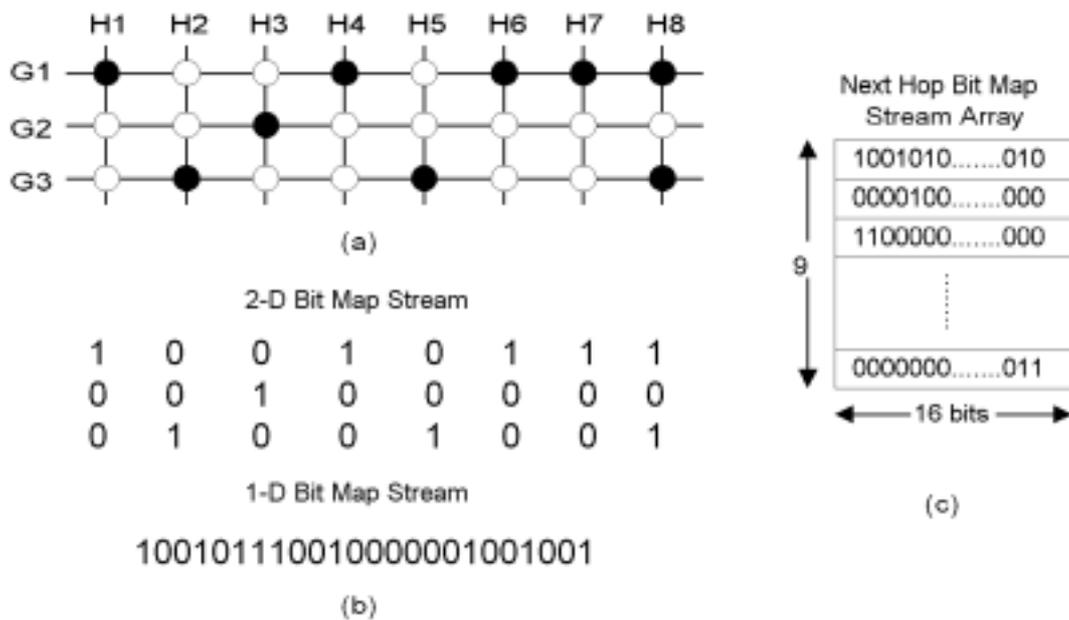


圖 0: (a) The grid to represent the existence information of the (group, source) pairs, (b) The pair bit map streams, (c) The next hop bit map stream array.

進行 Multicast Routing 路由選徑查詢時，先要根據封包標頭的 (Source IP, Destination IP) 資訊在這個二維空間棋盤狀對應表中尋找其對應的圓圈的幾何位置，接著再從棋盤最左上邊的第一個交會點開始計算到此對應位置前位元值為 1 的個數是多少，此數目即代表從棋盤狀對應表左上角第一個交會點算起有效的 (Multicast Group, Membership Source) 關係組數，將這個數目對應到存有輸出埠資訊的 Next Hop Bit Map Stream Array 陣列中的位置，如此便可快速取得所需要的路由資訊。為使路由速度更快，在查詢由來源位址和目的位址所組成的幾何位置資訊上，採用「內容定址記憶體 Content Addressable Memory (CAM)」的方式來實作，並將原本圖 0(b) 的二進位數字串列作一個簡單切割以利實作上匯流排的資料寬度需求，形成如圖 0 的系統架構圖；這個架構適用於管線運作方式，因此也可大幅提高整體路由查詢的資料輸出率(Throughput)。

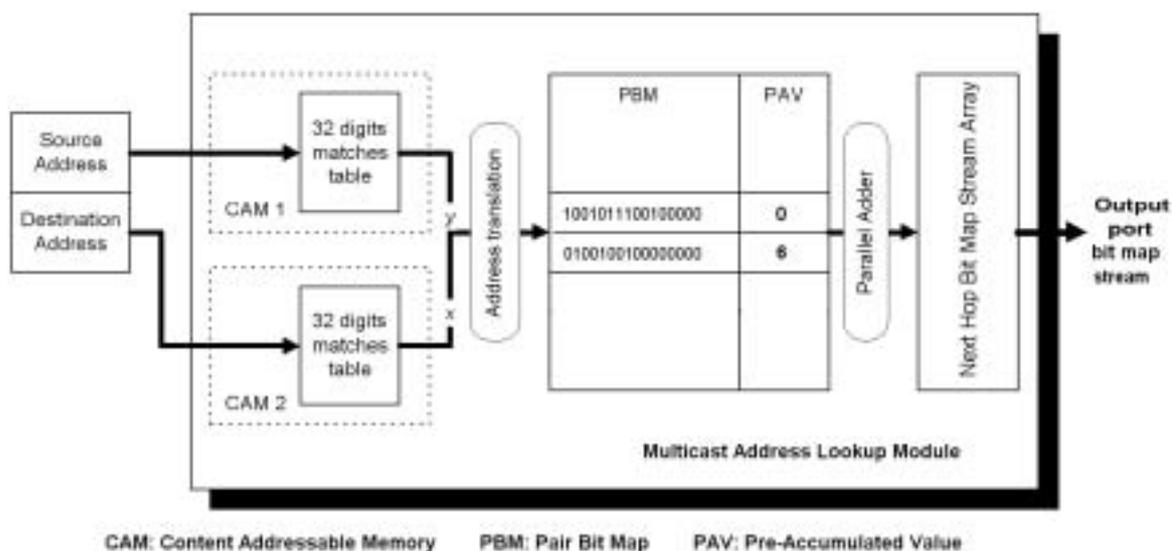


圖 0: The multicast forwarding table structure

2. MPLS 網路之 VC-Merge 機制的效能分析

在此，我們主要是設計一套有效的系統效能分析演算方法，針對 VC-merge 的 ATM-LSR 所需的緩衝器和 cell blocking 機率之間的關係進行分析，試著去探討具有 VC-merge 能力的 ATM-LSR 交換機需要比傳統的 ATM 交換機具備多少緩衝器資源，這其中包含了 Frame-level Interleaving 機制所需要的 ATM cell 重組緩衝器以及 ATM-LSR 原本既有的輸出緩衝器(Output Buffer, OB)。

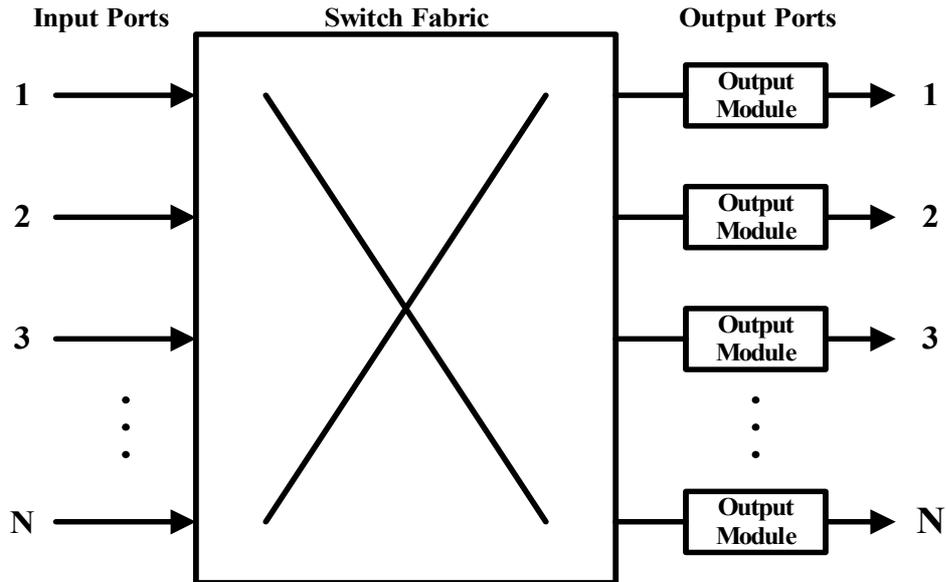


圖 0: Block diagram of a VC-merging capable ATM-LSR

圖 0 顯示一個具有 VC-merge 能力的 ATM-LSR 交換器架構;包含了 N 個輸入/N 個輸出的 ATM cell 交換單元及 N 個輸出模組(Output Module) 單元。圖 0 進一步顯示輸出模組的架構, 包含了 M 個重組緩衝器(Reassembly Buffer)、具 VC-merge 功能方塊、S 個依服務等級不同的輸出緩衝器(Output Buffer) 及 ATM cell 服務的輸出排序等。

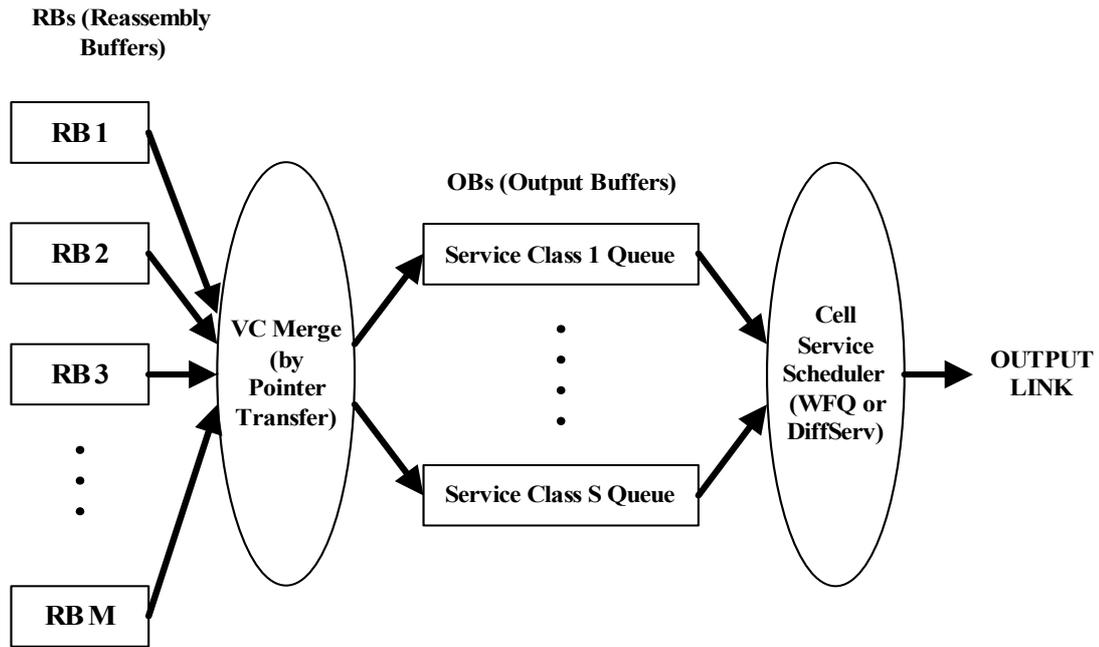


圖 0: Block diagram of the output module

圖 0(d) 為傳統不具 VC-merge 能力的 ATM-LSR 交換機形式，即便是欲前往相同目的網路或節點的 IP Route，其相對應的 ATM cell 在交換後仍是採用不同的 VPI/VCI 以區隔出其為分屬不同 IP Route 的訊務流，如此即便是對應至不同 IP Route 中 IP 封包的 ATM cell 有任意交錯傳輸的現象也沒有關係；圖 0(a)、(b) 及(c) 顯示三種不同型態的 VC-merge 形式：(a)是為 Full VC-merge，對於欲傳輸至相同目的網路或節點的 IP Route，將其所屬的 ATM cell 皆轉換成一致的 VPI/VCI 進行傳輸，(b)和(c)則為在進行 VC-merge 時，除了一致的目的網路或節點之外，額外考慮 IP Route 不同的服務品質或其他屬性，將欲前往相同目的網路或節點而且屬性相同的 IP Route，其所屬的 ATM cell 才會在交換後轉換成一致的 VPI/VCI 進行傳輸，此種 VC-merge 方式則稱為 Partial VC-merge。在此種 VC-merge 方式中，VC-merge 後不同 VPI/VCI 的 ATM cell 可以任意交錯傳輸沒關係，但是相同 VPI/VCI 的 ATM cell 仍是必須要遵循 Frame-level interleaving 機制，不可與對應至不同 IP Route 中某一 IP 封包的 ATM cell 交錯。圖 0 中的縮小數字表示起始的 VCI，底線代表 ATM EOM (End of Message) cell。例如 $\underline{5}_2$ 代表 EOM cell，原先 VCI=2 經過 ATM-LSR 交換機後轉成 VCI=5。

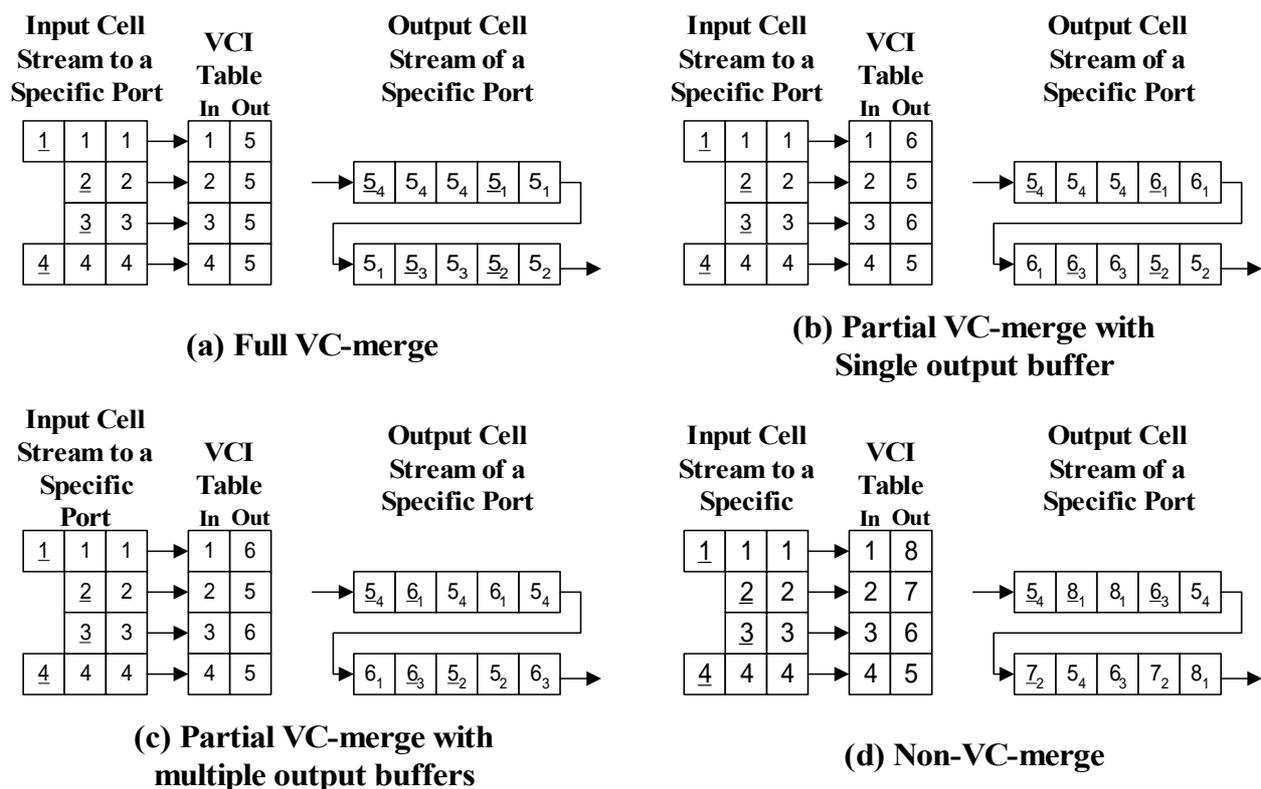


圖 0: Three types of VC-merging and one type of non-VC-merging

我們所設計的分析方法主要是參考論文[7]的方法並加以改進：在我們提出的系統中，放寬[4]的 Input Model，可允許 Cell-interleaving 的 Input Pattern，並且允許 Partial VC-merge 的運作方式。我們建立一個排隊理論數學模式 D-BMAP/D/1 來分析輸出模組的機率分佈。D-BMAP/D/1 的排隊理論模式為：

- M 個 IP Stream 可模擬成 M 個 ON-OFF sources，ON 的時間和 OFF 的時間為幾何分佈(Geometrical Distribution)；
- 在 ON 的時段裡，IP Stream 會產生 1 個 ATM cell 的機率為 r 。在 OFF 的時段裡，IP Stream 不會產生任何一個 ATM cell。參數 $(1/r)$ 代表 ATM cell interleaving 的程度；
- 輸出服務模式為每個 ATM time slot 處理完 1 個 ATM cell。

研究結果顯示，相較於論文[7] D-BMAP (Discrete – Batch Markovian Arrival Process)/D/1 的排隊理論模式，我們所設計的分析方法其計算複雜度可由 $O(M^4)$ 減少為 $O(M^2)$ ，其中 M 為所模擬的 ON-OFF Source 的個數，並且更進一步多引進了一個參數 r ，來描述 IP 封包的 interleaving 程度，使得數學模型更接近實際

的情況，所得的分析結果也較為接近實際。複雜度較低的結果，使得對較大的 Buffer Size 分析也能夠得到較高的 Cell Loss Prob.的準確度(尤其 Cell Loss Prob. 通常低於 10^{-6} 以下) 我們也利用了 Moment-Generation Function 的理論方法來近似 D-BMAP/D/1 的排隊理論模式，求得一個精確度蠻高的、接近前述所設計的分析方法結果的輸出緩衝器(Output Buffer) 中 ATM cell 數量分佈的數學近似方程式，可更快速計算 Cell Loss Prob.。分析及模擬結果顯示 ATM-LSR 交換機需要具備比傳統 ATM 交換機多 50-70%的緩衝器資源來支援 ATM cell 重新組合使用，以及避免 VC-merge 後無法在目的網路或節點處分離出不同 IP Route 訊務的問題。

圖 0 顯示透過數學分析及電腦模擬一個支援 VC-merge 功能的 ATM-LSR 其總緩衝器 overflow 機率分佈的結果。可以發現，我們的數學分析結果和電腦模擬結果相當一致。由圖 0 我們可得到以下結論：

1. 具有 VC-merge 能力的 ATM-LSR 交換機較傳統 ATM 交換機需要較大的緩衝器資源。如圖 0 中顯示在 overflow 機率為 10^{-5} 次方的假設下，傳統 ATM 交換機需要約 390 個 ATM cell 緩衝器；ATM-LSR 交換機需要至少 520 個 ATM cell 緩衝器。
2. 在細胞交錯 cell-interleaving 愈嚴重的情況下(r 參數愈小)，需要更多的重組緩衝器來重組 IP 封包，因此需要更多的緩衝器資源。在一般的 cell-interleaving 程度的訊務下，約需要比傳統 ATM 交換機增加 50%-70%的緩衝器資源。

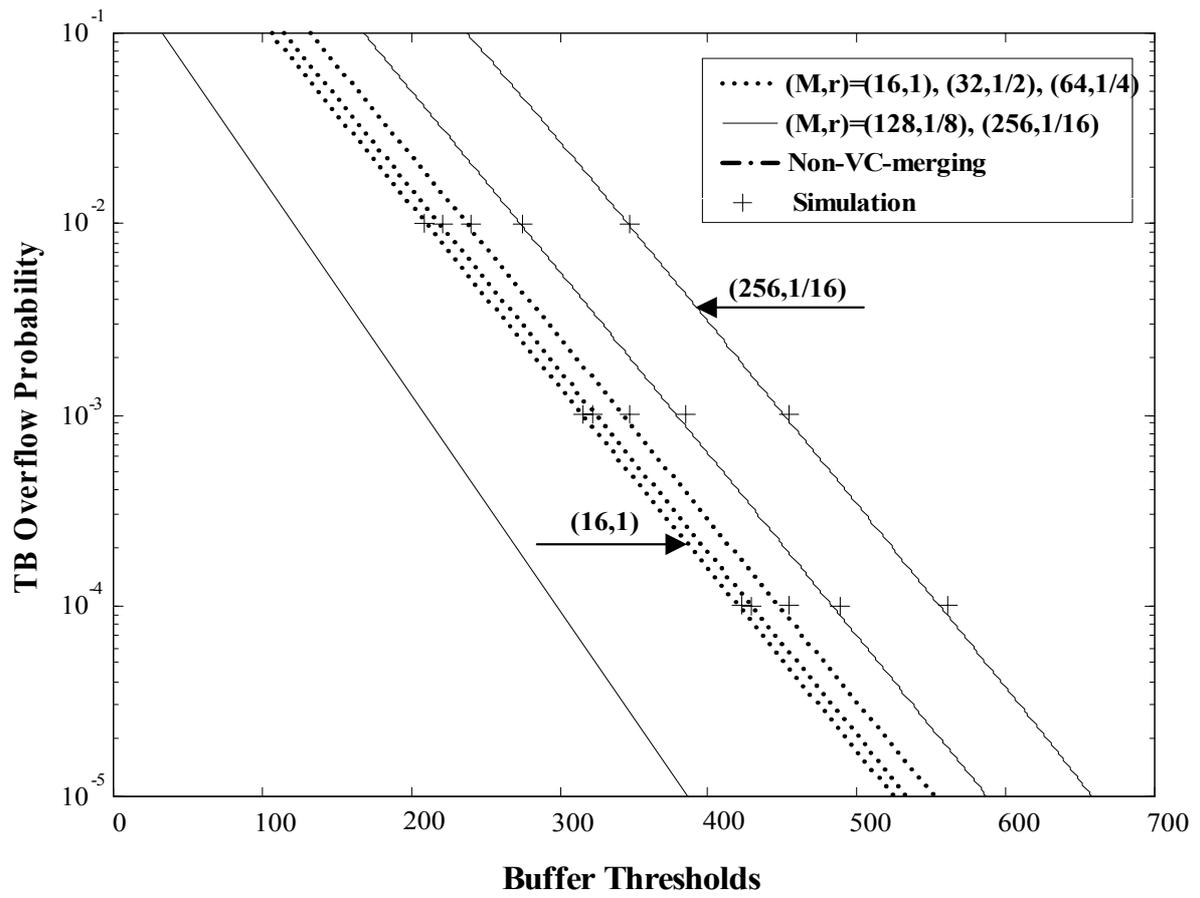


Fig. 0: Overflow probability of the total buffer versus buffer threshold

3. MPLS 網路之路徑保護(Protection)及快速回復(Path Recovery) 機制

這部分的研究主要在於提出一套有效的 MPLS 網路之路徑保護 / 回復機制，以便於 MPLS 網路傳輸路徑發生錯誤或損壞時還能夠維持部分基本的通訊，並可以快速而正確地恢復既有的通訊，降低 MPLS 網路上傳輸路徑錯誤或損壞所帶來的影響，減少封包遺失率，並期望能夠進一步達到動態負載平衡的附加效益，使系統資源做最佳的利用，如此便可有效的提高網路的資料輸出率 (throughput)。

Recovery Model	Backup Path Type	Backup Path Establish Point	Switch-over Processing	Recovery Time	Resource Utilization/Optimization
Re-Routing	Establish-On-Demand (Simple-Dynamic, Shortest-Dynamic)	After Fault	Local	Slowest	High
	Pre-Qualified			Medium	Medium
Protection Switch	1+1 (Haskin Algorithm, Makam Algorithm)	Before Fault	Head-end	Fast	Lowest
	1:1, 1:n, m:n*			Fast	Low

* m : n = (Backup path number) : (Working path number)

表 0: 兩種 Path Protection/Recovery Model 之比較

一個完整的路徑保護 / 回復機制應包含有幾個要素：單一或多條的工作路徑(Working path) 與備援路徑(Backup path)、路徑狀態監控 / 錯誤偵測和警告通知機制、以及路徑回復程序啟動時，切換到 Backup path 的機制。IETF 提出了兩種路徑保護 / 回復機制的運作模型(Model) [12, 9]，也可做為許多 Path Protection/Recovery 方法的兩大分類依據，其主要是根據 Backup path 建立的時間點做為區隔：如表 0 所示，在路徑發生錯誤之前即預先規劃並建立好 Backup path 的方式稱之為「Protection Switch Model」；相對的，在路徑發生錯誤之後才開始建立適當 Backup path 的方式便稱為「Re-Routing Model」。「Protection Switch Model」由於採取事先規劃並建立好 Backup path 的方式，因此當發生路徑錯誤時可以迅速地切換至 Backup path 以回復正常通訊，然而由於相關的系統頻寬資源也會是在建立 Backup path 時即被保留，等待路徑發生錯誤時可以立即切換專

