TWAREN 路由政策與架構：可精確管理路由之設計與實現

梁明章　　陳俊傑
財團法人國家實驗研究院國家高速網路與計算中心
E-mail: liangmc@nchc.org.tw、jjchen@nchc.org.tw

摘要

由於 TWAREN 的連線單位幾乎也是 TANet 的連線單位，多半被視為第二套骨幹出口，所傳送的路由會影響到每個連線單位的出口選擇，因此必須承擔調整管理路由的責任，再加上 IP space 屬於 TANet 的因素影響，有許多環境限制，導致路由管理難度很高。本文從討論各項環境因素引發的幾大路由政策議題開始，逐步分析困難點，並描述我們思考解法的理念過程，最後具體描述我們規劃完成的路由政策與架構，尤其是路由變換以及精確路由管控、路由管理系統、路由自動監控系統的概念，期能帶給學術網路網管人員一個具有參考價值的網路設計案例。

關鍵詞：Routing policy、路由政策、路由管理、BGP community。

1. 前言

TWAREN 包含 Production、Research、Optical network 共三套網路，其中 Production network 乃是負責日常一般用途的骨幹網路，對使用者而言，他必須是一套管理完善穩定的網路。對大多數的網管人員來說，網路架構越精簡就越容易管理，越不易出錯，而 TWAREN 的 Production network 就設備架構而言，其實是相當簡單的二階層設計，路由器數量也不多，在一般狀況下，採用單一 OSPF 或 ISIS domain 路由設計就已經是綽綽有餘。

然而天不從人願，許多外在的環境因素與政策因素使得簡單的設備架構必須配上一套極為複雜的路由政策。首先，TWAREN 本身沒有足夠的 IP 位址可以發配給連線單位是個很大的問題，以現今全世界 IPv4 捉襟見肘的情況，TWAREN 即使有再多的經費也很難申請到足以與 TANet 相當的 IP 位址數量，因此各連線單位只能使用 TANet IP 與 TWAREN 進行連線，導致 TWAREN 在對外交換路由時不能宣告連線單位的 IP。

其次，由於 TWAREN 在現階段與 Internet 交換的領域不廣，目前已經接洽好的僅有國外研網與微軟 MSN 等等，與 TANet 的 Internet 交換領域相比差距甚遠，因此連線單位自然會以 TANet 為 default route，TWAREN 做為第二出口。既然身為第二出口，TWAREN 須轉送所學得的 routes 給連線單位，且會影響各連線單位的路由行為，於是 TWAREN 就必須對這些影響負起責任，謹慎管理所有路由的轉送，該給則給、不該給要濾除，因此每新增一個連線單位，就有一個獨立不同的過濾規則要相應而生，如此一來將會違反「簡約即穩定」的狀況，使得 TWAREN 的路由政策變得困難重重。

為了解決上述困難，我們必須構思一套能符合各項環境與政治因素限制、盡量簡化每個過濾器設定、提高穩定度、降低出錯率、能提供客製化彈性的完整路由政策與架構，這就是本文往後章節所要詳述的內容。

2. TWAREN Production 設備架構

TWAREN Production Network 架構如圖 1 所示，以四個 12416p 為骨幹核心，每個 12416p 下掛數個 7609p 為 GIGAPOP，每個 7609p 又與 TANet2 7505 線路相連，但是 7505 卻是直接與 12416p 交換路由，7609p 對 7505 而言僅扮演 Layer2 的角色，GIGAPOP 架構圖如圖 2 所示。TWAREN 目前對外的主要出口多半連接在 MOE12008、TP12416p 以及 ASCC7609p。

因此，TWAREN 目前是以 Production Network 的骨幹，同時收納 TANet2 與 TWAREN 連線單位，TANet2 的連線單位以 T1、10Mbps 等低速連線單位為主，100Mbps 以上則接至 TWAREN 7609p 上，這些便是 TWAREN 的基本架構。

