行政院所屬機關因公出國人員出國報告書

(出國類別：實習)

實習 Internet 網路維運新技術研習報告

出國人　服　務　機　關　　職　　稱　　姓　名
中華電信數據分公司　副設計師　康崇原

出國地點：美國

出國期間:89 年 7 月 9 日至 89 年 7 月 22 日

報告日期：90 年 2 月 2 日

行政院研考會/省(市)研考會 編　號　欄

目　　錄

前言

Internet 從 1987 年商用化以來，由於電子郵件(E-mail)等創造性應用(Killer Application)之產生及全球資訊網上網(WWW)已成為人們工作及生活的一部份，因此商業化服務之 ISP 是新興業者與傳統電信業者全力角逐的戰場。

本公司營運之 HiNet 居目前國內商業 ISP 市場占有率之冠，實有必要吸取最新維運相關技術，以更加提昇網路品質鞏固市場競爭優勢，以因應即將開放服務之固網再次競爭態勢。

研習行程及課程

八十九年七月九日
去程

八十九年七月十日 ~ 八十九年七月十四日
Internet 網路維運新技術課程研習

八十九年七月十七日 ~ 八十九年七月二十日
Internet 網路維運新技術課程研習

八十九年七月二十一日 ~ 八十九年七月二十二日
返程

一、 Internet 路由架構

今日的 Internet 是一個分散式的架構，由全世界很多很多的 ISP(Internet Service Provider)，像是美國的 UUNET，Sprint，C&W 以及台灣的 HiNet，TANet，SEEDNET 等等經由網路交換接取點 NAP(Network Access Point)/IX(Internet eXchange)或以直接連接的方式互連而成，故又稱為網際網路，互連網等。

1.1 網路交換接取點 NAP/IX

Internet 的路由架構中 NAP/IX 扮演一個很重要的角色，NAP 的概念是由美國 FIX(Fedral Internet eXchange)與 IX(Commerical Internet eXchange)來的，目前美國 NSF(Nation Science Foundation)認可的 NAP 有:

Sprint NAP-Pennsauken，NJ

PacBell NAP-San Francisco，CA

Ameritech Advanced Data Services(AADS)NAP-Chicago，IL

MFS Datanet(MAE-East)NAP-Washington，D.C.

因為 ISP 之間需要互連使國內訊務不必繞道到美國再作交換，因此 Internet 發達的世界各國便陸續建置了其他的 NAP 或者稱為 IX(Internet eXchange)如倫敦的 LINX，阿姆斯特丹的 AMS-IX，日本的 JPIX，韓國的 KIX，新加坡的 STIX。國內由中華電信數據通信分公司於民國 86 年 11 月建置完成開放服務，命名為台灣網際網路交換中心 TWIX(Taiwan Internet eXchange)，目前計有 40 餘家 ISP 加入，交換之訊務量約為 640Mbps 左右。