用，造成系統資源的使用效率最低，因為需要預先保留做為備援專用的相當數量的資源在一般時候是閒置的。也由於必須耗費較多系統資源，因此 Backup path 多半僅在靠近通訊兩方的頭端(Head-end) 網路節點建立。「Re-Routing Model」的傳統方式（如 Simple-Dynamic 和 Shortest-Dynamic）是當路徑錯誤時才開始搜尋並建立適當的 Backup path，因此路徑回復（切換至 Backup path）的速度較慢，但也因為不事先保留資源所以可以讓系統資源做最有效的利用，而且可以在距離路徑錯誤當地(Local) 的最近或較近的節點（LSR）即進行路徑回復，不需要再回到頭端節點才來進行，對路徑錯誤的反應速度較快且靈活度佳。其在路徑錯誤的 Local 節點建立至通訊終端的 Backup path 的方式，也對於路徑回復後因路徑錯誤影響而需要重送的資料量減少。而分類上仍歸屬於「Re-Routing Model」的 Pre-Qualified 方式則其實是介於兩類方法間的機制，兼具兩者的優點，其關鍵在於：以 Re-Routing Model 為主，但是引用 Protection Switch Model 預先建立 Backup path 的觀念，在路徑錯誤發生之前先搜尋、規劃好可作為備援 Backup path 的數個候選路徑，一旦路徑錯誤發生時即能夠透過簡單的資源保留與設定的動作，便可迅速地建立 Backup path 並恢復原有的通訊。由於並未在事先建立 Backup path 並保留資源，僅是預先搜尋、規劃候選 Backup path 而已，因此資源仍可做充分的利用而不浪費，並且也可以預先在 Working path 沿線的各個 LSR 節點上分別搜尋、規劃至通訊終端節點的 Backup path，保有 Re-Routing Model 中在路徑錯誤的最近幾個 LSR 處即可以快速反應的優點。如此再配合上前述當錯誤發生時可節省搜尋時間並快速建立 Backup path 的優點，更加使 Path Recovery 的速度獲得更大幅度的提升。

著眼於兩種 Path Protection/Recovery Model 各有其優缺點，因此我們嘗試結合兩邊的優點，以屬性介於兩者之間的 Re-Routing Model 作法中 Pre-Qualified 類型的網路路徑錯誤復原演算法為主，提出一套高速而最佳的路徑保護 / 回復機制的�方法。然而在實際的操作上，由於 Pre-Qualified 方法對於 Backup path 的形式與數量並沒有明確的定義或限制，因此我們可以，也有必要根據 Pre-Qualified 方法的原則，設計、定義出一套適當並有效率的實際運作方式。如同之前內容所提，除了在通訊的起迄的兩終端 LSR 節點之間於通訊開始之前預先規劃 Backup path 之外，我們也在 Working path 沿線的各個 LSR 節點上分別搜尋至通訊終端節點的 Backup path，然而此種 Backup path 的搜尋動作並不需要在通訊開始之前就進行完畢，可以在通訊開始之後才進行，以避免過多通訊前的程序延遲了通訊開始的時間，而且也不要求每一個沿線 LSR 節點都必須要找到 Backup path。若發生路徑錯誤時其最近一 LSR 節點無法或尚未搜尋到 Backup path，便可通知其上游 (Up-stream) LSR 節點進行 Backup path 的建立與通訊的恢復；若此上游 LSR 節點亦尚未搜尋到適當的 Backup path，便可以再往其上游的 LSR 節點進行通知，直到有上游 LSR 已經找到 Backup path，或是到最上游的通訊終端 LSR 節點處為止（這裡必定有 Backup path，是通訊開始前即預先搜尋、規劃好的）。由於在搜尋到 Backup path 之後，並不需要在路徑錯誤發生前對其所需的資源加以保留，然

而又必須確保當路徑錯誤發生時 Backup path 的「可用性(Availability)」一路徑連線正常且仍有足夠資源可立即建立 Backup path，因此必須要有一套有效的 Backup path 可用性的監控、維護方法。最簡單的方式即是以一套週期性的信令(Signaling)方式，對於預先選擇、規劃好的 Backup path 的通訊路徑連線狀況以及剩餘資源情形進行持續的監控，一旦可用資源不足建立 Backup path 的需求，或在此路徑上亦同樣發生路徑錯誤，便重新搜尋可用的其他路徑作為 Backup path，以保持 Backup path 在 Working path 路徑錯誤時的可即時使用性。此外也可以進一步對於 Backup path 所需的資源先進行軟性保留(Soft Reservation)，亦即對 Backup path 所需要的網路資源有進行預先保留的動作，然而與 Protection Switch Model 不同的是，此保留的資源在 Backup path 尚未真正建立使用時，仍可允許被其它的通訊利用而不因此造成資源的浪費（特別是在有 traffic priority 的網路中，priority 等級低於目前連線的通訊而言），只有當 Working path 路徑發生錯誤而必須啟用 Backup path 時，便以優先使用權的身份將此預先軟性保留的資源收回，以快速而順利地建立 Backup path。在此方式中，Backup path 狀態監控機制僅需針對其路徑連線狀況進行偵測即可，省去可用資源方面的監控，如此可降低監控機制的複雜度與系統資源負擔(overhead)，也使所選定的 Backup path 穩定度增加，較不容易時常需要進行重新搜尋、更換（路徑），而 Backup path 也僅需要一個簡單的通知性(Notification) 信令經過 Backup path 上的所有 LSR 之後即可以馬上建立使用，有助於備援切換速度的提升。而唯一的代價是，當 Backup path 建立並進行切換時，對於原來位於 Backup path 上其他通訊的影響較大，尤其是 Traffic priority 等級低於此連線的通訊。另外，Backup path 的可用性監控、維護方法還可以根據單一節點所建立的 Backup path 數目發展出多種變形，例如：一開始即選定多個候選 Backup path，並對所有的候選 Backup path 都進行監控，確保有需要時至少有一個 Backup path 能立即使用；或是仍僅對其中一個進行監控，若剩餘資源不足或路徑錯誤再換另一個，如此只是節省重新搜尋的時間而已。一般來說，對於比較重要的 Backup path（例如如通訊開始之前，在兩通訊終端 LSR 節點之間必須預先規劃的 Backup path），可以採用軟性資源保留的方式，以較穩定的 Backup path 條件確保其立即可用性；而其餘的 Backup path 便可以採用最簡單的監控方式，即時可用性雖然會較 Soft Resource Reservation 方式稍差，也會有較多的監控資源負擔(overhead)，但是當路徑錯誤發生時，對於 Backup path 上其他通訊的影響則是最小的。

在確定 Backup path 的實際運作方式之後，我們設計採用網路第三層中之 OSPF 繞徑協定做為搜尋適當 Working path 與 Backup path 的基礎 [15, 16]，並利用 MPLS 中既有之 E-RSVP 通訊協定做為持續偵測 Backup path 線路狀態與資源使用情況的可用性監控機制 [18-20]。此外，還針對 Working path 著手，參考 Protection model 中 m:n 方式的觀念，利用「多路徑傳送」方式所帶來空間上的多樣性(Diversity)，來減少當錯誤發生時封包的遺失率，這也就是主動的 Path Protection 的能力，並同時也在此多條 Working path 上進行動態負載平衡(Load

balancing) 機制 [17]，充份利用 Working path 資源，提升網路資料輸出率 (Throughput)。

我們所設計的路徑保護 / 回復機制的運作步驟摘要如下：首先在通訊之前，先透過第三層網路的 OSPF 繞徑協定，配合我們對於通訊路徑品質（例如：頻寬、傳輸延遲、封包遺失率等）的需求，進行自通訊起始 LSR 節點處至通訊終端 LSR 節點處的 Working path 與 Backup path 的選擇。原則上我們是以 $N(N \geq 1)$ 條 Working path 與 $M(M \geq 1)$ 條 Backup path 做為通訊的基礎。接下來，便需要在所選擇的 M 條 Working path 上進行 Load balancing 的規劃：根據各 Working path 的 (Bottleneck) BW 與 Delay 來適當地分配其上傳輸的 Traffic 的比例，以確保能夠獲得最佳的通訊品質與 Throughput。

當規劃好 Load balancing 之後，便可以開始進行通訊資料的傳輸動作。而在通訊開始進行的同時，除了通訊起迄兩終端 LSR 節點之間 Backup path 的監控機制之外，位於通訊路徑 (Working path) 上各 LSR 處至通訊終端 LSR 節點處 Backup path 的搜尋與監控機制也開始運作：原則上我們會在 Working path 沿線的各 LSR 同樣以 OSPF 繞徑協定配合我們對於通訊路徑品質的需求，來找尋一條最佳的 Backup path，若能順利找到此 Backup path，便將該路徑資料存在 LSR 的 Database 之中，接下來則以監控機制確保其可用性，透過現有 MPLS 本身的 E-RSVP Signaling Protocol 以 Monitoring 的方法來代替實際上的資源保留與佔用，根據 RFC2205 的定義，E-RSVP 每隔 30 秒會重新檢查每個 LSR 節點與鏈路(Link) 的 State 狀況 [21, 22]，包含是否有 Node 或是 Link 被增加、移除或發生錯誤，以及其頻寬的使用與相關的 QoS 參數條件是否滿足需求等等。如此便可決定此 Backup path 是否可以繼續保留或需要重新搜尋，如結果為後者，則將再回到以 OSPF 搜尋 Backup path 的步驟。而當 Working path 發生路徑錯誤時，系統即在最近一個且 Database 具有 Backup path 紀錄的上游 LSR 處，進行 Backup path 的建立與訊務的切換。

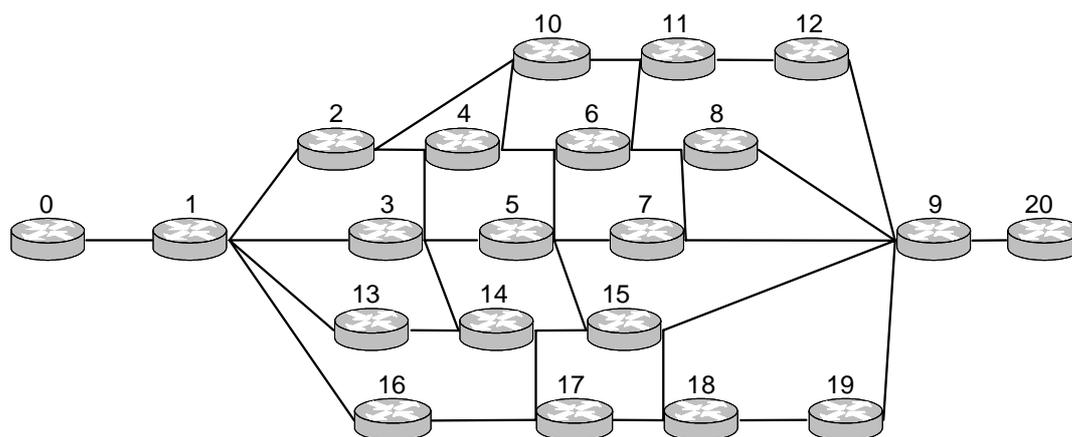


圖 0: Network configuration for NS-2 simulation

我們採用 NS-2 這一套網路模擬軟體平台，來對所設計的 Path Protection/Recovery 方法進行系統模擬與效能評估、驗證的工作，並與其他方法進行比較。模擬驗證時所採取的網路架構如圖 0 所示：通訊的起迄終端節點分別為 LSR0 與 LSR20，其中每個節點間的鏈路(Link) 頻寬為 10Mbps，並且在通訊的啟始節點 LSR0 處產生 5Mbps 的 CBR/UDP Traffic 進入網路中。圖 0(a) 是從 Working path 上幾處鏈路損壞(Link Failure)所導致的路徑錯誤狀況中，來比較各 Path Protection/Recovery 方法在封包遺失率上的表現。從圖上可以明顯的看出，我們所設計的方法 - E-Algorithm - 在 Packet Loss 上的確比其他的方法有較好的結果。因為我們以預先規劃的、離路徑錯誤處最近節點的 Backup path 資料庫的存取來取代重新搜尋 Backup path 的時間，因此 Packet Loss 便少很多；而 Makam 的方法則是必須透過 Signaling 方式告知頭端網路節點 LSR1 網路錯誤的狀況並進行 Backup path 的切換，因此一旦錯誤發生點離入口處的 LSR 愈遠，則通知的時間愈長，丟掉的 Packet 便愈多。Re-Routing Model 中的兩種方法 (Simple-Dynamic 和 Shortest-Dynamic) 因為是在路徑錯誤之後才搜尋 Backup path, 因此 Packet Loss 數量普遍較前兩者為高。Simple-Dynamic 則是因為在 Local 處尋找直接至目的地端的其他路徑做為 Backup path，因此路徑錯誤發生點離目的地端愈近，丟掉的封包數便愈少。在 Shortest-Dynamic 的方面則是因為利用 IP 層繞徑(Routing) 的方式將資料 Bypass 地繞過錯誤發生點至下一個 LSR 繼續傳輸，所以基本上與錯誤發生點沒有太大的關係，但是若路徑錯誤發生點離目的地端愈近，丟掉的封包數也是會減少。在圖 0(b) 我們可以發現，只有 Shortest-Dynamic 方式會因為採用 IP 層繞徑(Routing) Bypass 的方法而容易發生 Packet disordering 的封包亂序情況，而在最後則是因為網路 Topology 的關係才使得多數的方式都容易產生 Packet disordering。

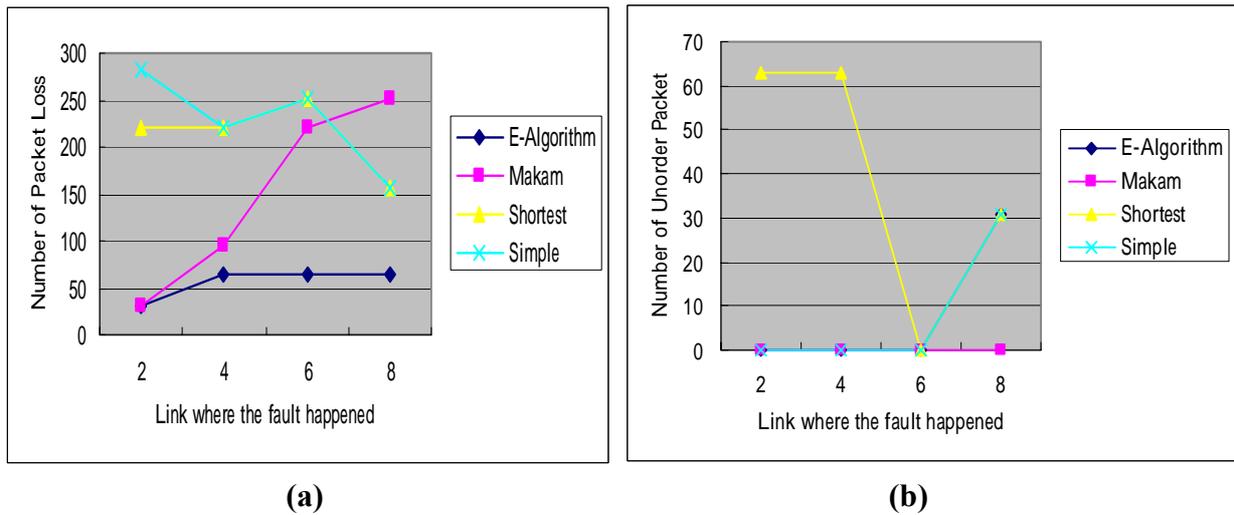


圖 0: Performance comparison on (a) Packet loss and (b) Packet disorder

圖 0(a) 顯示的是當 Working path 上 LSR5 與 LSR7 之間的 Link6 發生 Link Failure 前後，位於通訊終端處所見到的 Throughput 的「(時間) 暫態圖」。由此圖可以看到，Link Failure 發生在 0.8 秒的時候，而我們的機制在 Link Failure 導致的 Throughput 衰減量比起其他的方法都小許多，並不會使 Throughput 降到零，衰減速度也因為 Multipath 多重路徑的特性而較其他方法為緩，在錯誤發生到回復機制的運行過程中，我們依舊可以維持 High Throughput 的狀態，因而能在最短時間內即恢復正常的傳送資料輸出率。反觀其他的方法衰減速度較快，加上又都沒辦法很快完成接替備援的工作，導致 Throughput 皆會在某段時間內降到 0。因此在圖 0(b) 的平均 Throughput 比較圖看來，無論 Link Failure 發生於 Working path 的何處，我們的方法都可以獲得最高的 Throughput 值，而且這些值皆可維持在一個相當好而穩定的水準；相較之下，其他的方法則會因為 Link Failure 發生地點的不同而使 Throughput 的值有所差異。

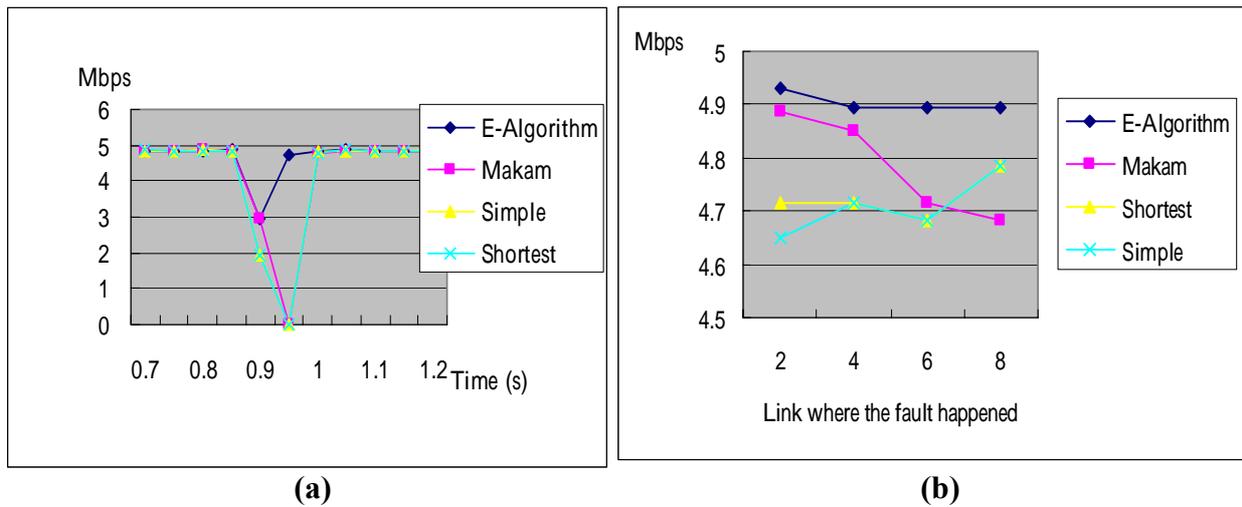


圖 0: Overflow probability of the total buffer versus buffer threshold

綜合上述的內容來看，我們在 Path Protection/Recovery 機制的設計上，利用每個 LSR 上預先規劃的 Backup path 的 Database 來幫助我們達成快速路徑回復的目的，取代了以往錯誤發生後才開始搜尋 Backup path 或是錯誤發生前先保留資源建立 Backup path 的缺點：藉由 Re-Routing Model 方法的使用，不但排除了 Protection Switch Model 此類作法中預先建立 Backup path 並保留資源而造成資源浪費的缺點，引入 Protection Switch Model 預先規劃 Backup path 的概念，加速了既有 Re-Routing Model 作法的 Recovery time，而且仍保有 Re-Routing Model 中在最近路徑錯誤的節點馬上進行切換的優點，因而能夠快速且有效地降低網路錯誤發生時之影響時間與所需付出的代價。然而此種 Path Protection/Recovery 機制需要一套有效的 Backup path 監控機制來確保 Backup path 可立即使用的特性，因此我們也配合 MPLS 網路既有的 E-RSVP Signaling Protocol 來進行 Backup path 狀態的監控。此外，同時選擇多條 Working path 進行資料傳輸，不僅享有多路徑傳送下 Path protection 的優點，即使有其中一條 Working path 發生問題，也不會在瞬間完全中斷資料的傳輸，維持通訊的可靠性與降低路徑錯誤的影響；同時可藉由適當的 Load balancing 機制的設計，獲得 Throughput 與傳輸品質提升的優點。最後，我們的方法可以讓 ISP 業者很自由而彈性地根據其服務品質的需求與使用者付費原則，評估採用不同複雜度的運作形式，在成本與演算法完整性之間權衡一平衡點。

4. DiffServ 網路中精確的訊務監控調節(Traffic Conditioner) 機制

這部分的研究主要在於提出一兼具訊務監控精確度與網路資源使用效率，並進一步達成 micro-flow 微訊務流之 **Cell Dropping Precedence** 標記公平性 (Fairness) 的訊務監控調節機制(Traffic Conditioner, TC)：在確保訊務特性符合 Traffic Profile 的規範下，精確地讓系統資源獲得充分且最佳的利用，並進一步保障其中各 micro-flow 連線所實際獲得的 QoS，如此才能夠提升 DiffServ 此網際網路 QoS 架構的可行性與使用效益，並做為其他 QoS 訊務控制機制(例如：CAC 連線允諾控制機制、Scheduling 排程控制、流量控制、壅塞控制等 Per-Hop Behavior)運作的基礎。IETF 在所建議的 DiffServ 網路 QoS 架構中，也如同 ITU-T 於 ATM 網路定義了 GCRA 的 UPC 訊務監控調節機制的參考模型一般，建議了三種訊務監控調節機制模型，分別是 Single Rate Three Color Marker (SRTCM) [34]、Two Rate Three Color Marker (TRTCM) [35] 與 Time Sliding Window Three Color Marker (TSWTCM) [36]。在初步的研究分析並參考過去我們在 ATM 網路上的發展經驗後，我們選擇以 TRTCM 方式的訊務監控調節機制做為基礎模型，來進行更高效能的 DiffServ 網際網路之訊務監控調節機制的發展與設計。

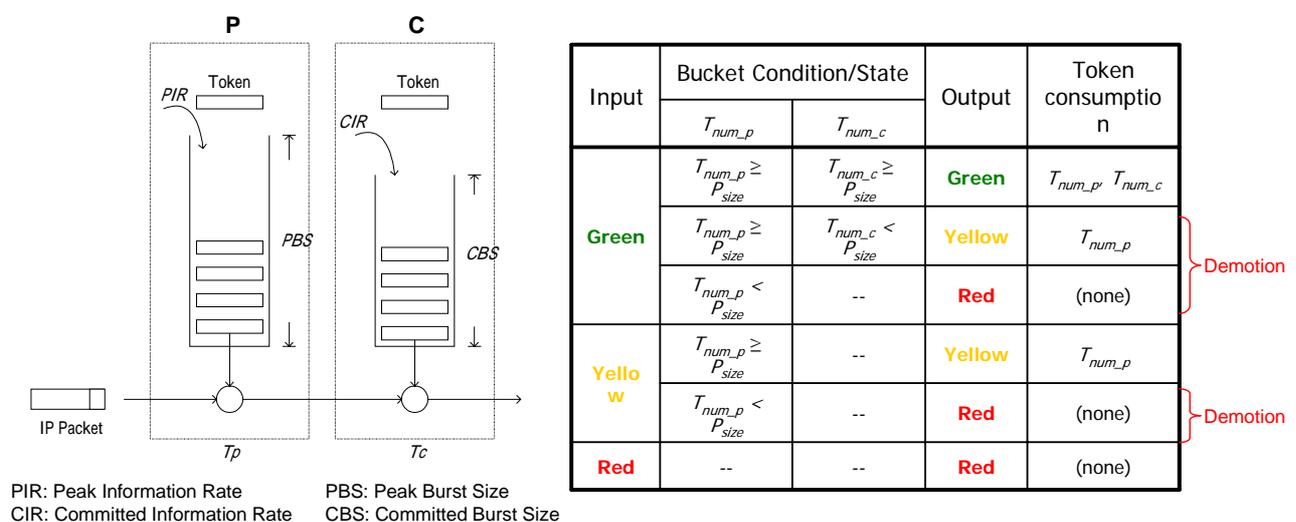


圖 0: TRTCM 之系統架構圖與運作邏輯表

如圖 0 所示，TRTCM 是藉由兩級的 Token Bucket 造成多階的訊務封包的合法性等級(Conforming Degree) (有別於傳統僅有合法與否的兩級判斷)，將進來的每一種服務等級(Service Class)的訊務封包皆根據其運作機制再區分並標記成 DiffServ 所定義的三種等級：綠色(Green)、黃色(Yellow)、紅色(Red)的封包出

