鏈路負載平衡演算法使用於企業網路環境之研究

楊千
交通大學資訊管理研究所

彭祖乙
交通大學資訊管理研究所

傅振華
國防大學國防管理學院國防資訊研究所

摘要

本研究針對使用於企業多條聯外路徑環境(multihoming)的負載平衡演算法作分析及比較。Multihoming 是一種提升企業網路運作穩定性的機制,其量測及選擇路徑的方式已經有 BGP 與 RON 等相關研究加以探討,但是 BGP 和 RON 需要藉由所連接的 ISP 額外提供大量路由訊息的交換；而使用線路負載平衡則是一種不需要 ISP 額外提供支援的運作機制，它已經被應用在許多企業網路環境之中，亦有許多商業化的產品可提供企業組織使用。

本研究以三個應用實例:「頻寬聚合」,「擁塞線路反應」及「斷線反應」進行探討。

本研究揭示了不同演算法的反應特性,可作為實務導入時演算法選擇的依據並可供相關演算法進一步的研究使用。

關鍵詞：網路可靠度、多路徑、負載平衡

A Study on Link Load-balanced Algorithms in an Enterprise Network

Chyan Yang
Institute of Information Management, National Chiao Tung University

Tsu-i Peng
Institute of Information Management, National Chiao Tung University

Chen-Hua Fu
Institute of Defense Information, National Defense Management College

Abstract

This study investigates load-balanced algorithms applied in multihoming environment for enterprise network. Multihoming is a mechanism to provide reliable connectivity of network. The path measuring and path selecting operations of multihoming networks using BGP and RON have been discussed in many studies, however, BGP and RON require ISPs to exchange extra large routing information. Link load-balancing is another mechanism; it does the path selecting and path measuring jobs well without ISP's extra support. Link load-balancing has already been applied in many enterprise networks. Currently, many commercial products are available.

The three practical applications: bandwidth aggregation, link congestion and link failure, would be analyzed and discussed in this study.

The responding behaviors of various algorithms are revealed, the results can be used for practical application and further studies.

Keywords: network reliability, multihoming, load-balancing

壹、導論

一個穩定的網路環境對於電子商務的內容服務或線上交易是非常重要的，它除了增加交易的可靠度亦提高使用者的滿意度。Multihoming 即是一種可為企業組織建立穩定網路運作的機制,它除了可以增加企業網路的效能及穩定度，同時具有節費的功效（Akella 2003 ; Goldenberg 2004）。

Multihoming 在企業網路環境的運作機制為使用多條聯外線路來傳輸企業的交通流量。企業並可以使用低廉且不需重新佈線的最後一哩線路（last miles）如 ADSL、Cable Modem、 Power link 及 Wimax 等來取代一條專線，用以減低線路成本。

目前有關 multihoming 網路環境路由的研究或實務應用主要包括： BGP（Rekhter 1995）、RON（D. G. Andersen 2001）及 Load-balance（FangluGu 2004, Akella 2004）三種方式；其中，BGP 需要終端客戶都與每一家和它連結的 ISP 交換 BGP 路由訊息，並能繞送相互的位址，由於這樣做的成本很高，因此只有大型的企業能負擔相關的產品及服務。RON 則是需要 ISP 提供特別的封裝通道（tunneling）來導引封包到不同的路徑，目前提供這樣的商用網路極少（akami 2005）。

Load-balance 則是將應用於伺服器負載平衡的觀念移植到網路鏈路的流量分配上，它會依據線路的負載，把流量以最少負載或權重的方式分配到所有的線路。Load-balanced 使用自訂的量測方式來選擇線路，因此不需要 ISP 的支援。負載量測方式有很多種:有使用最後一哩可用頻寬的量測，也有使用反應時間（response time）的方式。目前市場上已有許多鏈路負載平衡商品(Radware 2005; F5 2005; Deansoft 2005)可供選用。

本研究將分析線路負載平衡的運作機制，探討構成不同演算法的共同要素。然後以仿真模擬（emulation）的方式，來觀察不同的鏈路負載平衡鏈路負載平衡演算法對於「頻寬聚合」、「線路擁塞」及「斷線」這三個方面網路連線狀況的影響,並做效能的比較。

「頻寬聚合」、「線路擁塞」及「斷線」這三個方面的重要性在於：

頻寬聚合可以使用多條成本較低的線路取代昂貴的專線;對擁塞線路的適當處理將可以增加使用者的傳輸滿意度;而避開斷線的情形將可以增加交易的可靠度。

貳、文獻探討

一、ＢＧＰ及ＲＯＮ的方法

BGP 可以支援 multihoming 環境下的路由處理，在不同 ISP 位址的繞送問題上，BGP 可以使用 RFC 2260（Bates 1998）或 RFC1518（Rekhter 1993）所提的方法，它們主要是讓接續在同一家企業不同的 ISP 有能力透過自己的線路繞送其他家 ISP 的位

址。企業可以使用這些 ISP 所提供不同的 IP 位址。在路徑的選擇方式上，BGP 在 BGP4（Rekhter 1995）會依數個屬性的優先順序比較它們的數值，選擇最好的路徑，這些屬性數值來自於外部的設定或是最短路徑的計算。

BGP 相關的問題有許多的探討，如不同定址方式帶來了的路由成本及 BGP 在選擇路徑上不夠即時也不能反映實際流量的問題（Li 2003 ; Labovitz 2000）。在 multihoming 的環境，BGP 要交換整個路由表（routing table），如果所有的終端客戶都與相接連的多家 ISP 交換 BGP 的訊息，那麼這些 ISP 將會承受以 O（n^2）倍增的複雜度，因此實務上只會運用在較大型的企業。

RON 的方式主要針對 BGP 無法即時反應網路的變動，而且也無法準確的找到最佳的路徑的問題提出改善方案，它使用自主性高的量測方式，針對到達目的地的每一條路徑進行反應時間的量測。在 RON 的網域（network domain）裡面，假設有 N 個節點，那麼每個節點會對其他 N-1 個節點進行量測，其彼此量測的複雜度也將是 $O(n^2)$。

