行政院及所屬各機關出國報告

(出國類別：實習)

赴Telcordia實習下一代網際網路IP/DWDM網路控制與管理系統出國報告

服務機關：中華電信研究所
出國人 職　稱：助理研究員
　　姓　名：侯行方 林冠平 林尚亭
　　徐浩然 巫啟生
出國地區：美國
出國期間：92年11月16日至93年5月16日
報告日期：93年6月15日

系統識別號:C09205063

公 務 出 國 報 告 提 要

頁數: 34　含附件: 否

報告名稱:

實習下一代網際網路IP/DWDM網路控制與管理系統

主辦機關:

中華電信研究所

聯絡人／電話:

楊學文／03-4244218

出國人員:

侯行方　中華電信研究所　928B0專案研究計畫　助理研究員
巫啓生　中華電信研究所　928B0專案研究計畫　助理研究員
林尙亭　中華電信研究所　寬頻網路技術研究室　助理研究員
林冠平　中華電信研究所　無線通信技術研究室　助理研究員
徐浩然　中華電信研究所　寬頻網路技術研究室　助理研究員

出國類別: 實習

出國地區: 美國

出國期間: 民國 92 年 11 月 16 日 -民國 93 年 05 月 16 日

報告日期: 民國 93 年 06 月 15 日

分類號/目: H6／電信　／

關鍵詞: 下一代,網際網路,IP,DWDM,網路控制,管理系統

內容摘要: 隨著網際網路蓬勃發展，使用人數快速增加，使得網際網路面臨了頻寬不足的問題，網路塞車的情況愈來愈嚴重，使用者必須花更多的時間等待，以取得所需要的資料。這種情況若不加以改善，網路的品質勢必日趨惡化，終至無法忍受的地步，是以世界各國，莫不投入心力，以尋求解決之道。爲了研究如何解決網路塞車的情況，Telcordia Technologies與美國軍方研究機構DARPA (Defense Advanced Research Projects Agency)進行一項合作計畫，其目的在於一但發生戰爭，爲能迅速於戰區中佈署光纖網路，並且可以做快速的管理和控制網路以避免網路壅塞，而達成網路資源能充分運用且避免造成資源的浪費。因應此需求，Telcordia Technologies開發一套SuperNet NC&M System，讓各個不同domain之間的網路透過SuperNet NC&M System，可以快速調配全美國分屬於不同domain之光網路，並且加以管理與控制。本公司爲達成IP網路作更有效率的運用之目的，就Telcordia原有的SuperNet NC&M System作修改後，與Telcordia合作共同開發適合中華電信的「下一代網際網路IP/DWDM網路控制與管理系統」以解決網路塞車的情況。透過本案之執行將可了解美國主要公司廠商對於下一代網際網路IP/DWDM網路控制與管理系統，全光網路，IP over DWDM等技術之最新發展之應用狀況，作爲本公司日後擬定reconfigurable DWDM系統設備之參考。而有關IP topology design，traffic engineering，minimum packet loss migration等技術亦可應用於本公司日益增多之IP網路架構及其效能規劃。此外，Telcordia將交付下一代網際網路IP/DWDM網路控制與管理系統之軟體原始碼，將有助於本公司日後發展相關網管系統。本案原先規劃重點爲IP over re-configurable DWDM，此爲未來發展之趨勢，但經實際考量本公司目前DWDM網路以點對點組態爲主，因此，本案增加相關技術應用於ATM網路，如此一來，不僅可運用於本公司現存之ATM網路，亦可兼顧未來發展主流趨勢之IP/DWDM，使整個合作案更具實用性及前瞻性。

本文電子檔已上傳至出國報告資訊網

摘 要

隨著網際網路蓬勃發展，使用人數快速增加，使得網際網路面臨了頻寬不足的問題，網路塞車的情況愈來愈嚴重，使用者必須花更多的時間等待，以取得所需要的資料。這種情況若不加以改善，網路的品質勢必日趨惡化，終至無法忍受的地步，是以世界各國，莫不投入心力，以尋求解決之道。

為了研究如何解決網路塞車的情況，Telcordia Technologies 與美國軍方研究機構 DARPA (Defense Advanced Research Projects Agency) 進行一項合作計畫，其目的在於一但發生戰爭，為能迅速於戰區中佈署光纖網路，並且可以做快速的管理和控制網路以避免網路壅塞，而達成網路資源能充分運用且避免造成資源的浪費。因應此需求，Telcordia Technologies 開發一套 SuperNet NC&M System，讓各個不同 domain 之間的網路透過 SuperNet NC&M System，可以快速調配全美國分屬於不同 domain 之光網路，並且加以管理與控制。本公司為達成 IP 網路作更有效率的運用之目的，就 Telcordia 原有的 SuperNet NC&M System 作修改後，與 Telcordia 合作共同開發適合中華電信的「下一代網際網路 IP/DWDM 網路控制與管理系統」以解決網路塞車的情況。

透過本案之執行將可了解美國主要公司廠商對於下一代網際網路 IP/DWDM 網路控制與管理系統，全光網路，IP over DWDM 等技術之最新發展之應用狀況，作為本公司日後擬定 reconfigurable DWDM 系統設備之參考。而有關 IP topology design，traffic engineering，minimum packet loss migration 等技術亦可應用於本公司日益增多之 IP 網路架構

及其效能規劃。此外，Telcordia 將交付下一代網際網路 IP/DWDM 網路控制與管理系統之軟體原始碼，將有助於本公司日後發展相關網管系統。

本案原先規劃重點為 IP over re-configurable DWDM，此為未來發展之趨勢，但經實際考量本公司目前 DWDM 網路以點對點組態為主，因此，本案增加相關技術應用於 ATM 網路，如此一來，不僅可運用於本公司現存之 ATM 網路，亦可兼顧未來發展主流趨勢之 IP/DWDM，使整個合作案更具實用性及前瞻性。

目　錄

圖　例

1. 目的

當網際網路的流量呈現幾何倍數的成長，網際網路的基礎架構已演進為以高速路由器相連的可設定傳輸網路。這個網路基礎架構使得電信公司需要有隨時或接近於隨時來提供網路，保護連結或網路復原、流量管理、及支擾 VPN 的功能。

因為以 IP 為基礎的控制協定已經相當成熟，再加上營運經驗，及現有 IP 網路之相通性，因此產業界一般都喜歡以 IP 為基礎的控制協定來控制光纖傳輸網路。然而一般的可設定網路通常是以"線路交換"為主，而線路交換不是為 IP 協定所設計。NC&M 系統所面臨的挑戰主要來自如何利用傳輸網路的設定功能來提供可變網路的功能來滿足客戶的需求。為了讓 IP 的設定網路能夠成為商用化，控制及管理的一些議題必須首先被解決。

本案的主要重點在 IP/WDM 網路在靜態與動態上之設定與流量管理。靜態之管理主要在網路規劃及提供，而動態的管理主要在改變網路流量要求時網路的改變。

因著參與美國政府支援的尖端研發計劃，例如：NONET、NGI 及 SuperNet，Telcordia 發展出許多 IP 在光纖可設定網路的網路控制及管理系統(NC&M)。該計劃主要針對一些重要的控制及管理議題，特別著重在 IP 在可設定傳輸網路的流量管理，Telcordia 並建立了下列的技術專家：

- IP 可設定傳輸網路的服務及網路架構定義。
- 在 NC&M 系統中不同 IP 可設定傳輸網路的重要議題。
- 在一個可設定傳輸網路中可提供之優點及模擬分析。