去，並期望輸出後的訊務特性能夠符合系統對於各服務等級(Service Class) 的訊務所定義的 Traffic Profile 之規範。經過初步的驗證與分析程序後發現，TRTCM 的確是一個相當簡單而又實用的訊務監控調節方法，然而畢竟其只是一個參考模型（作法），所以在繼續深入探究之後發現仍有許多值得改進之處，其中一個重要的缺點為：當 TRTCM 運作在 Color-aware 模式下時，如果實際進入到 TRTCM 的訊務中，黃色封包與綠色封包訊務的總量違反 Traffic Profile 的協定超量使用，而且黃色封包訊務量對綠色封包訊務量的比值超過 Traffic Profile 中的比值時，則會因為共用第一級 Peak Rate Token Bucket 產生交互作用的關係，使得進來的合法綠色封包訊務量會嚴重受到黃色封包訊務量的影響，使得 Output 出去的綠色封包訊務明顯的減少且黃色封包訊務則超過 Traffic Profile 所定義之值，而不符合事先在 TRTCM 所設定的訊務合約。雖然在黃色與綠色封包訊務的總量上符合規範，但是讓許多綠色封包因為在第一級 Token Bucket 未有足夠的 Token 即被判定標記成違法的紅色封包，而將其原本該享有的 Token 轉讓給超量的黃色封包取得而達成黃色與綠色封包訊務總量的合法，實屬不恰當的設計，而且單獨的黃色封包訊務也會明顯違反 Traffic Profile 的規範。著眼於此，為了提供更精準的監測控制，對於上述的問題，我們基於「保護高 Conforming Degree 的綠色封包訊務不受 Conforming Degree 較低的黃色封包訊務影響」的原則，在原有的 TRTCM 運作機制作了細部的修改及調整，增加了綠色封包訊務的保護機制。綠色封包訊務保護機制的關鍵在於，允許第一級的 Token Bucket 進行「Token 預借」（或稱「Token 超量使用」）的機制：當綠色封包抵達時，若第一級 Token Bucket 中沒有足夠的 Token 數量可使用（←即上述被超量的黃色封包訊務所影響），但在第二級的 Committed Rate Token Bucket 中仍有足夠的 Token，則仍然將該綠色封包輸出為綠色封包，並在第一級的 Bucket 中進行 Token 的預借機制，也就是允許第一級的 Token Bucket 中的 Token 數量呈現負值。如此在上述的超量黃色封包訊務的情形發生時，將可以保障綠色封包訊務不受其影響而取得該享有的 Token，而超量的 Input 黃色封包訊務將會因為第一級 Token Bucket 的允許呈現負值狀態，而在一段時間後將其輸出的黃色封包訊務下降至符合 Traffic Profile 所規範的範圍內。由實驗數據中[39] 可得知，修改過後的 TRTCM 比原有的機制可以提供更精確而優良的訊務監測控制：在多種訊務組合情境下都可以維持並確保 Output 的黃色與綠色封包訊務的總量合法，同時也確保 Output 的綠色與黃色封包訊務必定分別符合其在 Traffic Profile 的規範，並保護綠色封包訊務能充分享有其應有的 Token 資源與網路頻寬；除此之外，所 Output 的訊務皆能夠呈現較原始 TRTCM 的設計更逼近理想值的最佳**訊務監控精確度**的狀態。

除了前述的綠色封包訊務保護機制外，在參考 [37] 此篇論文的内容後，我們也進一步進行「封包 Conforming Degree 升級(Promotion)」與「micro-flow 微訊務流公平性(Fairness)」的可行性研究與方法設計，完成一兼具優良的訊務監控精確度與網路資源使用效率，並可達成 micro-flow 微訊務流公平性(Fairness)的「增強型 TRTCM 流量監控調節器」。其中在「封包 Conforming Degree 升級

(Promotion)」機制的研究方面，最初的設計動機是由於：原始 TRTCM 或 SRTCM 等 DiffServ 所定義的訊務監控調節機制只是以確保 Output 訊務特性符合 Traffic Profile 的規範為最高原則，是故考量的功能僅有當訊務違反 Traffic Profile 時對其封包進行的 Conforming Degree 的降級(Demotion) 標記，然而此單調的功能卻可能導致當一訊務流經過愈多個監控調節器後，會有較高的機率使得愈多的封包因為經過的某些網路資源不足而被降級，如此一但當網路資源恢復或是充裕的時候，仍無法使訊務回復到其原有的特性與資源使用率；而且只要後來若又面臨到網路資源不足的情況，則便可能因為 Conforming Degree 不高而面臨被首先丟棄的情形。因此，我們希望藉由「封包 Conforming Degree 升級」的機制使得當網路剩餘資源夠多的時候，可以將 Yellow Traffic 升級為 Green 等級，甚至是將 Red Traffic 升級為 Yellow Traffic，如此除了可以彌補有些 Green Traffic 通過某些 Congestion 區域而被判定成 Yellow，而造成之後都不再能獲得完全服務品質保障的情形，也可以在系統資源足夠的情況下增加系統資源使用效率 (Utilization)。圖 0 所示即是整合增加「綠色封包訊務保護機制」與「封包 Conforming Degree 升級」機制後的 TRTCM 運作流程圖，主要的設計概念是在 Token Bucket 上設定一 Promotion 的 Threshold (如圖 0 上的 τ 所示)，一旦 Token 數量超過此門檻值即表示網路資源充足，便可以啟動封包 Conforming Degree 的升級標記動作。由實驗數據中[39] 證實，此增強功能的 TRTCM 的確能夠進一步使網路資源獲得充分的利用，得到最佳的 Traffic Throughput。

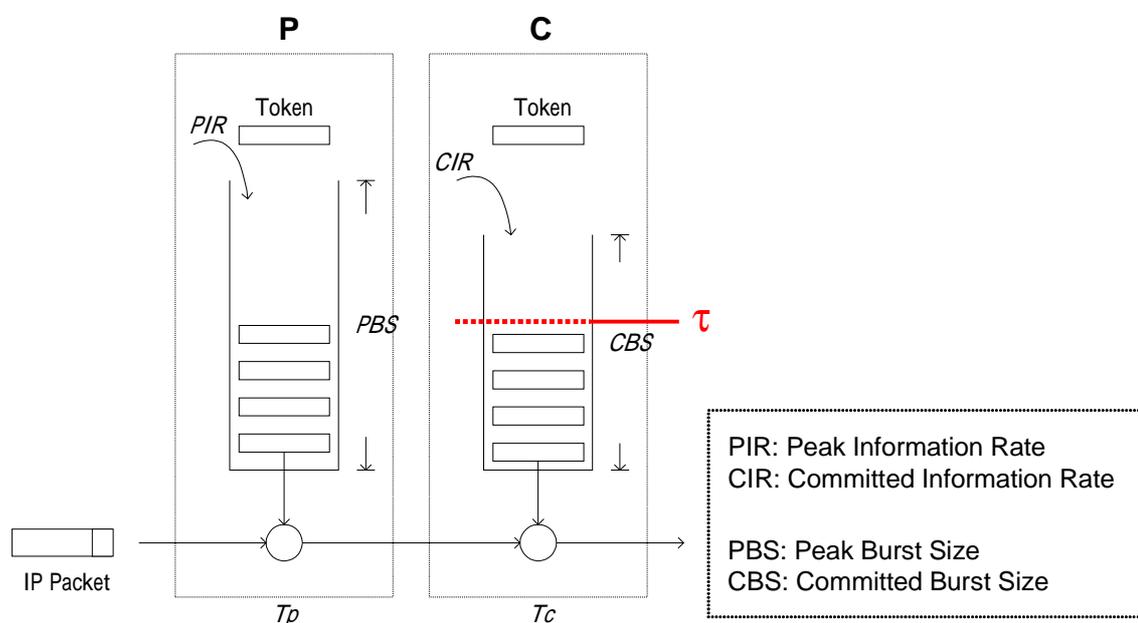


圖 0: 具備 Conforming Degree Promotion 功能之 TRTCM 系統架構圖

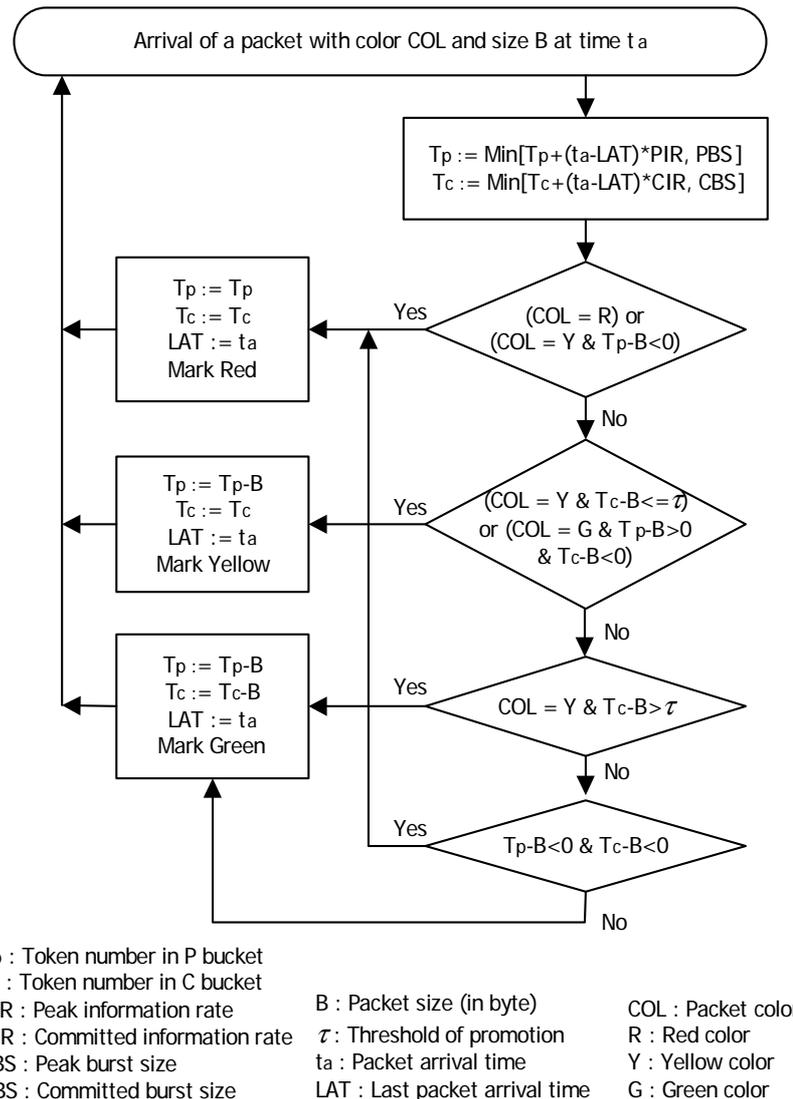


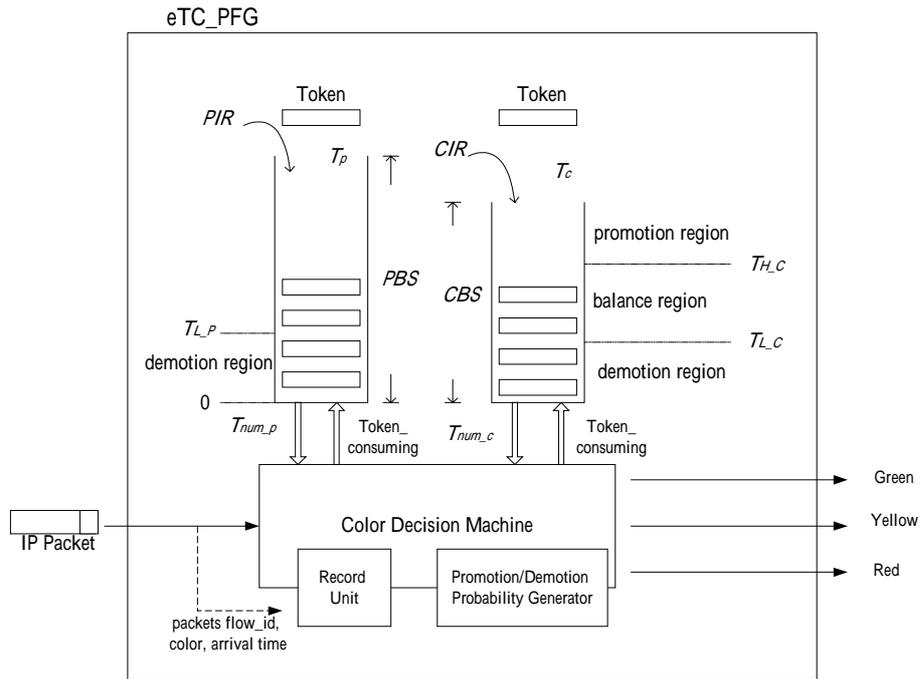
圖 0: 整合 Green Traffic Protection 功能與 Yellow Traffic Promotion 功能之 TRTCM 運作流程圖

除了前述透過新增的「綠色封包訊務保護機制」與「封包 Conforming Degree 升級機制」達成訊務監控精確度與網路資源使用效率的有效改進之外，我們也更進一步在 micro-flow 微訊務流 Conforming Degree 標記公平性(Fairness) 的機制方面進行研究與設計，以期在同一 Service Class 訊務流中的各個 micro-flow 之網路資源使用達到最佳的公平狀態，保障各連線所實際獲得的 QoS，並充分利用系統資源。由於不同的 Traffic Source 特性（例如週期性傳送或是網路擁塞敏感性 (Congestion Sensitive)）的交互影響下，使得在同一個服務等級(Service Class)中

的多個微訊務流，卻可能會產生封包的 Conforming Degree 標記不公平的情形，如此將使屬於同一 Service Class，原本應具備相近 QoS 的各訊務流的實際 QoS 表現差異過大，因而有必要針對 micro-flow 的 Conforming Degree 標記公平性進行控制與改進。實務上欲達成訊務監控調節標記的 Fairness，則是必須在進行封包 Conforming Degree 升級(Promotion) 或降級(Demotion) 等標記變更動作的時候，再根據所欲達成的 Fairness 目標，適當地決定是否真正執行該封包 Conforming Degree 升級或降級動作來達成。在參考相關論文資料並深入的探討後發現，在運作概念上屬於 Hard-decision 方式的 TRTCM 的 Conforming Degree 標記法並不容易做到標記公平性的功能，因此我們首先針對前述所發展出來的增強型 TRTCM，引進屬於 Soft-decision 方式的「隨機早期偵測(Random Early Detection, RED)」的概念與方法加以改進，在 Bucket 未達到極限狀態（空了 or 滿了）時即提前進行處理，根據目前在 Bucket 中的 Token 數量與其佔有空間的比例，以一機率成分來進行封包 Conforming Degree 降級或升級的標記，而非只是如過去傳統上僅根據目前 Bucket 中的 Token 數量或是與 Threshold 的高低關係，就進行標記；同時仍維持我們原本設計上於 Color-aware 運作模式下封包 Conforming Degree 升級(Promotion) 與綠色封包訊務流保護機制的功能：當 Bucket Size 低於某一 Lower Threshold 值時，即進入 Demotion-phase 運作模式，也就是訊務監控調節器原始的 Conforming Degree 降級標記功能，並啟動 RED 運作方式計算一 Demotion Prob. Factor，將進來的 Green 或 Yellow Packet 分別根據此 Demotion 機率值降級成 Yellow 或 Red Packet；當 Bucket Size 高於某一 Higher Threshold 值時，即進入屬於新增功能的 Promotion-phase 並仍採用 RED 運作方式計算一 Promotion Prob. Factor，將進來的 Yellow Packet 升級成 Green Packet。

在引入 RED 概念的運作方式改進後，便可以容易地進行 Conforming Degree 標記公平性機制的設計工作。我們所設計的 Conforming Degree 標記公平性機制主要的概念為：首先在訊務監控調節器上新增一訊務流量統計的附屬輔助裝置，配合 DiffServ 架構中定義在該訊務監控調節器前端的封包分類器(Packet Classifier) 來識別 Incoming 封包所屬的 micro-flow，持續計算並紀錄該訊務監控調節器所處理的 Service Class 訊務流中，每一 micro-flow 的各種 Conforming Degree 封包的流量統計值；接下來當系統進入 Demotion-phase 或 Promotion-phase，並根據 RED 概念的運作方式計算出一適用於該 Service Class 訊務流整體的 Demotion 或 Promotion 機率值時，並不是如同上述一般將此機率值直接作用在此時抵達的封包上，而是必須進一步根據前述輔助的流量統計裝置所計算的各 micro-flow 的流量狀況，以及所定義的公平性原則（在此我們以 Max-Min Fairness 為基本的 Fairness 原則），藉由適當的權重(Weighting) 調整該機率值，使抵達的各 micro-flow 的封包享有不同的升降級機率的手段，來達成各 micro-flow 的最高 Conforming Degree 的綠色與次 Conforming Degree 的黃色封包訊務流彼此間標記的公平性。圖 0 所示即為所設計的具備 Conforming Degree 標記公平性並整合「綠色封包訊務保護機制」與「封包 Conforming Degree 升級」

機制的增強型 TRTCM 的系統架構圖與其運作邏輯表。



Input	Bucket Condition/State		Output	Conditional probability	Token consumption
	T_{num_p}	T_{num_c}			
Green	--	$T_{num_c} \geq T_{L,C}$	Green	1	T_{num_p}, T_{num_c}
	$T_{num_p} \geq T_{L,P}$	$T_{num_c} < T_{L,C}$	Green	$1 - P_{demo_g}(j)$	T_{num_p}, T_{num_c}
	$T_{num_p} \geq T_{L,P}$	$T_{num_c} < T_{L,C}$	Yellow	$P_{demo_g}(j)$	T_{num_c}
	$T_{num_p} < T_{L,P}$	$T_{num_c} < T_{L,C}$	Green	$1 - P_{demo_g}(j)$	T_{num_p}, T_{num_c}
	$T_{num_p} < T_{L,P}$	$T_{num_c} < T_{L,C}$	Yellow	$P_{demo_g}(j) * (1 - P_{demo_y}(j))$	T_{num_c}
Yellow	$T_{num_p} \geq T_{L,P}$	$T_{num_c} \geq T_{H,C}$	Green	$P_{prom}(j)$	T_{num_p}, T_{num_c}
	$T_{num_p} \geq T_{L,P}$	$T_{num_c} \geq T_{H,C}$	Yellow	$1 - P_{prom}(j)$	T_{num_p}
	$T_{num_p} \geq T_{L,P}$	$T_{num_c} < T_{H,C}$	Yellow	1	T_{num_p}
	$T_{num_p} < T_{L,P}$	$T_{num_c} \geq T_{H,C}$	Green	$P_{prom}(j)$	T_{num_p}, T_{num_c}
	$T_{num_p} < T_{L,P}$	$T_{num_c} \geq T_{H,C}$	Yellow	$(1 - P_{prom}(j)) * (1 - P_{demo_y}(j))$	T_{num_p}
Red	$T_{num_p} < T_{L,P}$	$T_{num_c} \geq T_{H,C}$	Red	$(1 - P_{prom}(j)) * P_{demo_y}(j)$	(none)
	$T_{num_p} < T_{L,P}$	$T_{num_c} < T_{H,C}$	Yellow	$1 - P_{demo_y}(j)$	T_{num_p}
	$T_{num_p} < T_{L,P}$	$T_{num_c} < T_{H,C}$	Red	$P_{demo_y}(j)$	(none)
Red	--	--	Red	1	(none)

圖 0: 增強型 TRTCM 之系統架構圖與運作邏輯表

圖 0 是為進行模擬驗證時所採用的網路架構與訊務源模型 (分別是兩個 TCP 與四個 UDP 連線), 並在其中 A、B 兩處分別觀察訊務經過 Demotion 與 Promotion 作用後的狀態。模擬結果如 圖 0 所示, 無論是在哪一個觀察點所得到的 micro-flow 訊務流量分佈狀態, 都可以很明顯地發現, 增強型 TRTCM 在標記公平性方面皆較原始 TRTCM 有著顯著的改善: 面對不同特性的 Traffic Source 組合, 增強型 TRTCM 的 Output micro-flows 的流量分配較為平均, 保障了屬於 Congestion Sensitive 的 TCP 連線的網路資源與 QoS 不易受其他連線的影響, 並使原本流量較大的 UDP1 連線不會因而獨佔較多的網路資源, 而是與其他連線皆保持相近的標記與資源使用比例, 達成標記公平的目標。除此之外, 在圖 0(b) 的數據也顯示, 增強型 TRTCM 亦能夠由於較公平而精確的監控標記, 而獲得較佳的訊務 Throughput 值。