RON 架構以一種點對點（end-to-end）的量測方式做為尋找最佳路徑的方式，其量測的方式是每隔一段時間（約 14 秒鐘）在兩個節點相互交換 UDP 封包， RON 藉著這種量測提供三種的評估值: Latency、Loss 與 Throughput。每個節點對於量測其他節點的這三個評量值，會被存入一個屬於這個節點的傳輸效能資料庫（performance data base），而不同節點會透過鏈路狀態路由交換協定（link state routing protocol）交換傳輸效能資料庫，以便知曉其他節點路由狀況，進而進行拓璞的計算，藉著這些交換，RON 可以選擇以 latency、Loss rate 或 throughput 找到點對點（end-to-end）最佳的路徑。由於不同節點之間是以非同步的方式持續進行量測，因此反應問題路徑的時間比較快，比較 BGP 要數分鐘的時間，RON 可以在幾十秒內避開問題路徑。RON 的研究也顯示有 11%的傳輸降低了 40ms 的延遲(latency)。

RON 的量測成本，在 50 個節點相互量測的時候會有 33kbps 頻寬需求，而它的量測成本與網路節點數量呈現指數（exponential）相關的成長，因此當有 500 個節點可能會有數個 mbps 的量測頻寬需求，這對網路是一種負擔，亦限制 RON 架構運作的規模。

二、負載平衡(Load-Balance)的方法

BGP 和 RON 都需要和網路中其他的節點交換路由資訊，以得到傳輸的最佳路徑。Load-balance 則不需要 ISP 的支援。Load-balance 使用 NAT（Egevang 1994）的方式來做定址，它以 session 為單位來分配線路，同一個 session 會被導到同一條線路。Load-balance 選擇線路分配的方式，是依據每條線路的流量負載來分配流量，它採用類似伺服器叢集負載平衡的概念。在 RFC2391 中提到了伺服器負載平衡的運算方法（Srisuresh 1998），其中一個網路節點將扮演分配者角色的 LSNAT 路由器，它依據網路或伺服器的負載量及接續成本（access cost）來分配連線，並把該連線的目的地位址轉換為被指定的伺服器位址。當負載平衡的對象是線路而不是伺服器的時候，需要做一些調整。首先對伺服器負載和流量的統計，得轉換成是對線路的統計，其次，也不

需要做目的地 IP 位址的轉換；在相關的商品(如：Radware， F5， Deansoft)也都有類似 RFC2391 使用最少連線、最少流量及最快反應時間的演算法。

RFC2391 所提 Load-balance 連線分配的演算法包括有 Round Robin（RR）、 Least Loading First（LLF）、 Least Traffic First（LTF）、 Ping to find the Most Responsive host（PMR）與 LLF 及 LTF 加入權重分配(weighted)的變化形式 WLLF 及 WLTF。其中， LLF 的負載量（loading）是以連線的數量來評估，而 LTF 的流量（traffic）是以實際流出的每個封包位元大小累計來計算，PMR 會以 ICMP 來回的 RTT 值來評估。具權重形式的演算法，主要是將 session 數目和伺服器的服務能力做為權重值一起考量，以 WLTF 為例，假設伺服器 A 的是 3，伺服器 B 的權重值是 1，若同時有四個 session 來到，則此時伺服器 A 會分配到 3 個 sessions， 伺服器 B 會分配到 1 個 session。

參、鏈路負載平衡機制

在這個章節中，首先介紹鏈路負載平衡系統的運作架構，並針對目前台灣的 ISP 線路環境及經濟效益進行相關分析，然後探討構成線路負載平衡演算法的幾個共同構面，最後說明各種相關的鏈路負載平衡演算法。

一、運作架構

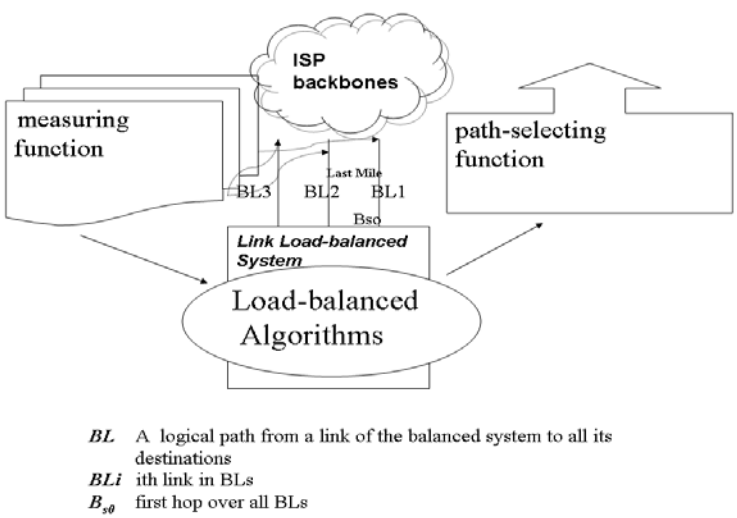


圖 1 multihoming 鏈路負載平衡系統示意圖

圖 1 是一個線路負載平衡系統的運作圖示，其中 BL_i 代表不同的 ISP 線路。B_{s0} 代表每條線路中的第一段跳躍（hop）。這個示意圖顯示線路負載平衡演算法涵括主要有兩個功能：量測（measuring）及線路分配（path selection）；量測功能可依流量的多寡及

反應時間的快慢來決定線路的好壞。線路分配功能可以使用最佳(Best)或權重(weighted)的方式來決定網路應用流量傳輸的線路。線路分配的單位通常是以交易連線(TCP/IP session)為主，每個連線會依線路量測的結果選擇某條 ISP 鏈路做為傳輸的線路。

連線方向可以區分為二種，一種是從企業網路內部經由負載平衡設備對外的(outbound)連線，另一種則是由外部鏈路經由負載平衡設備對內的(inbound)連線。對外的連線分配會使用 NAT 的方式來定址，使回覆的封包也從同一條線進入。對內的連線，會以網域名稱伺服器（DNS）解析不同線路的 IP 來導引連線進入不同的線路，由於這需要 DNS 階層式架構的支援，而控制權也不全然在負載平衡設備本身，因此本研究將以企業網路對外的連線為研究之主軸。