- 一個以 MPLS 為基礎的 IP 路由及 MPλS /WDM 架構以及其他不同模式來滿足營運需求的創新網路流量模式架構。
- 在 IP 網路組合改變中（因為光網路的重新設立），防止 IP 封包流失的專利技術。
- IP/WDM 網路的網路流量軟體模型，而這個軟體模型也能在 IP/ATM 的網路架構中。

雖然該計劃是著重 IP 在 MPλS 之上的可設立 WDM 網路，但是其結果可以被應用在 IP 在任何可設定的網路架構中。

綜觀 NGI NC&M 系統，其主要功能可為以下兩點：(一)可依據網路資料流量來設計較佳化之 IP 拓撲，以提升網路效能；(二)計算如何改變 IP 拓撲流程對整體網路影響最小，並降低因拓撲調整而產生之衝擊。其主要應用可從幾方面探討：

- 從網路資料流量來設計較佳化的 IP 拓撲：如此將使現有設備效能提升，延後或減少設備採購降低營運成本。
- 可根據計算出來之 IP 拓撲改變流程來重整網路：利用人工或自動化的方法，提供對網路做調整和調度一套較系統化的方法。
- 利用自動化的方法來改變 IP 拓撲：可加速網路的調整能力，使得隨選頻寬的服較易實現。

2. 過程

92 年 11 月 16 日：台北 → 美國紐澤西

92 年 11 月 17 日—93 年 5 月 14 日：Telcordia Technologies 研習

93 年 5 月 14 日—5 月 16 日：美國紐澤西 → 台北

於 92/11/16 深夜抵達美國東岸紐澤西州，打點生活上之必需品，包括住宿、交通、飲食等事務，並藉此熟悉附近環境，隨即投入此次為期長達半年的實習。經與計劃主持人蔡獻琨博士討論，研擬出此次實習內容的大綱，其主題與時程如下表所示：

編號	主題	完成日期
1	可設定網路之網路控制及管理架構的調查	2004/05/03
2	特定網路控制與管理架構的網管議題	2004/05/14
3	IP/WDM 網路流量工程的架構	2004/04/01
4	IP 網路形狀設計演算法	2004/04/08
5	最小影響之 IP 網路演進方式	2004/04/15
6	建立 IP/ATM 的相關應用軟體模型模(prototype)	2004/05/03

2.1. 「下一代網際網路控制與管理系統」計劃概述

「下一代網際網路控制與管理系統」是針對 IP 層執行 topology reconfiguration，它與下層 layer 2 or layer 1 是獨立的，只要在此網管介面中，訂義與下層的介面格式及參數，即可透過此介面向下層傳遞 reconfigure 後的 topology，讓整個 IP 網路重新佈署(Network Layers Consolidation)。

本計畫共有六份交付項目（Deliverable），各交付項目簡介如下：

- 計畫交付項目 1：可設定網路之網路控制及管理架構的調查。

  隨著網際網路的蓬勃發展，使用人數的快速增加，使得網際網路面臨了頻寬不足的問題，網路塞車的情況愈來愈嚴重，使用者必須花更多的時間等待，以取得所需要的資料。這情況若不加以改善，網路的品質勢必日趨惡化，終至無法忍受的地步。因此網際網路的基礎架構已演進為以高速路由器相連的可設定傳輸網路。目前有多種際網路控制系統架構，例如 the Optical Internetworking Forum User Network Interface (Optical Internetworking Forum 網路使用介面)， Resilient Packet Rings (抵擋性的封包環) 及乙太網路在SONET技術上包括使用 Link Capacity 調整法以及一般性框(Frame)的協定。這些架構能夠支援不同的網路技術（例如：WDM、SONET、ATM），不同的服務能力並需要不同的網路管理支援。

- 計畫交付項目 2：特定網路控制與管理架構的網管議題。

  在網路控制與管理架構的調查之後，我們預計會選擇兩個網路架構來進一步分析。這個交付項目將對特定的網路控制及管理架構提供更多有關網管上的可能影響。在這個交付項目中所要探討的題目包括：管理介面、不同網路之相通性、網路設定中網管系統所扮演的角色、錯誤與表現管理。

- 計畫交付項目 3：IP/WDM 網路流量工程的架構。

討論 Telcordia 以 MPLS 的 IP 路由以及 MPλS WDM 網路架構的網路流量架構。傳統的 IP 網路並不支援網路流量工程，這是因為傳統的 IP 網路是找最短路徑的緣故。然而在 IP 可設定光纖傳輸網路中網路流量工程可以採用下面 2 個方式來解決。在 IP 層應用 MPLS 及在光纖層中來做重新設定（reconfiguration）。每一個方式都有其優點與缺點。一個創新的網路流量架構，著重在二個不同層中的合作，如此彼此在網路流量工程中相互支援，而產生更好的效果。

- 計畫交付項目 4：IP 網路形狀設計矩陣。

這個交付項目包括 2 套網路形狀設計的矩陣，這些矩陣是在 IP/WDM 網路流量工程系統中所發展出來。但是這些矩陣也適用於 IP 的網路規劃。一個矩陣是將網路流量的需求在網路點及連結的限制下，設計出 IP 網路的形狀。另一個矩陣則上做微幅式的修正來動態的改進流量的路徑。

- 計畫交付項目 5：最小影響之 IP 網路演進方式。

這個交付項目著重在沒有封包遺失的 IP 技術演進，而這個技術已經在申請專利中。在網路的形狀改變後現有的 IP 路由協定最後會使的網路的路由呈現一個穩定的狀態。但是在這樣一個過程中，使用者的封包可能因為過渡中的黑洞及前置迴路而有遺失的情況。使用在這個交付項目的技術，在光纖傳輸層做重新設定時所產生的 IP 虛擬形狀的改變，可以使的現有的網路流量受到最小的影響。這個技術也可以讓 IP 網路管理者來進行

對使用者的網路維護。

- 計畫交付項目 6：建立 IP/ATM 的相關應用軟體模型模 (prototype)。

  Telcordia 會提供 IP/WDM 流量工程軟體模型的原始碼。中華電信選擇 ATM 網路元件與網管系統與 Telcordia 提供之工程軟體 prototype 相連接。Telcordia 與中華電信會一起合作來修正軟體模型以便使用在 IP/ATM 的網路中。

3. 實習內容紀要

3.1. DARPA NGI 計畫概觀

Telcordia Technologies 與軍方研究機構 DARPA (Defense Advanced Research Projects Agency 合作的一個計畫。目的為於戰區中快速佈署光纖網路，並且可以做快速的管理和控制這個網路；所以 Telcordia Technologies 就發展了 SuperNet NC&M System，讓各個不同 domain 之間的網路透過 SuperNet NC&M System，尤其在戰爭時，可以快速調配整個全美的分屬於不同 domain 的光網路，並且加以管理與控制。但是平時，此 domain 內的網路是自我管理與控制的。

Telcordia Technologies 是針對 IP 層做 topology reconfiguration，它與下層 layer2 or layer1 是獨立的，只要在此網管介面中，訂義與下層的介面格式及參數，即可透過此介面向下層傳遞 reconfigure 後的 topology，讓整個 IP 網路做重新的佈署 (Network Layers Consolidation)。DARPA NGI 系統架構如圖 1 所示。