1.2 NAP/IX 的架構

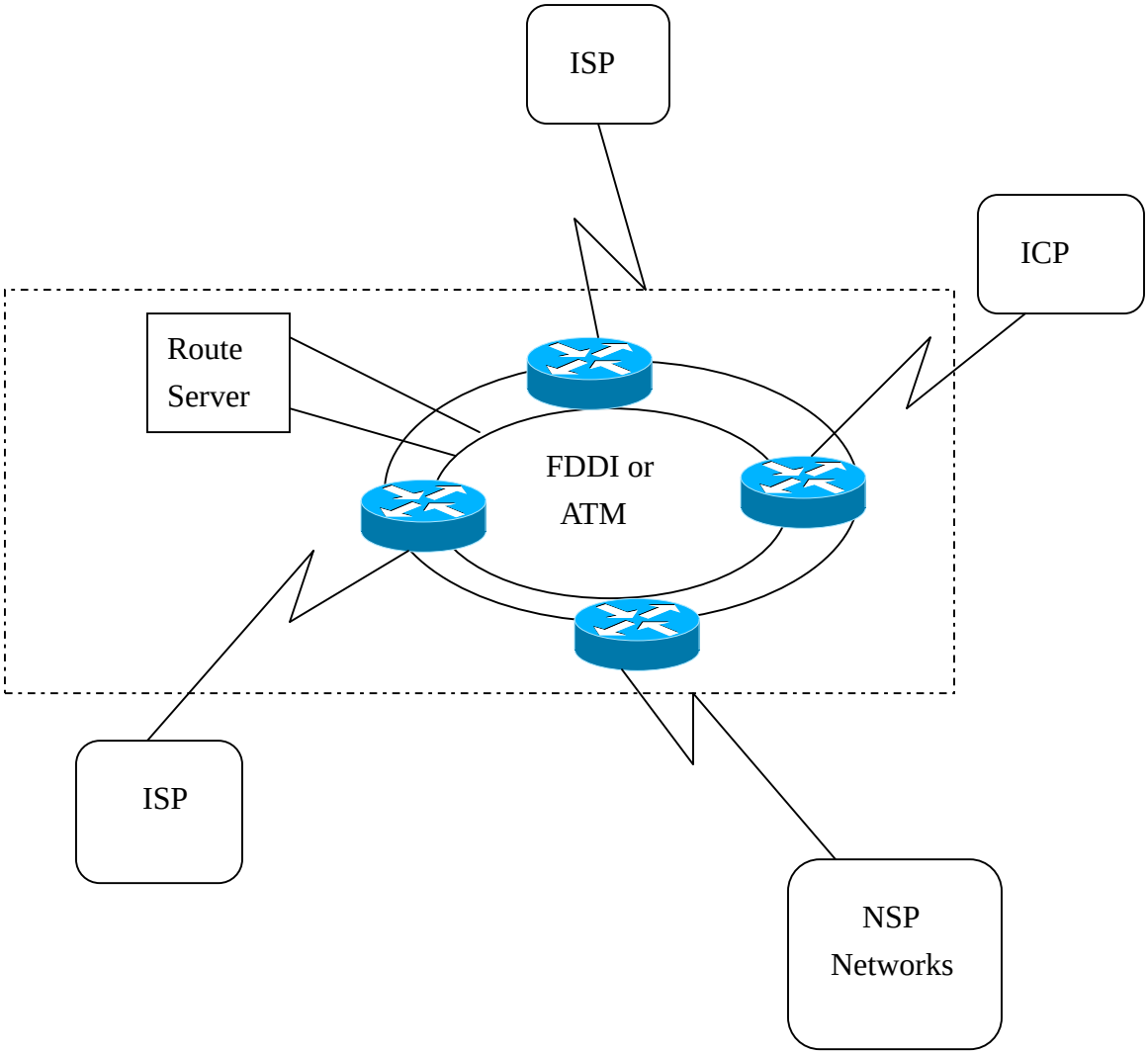
早期 NAP/IX 大部份是 FDDI Switch 或 ATM Switch，各 ISP 的路由器可以與它連接，以便互相交換訊務(如圖一所示)，此時 ISP 連接至 NAP/IX 之線路頻寬大都是 DS3(45Mbps)等級，然而 Internet 蓬勃發展後訊務急遽增加，大型 ISP 間訊務交換也提昇至 Gbps 級，而且 Gigabit Ethernet Switch 設備也大量上市，NAP/IX 於是紛紛提昇使用 Gigabit Ethernet Switch 以應付大量訊務交換所需。

1.3 對等互連(Peering)與付費轉接(Transition)

同一 NAP/IX 之 ISP 成員雙方在各自對市場競爭策略及其他因素考量下同意後可建立對等互連或付費轉接之關係，對等互連即兩方互不付費而讓屬於各自的訊務直接交換，這種情況較常發生在兩個規模大小差不多之 ISP 間或者互有需求之 ISP 與 ICP(Internet Content Provider)間，而最近非常興盛的設備共置(Co-location)業者也積極加入 NAP/IX 為其顧客尋找內容(content)之輸送通道，譬如美國最大的設備共置業者 EXODUS 公司即為一例，它參與了所有美國重要的 NAP/IX 並積極與各 ISP 建立 Private peering。