在擁有多條 ISP 線路的 multihoming 環境下，可藉由下列表示式(1)表達鏈路負載平衡機制的最佳化運作效能：最大資料傳送量、最小斷線率與最低成本。

$$
\begin{aligned}
&Max\left(\sum_i Throughput\left(li\right) - Throughput\left(L\right)\right)\\
&Min\left(\prod_i FailRate\left(li\right)\right)\\
&Min\left(\sum_i Cost\left(li\right)\right)\\
&subject\ \text{to}\\
&\prod_i FailRate\left(li\right) < FailRate\left(L\right)\\
&\sum_i Cost\left(li\right) < Cost\left(L\right)\\
&li\ is\ an\ existing\ choice\ in\ the\ market
\end{aligned}
\qquad (1)
$$

在表示式(1)中，L 係指可提供資料傳送量較高且成本高的聯外鏈路，而 l 係指資料傳送量較低但成本低高的聯外鏈路，透過鏈路負載平衡機制之運作，可利用具經濟效益且最穩定實際存在的線路組合，來滿足企業聯外網路所需之頻寬需求。在價格的組合上，以國內目前的 ISP 業者為例，以 T1 專線 1.544Mbps 的線路，每個月要數萬元的連線費用，就可以使用較低價的 ADSL 來取代，如表一列出了國內三家 ISP 業者 T1 專線及 ADSL 雙向 512kbps 線路的費用。

企業網路多條聯外鏈路可以經由同一家 ISP 業者或多家 ISP 業者連接網際網路，使用者從同一家 ISP 及不同家 ISP 來申請聯外鏈路，各有不同的好壞處；聯外鏈路均經由同一家 ISP，除了統一管理與維護較方便外，亦可爭取優惠價格，但線路的失敗率可能較高，不過同一家 ISP 業者也可以藉著從不同機房提供線路來提高最後一哩線路的可靠度；從不同家 ISP 申請聯外鏈路，主要可以降低線路失敗的風險。

表一 台灣主要 ISP 業者 T1 and 512 k ADSL 線路月費比較表

ISP 業者	T1 月費	512k ADSL 月費
中華電信	54,600	3,700
速博	145,000	3,700
台灣固網	144,000	3,700
		金額單位：新台幣 元

二、運作要素分析

本節將探討鏈路負載平衡機制運作二個共同構面：線路流量量測及線路選擇。

（一）線路流量量測構面

這裡探討流量量測的數個要素, 及與流量量測息息相關的「斷線偵測」

1.流量量測的要素

線路流量量測構面可區分為三個必要元素：量測的時間點、量測的方式和量測的距離。

(1)量測時間

線路流量量測的頻率越高，越能反映網路的流量變化情形，但相對量測花費的成本亦越多，也越消耗系統的資源。量測時間點的選擇可以為被動(passive)或主動(active)。被動的方式可以在每個封包或每個連線(TCP/IP session)到來的時候做量測; 主動的方式則是每隔一個時段對網路線路進行量測。一般來說，被動的方式是比較即時的，但是能否在連線到達時間(arrival time)的同時量測到當時網路的流量，則取決於量測的方式所花的時間。

(2)量測方式

量測的方式一般有三種形式: 連線數（session number）、流量統計及來回時間（round trip time）。連線數是目前還在進行連線的數目，連線數可以代表在固定的頻寬下，每個連線平均可以得到的服務。舉例來說，假設 A，B 線路的頻寬一樣是 R，而 A 線路有兩個連線，B 線路有一個連線，那麼 B 線路的可以得到的平均服務值（*R*）比 A 較大（*R/2*）。

流量統計主要是觀察過去的時間所使用頻寬的多寡，通常可以封包長度的累計來獲得，但是須考慮「已使用」頻寬是有時間性的，舉例來說，以上面 A、B 兩線路為例，假設 A 線路在五秒前已使用 *R* 個位元的頻寬， B 是在這一秒用掉了 *R/2* 個位元的頻寬，那麼 B 的線路可能還比較擁塞。

來回時間主要依發出某個封包到有回應的來回時間，如 ICMP 的 echo request 到 echo reply 或 TCP connection 的 SYN sent 到 SYN ack；一般而言，來回時間和量測的距離有一定的關係。

(3)量測距離

量測的距離係指從起始網域（source network）到目的網域（destination network）間所行經路由中某一特定節點與鏈路負載平衡設備之間的距離；圖二顯示從鏈路負載平衡設備到目的端之間數個不同的量測距離。

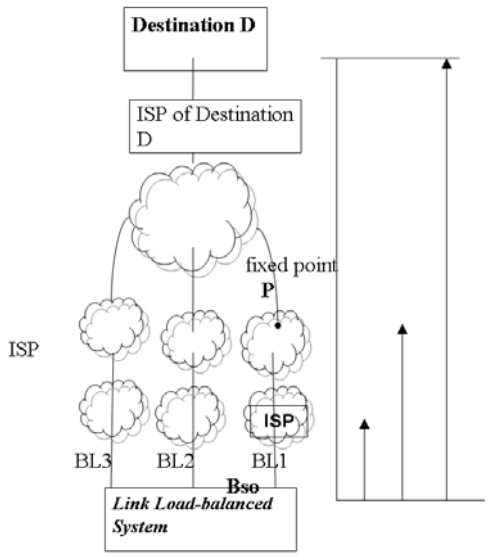


圖 2 鏈路負載平衡機制不同量測距離示意圖

由圖 2 所顯示各種不同的量測距離代表著不同的量測功能：

- B_{s0} 本身，在鏈路負載平衡設備上觀察流量資訊，如每條線路的連線數。
- B_{s0}-ISP，觀察最後一哩的流量，如用最後一哩的頻寬和使使用的頻寬來計算可用的頻寬。
- B_{s0}-P，在路由路徑中，對所有不同的目的端選擇一個會經過的定點來計算回應時間，這個定點可能是 ISP 一個出口路由器。
- B_{s0}-D，點對點進行回應時間的量測。