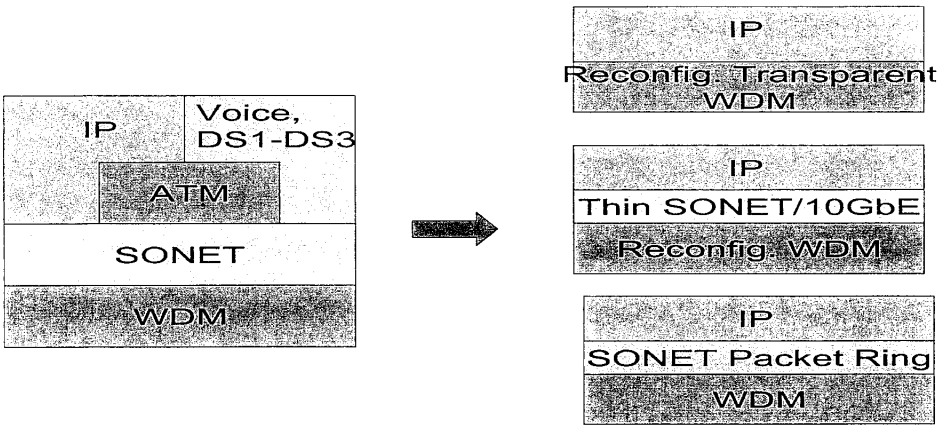

圖 1　DARPA NGI計畫之系統架構

3.2. Reconfigurable Control & Management 分析

由於 IP layer connectionless 的特性，使得 IP Network 常常造成壅塞，並且沒辦法達到 QOS。因此，必須 reconfig IP layer，並且分析 Reconfiguration 後會帶來的結果為何。考量之因素包括：

- 可否增加網路的可傳送流量(throughput)？
- 可否減少網路間傳送延遲？
- 可否針對 IP 層 fail 時，做有效的保護切換機制？
- 可否提供快速且準時的網路供裝？

分析 DARPA NGI 在提高網路 throughput 之表現，可得如圖 2所示之結果，顯示運用 DARPA NGI 系統於高負載或 skew 之訊務網路將可提高 40%至 60%之 throughput。

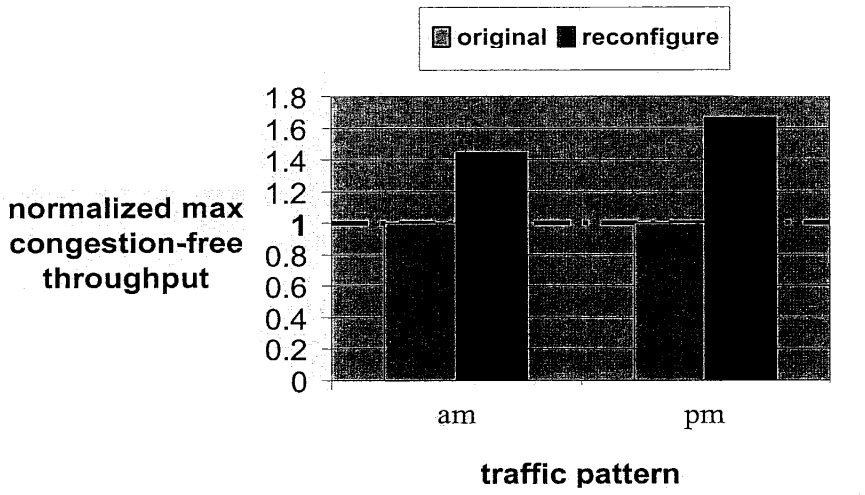

圖 2　DARPA NGI系統網路throughput分析

至於改善網路間傳送延遲方面，於高負載或 skew 之網路，經過 reconfiguration 後之網路將可改善 90%之端點至端點傳送延遲，其表現如圖 3 所示。

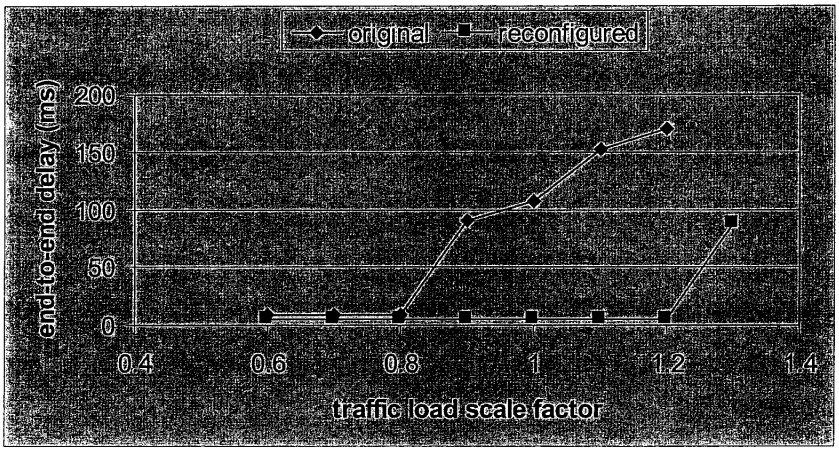

圖 3　DARPA NGI 於改善網路傳送延遲表現

當網路之 IP layer 發生 fail 時，DARPA NGI 系統之保護機制透過 reconfigure 之方式提高網路之 throughput 及降低端點至端點之傳送延遲。以圖 4 為例，當位於達拉斯（DL）之 router 發生 failure 時，NGI 系統將會進行 IP layer reconfiguration，其前後效能表現如圖 5及圖 6 所示，reconfigure 後之網路 throughput 將提高 30%，而且如果網路負載愈大，改善之情形愈明顯；而 reconfigure 之後原本發生 failure 之節點將不會再經過，因此將可有效地改善延遲之問題，且改善之幅度將可高達 80%以上。