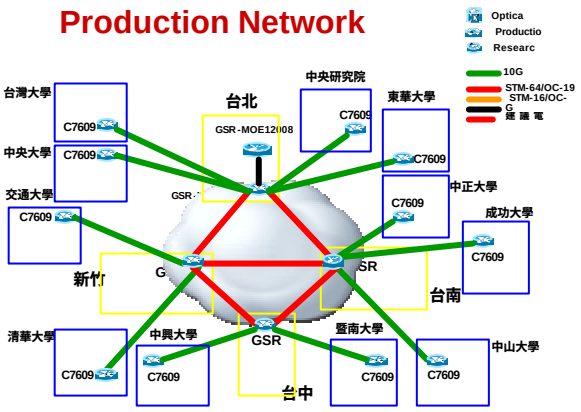

圖 1 TWAREN Production Network

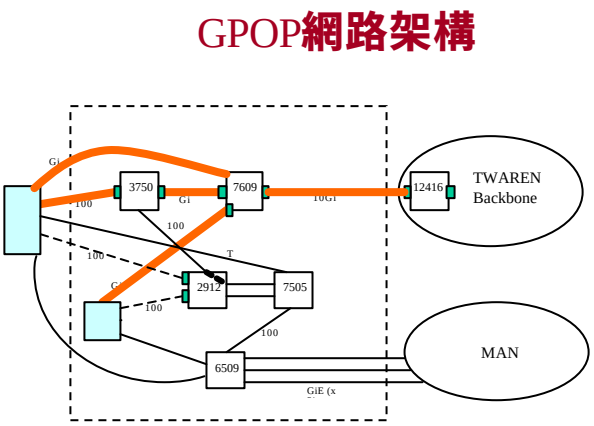

圖 2 TWAREN GIGAPOP 網路架構圖

3. 路由政策議題

TWAREN 在路由設計上需思考的議題主要為下列幾點：

3.1 需結合 TANet2

原有 TANet2 計畫並未停止且尚有連線單位，因此 TWAREN 必須接納 TANet2 之單位，然而有些單位在短期內無法將 T1 線路升級至高速頻寬，因此必須在保留 TANet2 舊設備的情況下將之併入 TWAREN 中。

3.2 無自有 IP 位址

TWAREN 的連線單位以學術單位為主，各單位的 IP 多半由教育部所配發，而 TWAREN 無法申請到足夠的 IP 可分配給各連線單位，因此連線單位只能以 TANet IP routes 與 TWAREN 交換路由，且大多數採用 IGP 交換，如果 TWAREN 將這些單位的 IGP routes 轉換成 BGP 送給其他 AS，將造成 Internet 上的混淆，相同 IP route 起源自多個相異 AS，結果可能被某些 AS 網管濾除掉。

3.3 連線單位速度差異過大

由於 TANet2 與 TWAREN 共用骨幹路由器，TANet2 的 T1 連線單位路由與 TWAREN 高速連線單位路由在骨幹路由會合交換，會導致某些校際之間的傳輸壅塞在 T1 低速連線上。

3.4 路由不對稱之避免

由於 TWAREN 目前跟 Internet ISP 接壤不多，因此連線單位多把 TWAREN 視為第二出口，將 TANet 作為 default，沒有接收 TANet routes，因此 TWAREN 送給連線單位的 routes 會影響各校的對外路由選擇，很容易就造成去回異路的不對稱問題。

3.5 連線單位特殊要求

有些連線單位之間有自己的專線，例如交大、清華、中央、陽明四校自有專線，可能會希望直接相連而不經由 TWAREN，又或某校有些特殊的需求，譬如不願意經過 TWAREN 到達某些目的地等等，這些類似客製化的特殊需求亦必須有能力解決。

3.6 管理複雜

由於 TWAREN 的連線單位幾乎同時也是 TANet 連線單位，而 TANet 被視為 default route，因此不需要將 Internet routes 送入骨幹給連線單位，故 TWAREN 必須將眾多 routes 送進骨幹並傳給連線單位，這些 routes 都會影響連線單位的出口路由選擇，因此 TWAREN 不只是骨幹內部必須流通許多 routes，更必須謹慎管控送給每個連線單位的路由，在維護與除錯上增加了許多成本與失誤發生的機率。

4. 路由政策設計與實現

根據前述之各項議題，我們將在本章節解釋我們的思考過程與設計理念。

4.1 區分骨幹路由與連線單位路由之三階層 iBGP domain

這個觀念是 ISP 早已行之有年，我們參考 Internet2 的美國著名研網 Abilene 的路由規劃[1]，將介面相關的路由與使用者相關的路由區分開來，如此可以避免因骨幹內設備介面 Up/Down 造成送給連線單位的路由產生頻繁的異動，甚至波動到 Internet routing，所以將介面相關路由以 IGP 交換而使用者路由以 iBGP 交換是個常見的方式。

我們考量到 IPv6 與 IPv4 必須並存，因此選擇 ISIS 作為 IGP[2][3]，所有 TWAREN router（包括國外兩部 GSR）都屬於同一個 ISIS domain，每個 router 都 redistributed connected，使得每個 router 都能知道其他 router 的每個 interface 如何到達，藉此提供一個不需要執行 full mesh iBGP 的基礎環境。