2.斷線偵測

線路的斷線檢測（link failure detection）可和線路流量量測分開或合併進行。

線路斷線檢測可藉由反應時間量測的方式進行，利用反應時間的量測結果判斷線路斷線與否；因為得不到反應時間的線路，通常代表有異常情況的發生；但是這和量測距離亦有關聯，如果使用 B_{s0}-P 的量測距離，那麼 P 點以外的異常都就無法量測。

此外，使用其他的量測方式，(如：連線數、流量統計）需要額外的斷線偵測機制，因為線路斷線時會造成流量減少的現象，這會使以連線數和流量統計量測方式誤以為線路傳輸狀況很好；額外的線路斷線偵測機制通常會針對幾個穩定的網路節點進行反應時間的量測，這些機制包括： ICMP echo request、UDP traceroute 或是直接使用 TCP 來連線到該網路節點。線路斷線的判定可藉由多個節點共同失敗的連線來判定線路斷線，它將會延長判定線路斷線的時間；亦可利用單一固定節點的連線失敗來判定，但有可能發生誤判的現象。如何取捨線路斷線判斷的速度與判斷的精準度是一個值得考量的議題。

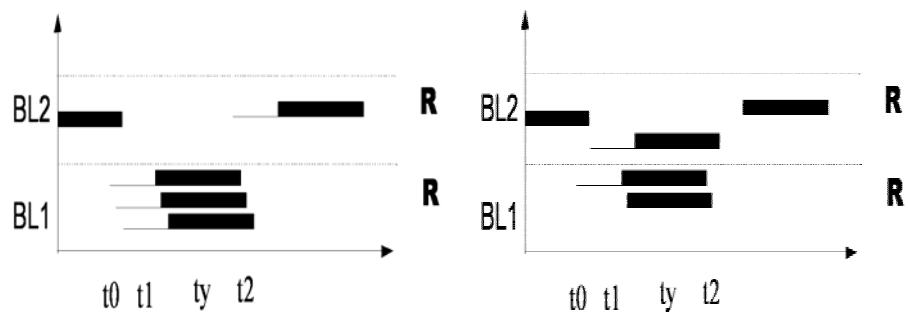
（二）線路選擇構面

就連線線路的選擇而言，通常鏈路負載平衡機制會選擇將流量分配至當時量測最好的線路，此一分配方式稱之為最佳（best）模式；但最佳模式可能造成「自我引發擁塞」的狀況，亦就是在同一時段內的連線都會被分配到量測最好的線路，進而造成該線路的壅塞；如圖 3(a)所示，在 t_0 這個時間點，BL_1 被量測為最好的線路，從 t_0 到 t_1 這個時段，三個連線都分配到這條線路，進而形成 BL_1 這條線路在 t_y 這個時間點產生壅塞的現象。等到 t_2 這個時間點，BL_2 才被量測為最好的線路，但是在 t_2 時間點之前 BL_2 線路的頻寬會因此被閒置。

為了避免這種情形，可以選擇另外一種權重（weighted）模式的分配方式。權重模式的線路分配方式會依每條線路的量測結果賦予一個權重值，然後依權重比例將當時所有連線需求依(2)式分配到不同的線路；如圖 3(b)所示，使用權重模式的方式來分配流量，在 t_1 這個時間點給與 BL1 的權重是 2 而 BL2 的權重是 1，這時候可以發現分配的方式會利用到 BL2 的頻寬。

$$Min(Si \ / \ Wi) \quad for\, i = 1..n \qquad (2)$$

S_i 是每條線路的累計連線數，W_i 是那條線路的權重值，權重可以依量測的結果來定其大小。



(a) 最佳模式　　(b) 權重模式

細線表示每個連線在開始建立時使用的流量還不大。BL_i 代表 balanced ISP link i

圖 3 鏈路負載平衡機制不同線路選擇方式示意圖

三、各種負載平衡演算法

前文提及 RFC 2391 的負載平衡方法包括：Round robin（RR）、Least loading first（LLF）、 Least traffic first（LTF）與 Ping to find the most responsive host（PMR），本研究將依據前面述及鏈路負載平衡的相關構面進行相關之分析並提出修訂方法。

(一)RR 與 LLF

RR 與 LLF 此二種演算法均可直接運用於鏈路負載平衡機制中，只是一般線路負載平衡設備商品會以 Least Connection First（LCF）（Radware, Deansoft）這個名詞取代 LLF，它會把連線分到目前擁有連線數最少的線路，可以用(3)式表達：

$$Min(SI_i) \quad for\ i = 1..n \qquad (3)$$

其中 SI_i 係指尚未完成傳輸作業連線的總數。

與 RR 的不同之處在於 LCF 使用還沒有結束的連線數目做為選擇線路分配的依據，而 RR 則是採用輪流分配的方式選擇線路並不會考慮每條線路上面的流量，然而只有還沒有結束傳輸的連線會對流量有影響。RR 和 LCF 還有另外一項特點，那就是它們不會產生如圖三（a）中「自我引發擁塞」的情形，因為 RR 和 LCF 不會把相鄰的連線分配到同一個線路。RR 會循序的分配線路,而 LCF 觀察式(3)可以發現是把式(2)的 W_i 設為 1 的特例,它也會用平均的方式分配。因此 RR 和 LCF 有和權重分配般平均分配線路的特性。

(二)LTF

LTF 可以延伸成兩類的線路演算法: Maximum Inbound/Outbound Remaining Bandwidth First （MIRBF/MORBF）和 Weighted Maximum （Inbound/Outbound）Remaining Bandwidth First（WMIRBF / WMORBF）。這樣的延伸改變主要有兩個考量因素：方向與可用頻寬；方向係因為LTF主要根據線路的流量做統計，但ISP線路在inbound和outbound的頻寬可能是非對稱的，而可用頻寬則是需考慮所統計的流量對當下頻寬的影響。

MIRBF 和 MORBF 是使用前面所提到流量統計的方法，在最後一哩線路使用不同的方向來把線路分配到有最大可用頻寬的線路。而 WMORBF 及 WMIRBF 則是考量了前面所提到的權重模式的分配分式，讓 MIRBF 和 MORBF 根據量測的結果與賦予的權重來分配每個連線。

（三）PMR

PMR 可以延伸成三種演算法: Fastest Round Trip Time to each Destination First (FRRTDF)和 Fastest Round Trip Time to a Fixed Node First (FRRTFNF) 及 Weighted Fastest Round Rrip Time to a Fixed Node First (WFRRTFNF).