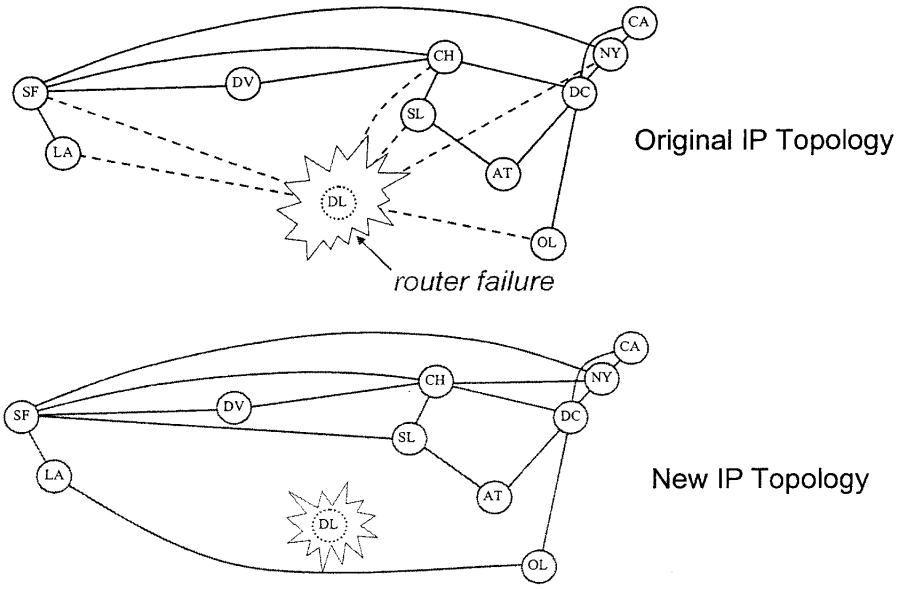

圖 4　網路發生hub failure後將進行IP layer reconfiguration

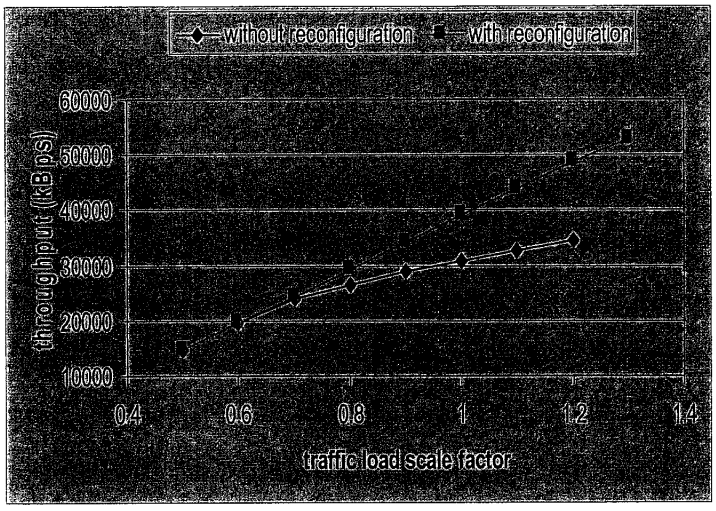

圖 5　hub failure後進行網路reconfiguration前後之throughput變化情形

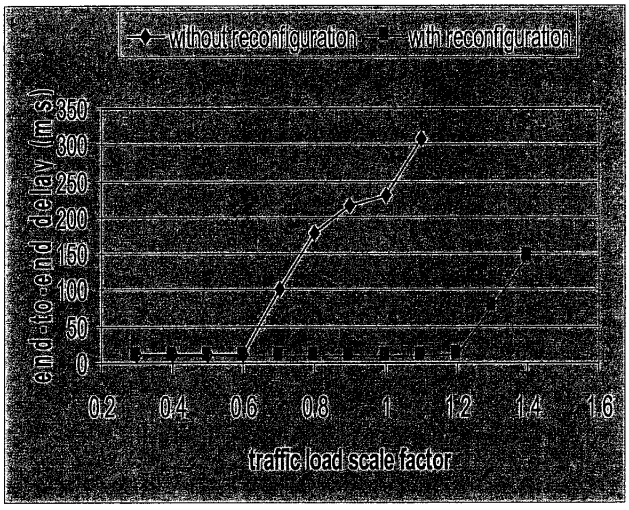

圖 6　hub failure後進行網路reconfiguration前後之end-to-end delay變化情形

3.3. IP/WDM 的網路架構下的訊務工程

本節探討重點在於 IP/WDM 的網路架構下的訊務工程(Traffic Engineering)。內容將會著重在 Telcordia Technologies 在 TE 的研究過程中，如何從控制面(Control Plane)中的 SIGNALING 去分析；並從 IP MPLS routing.signaling(Layer 3)及以 MPλS configurable WDM (Layer1)的架構下，來討論訊務工程的相關議題及研究。

在傳統的 IP 網路中，如果從控制面的角度來看，在 IP routing protocol 有幾種選擇：OSPF(Open Shortest Path First)、RIP(Routing Information Protocol)和 IS -IS(Intermediate System to Intermediate System)；但是這些 routing protocol 的基本意義都是 traffic-independent，但是事實上每個獨立的訊務都是互相影響的，所以基本傳統的 Routing Protocol 並沒辦法滿足訊務工程的條件。

然而 Telcordia Technologies 在 IP over configurable optical transport networks 的基礎架構下，透過兩種手段來達成訊務工程的目的：

- MPLS 應用在 IP 層
- 重新設定 PATH 應用在光網路傳輸層

基本上這兩層的訊務工程會是獨立的，而且各自有各自的優點及限制；而 Telcordia Technologies 的目的就是要追求除了分別在這兩個網路層中，各自達成 TE 的功能，並且在這兩層中的交互合作，探討以最佳的方法，來達成並建立一個全新附有完整訊務工程的平臺。

3.3.1. 背景

這整個概念的源由來自 Telcordia Technologies 在針對 DARPA (Defense Advanced Research Projects Agency)發展一個下一世代網際網路網路控制與管理系統，為了須要達到一些目的：

- 平時的網路是隸屬於各個民間的高速光纖網路
- 戰爭時，通信網路必須快速的結合在一起，並且是以整個美國為領域的大範圍結合，而不是地區性的網路結合
- 此網路控制與管理系統必須能在最短的時間內，控制管理整個分屬於不同網域的民間高速光纖網路

圖 7 的 SuperNet NC&M System 就是在戰爭發生時，能夠訊速控制及調度整個來自不同民間公司的網路的網管系統。

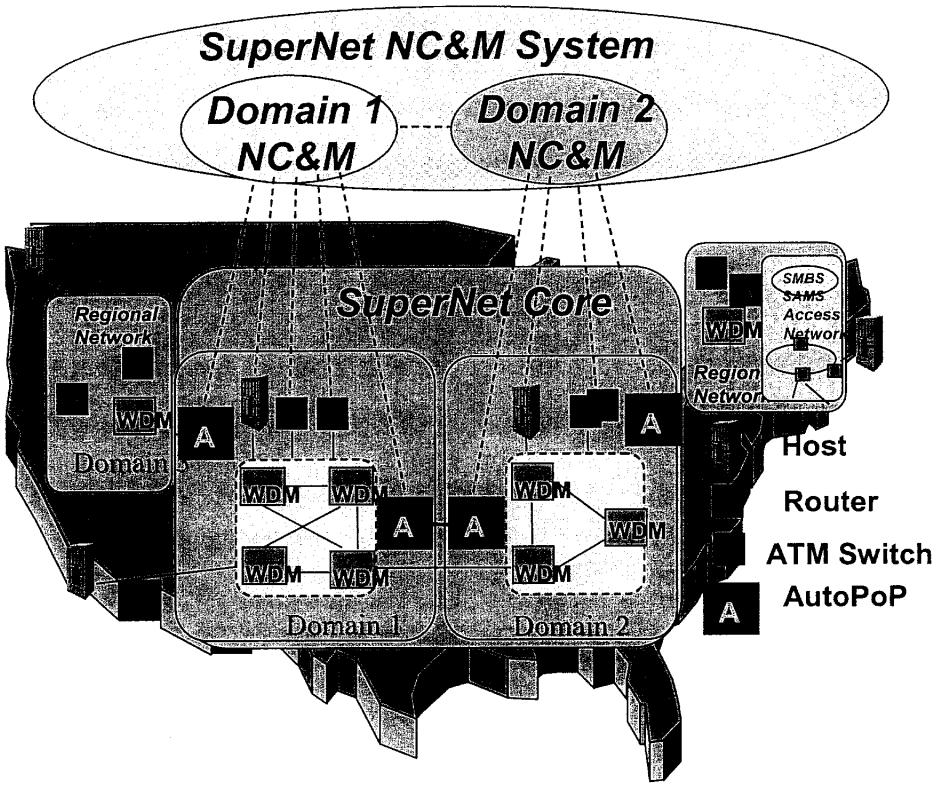

圖 7　SuperNet NC&M System

Telcordia Technologies 在發展的過程中，也找了相關隸屬於國家各個不同單位的網路，並在 Washington DC 設置一個測試網路平臺；這個平臺是以 ATM/SONET/MONET(Multi-wavelength Networking)為架構，此測試網路平臺的特性如下：

- 達到 ANSI HIPPI(High Performance Parallel Interface)的標準，也就是透過連結 super computer，能快速提供高速的核心網路