我們在 ISIS 的基礎平面上，建構一個階層式的 iBGP domain，以 MOE12008 為第一層 RR（Route Reflector），國外兩個 GSR 以及 TP12416 為其 client。第二層 RR 是 TP12416，而新竹、台中、台南三個 GSR12416 為其 client。第三層 RR 則是四個 GSR12416，各 GSR 之下的 7609 為其 client。以此三階層 RR 架構來達成 iBGP 的同步[4][5]。

我們將連線單位送進來的路由都轉換成 BGP，因為 TWAREN 的連線單位大多沒有自己的 ASN，或限於設備的能力不適合使用 EBGP 與

TWAREN 交換路由，故採用 IGP 的單位相當多，這些 IGP 一進入 TWAREN 7609 就立即轉換成 iBGP 流通於整個 iBGP domain 之中，以使之與 backbone IGP 區隔。我們以 BGP 形式統一路由的主要理由是因為 BGP 具備很多屬性可做出相當多的政策管控，且因為有這樣的轉換機制，我們無需限制連線單位的路由協定選擇，如此可具備高彈性與自由度[6]。

4.2 不對外宣告 TANet IP於本身 AS中的方法

我們必須將連線單位路由送給外界，卻又不能宣告於 TWAREN ASN7539 之中，且無法強制轉換 iBGP route 的 Origin ASN 為 MOECC 的 ASN1659，可行的方法就是把 MOE 送過來的 TANet superblock EBGP routes 轉送給外界，然而 TWAREN 連線單位並未包含整個 TANet，因此 Border Router MOE12008 必須能區分國外回流的傳輸目的地是否為 TWAREN 連線單位，如果是，則往 TP12416p 轉送南下，如果不是，就轉送往 MOE。為了達成此目的，我們在 TP12416 傳送 iBGP 給 MOE12008 的時候，將源自 IGP 的路由轉換成 EIGRP 才送給 MOE12008，如此一來，MOE12008 上面就擁有來自 MOE 的 TANet superblock EBGP routes 可轉送給外界，同時也擁有來自 TP12416 的 IGP（EIGRP）可得知連線單位的路由，就能達成分流的目的，並且解決了 TANet IP 不宣告在 ASN7539 的議題。

4.3 利用 BGP community 屬性精確管控路由

我們參考 Internet2 Abilene 研網 BGP community 的設計精神[7]，希望能藉由精密的 community 管控來解決前章節所提到的幾大議題，當某個連線單位的路由（可能是 IGP、EBGP、Private EBGP）一進入 7609 時，立即轉換成 iBGP，並且賦予一個 unique community number，對於其他 AS 進來的 Routes 也一樣處理，使得整個 iBGP domain 當中流動的每一行 iBGP routes 都帶有 community number[8]。

我們針對 community number 制訂了一套詳細的規則，限於篇幅，僅摘述一些基本規則介紹如表 1，以及一些使用中的號碼範例如表 2：

表 1 BGP community number 基本規則

*:1????	可轉送給國外與 TWAREN 國內連線單位，但不轉送給 *:2???? 的單位
:2????	只轉送給國外，不給 TWAREN 國內連線單位
*:3????	來自 TANET2 國外電路的 Routes
*:4????	來自中研院國外電路的 Routes
7539:*	以 IGP 連線的單位
*:?NMMM	N 為 POP 號碼，NMMM 組成四位數字編號，作為每個連線單位的唯一編號

表 2 BGP community number 基本規則

ASN:2????	有獨立 ASN，且只送國外不給國內，也不接國內路由（台大、交大）
ASN:1????	有獨立 ASN，國內外皆轉送（中山、清華、中央、成大、中正、東華）
7539:?????	無 ASN 的 TANet 單位，不能以 AS7539 送出國，會在 EBGP 出口時濾掉
*:4????	來自中研院的國外 Routes
*:3????	來自 TANET2 國外線路的 Routes
Private:?????	非 TANet 的連線單位（林業局、中央氣象局）
1659:20000	來自 MOE，完全轉送給國外，但不給 TWAREN 國內
18181:10000	來自國衛院
18414:10000	來自 NCHC
17713:19001	中山大學
7539:15001	中興大學
18185:24001	交通大學
65535:10001	林業所
*:?NMMM	NMMM 作為連線單位的 Unique Number

藉由每行 route 所攜帶的 community number，我們可以精確掌握其來源單位，並且在 community number 中規劃幾位數字來表達分類概念，相較於傳統的 IP prefix 表示法，我們大大簡化了 route_map 表達的複雜度，因為 TWAREN 必須接受連線單位的客製化要求，謹慎掌握送給連線單位的路由，導致每個連線單位可能都得使用獨立的 route_map filter，而精確的 community number 發配規則確實替我們簡化了相當多的困擾，並且提供了比 IP prefix 寫法更多的彈性運用，同時可接受連線單位量身定做的要求，假設交通大學因為與清華大學之間有專線相連，可以要求兩校之間不走 TWAREN，則我們可在交大處濾掉清大路由、在清大處濾掉交大路由，因為每行路由都有「身份標籤」，對於簡化路由對稱與否的調整有相當的幫助。對於前述 TANet2 T1 低速連線單位之議題也可以利用 community 群組區分使得 TWAREN 跟 TANet2 單位不會互相學習到對方的路由以避免高低速差造成的壅塞。