FRRTDF 和 FRRTNF 的差異主要在量測的距離。以前一節及圖 2 提到的量測距離，FRRTDF 使用(B_{s0}-D)即點對點的量測；FRRTNF 使用(B_{s0}-P)即選擇一個固定的點進行量測。FRRTDF 和 FRRTNF 都會使用反應時間最好的線路來分配連線。WFRRTFNF 則是依據 FRRTFNF 量測每條線路反應時間的結果賦予權重，然後使用權重模式方式來分配連線。

FRRTDF 演算法並沒有賦予一個對應的權重轉換方式，主要是因為它對每個目的地同時進行量測，如果要以權重模式進行線路分配則需得到每條線路的量測結果，其

成本相對比較高，因此只它使用最快反應回來的那一條線路進行線路分配的指定。

此外，此一類別演算法本身的量測結果即可做為判斷線路是否斷線的依據，可以用於線路斷線的偵測。

表二將上述三種類型的演算法加以整理與比較。

表二 線路負載平衡方法的分類表

名稱	量測距離	分配方式	量測方式	可獨自反應斷線
RR	B_{s0}	weight	連線數目	否
LCF	B_{s0}	weight	連線數目	否
MORBF / MIRBF	B_{s0}-ISP	best	流量統計	否
WMORBF / WMIRBF	B_{s0}-ISP	weight	流量統計	否
FRRTDF	B_{s0}-P	best	反應時間	可
FRRTFNF	B_{s0}-D	best	反應時間	可
WFRRTFNF	B_{s0}-P	weight	反應時間	可

肆、模擬結果及探討

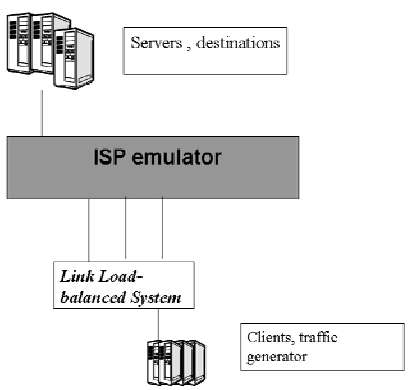


圖 4 模擬環境示意圖

圖 4 顯示本研究仿真模擬環境，其中 ISP emulator 具有數條 512k/512k 的 ISP 線路，並可以產生近端和遠端的塞車行為和斷線行為。這個仿真模擬環境會由用戶端產生 http 的下載流量到遠端的伺服器端，藉由不同的資料流量（traffic load）和不同的網路情況，本研究將從頻寬聚合，塞車反應及斷線反應來瞭解不同線路負載平衡演算法的運用及特性。

一、頻寬聚合

本研究以表一所列出國內 T1 和雙向 512k ADSL 不同費率做為頻寬聚合（bandwidth aggregation）模擬想定依據，期望能以最少的 512k 線路達成 T1 的效能，因此本研究選擇與 T1 頻寬相同的 3 條 512k ADSL 做為企業網路的聯外鏈路，並以 L(U,S)表示使用者持續產生的資料流量，其中 U 代表同時間的人數，S 代表每個使用者同時間的 session 數目。這些資料流量將以類似網頁下載傳輸的方式模擬，當同一個使用者在同一時段內有多個 session 即代表某一個網頁存在著多個資料物件（objects）將被傳送。由於焦點是在頻寬聚合，本研究將使用以最後一哩的流量統計的演算法 MIRBF 和 WMIRBF 做為比較分析的基準，如表二的分類,它們分別使用最佳模式與權重模式來做線路分配的方式。

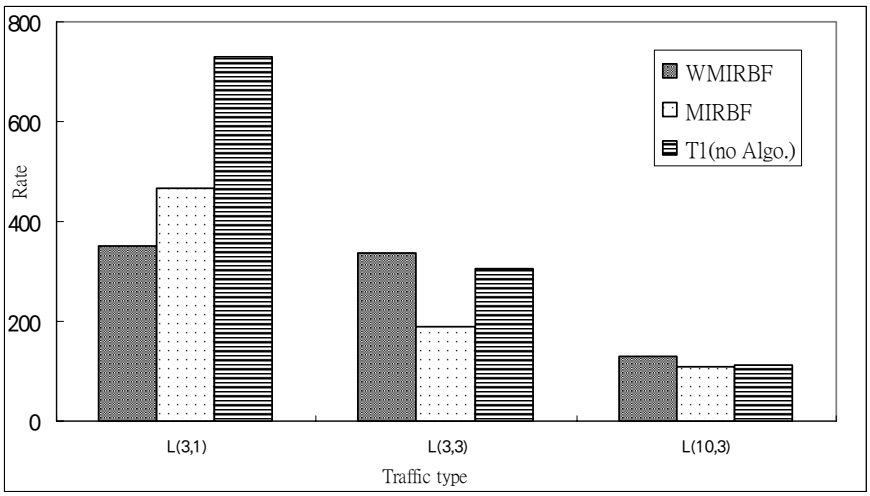


圖 5 頻寬聚合比較統計圖

依據圖 5 所顯示的模擬統計數據，可以發現下列現象：

- 當資料流量小的時候,鏈路負載平衡機制運用 MIRBF 演算法得到的結果是最好的。
- 當 session 數目增多的時候，如 L(3,3)和 L(10,3)，WMIRBF 演算法的效能就變得比較好，它的頻寬聚合效能也驅近於 T1 的效能。
- 當 session 數少的時候，第三節圖 3 中提到的「自我引發擁塞」情形並不嚴重，所以只有在 sessions 多的時候，以權重模式方式分配線路才會獲得比較好的效能。

此外，本研究亦發現如欲以 3 條 512k 線路達到 T1 線路，發揮聯外線路頻寬聚合的效能，其前提必須是 3 條 512k 聯外線路能夠同時傳送資料;對一個企業網路而言，同一時段內產生超過 3 個使用者產生 3 個連線需求數（如：L（3,3））是十分容易的。