- 光傳輸網路是透過 OXC(光交接機)來調整底層 PATH 的改變
- 在接取端，可接收雷達所傳送的訊號

3.3.2. IP/WDM Reconfiguration 的效益

在前幾個章節中已經有針對"Network Throughput"、"Latency"等分析，透過 IP 層的 RECONFIG 而達到第二層或是第一層 LINK 的改變，會帶來若干的好處，我們這次將針對如果整個 topology 由於網路流量的變動而做改變時，對於成本的分析。

Reconfigure 後的 WDM 層，我們可以從兩個因素來分析成本：

(1) 減少與 WDM 介接的介面卡數量：藉著減少路由所經過的 HOP 數目，可重新設定 PATH 的 WDM 可以減少介面卡數量的需求。

- H =平均加權 HOP 距離
- C = 在一部 ROUTER 中，平均接取的流量
- W = 波長頻寬
- **一個 router 中，介面卡的平均數量 =** $\frac{H \quad C}{W}$ **，效益 =**

$$\frac{H^f - H^r}{H^f}$$

對於 AT&T 來說，在 UNIFORM model 下的網路流量效益大約是 44% (1.8-1/1.8)。

(2) 增加 OXC(Optical Cross Connect 光交接機)數量：針對

8.16.32.64 等不同波長的 WDM 來說，OXC 增加的數量是非常少的。

另一方面，對於下一世代網際網路 IP/DWDM 網路控制與管理系統的管理成本來分析，可以增強下列的優勢：

- 可以簡化網路供裝的程序及對於網路流量的規劃更能掌控。
- 提高採用新世代網路技術的彈性。
- 減少約 40%的維運成本。
- 減少約 30%的網路擴充成本。
- 對於網路很壅塞，或是對於變動較大的網路流量時，可以改善網路速率增加 40%-60%並可減少 90%點對點的延遲。
- 在 IP 層，能做有效率的保護機制，如果重要的 hub 遇到重大的停擺，經過重新的 IP layer reconfigure path，可以使網路的傳送效能改善 30%，並且降低傳送的延遲達 80%：
  (a) 經過流量的分析，而有週期性的規劃時，可以提供快速的網路供裝，及避免網路壅塞。
  (b) 對於已鎖定的負載期間，能做有效的減輕負載(例如減少 TCP 的傳輸延遲。
  (c) 減少 router 介面卡的數量約 40%。
  (d) 網路的 link 可以保持平衡負載的狀況。

3.4. IP Topology Design Algorithm

3.4.1. 前言

Telcordia 的 IP topology design algorithms 在本案中扮演重要且關鍵的角色。這些演算法主要目的便是要建立起 IP 網路 topology reconfiguration 的藍圖。換句話說，一旦決定 IP 網路 topology 需要被調整（reconfigured），那麼我們心中可能會浮現一堆問號：topology 該調整成什麼模樣？換另一種說法：reconfigured 後的 topology 該滿足什麼條件？該建立什麼樣的數學模型以具體化描述上述問題？數學模型建立後，有沒有一種系統化的方法來進行運算與分析？

本文結構大致如下：首先定義兩個 cost function，前者的主要目標是極大化 throughput；後者的主要目標則是極小化 weighted hop-counts。我們接著談到兩類主要的操作模式："Desert Start" 和"Incremental Design"。基於前面兩個 cost function，Telcordia 提出 3 種 heuristic 的演算法：RDHP、DHP 及 RD。RDHP 與 DHP 的主要目的是要極小化 weighted hop-counts；而 RD 的主要目的是要極大化 throughput。接下來會以 AT&T IP Backbone 網路拓撲為例，模擬該網路經此 3 種 heuristic 演算法所計算出的 reconfigured IP 網路 topology。並且進一步地將不同的 traffic pattern 置入（load）這三種 topology，以圖表的方式，就 overall throughput 與 weighted-hop count 分別分析比較在不同 traffic pattern 及不同 IP topology 時的效能表現。最後提出相關的觀察。

3.4.2. IP Topology Design

在開始進一步說明細節之前，我們必須了解 IP topology design algorithms 在整個系統流程中的定位及所扮演的角色：一旦系統根據 traffic demand 判定原先的 topology 需被 reconfigured，那麼該把 topology 重新安排成什麼樣子，才能夠使整體網路的效能表現有所提升。

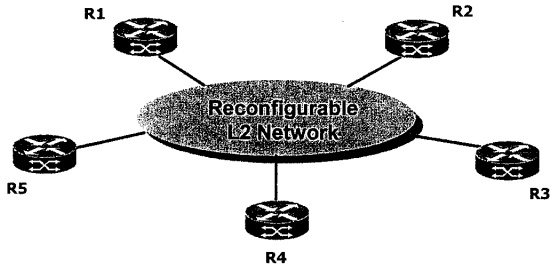

圖 8　IP Topology Design 示意圖

上圖的網路架構表示：IP router 在四周，透過核心 reconfigurable 的底層網路彼此相連。現在將問題以比較數學化的語言描述如下：

N：router 的集合

L：將 router 彼此相連的 link（IP trunk）的集合

D：表示由 router 彼此之間的 traffic demand 所構成的矩陣（matrix），matrix 中的 entry $D(i,j)$ 表示由 i 點到 j 點的 demand。

我們的目標如下：

(1) 極大化 throughput，意即：*Max* $\sum_{\forall(i,j)\in\Omega} X(i,j)$，或是

(2) 極小化 weighted hop-counts，意即 *Min* $\frac{\sum_{\forall(i,j)\in\Omega}(H(i,j)\times X(i,j))}{\sum_{\forall(i,j)\in\Omega} D(i,j)}$

其中，$X(i,j)$ 為由 i 點到 j 點的 throughput；$H(i,j)$ 為由 i 點到 j 點的 hop count。Ω 是節點對(node pair)的集合，即 $\Omega = \{(i,j)|i \in N, and, j \in N\}$。

當然在極大化 throughout 與極小化 weighted hop-counts 的過程中，任何做法必須受限於每個 router 可用 interface 的數量。

3.4.3. Mode of Operation

基本上，本次 deliverable 中 IP topology design 的操作模式可概分為兩大類。第一類稱為 Desert Start；第二類稱為 Incremental Design。

所謂 Desert Start，其做法主要的精神就是：網路中所有節點在最初（initial）狀態，彼此之間是沒有任何 link 相連的，一切從頭開始，亦即："start from 0 links"。因此所有 available link 皆納入 reconfiguration 機制考量的範圍內。所謂的 Incremental Design 指的是：某些特定的 embedded links 需要被保留；換句話說，只有非 embedded links 才可以被變更。這種做法特別適用於已有服務上線使用（in service）的環境，能夠相當有彈性地、動態地調整網路 topology。

3.4.4. Heuristics

Telcordia 針對前面所述的兩個 cost function，提出了三種 heuristic

的演算法。分別是：

- Residual Demand Hop-count Product Heuristic Algorithm (RDHP)
- Demand Hop-count Product Heuristic Algorithm (DHP)
- Residual Demand Heuristic Algorithm(RD)

3.4.4.1.1. Residual Demand Hop-count Product Heuristic Algorithm (RDHP)

RDHP 的目標是要極小化 weighted hop-counts，其做法的基本精神就是要把占權值最高的各項（dominant terms），將其 hop distance 降至最低。其步驟如下：（當然以下各步驟皆要滿足 nodal degree 的限制。所謂的 nodal degree 即包含 embedded link 在內的所有可用的 interface）

(1) 將 Demand matrix ***D*** 對稱化（symmetrization），得到一個新的 matrix，稱為 *sdemand_matrix*，亦即：