我們在幾個對外的入口處也盡可能地以 community number 區分群組，例如中國大陸、香

港、新加坡、美國、澳洲、日本等等，如此做的原因是因為台灣學術網路有三大對外連線單位「教育部、中研院、國網中心」，可以依照目的地的網路距離與速度來選擇出口，決定哪些該送給連線單位，而 community 比起 AS list 更能簡化這些區分。

對於路由的走向與 community 的區分，請參考圖 3 的概略圖介紹。

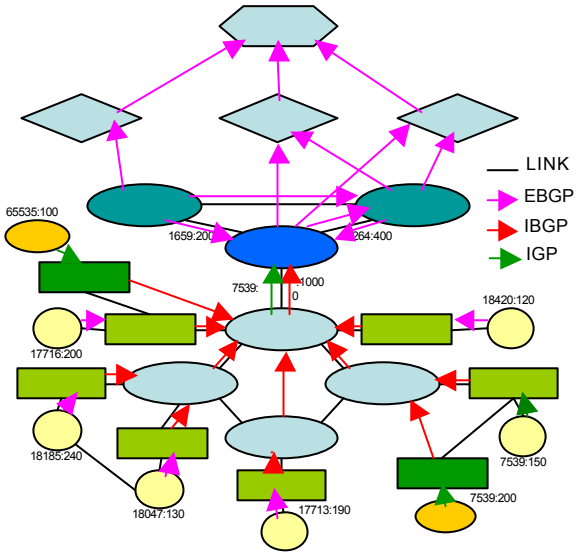

圖 3 TWAREN 路由架構圖

4.4 建立路由管理系統

由於設計了如此複雜的路由規則，自然要考量到管理方法，由於人力不足且人類容易出錯的特性，我們必須建立一套路由管理系統，以網頁程式結合資料庫，儲存各連線單位的聯絡資料、路由資訊、路由規則、特殊要求、甚至包含該單位的網路拓樸圖，工程師只要登錄這些資料，系統自動給予 uniqueue community number，未來工程師可用單位名稱、community number、route 等等來查詢，可以很快查明路由的資訊，在除錯上有相當大的助益。

5. 結論與未來研究

本文中描述 TWAREN 採用 ISIS 作為設備介面路由協定，藉此實現三階層式的 iBGP domain 架構來取代 full mesh 的 iBGP link 架構，將所有連線單位的路由在 7609 進入點即轉換成 BGP，貼上 unique community number 辨識屬性，使工程師更容易掌握每個連線單位的路由來避免非對稱路徑、特殊要求等等。在出口處則利用 IGP 與 EBGP 並存方式來達成不宣告 IP 的限制及分流的目的。

本文也論及路由管理系統的開發來輔助工程師管理複雜的單位 community 與 policy，在未來的工作上，我們會進一步延伸自動管理的概念，利用自動化的路由收集程式，將 iBGP domain 中的路由定時收集下來，配合路由管理系統的資料庫作分析，可以立即發現哪些連線單位的路由發生變化或錯誤，自動發出警訊通知操作人員或是工程師，如此一來，TWAREN 將可在使用者反應網路問題之前就發現狀況並且迅速處理。這一方面我們仍在開發中，等到完成時，將會另文討論系統細節，提供經驗與各網管人員參考。

參考文獻

[1] http://loadrunner.uits.iu.edu/upgrade/docs/routing-design-doc-0703.txt

[2] Callon, R., "Use of OSI IS-IS for Routing in TCP/IP and Dual Environments", RFC 1195, December 1990.

[3] Christian E. Hopps, Routing IPv6 with IS-IS, Draft-ietf-isis-ipv6-05, January 2003.

[4] T.Bates, R. Chandra, E. Chen., “RFC2796 BGP Route Reflection - An Alternative to Full Mesh IBGP”, April 2000.

[5] Y. Rekhter, T. Li., “RFC1771 A Border Gateway Protocol 4”, March 1995.

[6] Y. Rekhter, E. Rosen, “RFC3107 Carrying Label Information in BGP-4”, May 2001.

[7] http://www.abilene.iu.edu/bgpcommun.html

[8] R. Chandra, P. Traina, T. Li., “RFC1997 BGP Communities Attribute.”, August 1996.