二、線路壅塞反應

本節將探究鏈路負載平衡機制相關演算法對於網路壅塞的反應及其資料傳輸量的比較；模擬想定將在一條線路上(第二條線路，BL_2)產生網路壅塞的狀況，產生網路壅塞的方式係利用將上傳的頻寬(upload)減小，使上傳頻寬只有 5kbits，同時反應時間會從 50ms 變成 2000ms；在使用 HTTP 下載的情況下，上傳頻寬將影響 TCP 通訊協定 acknowledgement 的運作，進而影響下載的速度。網路壅塞的地點會先放在最後一哩線路，讓大部分的演算法都能偵測出流量的變化，接著讓網路壅塞的地點發生在最後一哩線路以外的地方，再來觀察演算法的反應。

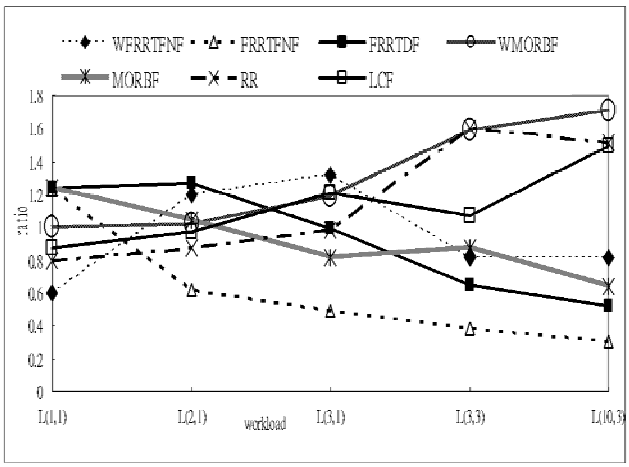
此外，對於使用者產生資料流量的模擬將如同上個小節一般，以 L(U,S)的形式表示，但模擬想定資料流量設定將會從 L（1,1）（即同一時段內一個使用者，一個 session 數）到 L（10,3）（即同一時段內十個使用者，三個 session 數）；研究嘗試收集從最小資料流量到較多資料流量的模擬數據，藉以觀察企業網路不同壅塞狀況下聯外線路被選擇的情形。

為了能夠清晰觀察鏈路負載平衡機制中相關演算法對於不同資料流量的資料傳送量，研究將以總平均資料傳送量（mean throughput）做為相關演算法模擬數據分析比較之基準；因此圖 6 是將總平均資料傳送量比例設定為 1，而總平均資料傳送量則是取所有演算法資料傳送量的平均數，透過此一方式的表達比較能夠明顯顯示出不同演算法在不同資料流量狀況下的資料傳送效能，亦於進行相關的分析與比較。

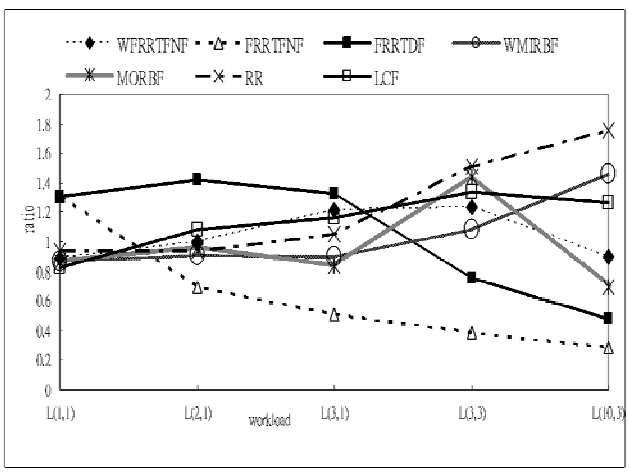
圖 6（a）顯示在最後一哩線路中及不同資料流量產生的狀況下，各個演算法的平均資料傳送量；觀察模擬所獲得的數據，可以發現下列現象：

- 當資料流量增加時，相關演算法的資料傳送量會隨之變化，部份演算法會上升，亦有部份演算法會下降。
- 當資料流量最小為 L（1,1）時，此時 FRRTDF，FRRTFNF 及 MORBF 的平均資料傳送量效能最佳，此三種演算法在資料流量最小（L（1,1））時為最佳之運作演算法；前述之討論，因為此三種演算法會找到量測最佳線路來傳送資料，其中 FRRTDF 及 FRRTFNF 是利用反應時間的方式來找尋最佳線路，而 MORBF 是利用上傳剩餘頻寬最佳模式來找尋最佳線路，所以此三種演算法均能避開塞車的線路。
- 當資料流量最小為 L(1,1)時，其他採用權重模式分配線路的演算法如：WMORBF、WFRRTFNF、RR 與 LCF，它們的平均資料傳送量效能則相對比較差。此一結果顯示：當企業網路資料流量少同時聯外鏈路又發生壅塞現象時，將流量以比較平均地分配在不同聯外鏈路傳送，鏈路負載平衡機制運作所獲得的平均資料傳送量效能相對較差。
- 當資料流量逐漸增加，使用權重模式分配線路的演算法，其所獲得的平均資料傳送量效能亦逐漸提升;因此，在企業網路資料流量為 L（3,1）的時候，WFRRTFNF 及 WMORBF 演算法的平均資料傳送量效能最好，因為此二種演算法具量測線路壅塞的能力，亦使用權重模式的線路分配方式。此外，在企業網路資料流量為 L（10,3）的時候，RR 與 LCF 演算法也呈現較好的平均資料傳送量效能，因為它們使用權重