*sdemand_matrix[i][j] = max (**D**[i][j], **D**[j][i])*

(2) 將對稱化後的 demand matrix 的 entry sdemand_matrix[i][j] 以由大到小的方式（descending order）排序，形成一個 flow vector F，在程式中，我們稱之為"dflow"。再以此 flow vector F 為基礎，建構一個最小的 spanning tree，其目的在於提供每一個節點的 initial connectivity。

(3) 將 traffic demand $D(i,j)$ 沿著 i 點到 j 點的 shortest path tree

置入現階段 incomplete topology 中。所謂"現階段"指的是在這個 iteration 的意思。在這一個步驟中，主要是要決定每一條 link 的 loading factor，稱為：truncation_factor。truncation_factor 等於 link capacity 除以所有經過（traverse）該條 link 的 traffic demand 的總合。

(4) 定義並計算 local_t_factor。local_t_factor 是沿著 i 點到 j 點所經過的全部 link，其 truncation_factor 的乘積。

(5) 定義並計算 residual demand matrix，rdemand_matrix：

rdemand_matrix[i][j]=demand_matrix[i][j]*(1-local_t_factor)

表示因為 link 的 overloading 而被 discard 的 demand traffic。

(6) 以此 incomplete topology 為基礎，用 Dijstra Algorithm 計算所有節點對（node pairs）之間的 hop distance。

(7) 計算 rdhp flow vector = symmetrized residual demand flow * hop count，並將 rdhp flow vector 的 entry 加以排序。

(8) 以 *rdhp flow vector* 為基礎，增加一條新的 link，以改進網路 topology。然後回到步驟(3)，直到沒有空的 interface。

3.4.4.2. Demand Hop-count Product Heuristic Algorithm (DHP)

DHP 的目標是要極小化 weighted hop-counts。其步驟除了 *dhp flow vector* 的計算之外，其餘皆與 RDHP 相仿，簡單說明如下：（當然以下各步驟皆要滿足 nodal degree 的限制）

(1) 將 demand matrix 對稱化（symmetrization），按照對稱化後的 demand matrix 的 entry $D(i,j)$ 大小排序，形成一個 flow vector F。再以此 flow vector F 為基礎，建構一個最小的 spanning tree，提供每一個節點的 initial connectivity。

(2) 將 traffic demand 置入現階段 incomplete topology 中，得出 residual demand matrix，rdemand_matrix。若 rdemand_matrix 大於 0，則繼續做步驟(3)。

(3) 以此 incomplete topology 為基礎，用 Dijstra Algorithm 計算所有節點對（node pairs）之間的 hop distance。

(4) 計算 dhp flow vector = symmetrized demand flow * hop count

(5) 以 dhp *flow vector* 為基礎，增加一條新的 link，以改進網路 topology。然後回到步驟(2)，直到沒有空的 interface。

3.4.4.3. Residual Demand Heuristic Algorithm (RD)

RD 的目標是要極大化 throughput。其步驟與 RDHP 相仿，但不考慮 hop distance。簡單說明如下：（當然以下各步驟皆要滿足 nodal degree 的限制）

(1) 將 Demand matrix 對稱化（symmetrization），按照對稱化後的 demand matrix 的 entry $D(i,j)$ 大小排序，形成一個 flow vector F。再以此 flow vector F 為基礎，建構一個最小的 spanning tree，提供每一個節點的 initial connectivity。

(2) 將 traffic demand 置入現階段 incomplete topology 中，計算 residual demand matrix。

(3) 利用 residual demand matrix 建構一個新的 rd flow vector。

(4) 以 *rd flow vector* 為基礎，增加一條新的 link，以改進網路 topology。然後回到步驟(2)，直到沒有空的 interface。

3.4.5. Performance Comparisons

我們在這節要談的是針對前述之 3 種 heuristics，進行 performance 的比較與分析。做法是首先將各種不同 pattern 的 random traffic 分別加入由這 3 種 heuristics 就 fixed network，如：AT&T IP backbone，所決定出來的網路 topology 中。接著計算以下兩個指標參數：

(1) Normalized Network Throughput

$$\frac{\sum_{\forall(i,j)\in\Omega} X(i,j)}{\sum_{\forall(i,j)\in\Omega} D(i,j)}$$

$D(i,j)$ 表示由 i 點到 j 點的 demand

(2) Weighted Hop-count

$$\frac{\sum_{\forall(i,j)\in\Omega} (H(i,j)\times X(i,j))}{\sum_{\forall(i,j)\in\Omega} D(i,j)}$$

我們的觀察如下：

(1) Throughout

a. 在 uniform traffic 的情形下，RDHP topology 的表現優於 DHP topology。當 traffic loading 低時，RD topology 有最佳的表現。當 traffic loading 增加時，RD topology 的表現會有惡化的現象。一般而言，traffic loading 重時，hop-count 數多的 demand 可能會消耗大量的 resource，這會大幅降低網路整體 throughput 的表現。這邊有一點需要注意：loading 會被所有的 demand 分攤，與哪一個 demand 先被 load 進來無關。

b. 在 lighted-skewed, skewed, 和 very skewed traffic 的情形下，RDHP topology 的表現優於 RD topology 和 DHP topology。當 traffic loading 增加時，DHP topology 的表現優於 RD。換言之，當 traffic loading 比較重時，如果該 topology 沒有考慮到 hop-count 數多的 demand，將會被消耗比較多的網路資源。

(2) Weighted Hop-count

a. 在 uniform traffic 的情形下，RD topology 的表現最差。當 traffic loading 低時，DHP topology 有最佳的表現。當 traffic loading 逐漸增加，RDHP topology 有最佳的表現。

b. 至於在 lighted-skewed, skewed, 和 very skewed traffic 的情形下，當 traffic loading 中等時，RDHP 的表現會優於 RD 及 DHP。在 traffic 是屬於 skewed 及 very skewed 的情形時，如果網路有輕微 overloaded 的現象，

reconfiguration 所導致的效能增益會相當顯著。

3.5. Maintain Routing Stability during Network Reconfiguration

3.5.1. 前言

當網路根據 IP Topology Design Algorithms 決定出新的 IP 網路拓撲之後，且確定 IP topology 必須改變時，如何 reroute 既存訊務使之不受到影響，或是將影響降至最低，即所謂之 Minimum Impact IP Topology Migration Schemes。一旦 IP 網路之拓撲發生變化，網路中的 router 將會執行 IP routing protocol，若採用 link-state routing protocol，如 OSPF，在網路進行 link-state 同步的過程中，無可避免地將產生 packet loss，其原因可歸納於兩方面：black hole 及 forwarding loss。Telcordia 所提出之 Minimum Impact IP Topology Migration Scheme 將可使 IP topology 必須改變而重新配置下層光傳輸網路時，將對既存訊務之影響降至最低。

3.5.2. Link-State Routing Protocol Review

首先回顧 link-state routing protocol 一些概念，並將重點放在 IP 網路 link-state 發生變化時之行為。仔細分析 link-state routing protocol 之動作，將發現 black hole 及 forwarding loop 將可能產生，如同圖 9 所示。造成 black hole 及 forwarding loop 之原因在於當網路拓撲發生變化之後，router 並無法在同一時間對網路持有相同一致的資訊，而產生 packet loss 或暫態迴圈。