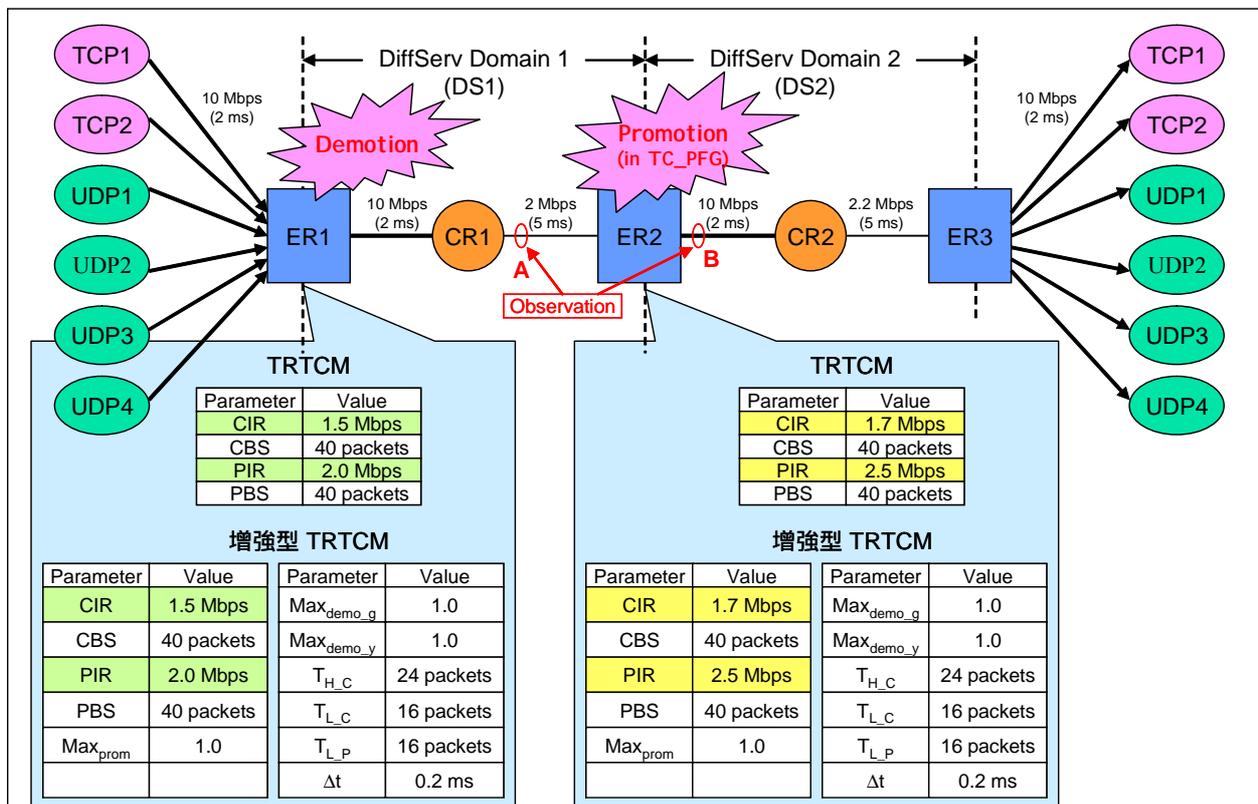
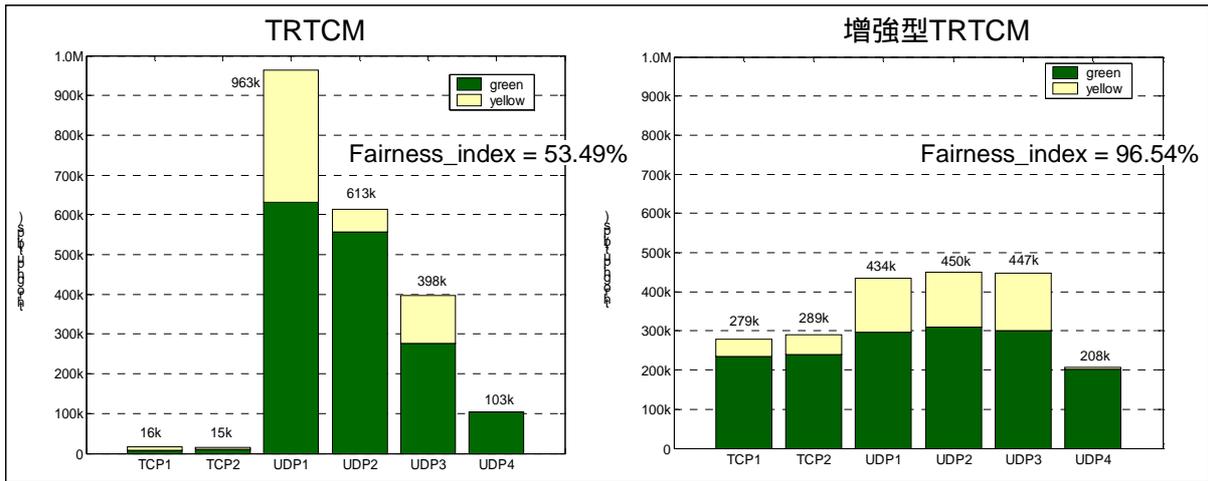
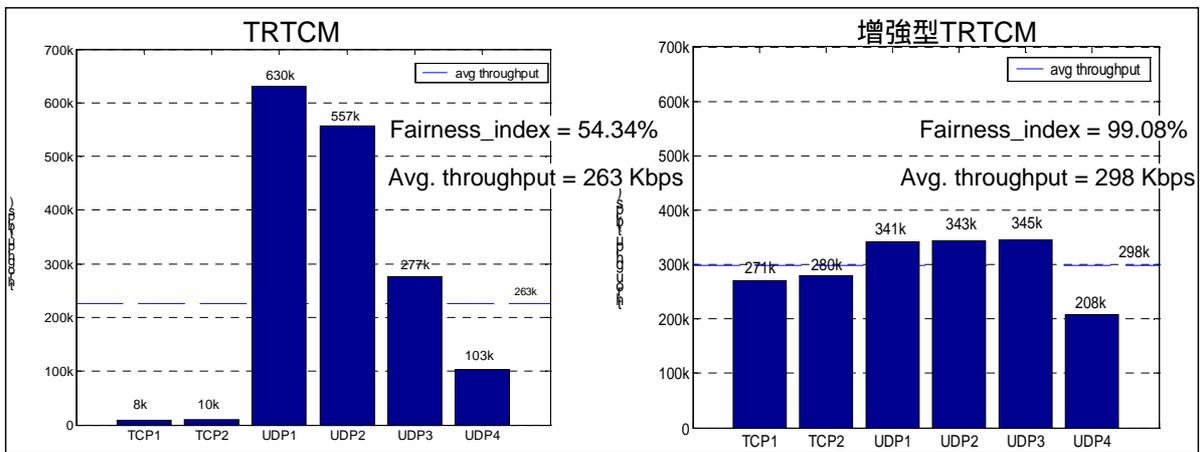


圖 0: 增強型 TRTCM 效能驗證之模擬網路架構與參數設定



(a)



(b)

圖 0: 增強型 TRTCM 與原始 TRTCM 方法之效能比較 : (a) 於模擬網路架構 A 點之訊務流量分佈 ; (b) 於模擬網路架構 B 點之訊務流量分佈

5. 高速 IP 封包分類(Packet Classification) 演算法

要達到具備服務品質保證或差異化服務(DiffServ) 品質的網際網路服務,其第一步驟就是要進行訊務流(Traffic Flow) 的區隔(Isolation), 如此才能針對不同的訊務流及其不同的服務品質需求施以相對應的訊務控制機制,達成其服務品質目標。而根據封包中各個欄位的資訊,對封包做不同服務等級分類的封包分類器即是在本質上並不提供訊務流識別功能的網際網路上實現訊務流區隔的主要機制。

在此我們設計了一個新的封包分類演算法,它能有效的將進入路由器的封包分類至所屬的過濾器,並針對該過濾器指定的動作採取行動。封包分類的動作基本上不外乎擷取封包中的資訊,與資料庫中的各個欄位做比對,以找出最佳比對的過濾器,並採取所指定的動作。目前已有相當多的論文在探討這樣延伸的問題,可以利用硬體的方式實做,也可利用軟體的方式進行模擬。在眾多方法中,令我們感興趣的是一個稱為朗訊位元向量(Lucent Bit Vector)的方式,這個方法原本是利用硬體的方式來探討封包分類的問題,但之後的一些論文則是開始利用軟體模擬的方式來研究該方法是否有改進的空間。由於此方法已經被路由器設計大廠朗訊採用,所以可以說是已被驗證可實際運用於目前的路由器中,如何改進這個方法便成為一個很有趣的問題。

我們著眼於朗訊位元向量方法耗費大量的儲存空間來維持位元向量,所以我們利用建立子樹的觀念來節省儲存位元向量的空間,不同於朗訊位元向量的方法,我們僅在子樹的根節點儲存位元向量(Bit Vector),將此子樹含有位元向量的所有資訊都整合起來儲存於根節點的位元向量中,如此一來,位元向量所耗費的儲存空間會隨著子樹中所包含子節點個數的增多而節省其儲存空間。之外,由於我們觀察到只利用這樣單純的觀念並沒有完全解決封包分類的問題,而會導致所謂錯誤比對(false match)的情形發生。為了解決錯誤比對的情況,我們在事先處理過濾器資料庫(filter database)時,會將資料庫做分析,並將封包分類分離成兩個階段來處理,一個採取原本位元向量的方式,另一個則採取線性搜尋的方法來比對。

以上的設計導因於我們的兩點觀察,第一,會發生錯誤比對的情形是由於我們採取子樹的概念來節省儲存空間,原本分屬於不同分支的資訊會因而整合到根節點中,使得在執行位元向量比對時找出錯誤的位置,所以我們在事先處理過濾器資料庫時,會把這些會發生錯誤比對的過濾器給分離成另一個資料庫,僅對於那些不會發生錯誤比對的產生位元向量,這樣可以得到的第一個優點是位元向量的長度大大減少,如此也是減少最差情形的記憶體存取次數;第二,事實上,在過濾器資料庫含有的過濾器數量很少時,利用複雜方式的演算法所得到的效能

並不會比簡單的線性搜尋方式來的快速。由於我們可以控制在第二階段要比對的過濾器數量大小，所以在第二階段利用線性搜尋的方式便可以達到極快的速度。以下即開始說明我們的方法是如何運作以及實驗的結果。

- 1) 我們的方法採用原本朗訊位元向量的方式來建立位元向量，其建立的方式是，若是過濾器資料庫中有 1000 個過濾器，那麼我們的位元向量的大小便是 1000 個位元(bit)，在我們的假設中，過濾器資料庫已經按照優先權的順序加以排序，越高優先權的過濾器放在資料庫的開始，相對應來說，越低位元位置所代表的便是越高的優先權，例如：位元 1 的優先權比位元 5 的優先權來的高。在進行位元向量的比對時，我們是利用交叉（intersection）的方式來比對，所以只要交叉比對出來的值等於 1，那麼該比對便可停止，此位元的位置也能直接對應到相對應的過濾器位址。
- 2) 為了節省儲存位元向量的位址空間，如前所述，我們利用建立子樹的觀念來建立位元向量，建立的方式如下：

```
GenerateSubtrieRoot(Cur_Node, Cur_Depth)
BEGIN
  IF ( ChildWithPrefix(Cur_Node) == Max_Prefix_Num )
    GenerateSubtrieRoot(Cur_Node);
  ELSE
    SubtrieConstructor(Cur_Node->left, Cur_Depth+1);
    SubtrieConstructor(Cur_Node->right, Cur_Depth+1);
  IF ( ChildWithPrefix(Cur_Node) >= Max_Prefix_Num )
    GenerateSubtrieRoot(Cur_Node);
  ELSE IF ( (Cur_Node == Trie_Root) &&
    (ChildWithPrefix(Cur_Node) !=
    Max_Included_Prefix_Num) )
    GenerateSubtrieRoot(Cur_Node);
END
```

利用從下而上（bottom-up）的建構方式，根據目前這個節點所包含的子樹中有多少個帶著 prefix 的節點以決定是否將此節點設成該子樹的根節點。藉由這樣的方式建構子樹的根節點，其目的是在有效的控制根節點維護的資訊量，也就是控制整合到根節點的位元向量的數目。整合的數目越多，位元向量中為 1 的位元便會增加，使得錯誤比對的機會大增，這個部分我們以第二階段的線性比對做為改進。

- 3) 建立完所有的根節點位置時，我們會給予每個根節點一個相對應的辨識碼，同時去修改過濾器資料庫，對於每個維度去更新相對應的辨識碼，接著搜尋整個過濾器資料庫，將所有維度辨識碼的集合相同之過濾器集中放置到第二階段處理，例如：以來源端位址的辨識碼為 2，目的端位址的辨識碼為 2，那麼我們便會去搜尋整個資料庫，將所有來源端位址的辨識碼是 2 和目的地位

址的辨識碼是 3 的集合起來，放置到第二階段的資料庫。若是某個辨識碼的集合僅有自己而沒有其他人與其相同，那麼此過濾器一定不會與其他的過濾器產生錯誤比對的狀況，所以我們將此過濾器放置在第一階段處理。

- 4) 根據步驟 3 處理的結果，我們將第一階段的過濾器資料庫依照位元向量的方式來建構其位元向量，在這裡可以注意到的是因為我們在事先處理時已經將原本的過濾器資料庫分成兩個部分來考量，所以此時在第一個階段的位元向量比起朗訊位元向量會短小許多，這也可以視為我們的方法在基本上比朗訊位元向量的方式擁有較少的記憶體存取次數。
- 5) 在實際執行我們演算法的部分，如同原本朗訊位元向量的方式，我們會針對每個維度都進行位址搜尋的步驟，這個部分可以利用現有的 IPv4 以及 IPv6 快速位址搜尋資料結構，來加快此步驟的速度。接著，找出每個維度所符合的子樹根節點後，我們首先對每個維度的根節點做一次記憶體存取，讀出存於該節點的辨識碼，檢查該辨識碼的集合是否存在第二階段的辨識碼集合中，若是，則我們可直接跳過第一階段的比對，直接進入第二階段做處理，可以這麼做的原因是，由於我們已經先做了事先處理的步驟，所以可直接在這裡便直接決定是否跳躍進第二階段；若否，那麼代表該辨識碼不在第二階段中，所以我們可以很明確的決定進入第一階段的位元向量比對。
- 6) 下面兩個圖表（圖 0 與圖 0）是我們的實驗結果之一，在這個結果中，我們的過濾器資料庫是兩個維度的資料庫，包含來源端位址和目的端位址兩個欄位。實驗中，我們假設路由器一次記憶體存取能夠讀取出 32 個位元，所以對於朗訊位元向量的方式來說，其最差的記憶體存取次數是（過濾器個數/一次記憶體存取的位元數），結果如圖四，可以發現我們方法的記憶體存取數量會因為子樹包含的節點數量增多而變小。在這裡我們會對最差的狀況有興趣的原因是，由於過濾器資料庫的型態會根據網路的特性而改變，以致於若探討平均的記憶體存取數時，其變動的因素遠大於最差的記憶體存取數，所以若是能提供一個在最差狀況下能夠表現優良的封包分類演算法，代表的是該演算法在平均的狀況下也會有不錯的效能。圖五是我們演算法所耗費的儲存空間，可以清楚的發現依然遠小於朗訊位元向量的儲存空間。

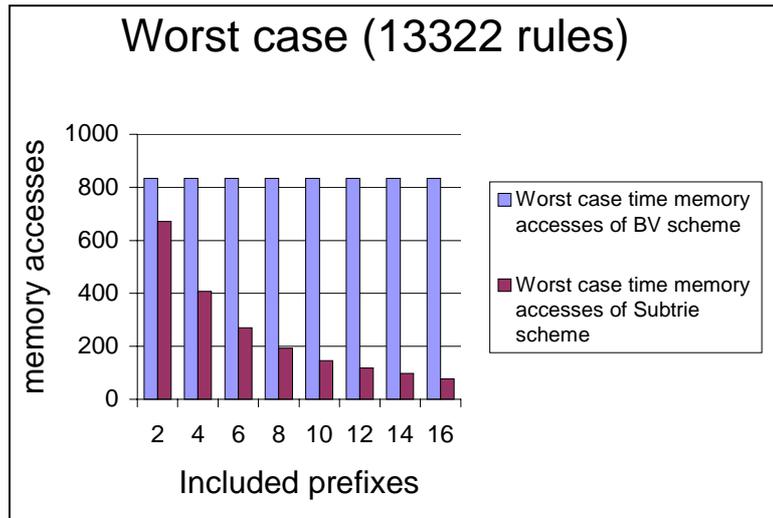


圖 0: 記憶體存取次數比較

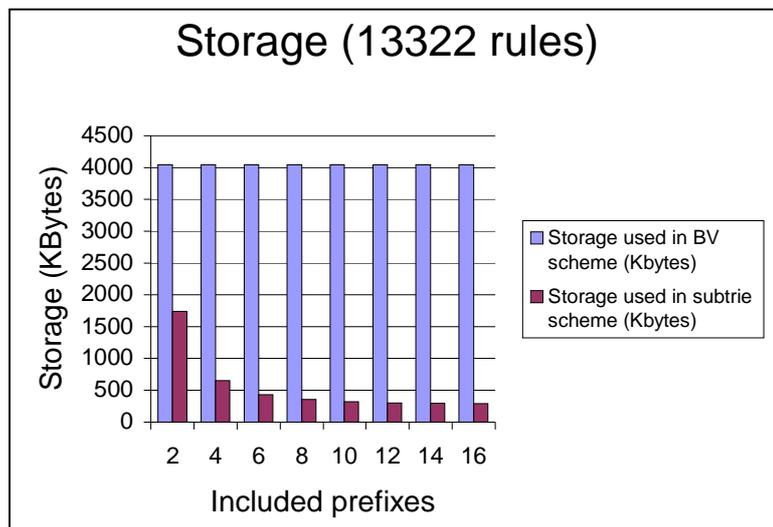


圖 0: 記憶體存取次數比較

6. 提供等比例封包延遲變異(jitter)之排程器(Scheduler)

我們著眼於傳輸資料流特性的不同而其所需的封包延遲變異(jitter)亦有所不同的需求上，希望能提供一解決方案。我們了解，現今網際網路架構的 best-effort 特性不足以提供這類的服務，Devrolis 提出了一個相對的差別服務架構，在這個架構下使用了比較簡單的方式來提供差別服務，基於這個相對的差別服務架構，我們提出了一個封包排程的機制，叫做 AJP 排程器，這個排程器可以在不同的傳輸資料流類別之中針對封包延遲的差異(jitter)提供成比例的差別服務，模擬實驗的結果顯示我們所設計的排程器的確符合我們的設計目標，也就是

對於長時間的封包延遲變異和短時間的封包延遲變異，排程器都能對於不同的競爭資料流類別提供成比例的差別服務，網路的管理人員可以直接透過設定差別服務係數來調整不同類別之間所獲得服務的等級差距，將 AJP 排程器整合進網際網路之後，提供類似互動式語音服務的想法就變得可行了。以下即開始說明我們的基本構想及所提出的方法：

- 1) 基於 Dovrolis 所提出的相對性差別服務架構，我們提出一個路由器內的封包排程機制，能夠對於不同類別的資料流量提供成比例式的差別服務。網路效能評量的標準有很多種，例如封包的延遲時間，封包遺失的機率和封包延遲的差異，在這個封包排程機制中，我們主要針對不同類別資料流的封包延遲變異來提供成比例式的差別服務。
- 2) 封包延遲的差異對於多媒體服務是相當重要的，特別是互動式的語音服務。這個封包排程機制依據不同類別的資料流優先權大小來進行封包的排程，首先我們將整個網路的資料流區分成不同的類別，假設一共有 N 個不同的類別，每個類別都有一個差別服務係數，這個差別服務係數是用來決定不同類別間所獲得服務的等級差距，網路管理人員可以藉由調整這些差別服務係數來改變不同類別間的服務等級差距。
- 3) 在路由器裡面，系統需要針對每個類別去記錄其平均的封包延遲變異，當路由器選擇某一個類別的排頭封包來服務時，路由器會計算這個封包與這個類別前一個被服務的封包之間的延遲差異，並更新系統所記錄的這個類別的平均封包延遲變異。更新的方式是採取成指數式的移動平均法，也就是按照下列的方程式來更新：

$$J_i(t) = (1 - e^{-(k-I)})J_i(t-1) + e^{-(k-I)}j_i^n$$

其中 $J_i(t)$ 為類別 i 新的平均封包延遲變異， $J_i(t-1)$ 為類別 i 舊的平均封包延遲變異， j_i^n 為類別 i 新產生的封包延遲變異。

- 4) 參數 K 決定舊的平均封包延遲變異與新產生的封包延遲變異所佔的比例，當 K 愈小時就代表新產生的封包延遲變異愈重要，系統可以很快反應資料流量的變動，但是會造成所算出來的平均封包延遲變異的變動幅度太大，使得在封包排程上會產生不穩定的狀態；當 K 愈大時代表舊的平均封包延遲變異愈重要，如此會使得系統在排程上會比較穩定，但是在反應資料流量的變動上就顯得比較緩慢。透過一些模擬的結果，我們認為 $K = 9$ 能夠在反應速度和系統穩定度上取得平衡的一個參數值，我們建議使用 $K = 9$ 做為系統的預設值。
- 5) 另一個參數 I 是一個亂數值， I 定義為新產生的封包延遲變異除以舊的平均封包延遲變異，雖然 $K = 9$ 適用於大多數的情形，但在不同的資料流量和系

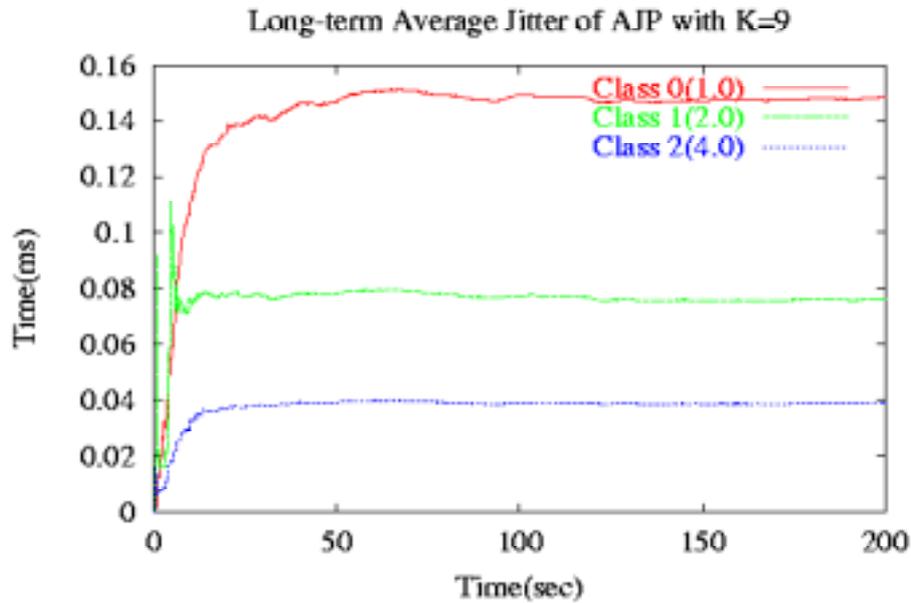


圖 0: 長時間的平均封包延遲變異

圖 0 展示了 AJP 排程器對於各類別在短時間平均封包延遲變異的表現，我們以 100ms 為一個計算點，分別算出類別 0 和類別 1 以及類別 1 和類別 2 的平均封包延遲變異比例，理論上這個比例是 2。由圖 0 我們可以發現，大部分的取樣點都圍繞在 2 附近，由此我們可以得知，AJP 排程器對於短時間平均封包延遲變異，也能維持一個合理的比例。

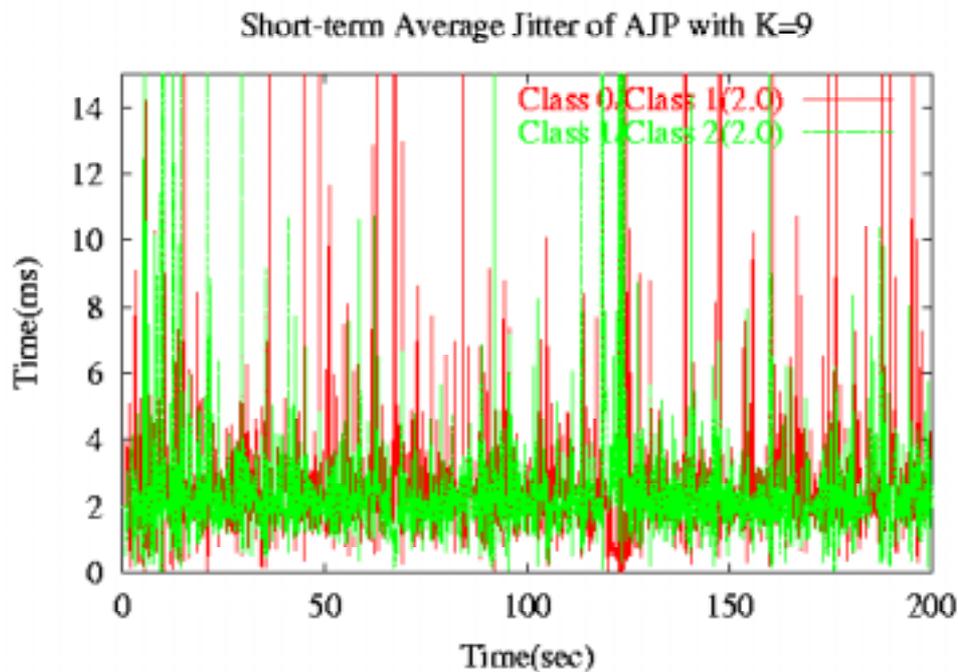


圖 0: 短時間的平均封包延遲變異

7. 具有自我類似性質之訊務特性研究與訊務源模型設計

7-1. 馬可夫程序(Markovian Process) 的自我類似性質探討

7-1-1. 研究動機

在所有關於網路自我類似性質的研究中，決定其成因似乎是最具根本重要性的。沒有一個對於訊務流量統計的正確的認識，網路設計上可能無法有效地紓解其壅塞情況。例如，如果網路的自我類似性質乃是肇因於現有網路規約，利用自我類似輸入訊務流來測試新的網路規約架構可能將導致錯誤的評估。

在文獻 [86] 中，Paxson 和 Floyd 發現 24 條網路流量紀錄中，由使用者啟動的訊務流，如 TELNET 和 FTP connection arrivals，相當適合以 Poisson processes 來加以描述並以之做為其訊務數學模型；然而，protocol-involved packet arrivals，如 SMTP、NNTP 與 FTP data transfers，卻更接近 *self-similar* processes [95] 的模式。此外，非自我類似模式，如 Poisson 模式，仍為 circuit-switched telecommunication traffics [96] 的典型訊務源模式。值得注意的是此時一個連結的開始與終止都是由用戶端所控制。雖然可能有多個因子影響在真實封包網路上所觀察到的自我類似行為，上述量測似乎建議聚合訊務流的自我類似現象與訊務源或端點在產生訊務流時受網路規約架構的影響有關。一個自然的猜想便是：受網路規約架構的影響產生的訊務流，如重傳 (*re-transmission*)，可能是整體流入訊務流的自我類似統計特性的主要原因之一。