模式的線路分配方式，雖然它們並不具有量測線路壅塞的能力。



6(a) 近端擁塞



6(b) 遠端壅塞

圖 6 鏈路負載平衡機制演算法對網路擁塞的反應效能

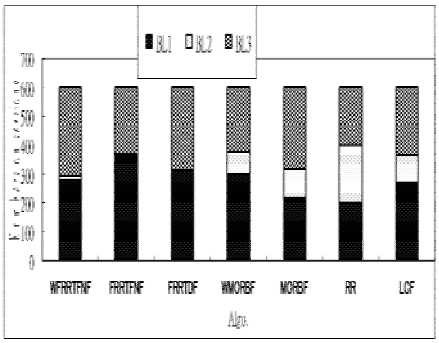
圖 6（b）顯示聯外線路壅塞發生在最後一哩線路之外的地方。使用最後一哩作為量測距離的演算法(如 MORBF,WMORBF)其平均資料傳送量效能與圖 6(a)相互比較，呈現較差的表現；使用最佳模式分配線路方式的 MORBF 在低交通量的時候（如：L(1,1)），其平均資料傳送量效能明顯的下降；此外，使用權重模式分配線路的演算法，如：WMORBF，在交通量增加的時候（L(3,1)），其平均資料傳送量亦未能顯現較佳之效能。

比較 6(b)及 6(a)在交通量最大的時候(L(10,3)),這時候使用權重分配方式的演算法(如 RR,LCF 和 WMORBF)則繼續保持較高的資料傳輸量。

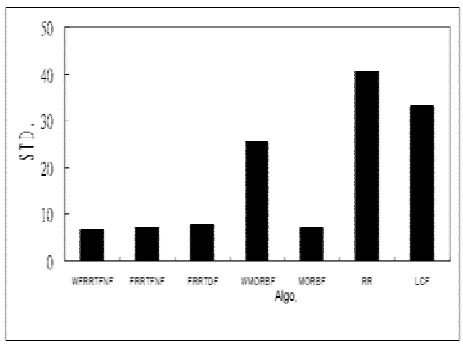
由述上的模擬結果，可以發現下列現象：

- 在交通量小的時候，演算法具備壅塞量測能力是提高資料傳送量的重要因素。
- 在交通量大的時候，演算法使用權重模式線路分配方式則是提高資料傳送量的重要因素；因為在交通量大的時候，藉由使用權重模式線路分配方式可充份利用較壅塞線路的可用頻寬，進而增加所有可供使用線路的頻寬，但這將帶來資料傳輸延遲(delay)的問題。

圖 7(a)顯示在最後一哩線路壅塞的情況下，不同演算法分配不同線路的比例,其中使用權重方式分配的演算法會把部分比例的線路分配到擁塞的第二條線路。在圖 7(b)使用每個 session 傳輸速率來計算絕對平均差,結果顯示使用權重模式分配線路方式的演算法，它們會有比較大的絕對平均差(RR、LCF 與 WMORBF)，這是由於某些 session 完成傳輸所花的時間較長,它們被分配到了擁塞的線路。



7(a)線路分配比例



7(b)絕對平均差

圖 7 分配到不同線路的比例統計圖

在電子商務環境中，不同網路應用有其不同的服務品質需求，因此企業網路鏈路負載平衡機制往往因其網路應用特性之不同而有不同的考量；如果是網路應用是屬於交易服務(如：網路下單)，那麼延遲可能容易造成網路發生問題;但是如果網路應用是提供點對點（peer-to-peer）的下載服務，那麼網路應用的資料傳送量就是比較重要的考量因素。研究亦發現既具有遠端擁塞量測能力又以權重模式分配線路方式的 WFRRTFNF 演算法可以在近端、遠端壅塞及不同的交通量得到較為均平的資料傳送量數值。

三、斷線

這節將觀察企業網路聯外鏈路斷線時，鏈路負載平衡機制中不同演算法的運作效能，研究模擬想定為：從用戶端連續產生 300 個 http 下載連線，並在一段時間之後在第二條聯外鏈路發生斷線的情況，並在一段時間之後，讓聯外鏈路線路斷線狀況能夠被偵測察覺。研究使用較快的判定線路斷線的偵測方式，即在同一個節點連續兩次偵測失敗就判定該條線路為斷線，此一判斷程序約需 6 秒的時間。研究將比較需要額外使用斷線偵測機制的演算法(如：RR 與 MORBF)和本身就可達成線路斷線偵測的演算

法(如：FRRTDF),如表二的分類；此外，亦將觀察使用權重模式分配線路的演算法(如：RR)和使用最佳模式分配線路的演算法(如：FRRTDF)的運作效能。圖 8(a)、8(b)與 8(c)分別為 RR、MORBF 和 FRRTDF 不同演算法在企業網路聯外鏈路發生斷線狀況下所量測每個 session 的資料傳送量，其中 T_f 是網路斷線的時間，T_d 是偵測到網路斷線的時間。

圖 8(a)顯示：在 T_f~T_d 這段時間，許多的 session 的資料傳送量是 0，這代表這些 session 被分配到斷線的線路。由於 RR 演算法分配到不同線路的機率是均等的，因此會有 1/3 的 session 會被分配到斷線的線路。圖 8(b) 顯示：MORBF 演算法在 T_f 時間點的附近的 session 都呈現傳輸失敗的情形，這代表這段時間正在進行傳輸的 session 都使用第二條已斷線的聯外鏈路，接下來有段時間 session 都成功地傳輸，然後接下來又所有 session 全數無法成功傳輸直到 T_d 時間點；由於 MORBF 演算法是使用最佳模式來分配線路，如同前面圖 3 所探討的狀況，以最佳模式分配線路會有一段時間內把所有 session 分配至當時量測流量最好的線路；因此在斷線一段時間之後，由於這條已斷線的鏈路幾乎沒有流量，會被最佳模式認為是最好的線路，所以把所有的連線都分配到此一斷線的線路。圖 8(c) 顯示：FRRTDF 演算法的量測時間是採被動式的方式，是以每個 session 來到的時間做為其量測的時間點；模擬結果顯示在 T_f 時間點的附近的 session 都呈現傳輸失敗的現象，這是因為雖然 FRRTDF 可以為所有 session 找到最佳路徑，但是正在進行傳輸的 session 則因無法再次進行線路的選擇，所以會發生 session 傳輸失敗的現象；接下來幾乎所有的 session 都呈現成功傳輸的現象，並不因有鏈路斷線而受到影響，這代表 FRRTDF 演算法本身即有能力避開斷線鏈路的功能。研究發現大部分的 session 資料傳送量均偏低，這是由於所有 session 的頻寬負擔需求都落在剩餘的兩條線上，因此甚至會有少數連線因傳輸超時(timeout)而連線失敗。但總體而言，FRRTDF 成功率還是最高，在 300 個連線中，RR、MORBF 與 FRRTDF 的失敗連線數分別為 45、87 及 29。