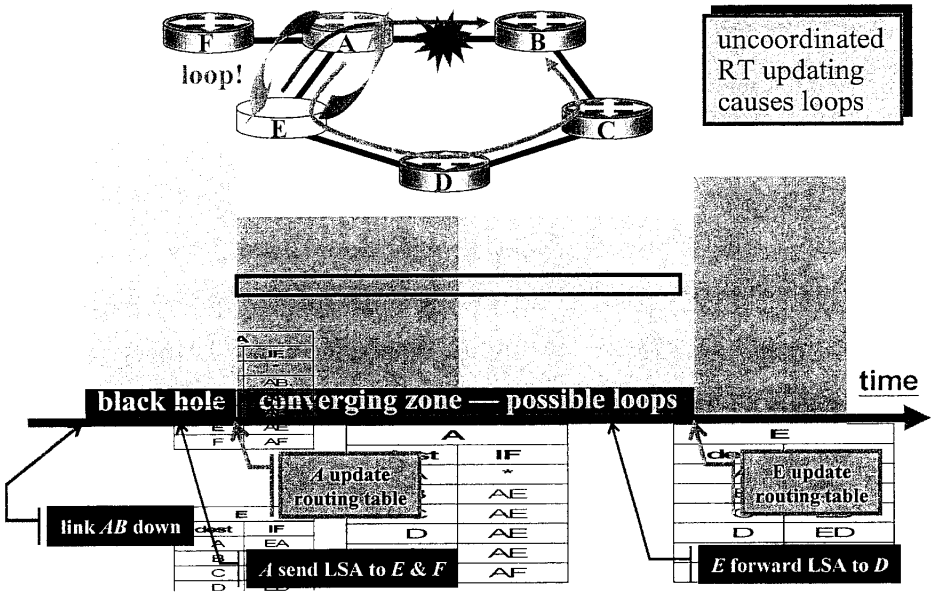

圖 9　black hole與forwarding loop可能發生之時機

為避免前述問題之發生，Telcordia 提出一種方法，即是將網路拓撲之變化分為兩類：新增及拆除。將網路拓撲之演變視為一連串之鏈路新增與拆除。Link 之新增部分由一般之 IP routing protocol 負責運算，而鏈路之拆除則由 Telcordia 一項獲得專利之演算法 LOOPRID 執行。以下章節即在闡述 LOOPRID 之概念以及 LOOPRID 之引進對於網路拓撲演進提供之助益。

3.5.3.　LOOPRID

3.5.3.1. LOOPRID 運作原理

LOOPRID 之功能為當網路拓撲之演變是依據事先之規劃，並且網路執行 link-state routing protocol 之條件下，藉由消除 black hole 及

forwarding loop之方式，可有效地將packet loss降至最低。以下用圖 10 為例說明 LOOPRID 運作情形。圖中網路運用 link-state routing protocol，節點 A～K 各點代表 router，而網路拓撲之演進將會拆除 AB 間之鏈路，通常鏈路皆為雙向性，在此先考慮其中一個方向，即由 A 至 B 之鏈路。

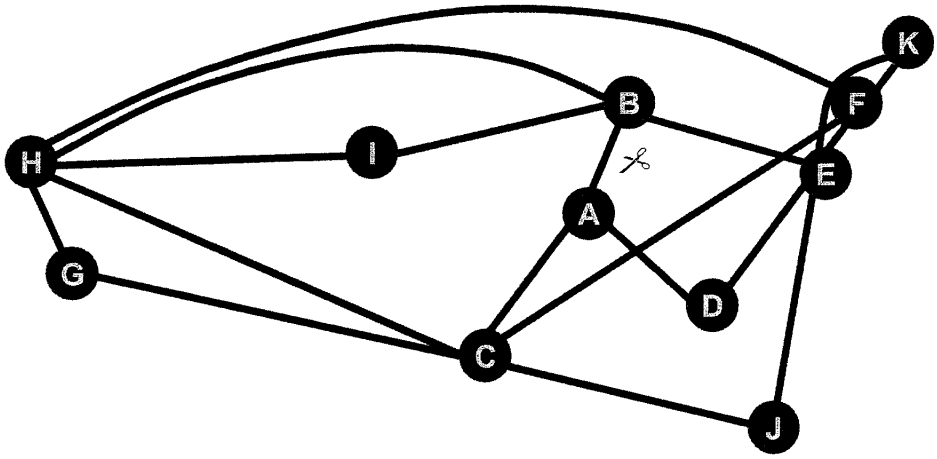

圖 10　　網路拓撲變化範例，AB間之鏈路將要拆除

首先，必須指出網路中哪些 router 之 routing table 將會因拓撲之變化而改變，以及如何更新 routing table 以避免 loop 之產生。對於一個具有 M 個 router 及 L 條 link 並且執行 link-state routing algorithm 之網路，而由 A 至 B router 間之鏈路將要拆除，在此定義以下名詞：

N：所有 routing table 將會改變之 router，即為受影響 router 所成之集合。

T：以 B 為 root 所架構出的 reverse shortest path subtree，其

中並包含 BA 分枝。

R：在 T 中所有節點所成之集合。

在此，Telcordia 提出以下兩個定理：

(1) Theorem 1：N $\subseteq$R

(2) Theorem 2：假設 S 為根據 hop count 為基準，以降冪方式排列抵達 B router 網路拓撲更新順序所成之數列，而此更新順序乃是預先計算(pre-computing)之後 routing table 變化之方式進行，則依據此順序 S 更改 router 之 routing table 將不會產生 loop。

利用上述之 Theorem 1 決定出受影響之 router 集合之後，接著決定該集合中 router 之 forwarding table 更新順序。在此，LOOPRID 的做法分為兩步驟：首先依據 breadth-first 的方式，及依據 hop count 決定更新順序，hop count 越大之 router 優先更新。接著，相同 hop count 之 router 更新順序由 vertex-label 為標準，以降冪方式 label 越大者先行更新。在考慮 AB bidirectional 鏈路將兩個方向各決定其受影響 router 及其 routing table 更新順序後將結果合併，可得如圖 11 之示意圖。

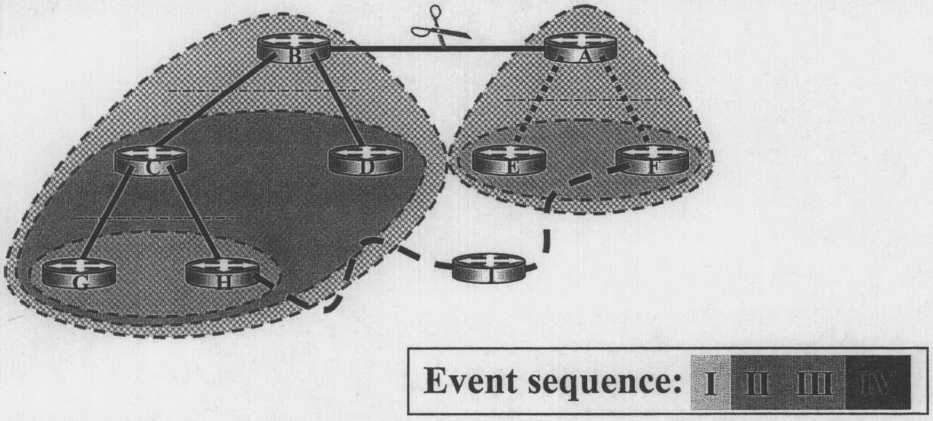

圖 11 更新受影響router之順序

3.5.3.2. SNMP Manager Controlled Update

當依據上述方式決定出受影響 router 以及變更順序之後，接下來就必須由 NGI 將所計算得之新的 forwarding table 透過 SNMP protocol 寫入 router 中以完成 forwarding table 之更新，並保留 distributed link-state database 之備分。