Private peering 指的是兩個 ISP 之間雙方直接建立連線，不經由 NAP/IX 之 Public Switch，甚至於是兩個 ISP 之 POP 間直接使用電路互連，此種情形多半是兩者間有大量訊務交換需要且雙方談妥某些協議下進行的。

付費轉接通常是小 ISP 付費給上游 ISP 由其代為轉接訊務至其他 ISP，也就是說小 ISP 成為大 ISP 之客戶關係，轉接之費用或與一般企業、個人用戶相同或有差異。



圖一 ：NAP/IX 實體架構

二、 網域間路由通信協定

2.1 自治區系統 AS

在 Internet 架構中每個 ISP 就是一個自治區系統(Autonomous System)，這個自治區系統是由執行相同路由策略(Routing Policy)之路由器群組而成(通常就是 ISP 本身)，而由 ISP 向 InterNIC、APNIC 等 Internet 註冊組織申請一個所謂自治區系統號碼(AS Number)來作為識別，譬如說美國 UUNET 為 AS 701，中華電信 HiNet 為 AS 3462，TANET 為 AS 1659 等。

2.2 IGP 與 EGP

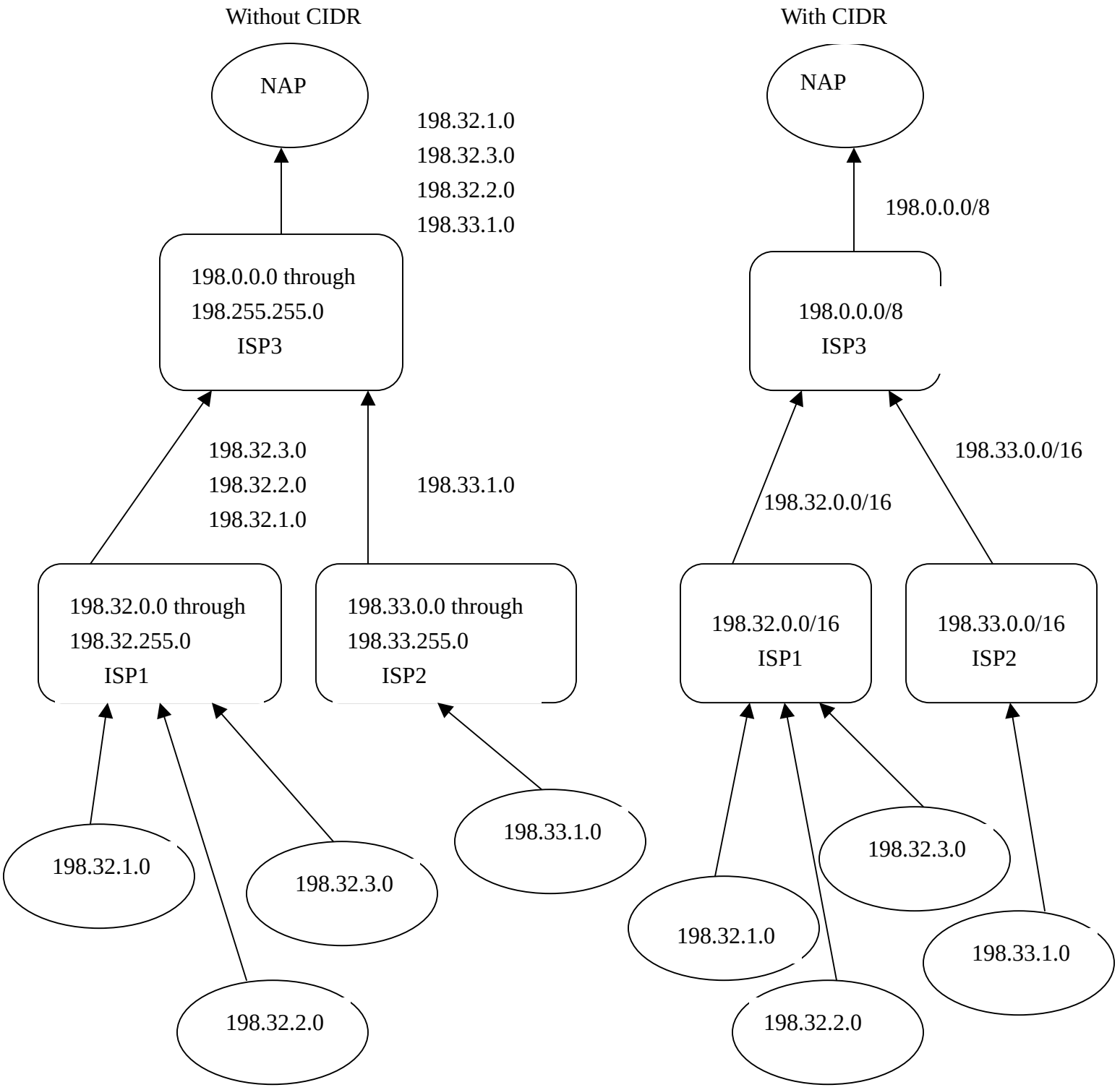
自治區系統的內部使用內部閘道通信協定 IGP(Interior Gateway Protocol)，像是 OSPF(Open Shorest Path Protocol)，IS-IS(Intermediate System to Intermediate System)，RIP(Routing Information Protocol)等，目前 ISP 均採用屬於鏈路狀態(Link state)通信協定之 OSPF 或 IS-IS，因該類通信協定之演算法具收斂較快，擴展性佳等優點。