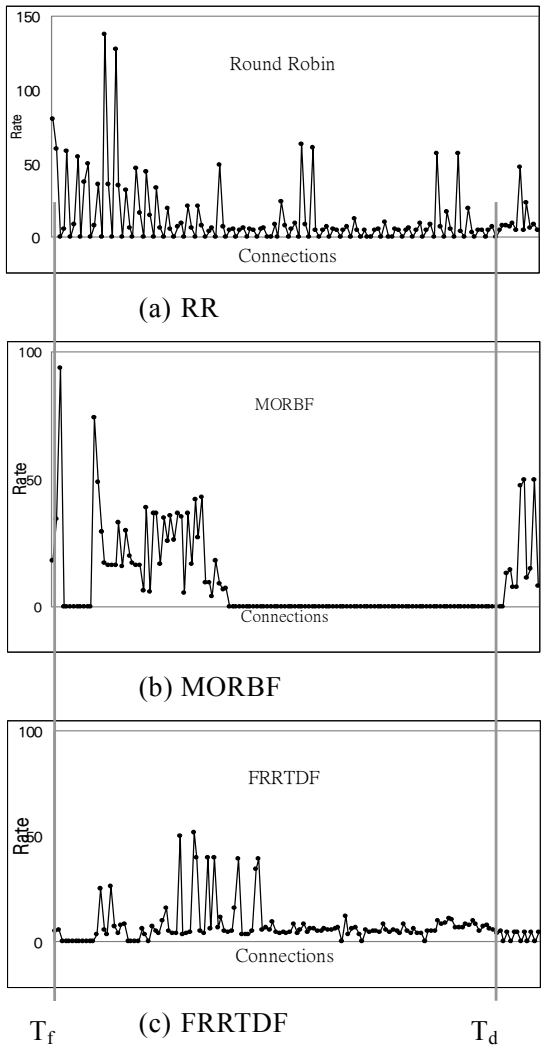


圖 8 不同演算法對於企業網路聯外鏈路斷線影響的交易連線資料傳送變化

連線成功對於電子商務而言固然是相當重要的，但是為了快速避開有問題的線路而使用大量的量測封包進行量測對網路也是一種負擔，如果電子商務的環境允許使用者藉著重新連結(re-connecting)而獲得成功的傳輸，那麼也可以使用對網路負擔較小的線路斷線偵測方法，這些方法反應線路有問題所需的時間是比較長的;使用者可以藉著重新連結的等待時間,讓反應時間較長的斷線偵測方法去除有問題的線路。

伍、結論

企業網路使用 multihoming 可以節費及提升網路的效能，而鏈路負載平衡機制不需要與接續的 ISP 交換大量的路由訊息，也不一定要進行大量的量測工作，是一種能夠適用於一般企業網路的可行方法。

在電子商務的網路環境裡面，交易品質和可靠度是非常重要的，藉由鏈路負載平衡機制之運作可以避開壅塞和斷線的路徑，提升整個交易環境的品質；同時，線路成本亦可透過該機制之運作達到降低的目的。此外，對於企業網路聯外鏈路使用率的提升，亦可藉由適合的鏈路負載平衡機制之運作達到，以增加企業網路聯外鏈路的投資效益。

本篇文章對於負載平衡演算法的分析，除了可以讓應用者瞭解其特性，以實際的需要來選擇相對應的演算方法，並可以做為進深負載平衡演算法的研究基礎來發展更好的方法。

致謝

感謝德恩資訊(Deansoft)在研究期間提供研究所需負載平衡測試設備。

參考資料

1. Akami， http://www.akamai.com/en/html/technology/overview.html,2005
2. A.Akella, B. Maggs, S. Seshan, A. Shaikh, and R.Sitaraman. “A measurement-based analysis of multihoming”, In Proc. of ACM SIGCOMM, August 2003.
3. A.Akella, S.Seshan ,A.Shaikh, ”Multihoming Performance Benefits: An Experimental Evaluation of Practical Enterprise Strategy”,USENIX 2004
4. C. Labovitz, A. Ahuja, A. Bose, and F. Jahanian, “Delayed Internet routing convergence”, In Proc. of ACM SIGCOMM '00, Stockholm, Sweden, pp. 175--187, 2000
5. D. Goldenberg, L.Qiu, H. Xie, Y.R.Yang and Y. Zhang, “Optimizing Cost and Performance for Multihoming”, in Proc. of the 2004 ACM SIGCOMM Conference, August 2004
6. Deansoft，http://www.deansoft.com.tw/Ehome.htm,2005
7. D. G. Andersen, H. Balakrishnan, M. F. Kaashoek, and R. Morris, “Resilient Overlay Networks”, in 18th ACM Symposium on Operating Systems Principles (SOSP), October 2001.
8. F5，http://www.f5.com,2005
9. F. Guo, J. Chen, W. Li, and T. Chiueh, “Experiences in Building a Multihoming Load Balancing System”, In Proc. of. IEEE INFOCOM, March 2004
10. K.Egevang and P. Francis, “The IP Network Address Translator(NAT)”, RFC 1631, May 1994
11. P. Srisuresh , D. Gan “Load Sharing using IP Network Address Translation”, RFC 2391, August 1998
12. Radware， http://www.radware.com
13. T. Bates, Y. Rekhter, “Scalable Support for Multi-homed Multi-provider Connectivity”, RFC2260, January 1998
14. W.Li, “Inter-domain Routing: Problems and Solutions, technical report”, State University of New York,Feb 2003.
15. Y. Rekhter, T. Li, “A Border Gateway Protocol 4 (BGP-4)” 1995, RFC1771, March 1995
16. Y. Rekhter, T. Li, “An Architecture for IP Address Allocation with CIDR”, RFC1518, September 1993