3.5.3.3. Interaction with Routing Protocol

當系統執行 LOOPRID 之後，即將拆除之鏈路上將不會承載任何訊務，而在將此鏈路拆除之後，因為 link-state 發生變化，網路將會自動執行 link-state routing protocol，並將會產生 LSA 並將之 flooding。因為經過 LOOPRID 之後已經將受影響 router 之 forwarding table 已預先改變，因此，這些 LSA 將不會造成任何 forwarding table 之變化。換句話說，即是 LOOPRID 已經”預先計算”出因 link-state 變化後受影響 router 之 routing table 將會更新的結果，(一般來說，router software 會

將 routing table 裡的 active route install 到 forwarding table，kernel maintain 一份 forwarding table 的 master copy，並將此 master copy 複製到 packet forwarding engine)，並已經事先將這些 forwarding table 改成與執行 link-state routing protocol 之後的結果一樣。因此也就避開了可能會產生的潛在迴圈。

3.5.4. 結語

據 Telcordia 提供之資料，LOOPRID 具備有以下特性與優點：

- LOOPRID 可以運用事先規劃的方法，將因網路拓僕變化產生鏈路拆除而產生之潛在迴圈消除，降低對於現存訊務衝擊。
- 採用此種策略將允許 Carrier 業者進行較頻繁之網路拓撲變動以因應客戶及市場需求，而無須顧慮拓撲變更所造成影響。
- 如此，將可運用於網路拓撲之 reconfiguration 及例行之網路維修。
- LOOPRID 之運用必須於執行 link-state routing protocol 之網路上；而 forwarding table 之更新將由 NGI 負責。

4. 心得與建議

基本上Telcordia的NGI NC&M系統是一種基於層疊模型 (Overlay model) 的訊務工程 (Traffic engineering) 解決方案。基本的精神是在維持原 IP 拓撲各網路節點介面數目及鏈路容量不變的前提下，透過其核心技術中的 3 種 IP 層的拓撲演算法，提出可以滿足網路訊務需求量的新 IP 拓撲。接著進一步透過網管面 (Network Management Plane) 對底層網路鏈路調整與調度，達到上層 IP 層拓撲改變的目的，而此一經過調整過的IP拓撲比起調整前的IP拓撲，可提高與發揮其網路運作效率，滿足既有的訊務需求量。

就訊務工程的做法而言，大致可分為三大類。

1. 基於IGP metric調整的訊務工程

2. 層疊模型解決方案

3. MPLS-TE解決方案

第一種與第三種做法並未更改既有之 IP 拓撲環境，其做法主要精神是在基於既有的 IP 拓撲之上，透過一些機制的運作改變訊務轉送的路徑。而第二種基於層疊模型的解決方案的重點就是可以透過底層網路的調整與安排，對上層 IP 層拓撲加以改變，進而影響並改進訊務轉送效率。

就目前觀之，IP-based的網路是現在及未來網路發展的趨勢。眾所皆知，傳統的 IGP 是基於最短路徑演算法（Shortest Path First

algorithm，SPF algorithm）做路由的選擇，將一些簡單的 metric 相加，計算出最佳化的路徑。早期的做法便是透過 NOC(Network Operation Center)裡面的一批網路管理員，不斷觀察網路流量與網路運作情形，透過其維運經驗與主觀的判斷力，藉由 IGP 鏈路 metric 的調整，企圖影響並調度訊務轉送路徑，以改進網路頻寬使用率不均的問題。當然這種做法目前已不多見。

隨著 ATM 與 Frame Relay 技術的發展與成熟，所謂的 IP over ATM 或 IP over Frame Relay 的層疊模型 (Overlay model)成為當時骨幹網路架設的風潮。但隨著 IP 技術的突飛猛進，這種在當年曾經是佈建骨幹網路主流的兩大技術，如今已趨式微。不過，在 IP over 傳輸網路或光網路的部分，層疊模型的解決方案仍然具備極為重要的應用與營運價值，如隨選頻寬，O-VPN 等。Telcordia 的 NGI NC&M 系統乃是基於層疊模型的架構，底層可以使用第二層的交換技術，也可以使用透過光交接連接的底層光交換技術。本系統其核心技術有二：一為 IP Topology Design Algorithms；一為 Minimum Impact IP Topology Migration Schemes，就是所謂 LOOPRID 的技術。由於其思維皆是基於 IP 層作為其考量的依據，因此在底層網路技術的使用方面具有極大的彈性，不需侷限第二層或第一層。

MPLS-TE (RFC2702) 是目前 IETF 所主導的訊務工程的做法。不變更既有的 IP 拓撲，而是在構建此既有 IP 拓撲的路由器上啟動相關的 MPLS 路由與信令機制，在控制面(Control Plane)方面直接賦予 IP 路由器路由與信令的智慧。在佈建一個基於 MPLS 技術的網路後，封包在資料面(Data Plane)可藉由標籤的交換，橫越基於 MPLS 技術的骨幹網

路。因此透過隧道的建立，改變訊務轉送的路徑，進一步改善網路鏈路頻寬使用率不均的問題，達到訊務規劃工程的目的。

在此，先不討論演算法本身的細節及模型，讓我們站在 30,000 呎的高空往下鳥瞰全局，筆者想提出一些思考的方向。首先，本系統所設計的新 IP 拓撲基本上僅針對訊務"量"的本身予以考量，訊務的種類及特性並未納入其考量的範圍。我們知道未來的多媒體網路比起最短路徑更重視服務等級的區分、網路壅塞的控制、網路鏈路故障的處理、整體網路頻寬使用率的提升等等。這些層面，本系統是否皆可以滿足，有待各位進一步思考。

再者，網際網路的訊務常有 burst 的現象，也就是說網路訊務負載常會因某些特殊原因而有可能在短時間內發生巨幅波動的情形，例如當有特殊的運動賽會或娛樂影視活動舉行時，都會使得網路上的訊務流突然大幅地增加，活動過後網路的訊務流又會大幅減少。如果 IP 拓撲在活動前夕為因應活動訊務量作出調整，而在活動過後又回復其原先之拓撲環境，短時間內對 IP 拓撲改變是否意味著不穩定的因子增加? 因為通常我們常談的電信級(carrier-grade)網路，一般來說是希望網路愈穩定愈好，愈可靠愈好。

目前數據網路的流量大都架於 IP 網路之上，客戶對於網路頻寬需求與日俱增，IP 網路寬頻化、高速化已是中華電信必須積極面對、因應的課題。為提高網路的傳輸頻寬之能力，中華電信已建設密集式分波多工(Dense Wavelength Division Multiplexing,DWDM)高速網路傳輸技術之骨幹網路，面對於 Traffic 變化劇烈的 IP 網路環境之下，我們更應

加速研究新一代網路傳輸及其相關技術如 GMPLS,RPR RING, Netxt GenerataionSDH 等技術，來提高轉送資料之速度，以及早因應用戶迴路開放後嚴峻的競爭環境。

在此同時，使用者亦有網路價格下降的強烈需求。流量需求增加，併同價格的降低，結果使得中華電信需要尋求一能提供大流量並具價格效益之解決之道。 而 TelcordiaNGI NC&M 適時提供一個解決此問題的工具，此工具能夠主動地分析 Demo Site 的網路拓撲結構與訊務資訊，透過改變網路的組態設定，來達到網路拓撲最佳化的目的，藉由訊務壅塞情況的改善，進而能使中華電信以更小的成本獲取到更大的利潤。

最後，筆者想引用 IETF 的哲學： “Let the market decide!”。就讓市場來決定技術吧！