自治區系統的外部即不同自治區系統間使用之路由通信協定稱為外部閘道通信協定 EGP(Exterior Gateway Protocol)，目前 Internet 上廣泛使用之 BGP4 即屬此類，外部閘道通信協定的發展目的除了控制 Internet 路由表(Routing Table)的擴充，同時它也提供了一個使 Internet 路由架構更為結構化的優點，將整個 Internet 路由網域劃分成個別的管理單位，就是自治區系統，而且每個自治區系統都可藉由設定邊界閘道通信協定(BGP4)之各種屬性(Attribute)值來執行各自的路由策略控制(Routing Policy Control)。

2.3 不分級之網域間路由 CIDR(Classless Interdomain Routing)

近年來由於 Internet 爆炸性成長，使得整個 Internet 網域路由表大幅增加，造成路由器記憶體很大的負擔，處理路由查表很複雜，幸好 Internet 在 1995~1996 年間採用了 CIDR 方法，使得整個 Internet 路由表的成長減緩。CIDR 和傳統分級 IP 把位址分級成 A/B/C 的方式不同，在 CIDR 中，IP 網路是以一個前導字(prefix)格式來表示，包含一個 IP 位址以及對這個位址從最左邊算起連續位元的一個指示值，例如 198.32.0.0/16，/16 就是一個指示值，指出網路遮罩從位址左邊算起的 16 個位元，也就是說和 198.32.0.0 255.255.0.0 是一樣的意義。這種表示法可以讓你將 198.32.0.0 的許多特定路徑(如 198.32.1.0 或 198.32.2.0 等等)都併入一個統合式的宣告裡，這種作用稱為整合(Aggregate)。

圖二顯示了整合的功效，本圖中右邊顯示使用 CIDR 的情形，ISP1 與 ISP2 各別對客戶子網路執行整合，ISP1 宣告 198.32.0.0/16 的整合，ISP2 宣告 198.33.0.0/16 的整合，同樣的方式 ISP3 也對 ISP1 與 ISP2 執行整合，只送出一個整合 198.0.0.0/8，比較起左邊未執行整合的網域，可見大幅減少全球 Internet IP 路由表的內容之功效。



圖二：分級定址與 CIDR 定址比較範例

2.4 超網路(Supernet)

當網路的 prefix 包含的位元少於網路自然遮罩的位元時，這個網路稱為一個超網路(Supernet)。例如 198.32.0.0/16，它的遮罩少於 C 級網路的自然遮罩(16<24)，所以它就是一個超網路 。

2.5 邊界閘道通信協定(Border Gateway Protocol Version 4)