前述猜想已由 Peha [97] 在數值模擬上確認。在他的研究中，他發現縱使是以傳統的 Poisson 封包抵達方式，一個簡單的重傳機制亦可導致聚合訊務流在工程上有意義的各個尺度上表現出自我類似的性質。更進一步地，他發現傳統上用來降低壅塞的技術，在壅塞發生時，會增強訊務流的自我類似性質，並可能反而延長壅塞時間。這促使我們去尋求一種理論上的相關解釋。

在此研究方向上，我們首先發現 Markovian 模式乃是此類簡單網路架構的共同模式。特別地，網路的流入量可以被模擬為一個先前系統狀態的隨機函數，而此系統狀態可以用系統中被阻塞的封包數來描述。而系統的下一時刻的狀態乃由現有的流入量及此刻的系統狀態來決定。因為此狀態相依的隨機流入量函數可一般視為時間穩態的，因而可簡化此系統為一個一階的 Markov Process。

我們接者注意到在 Markovian 架構與現有的自我類似定義上，存在兩個落差。首先，在傳統的 exact 與 asymptotic 離散時間二階自我類似定義上，廣義穩態總是基本條件 [91]，此時自相關函數將是一個只與時刻差有關的函數。但此

廣義穩態條件在一般的 first-order Markov-modeled network 中卻不能一般地假設成立。雖然一般在模擬上決定網路流量的自我類似參數時，常假定所有可能的非穩態行為都是只發生在短時間內，當模擬時間範圍夠長時，都可以忽略。但此種假定的正當性仍值得懷疑。

其次，利用對初始分佈的適當選擇，可使一個齊次 Markov 過程成為穩態過程。然而，在模擬上最常見的初始分佈設定卻是空佇列設定。這兩種對初始分佈的要求通常不會一致，特別當空佇列設定無法產生真正穩態的系統時。

因此，我們特別提出了一個針對 Markov 過程，結合初始機率分佈的自我類似推廣性定義。如果系統的初始機率分佈採用 Markov 過程的穩態分佈，則此一定義和傳統的二階自我類似定義是一致的。此定義如同傳統的二階自我類似定義一樣，可與 Markov-modeled network 的諸參數具體的結合。

7-1-2. 主要結果和討論

設有 X_1, X_2, X_3, \dots 為一個 first-order Markov process 具有 stationary transition probability $\mathbf{T} = [p_{ij}]$ ，在此， $p_{ij} = \Pr(X_2 = x_j | X_1 = x_i)$ ，且 $\{x_1, x_2, x_3, \dots\}$ 為 Markov process 的狀態空間。我們設狀態空間大小事有限或可數的，並且， $\{X_i\}_{i=1}^{\infty}$ 的 t th order transition probability 為 \mathbf{T}^t 。於是自相關函數（autocovariance function for the initial probability \mathcal{P} ）便是

$$b(t, \pi) = \mathcal{P}^T \mathbf{X} \mathbf{T}^t \mathcal{P} - \mathcal{P}^T \mathcal{P} \mathcal{P}^T \mathbf{T}^t \mathcal{P},$$

在此，上標“T”表示取 matrix 的轉置 matrix，而 \mathbf{X} 為一個對角矩陣，其對角元素為狀態參數 x_1, x_2, \dots 且 $\mathcal{P}^T = [x_1, x_2, \dots]$ 。由上述關於一階 Markov 過程的自相關函數的式子可以看出，其同時倚賴時刻差與 X_1 上的初始機率。我們於是定義 asymptotic self-similarity for first-order Markov processes 如下：

Definition 7-1-2-1 (Asymptotic self-similarity for Markov processes) 一個離散時間的 Markov process X_1, X_2, X_3, \dots 是 *asymptotic second-order self-similar* 且具參數 H 的條件為 (在此 $1/2 < H \leq 1$ ，且 \mathcal{P} 為初始機率)

$$\lim_{j \rightarrow \infty} \frac{b_{mj}(0, \mathcal{P})}{b_j(0, \mathcal{P})} = m^{-2(1-H)},$$

對於 $m \in \{1, 2, 3, \dots\}$ 。由上述定義，我們可以得到

$$\begin{aligned}
b_m(0, \boldsymbol{\pi}) &= \frac{1}{m^2} \boldsymbol{\rho}^\top \left(\sum_{j=1}^{m-1} \mathbf{T}^j \right) \mathbf{X} \boldsymbol{\rho} \\
&\quad - \frac{1}{m^2} \boldsymbol{\rho}^\top \left(\sum_{j=0}^{m-1} (\mathbf{T}^\top)^j \boldsymbol{\rho} \boldsymbol{\rho}^\top \mathbf{T}^j \right) \boldsymbol{\rho} \\
&\quad + \frac{2}{m^2} \boldsymbol{\rho}^\top \left(\sum_{i=1}^{m-1} \sum_{j=0}^{m-i-1} \mathbf{T}^j \mathbf{X} \mathbf{T}^i \right) \boldsymbol{\rho} \\
&\quad - \frac{2}{m^2} \boldsymbol{\rho}^\top \left(\sum_{i=1}^{m-1} \sum_{j=0}^{m-i-1} (\mathbf{T}^\top)^j \boldsymbol{\rho} \boldsymbol{\rho}^\top \mathbf{T}^{j+i} \right) \boldsymbol{\rho}.
\end{aligned}$$

如果 transition probability matrix 為 *simple* [98, Sec. 5.7]，在此 \mathbf{T} 能被寫成 $\mathbf{T} = \mathbf{S} \boldsymbol{\Lambda} \mathbf{S}^{-1}$ ，上標 “-1” 表示 the matrix inverse operation，且 $\boldsymbol{\Lambda}$ 為 the diagonal matrix 具有對角元素為 \mathbf{T} 的 eigenvalues，並且 \mathbf{S} 的各行為 \mathbf{T} 的 eigenvectors。例如，一個二狀態的 Markov Process 的情況，我們可得：

$$\left(\sum_{j=1}^{m-1} \mathbf{T}^j \right) = \mathbf{S} \begin{bmatrix} m & 0 \\ 0 & \frac{1-\lambda^m}{1-\lambda} \end{bmatrix} \mathbf{S}^{-1}$$

和

$$\left(\sum_{j=0}^{m-1} (\mathbf{T}^\top)^j \boldsymbol{\rho} \boldsymbol{\rho}^\top \mathbf{T}^j \right) = (\mathbf{S}^{-1})^\top \begin{bmatrix} m & \frac{v(1-\lambda^m)}{1-\lambda} \\ \frac{v(1-\lambda^m)}{1-\lambda} & \frac{v(1-\lambda^{2m})}{1-\lambda^2} \end{bmatrix} \mathbf{S}^{-1}$$

和

$$\left(\sum_{i=1}^{m-1} \sum_{j=0}^{m-i-1} \mathbf{T}^j \mathbf{X} \mathbf{T}^i \right) = \mathbf{S} \begin{bmatrix} h_{11} \frac{m(m-1)}{2} & h_{12} \lambda \left(\frac{m}{1-\lambda} - \frac{1-\lambda^m}{(1-\lambda)^2} \right) \\ h_{21} \left(\frac{m}{1-\lambda} - \frac{1-\lambda^m}{(1-\lambda)^2} \right) & h_{22} \lambda \left(\frac{\lambda(1-\lambda^m)}{(1-\lambda)^2} - \frac{m\lambda^m}{1-\lambda} \right) \end{bmatrix} \mathbf{S}^{-1}$$

和

$$\left(\sum_{i=1}^{m-1} \sum_{j=0}^{m-i-1} (\mathbf{T}^\top)^j \boldsymbol{\rho} \boldsymbol{\rho}^\top \mathbf{T}^{j+i} \right) = (\mathbf{S}^{-1})^\top \begin{bmatrix} \frac{m(m-1)}{2} & v \left(\frac{m}{1-\lambda} - \frac{1-\lambda^m}{(1-\lambda)^2} \right) \\ v \left(\frac{m}{1-\lambda} - \frac{1-\lambda^m}{(1-\lambda)^2} \right) & v^2 \left(\frac{m}{1-\lambda^2} - \frac{1-\lambda^{2m}}{(1-\lambda^2)^2} \right) \end{bmatrix} \mathbf{S}^{-1},$$

在此 h_{ij} 為 the component of $\mathbf{H} = \mathbf{S}^{-1} \mathbf{X} \mathbf{S}$ ，locating at i th row and j th column，而 v 為 the second component of $\mathbf{S}^\top \boldsymbol{\rho}$ 。我們於是能測試此一 Markov process 的 degree of asymptotic self-similarity：

$$\lim_{m \rightarrow \infty} m^\beta b_m(0, \frac{0}{h}) = \begin{cases} \infty, & \beta > 1; \\ 0, & 0 \leq \beta < 1, \end{cases}$$

和

$$\begin{aligned} \lim_{m \rightarrow \infty} m \cdot b_m(0, \frac{0}{h}) &= E[X_1^2] - \frac{(ax_2 + bx_1)^2}{(a+b)^2} \\ &+ \frac{2ab(1-a-b)(x_1 - x_2)^2}{(a+b)^3} \\ &- \frac{2(a\pi_1 - b\pi_2)^2(x_1 - x_2)^2}{(2-a-b)(a+b)^3} \\ &- \frac{2(a\pi_1 - b\pi_2)(ax_2 + bx_1)(x_1 - x_2)}{(a+b)^3}, \end{aligned}$$

在此 $a = \Pr\{X_2=x_2|X_1=x_1\}$ 和 $b = \Pr\{X_2=x_1|X_1=x_2\}$ 。由上述的式子，我們推論具有 *simple transition probability matrix* 的 Markov process，其區塊平均的變異大致延續到區塊大小的程度，雖然 degree of asymptotic self-similarity 不受初始分佈的影響，但 time-finite self-similarity 卻高度受到初始分佈的影響。

7-2. 自我類似特性之網路封包的產生

7-2-1. 研究動機

一個通訊系統是否良好地運作取決於其使用者對通訊品質的可靠度的評量上。例如，目前有線電話系統的高評價，乃是因其提供用戶可靠的交換線路連結。為了確保一個系統的可靠性，必須在其部署之前預先進行若干測試。這些測試必須適當地設計以使部署後的系統效能為可預測的。因此，需要產生實驗性的訊務流來逼近真實的訊務流。例如，由 Markov 型模式中可導出著名的 Erlang B 及 Erlang C 公式，能成功地象徵使用者的行為，準確地預測總體的撥打遭阻塞及佇列的機率。由此可知一個訊務流模式對通訊系統測試的重要性。

迄今已有許多合成自我類似訊流資料的方法。在 [99] 中，Paxson 利用對分數高斯雜訊的訊流功率頻譜 fitting 來合成自我類似訊流。Lau 等人 [84] 提出一個稱為 random midpoint displacement 的演算法來產生自我類似訊流軌跡。我們隨即注意到這些方法有兩個缺點。第一，需要的訊流長度必需在產生訊流前預先決定，因而當需要產生一個較長的訊流序列時，必需從頭重新產生全部的序列。換言之，訊流資料不能被即時地產生。其次，這些訊流產生器會生成負的整數值，這對任何的抵達封包串列都是不合理的。更重要地，真實量測網路訊流雖在超越工程可控管的尺度上展現出自我類似的特性，但在極大尺度上仍為非自我類似的。因而，使用一個極限尺度上仍自我相似的模式去作系統效能評估是值得懷疑的。相反地，產生一個具有可控制最大自我類似聚合度的訊流可能更接近真實網

路的情形，以上諸點使我們發展出了可克服的產生模式。

7-2-2. 濾波器式自我類似訊務流產生器

我們的序列產生器的基本要點是基於功率頻譜 fitting。令 $S_y(w)$ 為離散時間隨機序列 $Y[n]$ 的功率頻譜， $S_x(w)$ 為離散時間隨機序列 $X[n]$ 的功率頻譜， $Y[n]$ 是把 $X[n]$ 輸入到濾波器 $H(w)$ 所得到的輸出。則有 $S_y(w) = |H(w)|^2 S_x(w)$ [87]。因此，如果我們設計一個濾波器 $H(w)$ ，使得 $|H(w)|^2$ 逼近自我類似過程的功率頻譜，那麼當輸入 $X[n]$ 為 i.i.d. 時，輸出自然將變成自我類似。

一個具自我類似參數 H 的恰二階自我類似過程的自變異函數為

$$(c/2)[|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}] \text{ for some constant } c > 0$$

且其功率頻譜 $F_H(w)$ 為

$$c |1 - e^{-jw}|^2 \sum_{k=-\infty}^{\infty} |w + 2\pi k|^{-1-2H} \text{ for } -\pi \leq w < \pi。$$

利用若干近似，可得 $F_H(w) \approx c |1 - e^{-jw}|^{-2H}$ for $-\pi \leq w < \pi$ 。即問題回歸到尋找一個適切的濾波器，使得對具有平均值 λ 的 Poisson i.i.d. 的輸入產生

$$\begin{aligned} S_y(w) &= |H(w)|^2 S_x(w) \\ &= \lambda |H(w)|^2 = \lambda |1 - e^{-jw}|^{-1-2H} \end{aligned}$$

這樣的濾波器可為 $h[n] = \Gamma(n+H-0.5)/[\Gamma(n+1)\Gamma(H-0.5)]$ for $n \geq 0$ ，在此 $\Gamma(\cdot)$ 表示 Euler gamma function。

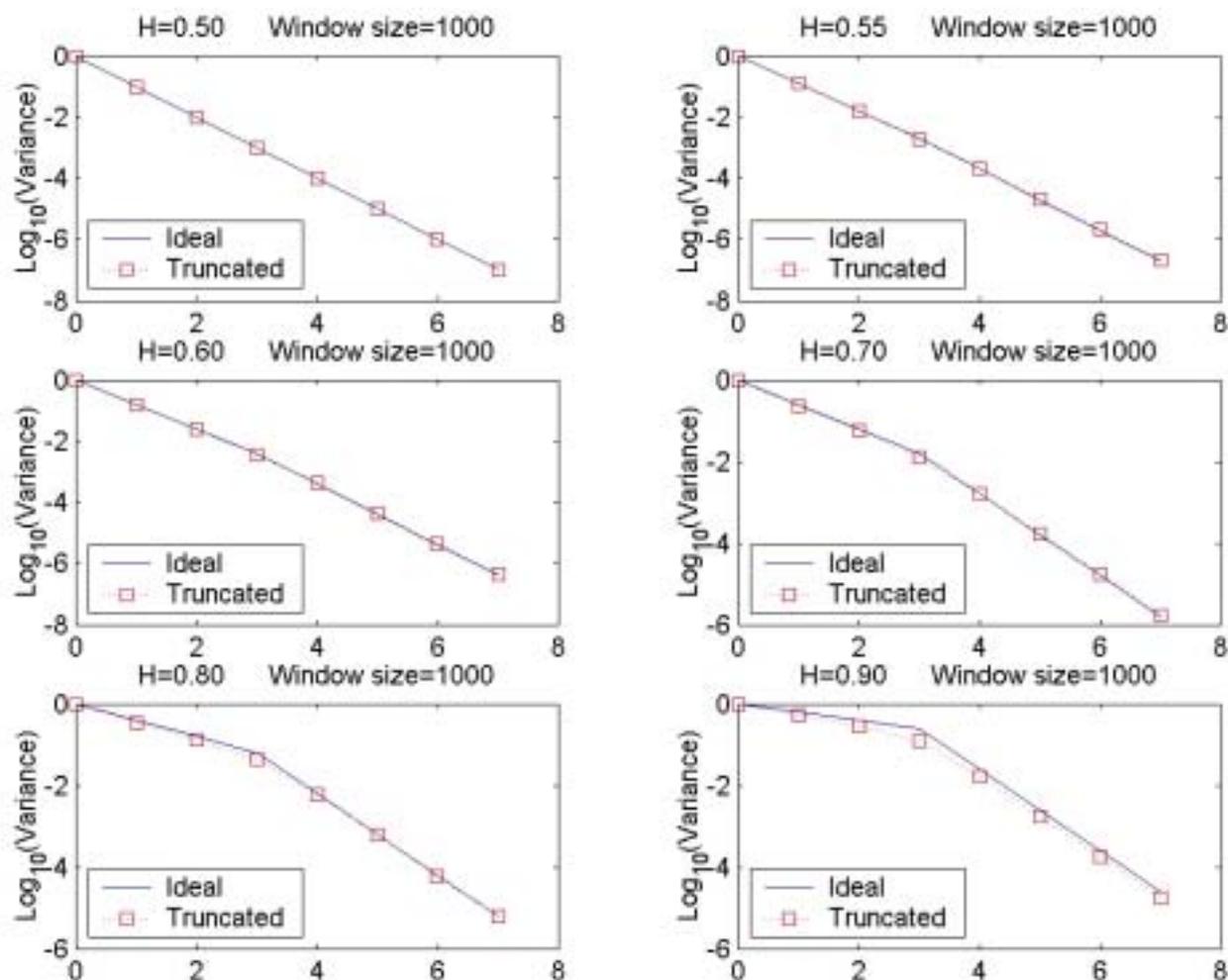


圖 0: Variance-Time Analysis for $W = 10^3$. The slope of the blue line is equal to $2H - 2$ for $m \leq W$, and -1 for $m > W$.

由於濾波器 $h[n]$ 具有無限長的脈衝響應 (Impulse Response)，而其係數呈冪次率下滑，因而在實現上必須加上一長度為 W 的截斷窗戶 (Truncation Window)，即當 $n > W$ 時則 $h[n] = 0$ 。令具長度 W 截斷窗戶的 m -aggregated 序列的變異數為 $C_m(0; W)$ 。下面圖 0 舉出了 $\log[C_m(0; 10^4)]$ 與 $\log[m]$ 之間的關係。我們發現對 $0 \leq \log_{10}[m] \leq \log_{10}[W]$ ，具斜率 $2\tilde{H} - 2$ 的直線吻合 $\log[C_m(0; 10^4)]$ 對 $\log[m]$ 的曲線。濾波器輸出結果的 H 表列在表 0。由這些資料，我們發現當平均窗戶長 m 小於或等於截斷窗戶長 W 時，輸出結果的 H 傾向略小於目標值 H ，儘管偏差率，定義為 $(\tilde{H} - H)/H$ ，於各種情況下是可接受地小。因此，對小的 H 值而言，自我類似的逼近是更為精準。

對 $m > W$ 的情況， $C_m(0; W)$ 能表示為 $A_H(W)m^{-1} - B_H(W)m^{-2}$ ，且有

$$\frac{\partial \log[C_m(0;W)]}{\partial \log(m)} = -1 + \frac{B_H(w)/A_H(w)}{m - B_H(w)/A_H(w)}$$

此時，自我類似的程度由下面的比率來決定：

$$\frac{B_H(w)}{A_H(w)} = \frac{(H-0.5)}{H(H+0.5)} \left(W + \frac{(H-1.5)}{2} \right) - O(W^{1-2H})$$

在此 $O(\cdot)$ 為 big- O 記號。

此外，如果我們想要產生一個輸出結果 $H' \in (0.5, 1)$ (接近目標值 H)，平均窗戶長 $m' > W$ ，則需要

$$-1 + \frac{B_H(w)/A_H(w)}{m' - B_H(w)/A_H(w)} > -\beta = 2H' - 2$$

此意謂

$$W < m' < \left(\frac{2H'}{2H'-1} \right) \frac{B_H(w)}{A_H(w)}$$

這表明截斷模型的輸出的自我類似度可展延至

$$\begin{aligned} & \left(\frac{2H'}{2H'-1} \right) \frac{B_H(w)}{A_H(w)} \\ &= \left(\frac{2H'}{2H'-1} \right) \left[\frac{(H-0.5)}{H(H+0.5)} \left(W + \frac{(H-1.5)}{2} \right) - O(W^{1-2H}) \right] \end{aligned}$$

因為對 $W > 0$ 及 $0.5 < H < 1$

$$m' > W > \frac{(H-0.5)}{H(H+0.5)} \left(W + \frac{(H-1.5)}{2} \right) > \frac{B_H(w)}{A_H(w)},$$

只要 $H' = H$ 。由上述不等式及 $m' > W$ 可得

$$(H-0.5)W < -\frac{1.5-H}{2} - \left(\frac{2H}{2H-1} \right) O(W^{1-2H}),$$

與 $W > 0$ 相違背。我們於是推得，如果要得到精確的自我類似 (即 $H'=H$)，那麼沿展輸出端的自我類似到超過截斷窗戶長是不可能的。由圖 0 可知，我們更

進一步由數值結果顯示，縱使 H 被允許略小於 H ，上述結論仍然成立。

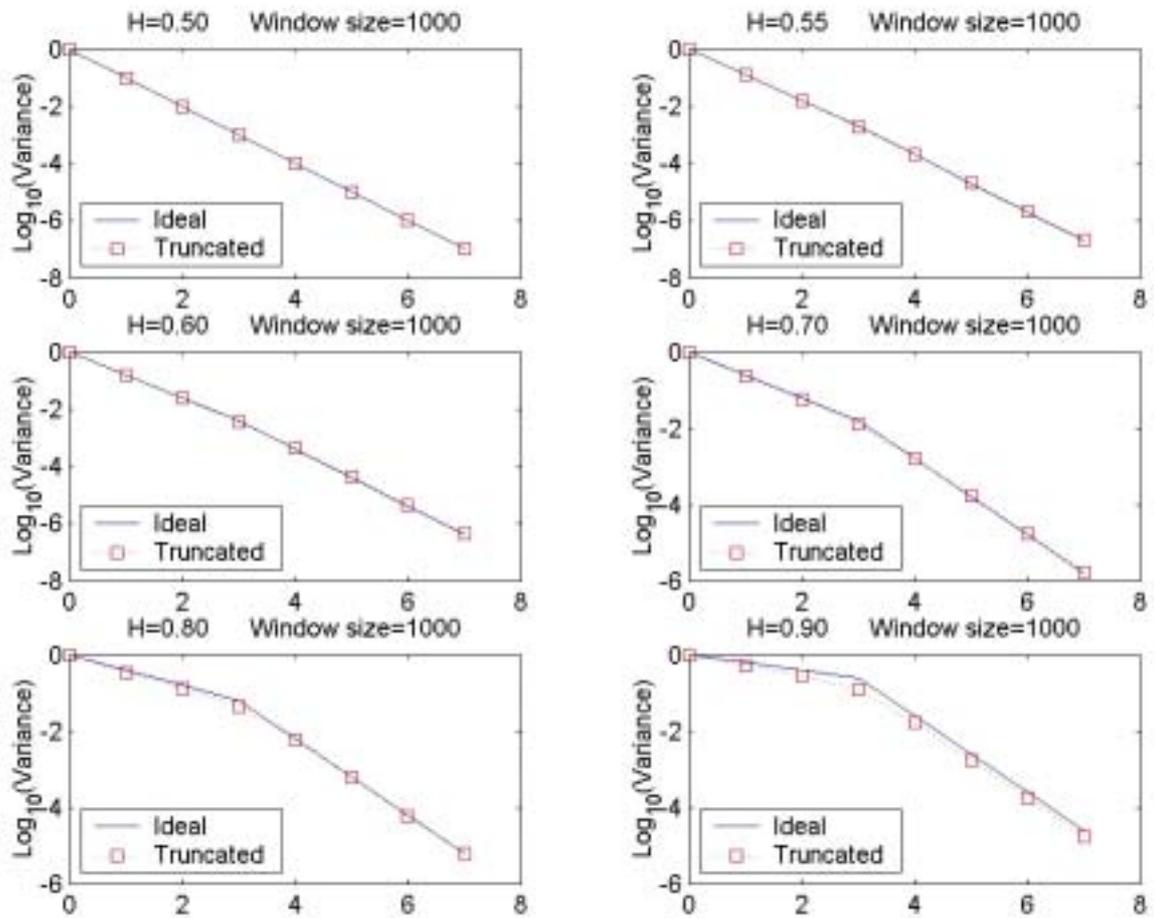


圖 0: Variance-Time Analysis for $W = 10^3$. The slope of the blue line is equal to $2H - 2$ for $m \leq W$, and -1 for $m > W$.