BGP 通信協定歷經了好幾個階段的改進，到了 BGP4 成為第一個可處理不分級之網域間路由整合(CIDR)及超網路(Supernetting) 的版本。

BGP 根據與相鄰 BGP 之間交換的訊息構建成一個自治區系統圖(AS graph)，這個指引路由方向的自治系統圖形環境也稱為樹(Tree)，對 BGP 來說，整個 Internet 就是一份 AS 的地圖，每個 AS 都有一個 AS 編號可供辨識，兩個 AS 之間的連線就構成路徑(Path)，集合路徑資料便組成到達某個目的地的路由(Route)。

2.5.1 BGP 運作方式

BGP 是一種路徑向量(Path Vector)通信協定，BGP 的路由訊息中載送了一連串的 AS 編號，指出每一條路由所經過的每一段路徑。

BGP 用 TCP 作為它的傳輸通信協定(port 179)，也就是說所有傳輸的可靠性如資訊重送等都由 TCP 負責處理，BGP 本身不需另行設計這些功能。

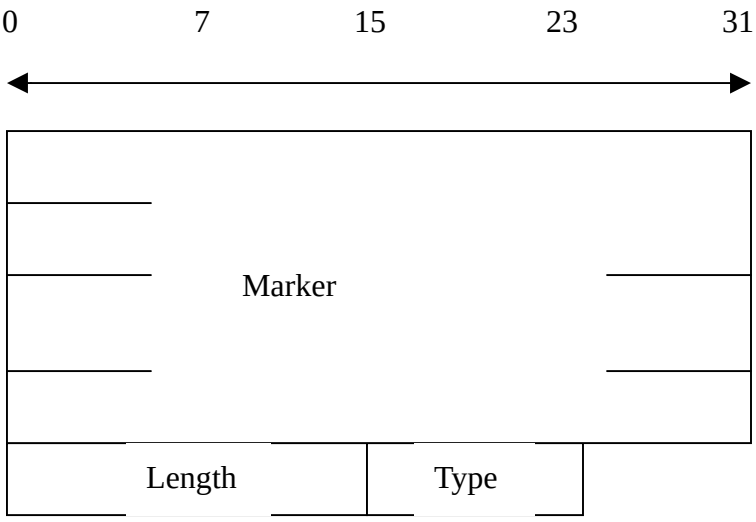
2.5.2 同伴(peer)

兩個 BGP 路由器之間形成傳輸通信協定的連線，稱為同伴(peer)或是鄰居(neighbor)。

2.5.3 BGP 訊息的表頭格式

BGP 訊息的表頭格式是 16 位元組的 marker 欄位，接著是兩位元組的長度欄位和一個位元組的型態欄位，圖三說明 BGP 訊息表

頭的基本格式。



圖三 ： BGP 訊息表頭格式

型態(Type)欄位說明訊息的種類，有下列幾種

- OPEN
- UPDATE
- NOTIFICATION
- KEEP ALIVE

2.5.4 BGP 路徑屬性(Path Attributes)

BGP 的路徑屬性是一組參數，用來記錄與路由有關的資訊，例如路徑資訊，對某路徑的偏好程度，路徑的下一站，及整合資訊等，這些參數使用在 BGP 的路徑過濾與路由優先決定程序中，每個 UPDATE 訊息都有一串不固定長度的路徑屬性，路徑屬性的可變長度格式是:

<屬性型態、屬性長度、屬性值>

屬性型態為兩個位元組的欄位，包含一位元組的屬性旗標與一位元組的屬性型態碼，圖四顯示屬性型態欄位的格式。

屬性型態碼(Attribate Type Code)位元組中含屬性代碼，目前定義的屬性如下:

1 – ORIGIN (Well-known mandatory, Type code 1)

2 – AS_Path (Well-known mandatory, Type code 2)

3 – NEXT_HOP (Well-known mandatory, Type code 3)

4 – MULTI_EXIT_DISC (Optional nontransitive, Type code 4)

5 – LOCAL_PREF (Well-known discretionary, Type code 5)

6 – ATOMIC_AGGREGATE (Well-known discretionary, Type code 6)

7 – AGGREGATOR (Optional transitive, Type code 7)

8 – COMMUNITY (Optional transitive, Type code 8, Cisco_defined)