Window Length = 10^4		
Targeted H	Resultant H'	Deviation
0.5001	0.5000961	-7.7984E-006
0.55	0.5481199	-0.0034
0.6	0.5961926	-0.0063
0.7	0.6909921	-0.0129
0.8	0.7809945	-0.0238
0.9	0.8599458	-0.0445

表 0: The resultant H' versus the targeted H . Deviation = $(H' - H) / H$

7-2-3. 結果和討論

在此報告中，基於濾波器理論，我們提出了一個用於自我相似訊流生成的新模式。此模式是具有可調的暴流(burstiness) 及相關係數(correlation) 的長程相依(Long Range Dependence) 模式。此模式具有極少的輸入參數，事實上，只需要三個參數。 H 是自我類似參數，控制合成訊流的暴流及自我相關性質， λ 合成訊流的平均流量，而 W 定義了濾波器截斷窗戶長度以及合成訊流的最大自我類似聚合尺度。雖然濾波器長度 W 限制了正確保持自我類似的最大聚合尺度，但此恰與量測到的網路訊流行為相吻合，即自我類似的性質僅保持到一個超越實際可控管的範圍，但當聚合尺度更大地增加時仍然消逝。此模式尚有其他優點，例如即時地產生訊流與必產生非負整數值來代表網路封包。

我們利用對其 variance-time 關係的數學分析及對 V-T plot, R/S plot 及 periodogram plot (如表 0 所示) 的統計測試驗證我們濾波器基礎的模式。而我們推論我們的模式可在高的精確度下合成具自我類似參數 H 的隨機程序。

Window Length = 10 ⁴			
<i>H</i> used	<i>H</i> resultant (V-T Plot)	<i>H</i> resultant (R/S Plot)	<i>H</i> resultant (Periodogram)
0.5001	0.4913099	0.5423777	0.5149618
0.55	0.5243788	0.5839482	0.5433228
0.6	0.5661478	0.6248291	0.5953569
0.7	0.6860798	0.6991949	0.6902056
0.8	0.7558080	0.7792713	0.7968477
0.9	0.8662405	0.8784192	0.8822255

表 0: *H* used versus *H* resultant

7-3. 自我類似特性之訊務流在資訊理論上的特徵

7-3-1. 主要結果

(A) 序言

一個隨機程序，如果其邊際平均與自相關函數在時間移動下不變，就稱為 wild-sense stationary (WSS)。一個 WSS 隨機程序，如果其自相關函數為冪次遞減 (decreases with power law)，就稱為 asymptotic second-order self-similar (Boris Tsybakov and Nicolas D. Georganas [91])。一般 asymptotic second-order self-similar process 似乎不具有明顯的資訊理論上的意義，然而，我們發現在我們提出的濾波器式自我類似訊務產生器輸出的自我類似訊務流，其不同時刻流量的相互關係資訊 (mutual information between two different instant values) 亦為冪次遞減。此外，我們亦發現任何二值 WSS 或 Gaussian process 不同時刻流量的相互關係資訊約為其自相關係數平方的一半。

(B) Nearly Independent

兩個隨機程序 X_1, X_2, \dots, X_n 與 Y_1, Y_2, \dots, Y_n 被稱為 *nearly independent* 如當 n 趨近無限大， X_n 與 Y_n 的相關係數趨近於零，且

$$P_{X_n, Y_n}(x, y) - P_{X_n}(x)P_{Y_n}(y) = \rho_n \left(\frac{x - \mu_{X_n}}{\sigma_{X_n}} P_{X_n}(x) \right) \left(\frac{y - \mu_{Y_n}}{\sigma_{Y_n}} P_{Y_n}(y) \right) + o(\rho_n),$$

在此， ρ_n 為 X_n 與 Y_n 的相關係數，即

$$\rho_n = E \left[\left(\frac{x - \mu_{X_n}}{\sigma_{X_n}} \right) \left(\frac{y - \mu_{Y_n}}{\sigma_{Y_n}} \right) \right], \quad \mu_{X_n} = E\{X_n\}, \quad \mu_{Y_n} = E\{Y_n\},$$

$$\sigma_{X_n} = E\{(x - \mu_{X_n})^2\}, \quad \text{and} \quad \sigma_{Y_n} = E\{(y - \mu_{Y_n})^2\}.$$

(C) 不同時刻的流量之相互關係資訊(Mutual Information)

Proposition 1 : 對兩個 nearly independent 隨機程序 X_1, X_2, \dots, X_n 和 $Y_1, Y_2, \dots, Y_n, X_n$ 與 Y_n 的相互關係資訊約為

$$I(X_n; Y_n) = \frac{1}{2} \rho_n^2 + o(\rho_n^2)$$

在此我們先省略證明。

(D) 兩個 Gaussian random variables 所造成的序列(Sequence)之相互關係資訊

如 X_1, X_2, \dots, X_n 與 Y_1, Y_2, \dots, Y_n are 為兩個 nearly independent 隨機程序, 且每對 (X_n, Y_n) 都具二維聯合 Gaussian 隨機分佈, 則 X_n 與 Y_n 的相互關係資訊約為 [80],

$$I(X_n; Y_n) = -\frac{1}{2} \log(1 - \rho_n^2) = \frac{1}{2} \rho_n^2 + o(\rho_n^2)$$

如果 X_n 為 asymptotic second-order self-similar Gaussian process 且具參數 $H = 1 - (\beta/2)$, $0 < \beta < 1$, 且 $Y_n = X_{n+m}$, 則

$$\rho_n = g(k) = \frac{1}{2} \{(k+1)^{2-\beta} - 2k^{2-\beta} + (k-1)^{2-\beta}\} \cong H(2H-1)k^{-\beta}, \text{ as } k \rightarrow \infty,$$

且

$$I(X_n; X_{n+k}) = -\frac{1}{2} \log(1 - g(k)^2) \cong \frac{1}{2} g(k)^2 \cong 2H^2(H - \frac{1}{2})^2 k^{-2\beta}, k \rightarrow \infty.$$

(E) 濾波器輸出端的相互關係資訊

在過去兩年, 我們提出一個的濾波器式自我類似訊務產生器來產生自我類似訊務流, 其不但可即時地生成訊務, 亦能指定所要的自我類似參數 H 。此濾波器式自我類似訊務產生器之脈衝響應為 $h[n] = \Gamma(n+H-0.5)/[\Gamma(n+1)\Gamma(H-0.5)]$ for $n \geq 0$, 在此, $\Gamma(\cdot)$ 表示 Euler gamma function, 其定義為 $\Gamma[n] = \int_0^\infty t^{n-1} e^{-t} dt$, 且濾波器轉換函數為 $H(w) = (1 - e^{jw})^{0.5-H}$ 。當一個 uncorrelated wide sense stationary sequence $Z[n]$ 被輸入濾波器, 我們已證明了濾波器的輸出 $X[n]$ 將為

asymptotically second-order self-similar 且其自相關函數 $b(n)$ 為幕次遞減。注意，在此 $h[n] \geq 0$ for $n \geq 0$ 。定義一個新濾波器如下：

$$p[i] = \frac{h[i]}{\sum_{i=0}^{M-1} h[i]} = (H - \frac{1}{2}) \frac{\Gamma(M)}{\Gamma(M + H - \frac{1}{2})} \frac{\Gamma(i + H - \frac{1}{2})}{\Gamma(i + 1)}$$

for $M - 1 \geq i \geq 0$, and $p[i] = 0$ for $i > M - 1$ or $i < 0$ 。如果一個 i.i.d. 序列 $Z[n]$ 被輸入濾波器 $p[i]$ ，且輸出序列為 $X[n] = \sum_{k=0}^{M-1} p[k]Z[n-k]$ ，則我們可發現：

$$I(X[n]; X[n+k]) \cong \frac{1}{2} \frac{(\sum_{i=0}^n h[i]h[i+k])^2}{(\sum_{i=0}^n h[i]^2)(\sum_{i=0}^{n+k} h[i]^2)} \cong 2H^2(H - \frac{1}{2})^2 k^{-2\beta}, \text{ as } k \rightarrow \infty, M \rightarrow \infty.$$

(F) 兩個二值隨機變數的相互關係資訊

如 X_1, X_2, \dots, X_n 和 Y_1, Y_2, \dots, Y_n 為兩個二值隨機程序，且每對 (X_n, Y_n) 皆為兩個二值隨機變數，則 X_n 與 Y_n 的相互資訊約為，

$$I(X_n; Y_n) = h_b(\mu_{Y_n}) - \mu_{X_n} h_b(\mu_{Y_n} + \frac{r_n}{\mu_{X_n}}) - (1 - \mu_{X_n}) h_b(1 - \mu_{Y_n} + \frac{r_n}{1 - \mu_{X_n}}),$$

在此， $h_b(\cdot)$ 為 binary entropy function, $\mu_{X_n} = E\{X_n\}$, $\mu_{Y_n} = E\{Y_n\}$, 且 $r_n = E\{(x - \mu_{X_n})(y - \mu_{Y_n})\}$ 。如當 n 趨近無限大, r_n 趨近於零, 則我們可發現 $I(X_n; Y_n) \cong \rho_n^2 / 2$ 。特別地，如 X_n 為任何二值 asymptotic second-order self-similar 隨機程序且具參數 $H = \bar{1} (\beta / 2)$, 在此 $0 < \beta < 1$, 則

$$I(X_n; X_{n+k}) \cong \frac{1}{2} \rho_k^2 \cong 2H^2(H - \frac{1}{2})^2 k^{-2\beta}, \text{ as } k \rightarrow \infty,$$

亦為幕次遞減。

7-3-2. 評論與延伸

我們的研究成果已經成功地顯示 asymptotic second-order self-similar 隨機程序在不同時刻流量的相互關係資訊(mutual information between two different instant values) 亦為幕次遞減。我們也展示了三個不同的 nearly independent 的序列，分別是我們的濾波器式自我類似訊務產生器的輸出端、自我類似的 Gaussian

隨機程序、以及二元值(binary) 的 asymptotic second-order self-similar 隨機程序。然而,自我類似訊流的兩個聚合區塊間的相互關係資訊的漸近行為仍是一個未解決的問題,可做為未來一個可行且深具挑戰性的研究課題。

8. 支援 QoS 的無線媒體多重接取(Multiple Access)競入技術

無線區域網路(WLAN)通訊協定是目前一個相當重要的無線網路技術，同時也將在下一代網路領域扮演一個相當重要的角色。它主要的特色在於設計簡單、彈性很大及建置費用合理。可提供人們不論在公司、學校、商場、或家中都可以很簡單的方式連上網際網路，享受多媒體的通訊服務。但是，多媒體應用軟體在傳輸時通常需要網路提供服務品質保證(QoS)才能達到其需求，例如：頻寬的保證、最小延遲的保證、最低錯誤率的保證等。在無線環境下因為有雜訊大及不同的實體層(PHY)溝通等問題，所以若在 WLAN 中保證 QoS 是個很有挑戰性的研究課題。

在無線接取網路上如何設計一個能滿足有效接取及 QoS 各項要求的 MAC 層是目前一個很大的研究課題。為了能在無線網路上能充份利用各式資源並對不同的訊務源提供其所需要的 QoS 保證，我們將探討如何建立具頻寬分配且以 QoS 為導向的 MAC 層通訊協定。我們在過去一年中設計一個適用於無線網路上的六回饋競入存取機制(HFCA)，可使控制封包訊務能有最大的 throughput。目前我們的研究方向為以 HFCA 之設計概念為主再加上 QoS 的考量，將之應用於 WLAN 上。以下我們針對 HFCA 之基本設計作解說。

HFCA 是以二階段的方式進行競入碰撞解析法則。在第一階段中，每一個要傳送的控制封包訊務會以某個機率來選擇是否要在下一個 slot 中傳送，這種做法會使在下個 slot 中發生碰撞的控制封包訊務個數有相當大的機率會小於 5；而到第二個階段則依序解析這群機率小於 5 的碰撞。接下來我們介紹六回饋競入存取機制(HFCA)的運作。其設計的原理是以機率的方式將同一時間欲連上網路的 MT 數量減少，並利用高效率的六回饋碰撞解析機制來將此網路擷取動作完成。這六回饋碰撞解析機制包含了六種不同的回饋：無人使用(idle)、成功(success)及二至六個 MT 訊號碰撞回饋。以下我們將先詳細介紹其運作方法，接著我們再以一個實例來說明並驗證其優缺點。

HFCA 的運作（如圖 0 所示）可分為允入(admission) 及解析(resolution)二個階段來進行。當一群有資料要傳送的 MT 先經過允入程序的篩選成較小的群體後，再經由解析程序一一解出，以完成通訊。這兩階段的過程一直反覆執行直到能滿足最小 blocking probability 的 QoS 保證，或是達到先前所設定的最大訊寬長度時才停止。對於在 HFCA 運作過程中才新進的使用者要等到目前正執行 HFCA 工作的 MT 們都完成傳輸後才可再行加入傳輸工作的行列。這二階段的 HFCA 演算法如圖 0 所示。

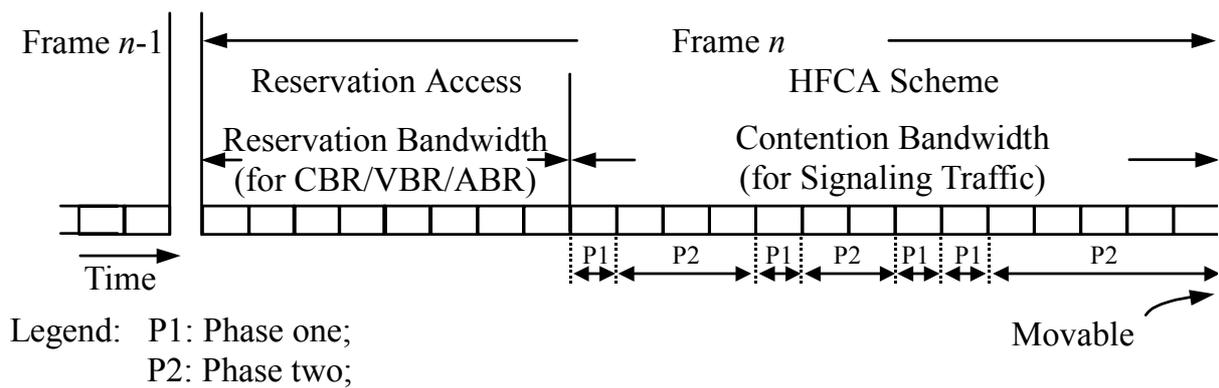


圖 0: HFCA 的運作

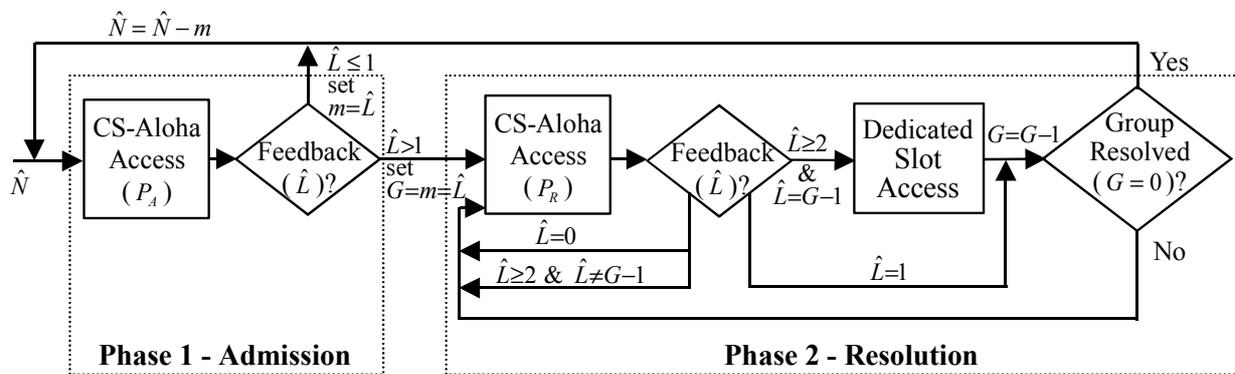


圖 0: HFCA 的演算法

假設有一群 MT 欲進行網路擷取的動作，並且其原始數量大小為 N ，在 AP 有個訊務預測程式預測其大小為 \hat{N} ，會將它預測的值 \hat{N} 透過無線網路廣播出去。在允入階段(admission phase)中，每個想擷取網路的 MT 都會收到來自 AP 的通知網路目前欲傳送資料的 MT 數量大小可能為 \hat{N} ，接著每個 MT 以類似 Controlled Slotted-Aloha (CS-Aloha)通訊協定的方式，自行擷取下個 slot，並依據 AP 通知的 \hat{N} 來決定允入機率， $P_A(\hat{N}) = \kappa_{opt} / \hat{N}$ 。 κ_{opt} 是個會使網路傳輸效能最佳化的常數，我們以嚴謹的理論推導來得到約為 $\kappa_{opt} \approx 1.52$ 。這群被選中的 MT 將會進行第二階段解析的功能。在這第二階段進行時，我們需要無線網路硬體支援碰撞大小的偵測，即當碰撞發生時在無線網路實體層能回報 MAC 層此碰撞中有多

少 MT 參與，我們將這個功能稱之為 PMER（如圖 0 所示）。再回到第一階段，每個有資料要傳送的 MT 經過允入機率的篩選後，或有部份的 MT 可以獲准開始傳送，此時 PMER 回覆其推估參與碰撞的 MT 數目 \hat{L} 將之回傳給 MAC 層。若 $\hat{L} \leq 1$ 則這個兩階段的工作便告一段落，準備重新回到第一階段；若 $2 \leq \hat{L} \leq 5$ 則開始進入第二階段。HFCA 收到 \hat{L} 之後，廣播給整個無線網域，每個 MT 依據這個 \hat{L} ，可以自行設定第二階段群組個數為 m 。值得一題的是若 $\hat{L} > 5$ 則因超過 PMER 所能解析的能力 PMER 將會回傳 0，視同沒有人傳送，這樣會造成一個 slot 的頻寬浪費，但是由機率的角度來看，這種情形並不常發生。

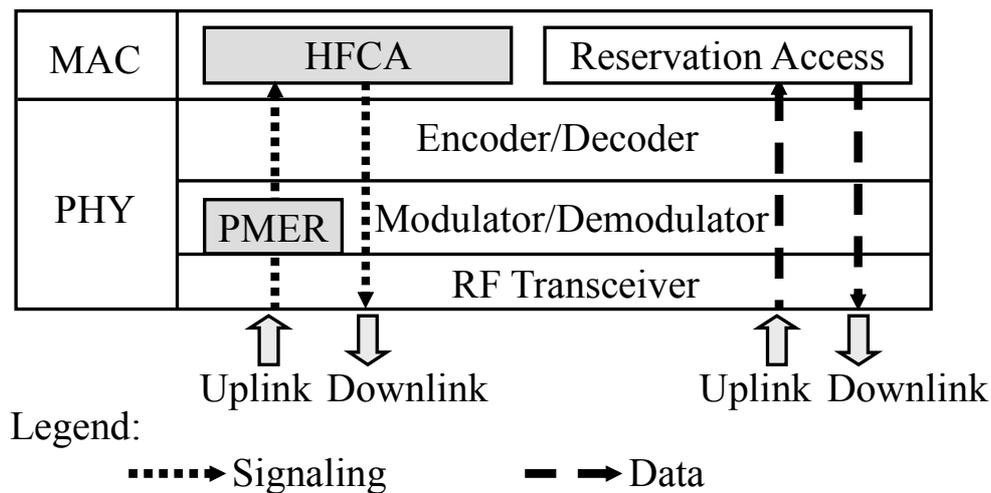


圖 0: HFCA 與 PMER 之關係圖

在第二階段，每個被允入的 MT 是否可以在下一個 slot 傳送，將依據目前還在第二階段的群組大小而決定。在第一階段結束時，AP 會傳送前一次碰撞的大小 \hat{L} 給所有的 MT，MT 自行維護一個群組大小的值 m ，之後所有未傳送的 MT 皆以 m 的大小來自行決定是否在下一個 slot 中傳送資料。其決定的演算法 CS-Aloha 通訊協定，其允入機率為 $P_r(G) = 1/G$ ，其中 G 為目前各 MT 所維護的群組大小(第二階段開始時 $G=m$)，在傳送完成時，PMER 會再發送新的 \hat{L} ，各個 MT 將依據接收到 \hat{L} 的值而改變其 G 值。 \hat{L} 與 G 的關係如下：若 $\hat{L} = 1$ 表示成功，則 $G=G-1$ ；若 $\hat{L} \geq 2$ 表示碰撞發生但 $\hat{L} \neq G-1$ 則 G 不變；若 $\hat{L} \geq 2$ 且 $\hat{L} = G-1$ 表示沒有參與碰撞的 MT 只有一個，此時在下一個 slot 強迫那個沒有送資料的 MT 開始送並將其它已送的 MT 禁止送一個 slot。如此反覆下去直到 G 為零時才停止第二階段的工作（即碰撞已全部解開）。

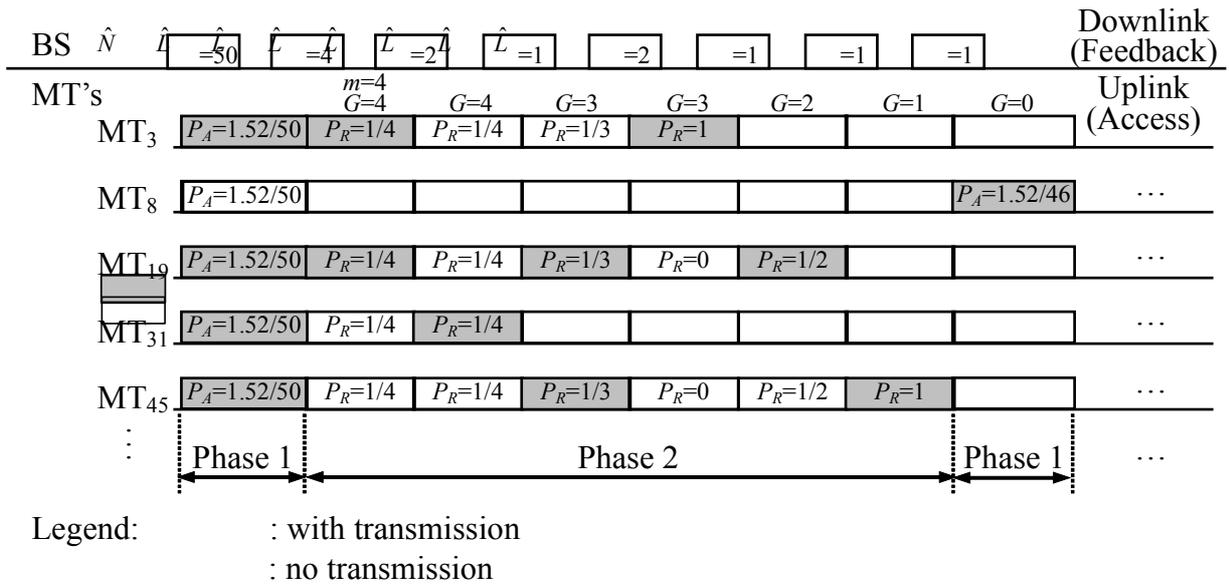


圖 0: HFCA 的實例

以下介紹一個 HFCA 運作的例子 (如圖 0 所示)。在這個例子中，我們假設此無線區域網路有 50 個 MT 想要傳送資料 ($\hat{N} = 50$)。在第一階段中，每個 MT 依據 AP 所傳送來的 $\hat{N} = 50$ 來決定 CS-Aloha 的允入機率 $P_A = 1.52/50$ 。假設有 4 個 MT 發送成功，則 PMER 將碰撞內含的 MT 數 $\hat{L} = 4$ 交給 HFCA，並由 HFCA 廣播給所有的 MT，同時開始以群組大小 $m = 4$ 進入第二階段。在第二階段中，這 4 個發送過資料的 MT，依據得到的 m 自行計算 G ，並套入 CS-Aloha 的允入機率式 $P_R = 1/4$ 中。假設在這個 slot 只有一個 MT (MT₃₁) 傳送資料，則 PMER 會偵測到 $\hat{L} = 1$ ，同時其它沒有送資料的 MT 會更新群組大小 $G = 4 - 1 = 3$ ，接著再以 $P_R = 1/G = 1/3$ 開始進行 CS-Aloha 的動作。假設在接下來的 slot 中有兩個 MT 同時傳送資料，PMER 回傳 $\hat{L} = 2$ ，因為 $\hat{L} = G - 1$ 所以我們將這回沒有傳資料的 MT 在下一個 slot 強迫傳送，這回有送資料的 MT 在下一個 slot 暫停傳送。如此，我們又可以解出一個 MT，使得群組大小又減少一個。如此反覆直到 $G=0$ 表示此群已解完，則開始下一個新的第一階段 $\hat{N} = 46$ 。

由於 HFCA 所在之環境與 802.11 WLAN 有相當的差異，採用以 HFCA 為基礎的方法面臨的最大問題為不知如何將此回饋告知每個 MT 及能動態分配不同的允入機率給需要不同 QoS 的 MT。當然原本 HFCA 是無法直接應用於 802.11 WLAN 上。然而我們將採取以 HFCA 本質之特性來對加強對 802.11 MAC 層的設計，使其能達到高網路效能(throughput)、低存取延遲(access delay)、及提供各種差異性服務(DiffServ)。

三、參考文獻

- [1] F. Baker, Ed. "Requirements for IP version 4 routers," *IETF RFC 1812*, June 1995.
- [2] P. Gupta, S. Lin, and N. McKeown, "Routing lookups in hardware at memory access speeds," *Proc. IEEE INFOCOM'98*, San Francisco, USA, pp. 1240-1247, Mar. 1998.
- [3] N. Huang, and S. Zhao, "A novel IP-routing lookup scheme and hardware architecture for multigigabit switching routers," *IEEE Journal on Selected Areas in Communications*, Vol. 7, No. 6, pp. 1093-1104, June 1999.
- [4] Michigan University and Merit Network Internet Performance Measurement and Analysis (IPMA) Project. Available WWW: <http://nic.merit.edu/~ipma/>
- [5] B. Lampson, V. Srinivasan, and G. Varghese, "IP lookups using multiway and multicolumn search," *IEEE/ACM Trans. Networking*, Vol. 7, No. 3, pp. 324-334, June 1999.
- [6] P. C. Wang, C. T. Chan, and Y. C. Chen, "A fast IP routing lookup scheme," *IEEE Commun. Letter*, vol. 5, pp. 125-127, Mar. 2001.
- [7] I. Widjaja and A. I. Elwalid, "Performance issues in VC-merging capable switches for multiprotocol label switching," *IEEE J. Sel. Areas Commun.*, vol. 17, pp. 178-1189, June 1999.
- [8] V. Makam, C. Huang, K. Owens and V. Sharma, "MPLS: Much Potential Leading to Somewhere? An Assessment of QoS and Protection in MPLS," Proceedings of MPLS'99, Paris, June, 1999.
- [9] S. Makam, V. Sharma, K. Owens, C. Huang, F. Hellstrand, J. Weil, L. Andersson, B. Jamoussi, B. Cain, S. Civanlar, and A. Chiu, "Framework for MPLS-based recovery," *Internet Draft*, Aug. 2000.
- [10] C. Huang, V. Sharma, S. Makam and K. Owens, "Extensions to RSVP-TE for MPLS Protection," IETF contribution, work in progress, June, 2000.
- [11] C. Huang, V. Sharma, S. Makam and K. Owens, "A Path Protection/Restoration Mechanism for MPLS Networks," IETF contribution, work in progress, July, 2000.
- [12] K. Owens, V. Sharma, S. Makam, B. Mack-Crane, and C. Huang, "A path protection/restoration mechanism for MPLS network," *Internet Draft*, Jan. 2002.
- [13] C. Huang, V. Sharma, K. Owens and S. Makam, "Building Reliable MPLS Networks Using a Path Protection Mechanism," *IEEE Communications Magazine*, pp. 3-8, March 2002.
- [14] V. Sharma, and F. Hellstrand, "Framework for Multi-Protocol Label Switching (MPLS)-based Recovery," *IETF RFC-3469*, Feb. 2003.
- [15] C. Villamizar, "MPLS Optimized Multipath (MPLS-OMP)," *Internet-Draft*, 1999.
- [16] C. Villamizar, "OSPF Optimized Multipath (OSPF-OMP)," *Internet-Draft*,

1999.

- [17] K. Long, Z. Zhang, and S. Cheng, "Load balancing algorithms in MPLS traffic engineering," *IEEE Workshop on High Performance Switching and Routing (HPSR) 2001*, pp.175-179, 2001.
- [18] L. Berger, D. Gan, G. Swallow, P. Pan, F. Tommasi, and S. Molendini, "RSVP refresh overhead reduction extensions," *IETF RFC 2961*, April 2001.
- [19] D. Awduche, L. Berger, D. Gan, T. Li, V. Srinivasan, and G. Swallow, "RSVP-TE: extensions to RSVP for LSP tunnels," *IETF RFC 3209*, Dec. 2001.
- [20] D. Awduche, A. Hannan, and X. Xiao, "Applicability statement for extensions to RSVP for LSP-tunnels," *IETF RFC 3210*, Dec. 2001.
- [21] J. M. Chung, "Analysis of MPLS traffic engineering," *Proc. IEEE Midwest Symposium on Circuits and Systems 2000*, Vol.2, pp.550-553, 2000.
- [22] P. Brittain, A. Farrel, "MPLS traffic engineering: A choice of signaling protocols," *Data Connection Limited*, Jan. 17, 2000.
- [23] R. Braden, D. Clark, and S. Shenker, "Integrated Services in the Internet Architecture: an Overview," *IETF RFC-1633*, June 1994.
- [24] J. Wroclawski, "Specification of the Controlled-Load Network Element Service," *IETF RFC 2211*, September 1997.
- [25] S. Shenker, C. Partridge, and R. Guerin, "Specification of Guaranteed Quality of Service," *IETF RFC 2212*, September 1997.
- [26] S. Shenker and J. Wroclawski, "General Characterization Parameters for Integrated Service Network Elements," *IETF RFC 2215*, September 1997.
- [27] S. Shenker and J. Wroclawski, "Network Element Service Specification Template," *IETF RFC 2216*, September 1997.
- [28] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "An Architecture for Differentiated Services," *IETF RFC 2475*, Dec. 1998.
- [29] V. Jacobson, K. Nichols, and K. Poduri, "An Expedited Forwarding PHB," *Internet Draft, draft-ietf-diffserv-phb-ef-01.txt*, November 1998.
- [30] V. Jacobson, K. Nichols, and K. Poduri, "An Expedited Forwarding PHB," *IETF RFC 2598*, June 1999.
- [31] J. Heinanen, F. Baker, W. Weiss, J. Wroclawski, "Assured Forwarding PHB Group," *Internet Draft, draft-ietf-diffserv-af-03.txt*, November 1998.
- [32] J. Heinanen, F. Baker, W. Weiss, and J. Wroclawski, "Assured Forwarding PHB Group," *IETF RFC 2597*, June 1999.
- [33] K. Nichols, S. Blake, F. Baker, and D. Black, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers," *IETF RFC 2474*, Dec. 1998.
- [34] J. Heinanen, and R. Guerin, "A Single Rate Three Color Marker," *IETF RFC 2697*, Sept. 1999.
- [35] J. Heinanen and R. Guerin, "A Two Rate Three Color Marker," *IETF RFC 2698*, Sept. 1999.

- [36] W. Fang, N. Seddigh, and B. Nandy, "A Time Sliding Window Three Colour Marker (TSWTCM)," *IETF RFC 2859*, June 2000.
- [37] F. Wang, P. Mohapatra, S. Mukherjee, D. Bushmitch, "A random early demotion and promotion marker for assured services," *IEEE Journal on Selected Areas in Commun.*, vol.18, no.12, pp. 2640-2650, Dec. 2000.
- [38] Keng-Ming Huang (黃鏗銘), "An Effective IP Address Lookup Scheme: Unicast and Multicast (高效能之 IP 位址查詢技術：單點廣播與多點廣播)," M.S. thesis (碩士論文), Department of Communication Engineering, National Chiao Tung University, Taiwan, 2001.
- [39] Ching-Shi Huang (黃慶喜), "Further Investigation of TS-UPC and New Metering Algorithms for DiffServ (訊務塑型器-使用參數控制的深入研究與差異性服務網路新的量測機制)," M.S. thesis (碩士論文), Department of Communication Engineering, National Chiao Tung University, Taiwan, 2001.
- [40] Po-Han Chen (陳柏翰), "Traffic Reduction and Path Recovery Algorithm for Real-time Services(即時性服務之流量縮減與路徑復原演算法)," M.S. thesis (碩士論文), Department of Communication Engineering, National Chiao Tung University, Taiwan, 2002.
- [41] Yung-Hung Cheng (鄭永宏), "The Traffic Conditioner improving Bandwidth Allocation for DiffServ Networks (在差異性服務網路中具有改善頻寬配置功能的訊務調節器)," M.S. thesis (碩士論文), Department of Communication Engineering, National Chiao Tung University, Taiwan, 2002.
- [42] Ning-You Yan(顏寧佑), "Packet Classification and Traffic Marking Algorithms for DiffServ Networks (差異性服務網路中封包分類與訊務標示方法之討論)," M.S. thesis (碩士論文), Department of Communication Engineering, National Chiao Tung University, Taiwan, 2003.
- [43] P. C. Lin and C. J. Chang, "A Priority TCAM IP-Routing Lookup Scheme," *IEEE Communication Letters*, Vol. 7, No. 7, pp. 337-339, July 2003.
- [44] K. M. Huang and C. J. Chang, "A Fast Multicast IP-Routing Lookup Scheme," *IEEE Communications Letters*, Vol. 7, No. 3, pp. 133-135, March 2003.
- [45] P. C. Lin and C. J. Chang, "Analysis of Buffer Requirement for ATM-LSRs with Partial VC-Merging Capability," *IEICE Transactions on Communications*, Vol. E-85B, No. 6, pp. 1115-1123, June 2002.
- [46] P. H. Chen and C. J. Chang, "An Enhanced Path Recovery Algorithm for MPLS Networks," *Proc. of National Symposium on Telecommunications (NST'2002)*, Nantou, Taiwan, Dec. 2002, Session. NET-1-6.
- [47] C. J. Chang, Y. H. Cheng, and L. F. Lin, "The Traffic Conditioner with Promotion and Fairness Guarantee Schemes for DiffServ Networks," *Proc. of ICC 2003*, Anchorage, Alaska, USA, May 11-15, 2003, Vol. 1, pp. 238-242.
- [48] C. Dovrolis, D. Stiliadis, and P. Ramanathan, "Proportional Differentiated Services: Delay Differentiation and Packet Scheduling," *ACM Sigcomm99*, pp.109-120.
- [49] T. Li, and Y. Rekhter, "A Provider Architecture for Differentiated Services and

- Traffic Engineering," *RFC2430*, October 1998.
- [50] H. Zhang, and D. Ferrari, "Rate-controlled static priority queueing," *IEEE INFOCOM '93*, Apr. 1993, pp. 227-236.
- [51] V. Shinivasan, S. Suri, and G. Varghese, "Packet Classification using Tuple Space Search," *ACM SIGCOMM99*, pp.135-146.
- [52] C. Dovrolis and P. Ramanathan, "A Case for Relative Differentiated Services and the Proportional Differentiation Model," *IEEE Network*, Vol. 13, No. 5, Sept./Oct. 1999, pp. 26-35.
- [53] S. Floyd and V. Paxson, "Difficulties in simulating the internet," *IEEE Trans. Networking*, vol. 9, no. 4, pp. 392-403, Aug. 2001.
- [54] Chen-Nee Chuah, A Scalable Framework for IP-Network Resource Provisioning through Aggregation and Hierarchical Control, *Ph.D. thesis*, University of California at Berkeley, 2001.
- [55] T. Ngo-Quynh, H. Karl, A. Wolisz, and K. Rebersburg, "Relative jitter packet scheduling for differentiated services," in *Proc. of 9th IFIP Conf. on Performance Modelling and Evaluation of ATM & IP Networks 2001*, Budapest, Hungary, June 2001, pp. 139-151.
- [56] Thyagarajan Nandagopal, Narayanan Venkitaraman, Raghupathy Sivakumar, and Vaduvur Bharghavan, "Delay differentiation and adaptation in core stateless networks," in *Proceedings of the 2000 IEEE Computer and Communications Societies Conference on Computer Communications (INFOCOM-00)*, TelAviv, Israel, Apr. 26-30 2000, pp. 421-430, IEEE.
- [57] Constantinos Dovrolis, Dimitrios Stiliadis, and Parameswaran Ramanathan, "Proportional differentiated services: Delay differentiation and packet scheduling," in *SIGCOMM99*, 1999, pp. 109-120.
- [58] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "RFC 2475: An architecture for differentiated services," Dec. 1998, Status: PROPOSED STANDARD.
- [59] J. Wroclawski, "RFC 2211: Specification of the controlled-load network element service," Sept. 1997, Status: PROPOSED STANDARD.
- [60] S. Shenker, C. Partridge, and R. Guerin, "RFC 2212: Specification of guaranteed quality of service," Sept. 1997, Status: PROPOSED STANDARD.
- [61] Leonard Kleinrock, *Queueing Systems, Volume II: Computer Applications*, Wiley inter-science, New York, 1976.
- [62] Network Simulator 2, Available: <http://www.isi.edu/nsnam/ns/>.
- [63] V. Srinivasan, G. Varghese, S. Suri, and M. Waldvogel, "Fast and Scalable Layer Four Switching," *Proc. ACM SIGCOMM '98*.
- [64] M. Waldvogel, G. Varghese, "Scalable High Speed IP Routing Lookups," *Proc. ACM SIGCOMM '97*, Sept. 1997, pp. 25-36.
- [65] B. Lampson, V. Srinivasan, and G. Varghese, "IP Lookups Using Multiway and Multicolumn Search," *Proc. IEEE INFOCOM '98*, Apr. 1998, pp.1248-56.
- [66] T. V. Lakshman and D. Stidifialis, "High Speed Policy-based Packet Forwarding Using Efficient Multi-dimensional Range Matching," *Proc. ACM SIGCOMM*

'98, Sept. 1998.

- [67] A. Feldman and S. Muthukrishnan, "Tradeoffs for Packet Classification," *Proc. IEEE INFOCOM '00*, Mar. 2000, pp.397-413.
- [68] P. Gupta and N. McKeown, "Packet Classification on Multiple Fields," *Proc. ACM SIGCOMM '99*, Sept. 1999.
- [69] F. Baboescu and G. Varghese, "Scalable Packet Classification," *Proc. ACM SIGCOMM '01*, Aug. 2001.
- [70] Ji Li, Haiyang Liu, and Karen Sollins, "Scalable Packet Classification Using Bit Vector Aggregating and Folding," *Proc. ACM SIGCOMM '02*, Aug. 2002.
- [71] V. Srinivasan, S. Suri, and G. Varghese, "Packet Classification using Tuple Space Search," *Proc. ACM SIGCOMM '99*.
- [72] P. Gupta and McKeown, "Classification Using Hierarchical Intelligent Cuttings," *Proc. Hot Interconnects VII*, Aug. 1999.;also available in *IEEE Micro*, vol. 20. no. 1, Jan./Feb. 2000, pp. 34-41.
- [73] Abilene NetFlow Nightly Reports, <http://www.itec.oar.net/abilene-netflow>
- [74] 郭國承, "利用濃縮的位元向量之可擴充性的封包分類演算法," 國立交通大學資工系碩士論文, Jun. 2003.
- [75] 陳志暉, "一個提供比例式延遲差異服務的封包排程演算法," 國立交通大學資工系碩士論文, Jun. 2002
- [76] 李春良, "在網際網路上提供具擴充性服務品質之研究," 國立交通大學資工系博士論文, Oct. 2001
- [77] F. Akujobi, I. Lambadaris, R. Makkar, N. Seddigh, and B. Nandy, "BECN for congestion control in TCP/IP networks: study and comparative evaluation," *Proc. IEEE GLOBECOM'02*, pp. 2588-2593, Nov. 2002.
- [78] Yi-Cheng Chan, Chia-Tai Chan, and Yaw-Chung Chen, "An Enhanced Congestion Avoidance Mechanism for TCP Vegas," *IEEE Comm. Letters*, Vol. 7, No. 7, pp.343-345, July 2003
- [79] Mark E. Crovella and Azer Bestavros, "Self-similarity in world wide web traffic: Evidence and possible causes," *IEEE/ACM Trans. Networking*, vol. 5, no. 6, pp. 835-846, Dec. 1997.
- [80] Will E. Leland, Muard S. Taqqu, Walter Willinger and Daniel V. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Trans. Networking*, vol. 2, no. 1, pp. 1-15, Feb. 1994.
- [81] J. M. Peha, "Retransmission mechanisms and self-similar traffic models," *IEEE/ACM/SCS CNDSMS*, pp. 47-52, Jan. 1997.
- [82] J. Beran, R. Sherman, M. S. Taqqu, and W. Willinger, "Long-range dependence in variable-bit-rate video traffic," *IEEE Trans. Commun.*, pp. 1566-1579, 1995.
- [83] D. E. Duffy, A. A. McIntosh, M. Rosenstein and W. Willinger, "Statistical analysis of CCSN/SS7 traffic data from working subnetworks," *IEEE JSAC*, pp. 544-551, June 1994.
- [84] W.-C. Lau, A. Erramilli, J. Wang, and W. Willinger. "Self-similar traffic

- generation: the random midpoint displacement algorithm and its properties," *ICC'95*, pp. 466-472, June 1995.
- [85] K. Meier-Hellstern, P. E. Wirth, Y. L. Yan and D. A. Hoeflin, "Traffic models for ISDN data users: Office automation application," *In Proc. 13th ITC*, Copenhagen, Denmark, pp. 167-172, 1991.
- [86] V. Paxson and S. Floyd, "Wide area traffic: The failure of Poisson modeling," *IEEE/ACM Trans. Networking*, pp. 226-244, June 1995.
- [87] B. B. Mandelbrot, "A fast fractional Gaussian noise generator," *Water Resources Research*, vol. 7, pp. 543-553, 1971.
- [88] B. B. Mandelbrot and J. W. Van Ness, "Fractional Brownian motions, fractional noises and applications," *SIAM Rev.*, vol. 10, pp. 422-437, 1968.
- [89] B. B. Mandelbrot and J. R. Wallis, "Computer experiments with fractional Gaussian noises," *Water Resources Research*, vol. 5, pp. 228-267, 1969.
- [90] Mandelbrot B.B., *The Fractal Geometry of Nature*, pp. 17-18, San Fransisco: W.H. Freeman and Co., 1982.
- [91] Boris Tsybakov and Nicolas D. Georganas, "Self-similar processes in communications networks," *IEEE Trans. Inform. Theory*, vol. 44, no. 5, pp. 1713-1725, Sep. 1998.
- [92] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, 2nd ed. San Francisco, CA: Holden Day, 1976.
- [93] C. W. J. Granger and R. Joyeus, "An introduction to long-memory time series models and factional differencing," *J. Time Series Anal.*, vol. 1, pp. 15-29, 1980.
- [94] J. R. M. Hosking, "Fractional differencing," *Biometrika*, vol. 68, pp. 165-176, 1981.
- [95] Gennady Samorodnitsky and Murad S. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*, Chapter 7, Chapman & Hall: New York, 1994.
- [96] V. Frost and B. Melamed, "Traffic modeling for telecommunications networks," *IEEE Commun. Mag.*, vol. 33, pp. 70-80, Mar. 1994.
- [97] Jon M. Peha, "Retransmission mechanisms and self-similar traffic models," *IEEE/ACM/SCS Communications Networks and Distributed Systems Modeling and Simulation Conference*, Phoenix, Arizona, pp. 47-52, Jan. 1997.
- [98] Jack L. Goldberg, *Matrix Theory with Applications*, New York: McGraw-Hill, 1991.
- [99] V. Paxson, "Fast, approximate synthesis of fractional Gaussian noise for generating self-similar network traffic," *Computer Communication Review*, pp. 5-18, Oct. 1997.
- [100] Y. Kwok, and V. Lau, "A Quantitative Comparison of Multiple Access Control Protocols for Wireless ATM," *IEEE Trans. Vehicular Technology*, vol. 50, no. 3, May 2001, pp. 796-815.
- [101] S. Jangi, and L. Merakos, "Performance Analysis of Reservation Random Access Protocols for Wireless Access Network," *IEEE Trans. Comm.*, vol. 42, no. 2/3/4, Feb./March/April 1994, pp. 1223-1234.

- [102] M. Yuang, and P. Tien, "Multiple Access Control with Intelligent Bandwidth Allocation for Wireless ATM Networks," *IEEE J. Select. Areas Commun.*, vol. 18, no. 9, Sept. 2000, pp. 1658-1669.
- [103] L. Lenzini, M. Luise, and R. Reggiannini, "CRDA: A Collision Resolution and Dynamic Allocation MAC Protocol to Integrate Data and Voice in Wireless Networks," *IEEE J. Select. Areas Commun.*, vol. 19, no. 6, June 2001, pp. 1153-1163.
- [104] I. Akyildiz, et al., "Medium Access Control Protocols for Multimedia Traffic in Wireless Networks," *IEEE Network*, July/Aug., 1999, pp. 39-47.
- [105] N. Passas, et al., "Quality-of-Service-Oriented Medium Access Control for Wireless ATM Networks," *IEEE Comm. Mag.*, vol. 35, no. 11, Nov. 1997, pp. 42-50.
- [106] M. Ivanovich, M. Zukerman, and F. Cameron, "A Study of Deadlock Models for a Multiservice Medium Access Protocol Employing a Slotted Aloha Signalling Channel," *IEEE/ACM Trans. Networking*, vol. 8, no. 6, Dec. 2000, pp. 800-811.
- [107] W. Jeon, D. Jeong, and C. Choi, "An Integrated Services MAC Protocol for Local Wireless Communications," *IEEE Trans. Vehicular Technology*, vol. 47, no. 1, Feb. 1998, pp. 352-364.
- [108] B. Paris, and B. Aazhang, "Near-Optimum Control of Multiple-Access Collision Channels," *IEEE Trans. Comm.*, vol. 40, no. 8, Aug. 1992, pp. 1298-1309.
- [109] L. Georgiadis, and P. Papantoni-Kazakos, "A Collision Resolution Protocol for Random Access Channels with Energy Detectors," *IEEE Trans. Comm.*, vol. 30, no. 11, Nov. 1982, pp. 2413-2420.
- [110] J. Mosley, and P. Humblet, "A Class of Efficient Contention Resolution Algorithms for Multiple Access Channels," *IEEE Trans. Comm.*, vol. 33, no. 2, Feb. 1985, pp. 145-151.
- [111] M. Yuang, and B. Lo, "QMAC: A QoS-Guaranteed MAC Protocol with Dynamic Granularity Control for Local Wireless ATM Networks," *Proc. IEEE ICC, 2001*, pp. 2359-2364.
- [112] Y. Su, and J. Chen, "Carrier-to-Interference Ratio Measurement using Moments or Histograms," *IEEE Trans. Comm.*, vol. 48, no. 8, Aug. 2000, pp. 1338-1346.
- [113] S. O. Rice, "Probability Distributions for Noise Plus Several Sine Waves- the Problem of Computation," *IEEE Trans. Comm.*, vol. 22, June 1974, pp. 851-853.
- [114] A. S. Rosenbaum, "PSK Error Performance with Gaussian Noise and Interference," *Bell Syst. Tech. J.*, vol. 48, Feb. 1969, pp. 413-442.
- [115] R. Rom, and M. Sidi, *Multiple Access Protocols-Performance and Analysis*, Springer-Verlag, 1990.
- [116] S. Ghez, S. Verdu, and S. Schwartz, "Stability Properties of Slotted Aloha with Multipacket Reception Capability," *IEEE Trans. Automat. Contr.*, vol. 33, no. 7, July 1988, pp. 640-649.

四、計畫成果自評

本總計畫的研究主題是在寬頻網際網路的環境中，為達成頻寬與 QoS 服務品質保證的各項端對端訊務控制或運作機制之研究設計，針對寬頻網際網路中的路由選徑（包含 Unicast Routing 與 Multicast Routing）寬頻傳輸技術與多項 QoS 相關之控制運作機制，皆提出有效、且效能更佳的设计或解決方案，並著眼於無線行動通訊的發展趨勢，將實體網路相關的控制技術延伸至無線通訊的環境，設計支援 QoS 的無線媒介存取控制技術(MAC) 以及頻寬分配機制；此外，亦對網際網路中具有自我類似(Self-Similar) 性質的真實網路訊務進行研究，瞭解其特性和分析其成因，並提出一有效率的 Self-similar 訊務源模型與訊務產生方法。在三年的執行時間中，計畫參與人員分別針對各項研究主題的各個層面進行探討，除了瞭解現有方法與研究的不足之處，也追蹤並研讀最新發表的相關研究論文或技術標準，進而提出相對的改進方案。

在高速路由選徑機制方面，我們分別發展出「階層式分群解析架構」與「TCAM-based 架構」兩種適合於硬體邏輯電路實現的路由選徑方法，對於每一個 IP 封包的路由查詢，可達成平均在 1 至 2 次的記憶體存取(Memory Access) 動作內即可以獲得路由結果，大幅提昇網際網路中路由選徑的處理速度，而配合硬體邏輯電路的運作方式，可以滿足 Gigabit 超高速網路環境下以及未來更寬頻的網際網路應用的需求；並且可以將相關電路整合成一硬體路由搜尋引擎的模組形式，更方便商用化路由功能產品的使用與成本的降低。同時透過適當的運作架構與資料結構的安排和設計，也兼具所需記憶體容量小的優點，此優點將有助於進一步將前述硬體路由搜尋引擎轉化為系統晶片(SoC) 的可行性(因為已經可以將運作所需的記憶體一併整合入單一晶片中)，或是在傳統的路由器架構中能夠以相同的記憶體容量支援更大規模網際網路(例如：骨幹級網路)的訊務路由工作。另外我們也對於可降低網路訊務流量的網路「群播(Multicast)」機制所需的特殊路由方式—群播路由(Multicast Routing) 進行研究設計，在瞭解 Multicast Routing 的關鍵技術是為一個多欄位資料完全相符(Exactly Match) 的比對搜尋機制後，我們提出以 CAM 此一支援資料完全相符比對搜尋的記憶體為基礎的高速群播路由選徑方法，可在 3 次的記憶體存取動作便完成一筆群播路由的查詢，再加上管線式(Pipeline) 運作架構的可行性，也同樣能使得平均查詢次數降至一次的記憶體存取次數。

在寬頻傳輸技術方面，我們著眼於 MPLS 網路的相關議題。首先，對於以 ATM 技術為基礎的 MPLS 網路之 VC-merge 機制，提出 Partial VC-merge 運作模式下 ATM-LSR 交換路由器記憶體容量需求之快速且精確的分析方法：不僅較傳統分析方式降低分析過程的複雜度（可由 $O(M^4)$ 減少為 $O(M^2)$ ），也因此能夠對較大 Buffer 的分析得到較高的 Cell Loss Prob.的準確度；而我們所採用的 Input

訊務的數學模型更接近於實際的情況，所得的結果也較為接近真實，所以此分析方法極適合用於評估實際 ATM-LSR 交換路由器所需的記憶體緩衝區容量(Buffer Size)，讓交換路由器的系統資源做更有效率的應用。其次，我們也提出一 MPLS 網路路徑損壞或錯誤發生的保護及快速回復(Recovery) 機制，讓傳輸路徑發生錯誤或損壞時還能夠維持部分基本的通訊，並可以快速而正確地地恢復既有的通訊，降低高速 MPLS 網路上傳輸路徑錯誤或損壞所造成的影響，減少封包遺失率，並能夠進一步達到動態負載平衡的附加效益，使系統資源做最佳的利用，因而有效提高 MPLS 網路的資料輸出率(throughput)。

在支援 QoS 保證的訊務處理與控制機制方面，我們著重在 IETF 所建議的 DiffServ 網際網路 QoS 架構之基礎關鍵性元件的研究與設計，包含封包分類器、訊務監控調節器、以及封包排程器。在封包分類器(Classifier) 的研究與設計方面，網際網路差異化服務的第一步驟就是封包的分類，根據封包中各個欄位的資訊對封包做不同的服務等級分類。在這個領域中，我們提出一個可實際使用在未來路由器設計的封包分類演算法，這個演算法可以依據實際網路的需求和路由器本身的硬體限制，調整演算法以得到最佳的效能，同時適應各種不同的分類器特性。在訊務監控調節器(Traffic Conditioner) 的研究設計方面，提出一兼具訊務監控精確度與網路資源使用效率，並進一步達成 micro-flow 微訊務流之 **Cell Dropping Precedence** 標記公平性(Fairness) 的訊務監控調節機制：在確保訊務特性符合 Traffic Profile 的規範下，精確地讓系統資源獲得充分且最佳的利用，並進一步保障其中各 micro-flow 連線所實際獲得的 QoS，如此才能夠**提升 DiffServ 此網際網路 QoS 架構的可行性與使用效益**，並做為其他 QoS 訊務控制機制（例如：CAC 連線允諾控制機制 Scheduling 排程控制 流量控制 壅塞控制等 Per-Hop Behavior）運作的基礎。另外，在封包排程器(Scheduler) 的研究設計方面，我們也針對一個較不容易讓人注意到的差異化服務指標—「封包延遲變異(Packet Delay Variation 或 Jitter)」進行研究，提出一個稱為 AJP 的排程器，此排程器可以使不同類別資料流量的平均封包延遲變異達到成比例式差異的差別服務。此外，著眼於無線區域網路(WLAN) 和行動通訊科技產業的蓬勃發展，未來網際網路亦將延伸至無線通訊技術的網路系統中，因此無線通訊網路中具備 QoS 的網際網路存取技術也成為本計畫在 QoS 相關機制研究的一項重要課題。對此，我們設計一個持續穩定且有高效能並支援 QoS 的無線媒體多重接取(Multiple Access) 競入技術，能夠使得競入訊號能以極高效率進入系統，同時保持區域網路的最大效能，並進一步於其上設計出一個動態頻寬分配機制。此外還能夠配合前述所發展的多項 QoS 訊務控制機制，真正達到無線通訊環境中具 QoS 保證的網際網路（端對端）應用或服務的目標。

在具有自我類似性質的真實網路訊務的成因分析與訊務源模型設計方面，主要的研究著重在三個課題：(1) 探討並解釋在原本 Markov-modeled 的網路訊務(Network Arrival) 中如何產生具有長程相依特性(Long-Range Dependence) 的

自我類似性質之網路訊務；(2) 具備自我類似性質之訊務流其合成方法的設計；(3) 分析自我類似訊務流在資訊理論上的特徵。透過子計畫三的研究，我們分別完成三個相對應的研究成果：(1) 在 Self-similar 網路訊務的成因探討上，驗證了聚合訊務流的自我類似現象與其訊務源或端點在產生訊務流時受網路規約 (Protocol) 機制運作的影響有關：此種受網路規約影響，特別是重傳 (Re-Transmission) 機制，產生的訊務流，應是整體流入訊務流具備自我類似統計特性的主要原因之一。此外我們也證明了有限時間內的自我類似效應的存在性及其在 Markov-modeled network 中的定義；(2) 在自我類似訊務源模型與訊務產生方法的設計方面，透過信號濾波器(Filter) 的理論與形式，我們提出一個具備即時性訊務產生功能，且產生訊務量必為正值的自我類似訊務源模型與訊務產生器。除了可以即時產生具備自我類似特性的訊務，方便各種系統模擬與效能評估的應用之外，更重要的是，其產生出來的訊務流較其他方法的結果更為接近真實網路中自我類似訊務的另一項特徵：在工程可控管的尺度或更高一些的尺度上展現出自我類似的特性，但在極大尺度上仍為非自我類似的。如此也將使其應用於系統模擬與效能評估時的結果更接近於真實。我們將此結果應用在前述各項寬頻傳輸技術與 QoS 控制機制的驗證上，則有助於這些技術與控制機制於真實網路的可行性；(3)從多種自我類似訊務流的分析結果中，自相互關係資訊皆呈現 power law decreasing 指數下降的形式，確定了自我類似訊務流在資訊理論上的特徵 - 自相互關係資訊是為 power law decreasing 的形式。此結果將可用於訊務流自我類似性質的檢驗與其自我類似性(Self-similarity) 程度的指標。

綜合這三年的研究成果，我們認為研究內容與原計畫緊密結合，已達成預期的計畫目標。此外，本計畫的多項研究成果已整理成相關的學術論文投稿，且為國際性學術期刊所接受並刊出，或為國際會議所接受並於相關議程中進行發表，這除了是對於本計畫研究成果的肯定之外，並且也提供我們能夠與相關領域中的研究同道或先進有充分互動、交流的機會。另一方面，我們的研究成果也具備相當高的實用價值，除了擬進行專利的申請外，亦可進一步與產業界合作或技術移轉進行實際產品的應用與開發。綜此觀之，我們於本計畫執行的研究成果兼具學術成就與實用價值，可謂相當成功而豐富。

國際合作研究計畫國外研究報告書

計畫名稱：寬頻網際網路端對端技術之研究

(End-to-End Techniques for Broadband Internet)

合作機構：加拿大 渥太華 Carleton 大學寬頻網路實驗室

執行時間：2002 年 7 月 26 日 2002 年 8 月 25 日

我方研究人員：張仲儒 教授（第一子計畫暨總計畫主持人）

陳耀宗 教授（第三子計畫主持人）

林立峰、詹益禎（博士班研究生）

對方計畫主持人：Prof. Changcheng Huang（黃長城 教授）

Prof. Ioannis Lambadaris

國際合作研究計畫之緣起和必要性：

本國際合作研究計畫案，是於原三年期國科會整合型計畫「寬頻網際網路端對端技術之研究(End-to-End Techniques for Broadband Internet)」執行的第二年度，進行計畫變更為國際合作研究計畫，與加拿大渥太華 Carleton 大學工學院的寬頻網路實驗室(BNL) 合作，共同投入下一代支援服務品質(QoS) 之寬頻網際網路相關技術的研究。

在我們原本的三年期整合型計畫中，主要的目標即是針對達成寬頻網際網路所需的有線或無線寬頻傳輸技術，以及頻寬與 QoS 服務品質保證的各項端對端訊務控制或運作機制進行研究設計，提出有效、且效能更佳的解決方案，或是從事其效能分析評估與改進的工作；此外，亦以一子計畫投入網際網路中具有自我類似(Self-similar) 性質之真實網路訊務的研究，瞭解其特性和分析其成因，如此將有助於 QoS 訊務控制機制或通訊協定於實務應用上的最佳化設計。此子計畫並擬提出一有效率的 Self-similar 訊務源模型與訊務產生方法，可做為所開發的各項控制機制或架構之效能模擬分析、評估之用，或進一步做為其設計的考量或改進的依據。

而 Carleton 大學寬頻網路實驗室在其實驗室計畫主持人 Prof. Changcheng Huang（黃長城教授）與 Prof. Ioannis Lambadaris 兩位教授的指導和帶領下，過去在寬頻網路之訊務源模型、快速（網路）模擬技術、壅塞控制機制、與光纖網路之架構與傳輸技術等方面已有相當豐富的研究成果，甚至在其中的 Self-similar 自我類似特性的訊務與快速網路模擬技術上更是有世界領先的卓越表現。近幾年來，該實驗室研究團隊則開始投入網際網路上的服務品質保證相關技術的研究，在支援 QoS 服務品質保證的訊務控制機制、網際網路群播(Multicast) 技術的應

用和相關運作控制機制的開發和設計、甚至是未來全光化高速光網路與現有 IP 網際網路之間的網路互連運作(Interworking) 或網路整合的問題都加以探討並提出解決方案，其中亦延續在 Self-similar 特性訊務源的研究(成果)，輔助各項控制機制或架構的設計，確保達到實務應用上的最佳效能。

由此觀之，雙方研究團隊在相關研究領域上均有相當且相近的背景與實力(經驗)，同時對於目前和未來規劃的研究課題上亦有共同的目標與興趣，因此非常適合以共同研究合作的方式交流彼此的經驗和成果，除了可促進雙方在既有研究題目的進度與研究領域的深度之外，也可由不同視野和經驗的激盪，達到提升雙方在其他相關研究領域的廣度的雙贏效益。此外，Carleton 大學系統及電腦工程學系寬頻網路實驗室，長期與國際通訊大廠 Nortel Networks 或其他業界公司有密切的合作關係，在寬頻網路訊務控制機制、訊務模式、與光纖通訊網路等已有多年的設計實作心得及研究成果。著眼於本研究團隊大多的研究成果僅止於理論分析或電腦模擬驗證，較缺乏實際系統或網路環境的試煉，因此期望透過本計畫雙方的合作研究，能達到理論與實務並重，以及實際問題思考經驗與理論推導另類方向的激盪討論。

國外研究之執行：

於本三年期整合性計畫執行之第二年度開始，進行計畫研究人員赴國外合作單位訪問研究之行程，並以兩個年度共四位教授與四位博士班研究生的人次來執行。第一年度的訪問研究於 2002 年 7 至 8 月間順利執行完畢，由張仲儒(子計畫一暨總計畫主持人)與陳耀宗(子計畫三主持人)二位教授帶領博士班林立峰和詹益禎二位研究生，赴加拿大渥太華 Carleton 大學進行為期一個月的國外(出訪)研究交流行程。然而由於研究主力仍是以學生部分的交流為主，因此二位教授的部分實際僅規劃十天的行程，主要是希望能夠透過雙方計畫主持人會面的機會，擬定此次國外研究的實際進行方式，並討論後續在雙方跨國合作計畫上的執行模式、細節與計畫預期成果。待二位教授返國後，兩位博士生仍留在合作的寬頻網路實驗室與該實驗室成員共同研究，學習並交流彼此的經驗與成果，至一個月後才結束此次國外研究行程返國。第二年度(原整合型計畫第三年度)的國際合作研究部分，亦分別規劃有兩位教授與兩位博士班研究生進行國外(訪問)研究的行程，由於原訂於本年度出訪的陳伯寧教授(子計畫二主持人)臨時無法成行，在行文國科會變更出訪教授名單後，則仍維持由第一年度的張仲儒與陳耀宗二位教授帶領另二位博士生來進行。第二年度所核准的研究時程較第一年度增加為三個月，但教授部分仍維持十天，除了用於雙方目前研究進度與成果的報告外，也仍希望透過見面的機會，進行上一年度計畫執行問題的檢討，同時有鑑於此年度已為計畫執行的最後一年度，故也期望能針對計畫結束後，如何延續此一國際合作研究關係、維持雙方持續而穩定的學術交流，討論出可行的進行方式。

然而原擬於 2003 年 4、5 月間出發的第二年度國外研究行程，卻因為當時國際間爆發了 SARS 的流行，而我國亦成為 SARS 疫區，且加拿大方面升高對我防疫與檢疫的標準，所有入境人員皆必須進行十天自我或強制隔離的安全防疫措施，相關手續繁瑣而不便，在與加方 Carleton 大學計畫主持人聯繫討論後，決定取消第二年度國際合作計畫的國外研究行程，並行文國科會告知此一訊息且同時申請變更計畫執行。

由於 SARS 的影響，導致本國際合作研究計畫之國外研究部分僅於原整合型計畫之第二年度執行，而且所核准的研究時程較短（教授十天，博士班研究生一個月），為了在短時間內能有一定的成果呈現，在雙方計畫主持人會議討論後決定採取的策略是，先針對各自手邊已略具雛形的研究進行報告及討論，在對彼此的研究專長與目前正在進行的研究題目有所瞭解後，再由雙方計畫主持人討論決定具體的合作研究題目，原則是採取雙方皆有興趣也最好是正在進行中的研究題目為主。因此接下來便以兩天的時間進行雙方的研究專題報告，題目的範圍涵蓋了網路標準的二、三，四層。

在我方的報告中，由陳耀宗教授與其博士班研究生詹益禎提出一個關於網際網路 TCP 通訊協定於網路壅塞狀態下的問題，即是當 TCP 的 ACK 路徑上若發生擁塞，會因為對於整個通訊連線的 RTT 值的估算錯誤而做出不恰當的 Congestion Window 的調整，如此將對於 TCP 的控制效能有嚴重的影響。理論上 ACK 路徑上的擁塞不應該造成 Data Packet 路徑上的效能減低，然而實際上的運作結果的確會發生這樣的問題，特別是對於一個新提出的 TCP 版本「Vegas」的影響最為嚴重，因此當前正著手進行 TCP Vegas 上此一課題的研究，以期提出簡單而有效的解決方案。而張仲儒教授與博士班研究生林立峰則針對本身實驗室過去在寬頻網路上各項支援 QoS 服務品質保證的訊務控制機制的研究，以及本整合型計畫第一年度的研究進度與成果，連同目前正在研究的課題，進行介紹和說明，主要是著重在連線允諾控制(Call Admission Control, CAC) 機制、訊務監控調節(Traffic Conditioner) 機制、採用硬體邏輯電路架構為主的高速路由選徑方法、以及 MPLS 寬頻傳輸技術網路之路徑錯誤保護暨快速回復機制 等方面。而有鑑於 Carleton 大學方面的計畫主持人黃長城教授在 MPLS 網路之可靠性研究領域亦長期投入，並已有相當成熟的研究經驗和成果 [8-13]，其中包括參與網際網路標準組織 IETF 在相關技術的建議標準—RFC-3469 [14] 的制訂工作，以及恰於當年度國際期刊 IEEE Communication Magazine 中有相關的、MPLS 網路之通訊路徑回復機制的論文發表 [13]，因此我們也特別針對此方面的研究進度做較為深入的報告，以期雙方能在此問題上有進一步探討與合作的機會。而 MPLS 網路之路徑回復機制的起因與問題主要是在於，雖然藉由 MPLS 技術的提出的確適時為網際網路提供了高速且低延遲的訊務傳輸能力，但相對的，當高速的 MPLS 網路發生傳輸路徑錯誤或損壞的時候，往往也會造成更嚴重的影響（例如更大量的資料遺失），尤其對於即時性服務而言更是如此。因此有必要提出一套

有效的路徑保護(Protection) 或回復(Path Recovery) 機制，以便當 MPLS 網路傳輸路徑發生錯誤(failure) 或損壞的時候，還能夠維持部分基本的通訊，並可以快速而正確地恢復既有的通訊，降低 MPLS 網路上傳輸路徑錯誤或損壞所帶來的影響，減少封包遺失率，如此便可有效的提高網路的資料輸出率(throughput)。

在 Carleton 大學方面，則是由計畫主持人及寬頻網路實驗室其他研究成員進行實驗室研究成果介紹，以及幾項進行中或已完成之研究專題報告，包含：採取分波多工(Wavelength Division Multiplexing, WDM) 技術之光纖網路中，資料傳輸用光波長的最佳化分配機制、IEEE 802.17 標準的 RPR 環狀光網路中，光網路節點之資料傳輸多工緩衝器的效能分析與設計、以及網際網路中 TCP 壅塞控制機制 等方面的具體研究成果。其中特別值得注意的是，在 TCP 通訊協定的壅塞控制機制方面，研究生 Frank Akujobi 提出了一個有別於傳統 TCP 壅塞控制的方式，不再只是單純地由網路上一通訊連線的兩個端點來進行而已，而是希望可以利用連線的封包路徑上所通過的路由器的協助，能夠進一步提高 TCP 的效能。其作法是，當路由器發現網路開始壅塞時，路由器便透過主動發出 ICMP 訊息(BEEN) 的方式，通知經過它的所有 TCP 連線的 Source 端網路已開始壅塞，使 TCP 的 Source 端及早對壅塞做出反映。經過不斷的討論及修正，最後經由網路模擬器(NS) 的模擬，結果證實達到了預期的效果。此研究結果已發表在國際會議論文中，如 [77] 所示。

根據前述專題報告會議的交流與討論的結果，雙方決定以「TCP 通訊協定之壅塞控制機制」和「MPLS 網路之路徑錯誤保護暨快速回復機制」此兩項議題為此次合作研究的主要項目，但也保持在其他研究議題方面的討論和經驗交流。

國外研究之具體成果：

在赴 Carleton 大學寬頻網路實驗室進行為期一個月的合作研究行程，以及返國後我研究團隊與對方持續透過 E-mail 進行討論後，在上述兩項具體的合作研究課題上皆獲得初步的研究成果。在「TCP 通訊協定之壅塞控制機制的設計」方面，針對 TCP Vegas 在 ACK 路徑壅塞時的效能下降問題，擬採取的解決方式是，在 TCP 連線的 receiver 端對每一個 ACK 打上時間標記(timestamp)，source 端經過某種運算，分別計算出封包在 Data Packet 路徑上以及 ACK 路徑上的佇列延遲時間(queueing delay) 以瞭解網路壅塞的狀況，藉著區別出網路擁塞的方向後，我們所提出的控制方法便可透過較正確的 RTT 值估算而有效的改進 TCP Vegas 在 ACK 路徑上發生擁塞時的效能，提升訊務流量(Throughput)。此研究結果已發表在期刊論文中，如 [78] 所示。在「MPLS 網路之路徑錯誤保護暨快速回復機制」方面，我們提出一個結合「Re-Routing Model」與「Protection Switch Model」兩種方式的優點高速而最佳的路徑保護 / 回復機制的�方法，利用每個 LSR

上預先規劃的 Backup path 的 Database 來幫助我們達成快速路徑回復的目的，取代了以往錯誤發生後才開始搜尋 Backup path 或是錯誤發生前先保留資源建立 Backup path 的缺點：藉由 Re-Routing Model 方法的使用，不但排除了 Protection Switch Model 此類作法中預先建立 Backup path 並保留資源而造成資源浪費的缺點，引入 Protection Switch Model 預先規劃 Backup path 的概念，加速了既有 Re-Routing Model 作法的 Recovery time，而且仍保有 Re-Routing Model 中在最近路徑錯誤的節點馬上進行切換的優點，因而能夠快速且有效地降低網路錯誤發生時之影響時間與所需付出的代價。而且在原訂的研究目標之外，我們還能夠透過所設計的機制具備多路徑傳輸的能力達到動態負載平衡的附加效益，使系統資源做最佳的利用，如此可再更進一步提高網路的資料輸出率(throughput)。目前此部分的研究成果已整理為論文投稿 [46]，並已在 2002 年於暨南大學所舉辦的全國電信會議(NST 2002) 中進行發表。

而在上述的兩項具體合作研究成果之外，在本計畫其他的研究項目中，有許多具有實用價值的設計亦有對方的協助，達成國際合作計畫預期的理論與實務並重的目標！

國外研究心得與建議：

經過短短四星期的交流及合作討論，合作雙方藉由腦力激盪的過程，都獲得了新的研究思維，也認識到對方研究態度上的長處，對於本研究團隊日後的研究應有不小的正面助益。同時藉著此次合作研究關係的建立，將有助於未來在長距離大型網際網路（特別是跨國性長距離網際網路）上通訊與訊務控制相關機制的分析與研究的發展，雙方可實際建立一跨國性網際網路實驗平台，彼此互為對方在此實驗平台兩端的研究助理，將所發展的研究成果在此實際的跨國性大型網際網路上進行測試、驗證，所得結果將可進一步改善我們的設計，亦可對於我們研究成果的正確性與實用性提供強而有力的證明。

由於此計畫是本研究團隊第一次的國際性合作研究案的執行，經驗仍不太足夠，所以在實際出訪時間的選擇上恰好正是當地暑假的開始，因此留在學校的學生人數並不多，互相交流機會比較少，再加上此期間也多為教授們的休假、休息時間，因此在原本合作的教授外，也未能再和更多當地的教授認識與交流，為本次國外合作較為可惜之處。另外在核准出訪的時間長度方面亦顯不足，許多研究交流與成果多有賴於返國後繼續透過 E-mail 進行後續的討論來達成，雖然這也可以是一種計畫執行的方式與考驗，不過面對面的對談、討論仍是比較有助於問題的釐清，對於研究工作而言是比較有效率的方式。期望在未來的國際合作研究計畫裡可以針對這些部分加以改進，延長出訪國外研究的時間，建議至少為三個月的時程，避免過短的時間將可能導致赴國外研究人員在開始習慣當地的生活

模式與作息規律，並可以進入實質的研究工作時即屆返國日期並必須暫停研究工作的缺點了。而較長時間的合作與交流，也較有助於建立更佳深厚而長久的友誼與合作研究關係。