9 – ORIGINATOR_ID (Optional nontransitive, Type code 9, Cisco_defined)

10 – Cluster List (Optional nontransitive, Type code 10, Cisco_defined)

11 – Destination Preference (MCI-defined)

12 – Advertiser (Baynet-defined)

13 – rcid_path (Baynet-defined)

255 – Reserved for development

2.5.5 EBGP 與 IBGP

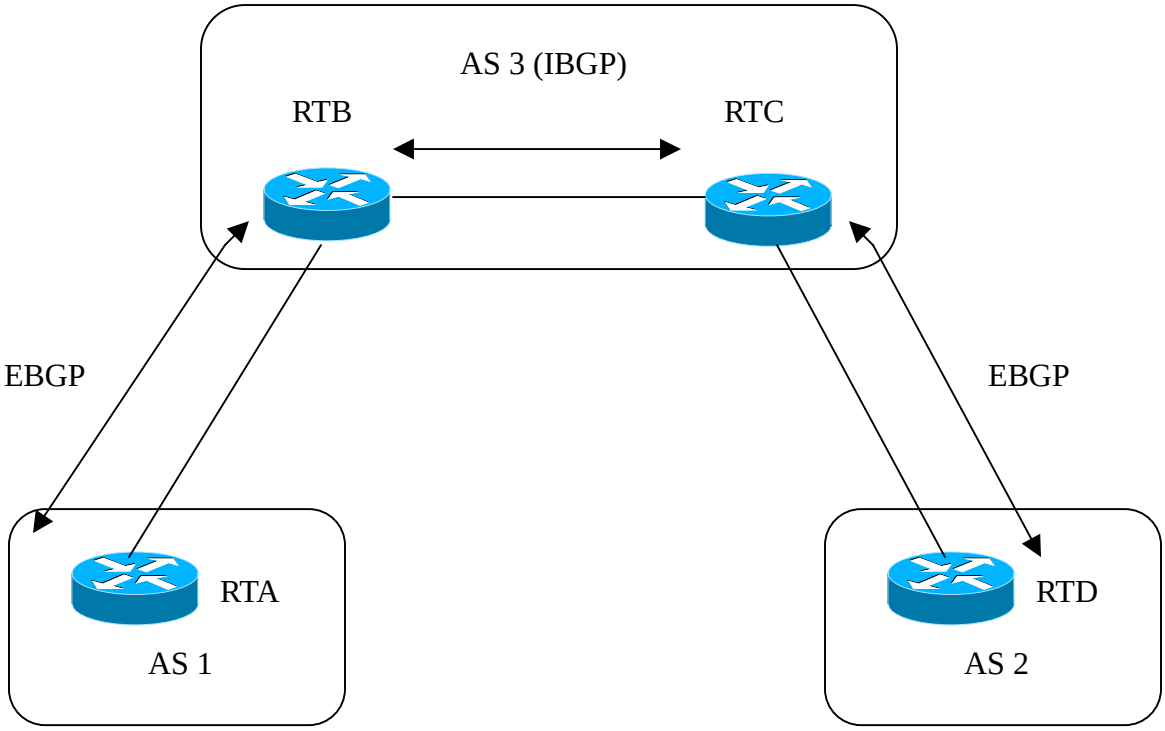
雖然 BGP 主要是使用於自治區系統 AS 之間，提供網域間無迴圈的一種架構，但是 BGP 也可以在自治區系統 AS 之內使用。

兩個路由器之間的 BGP 連線，稱為同伴連線(peer connoction)，可以在同一個 AS 之內建立，這種情況下的 BGP 稱為內部 BGP(IBGP)，

同伴連線也可以在不同 AS 的兩路由器之間建立，這種 BGP 稱為外部 BGP(EBGP)。

圖五中 RTA 與 RTB 分屬 AS1 及 AS2，這種情況下建立之 BGP session 稱為 External BGP(EBGP)。

RTB 與 RTC 之 peer connection 在同一 AS3 之內建立，這種情況下的 BGP 稱為 Internet BGP(IBGP)。



圖三 ： EBGP 與 IBGP 說明範例

2.5.6 BGP 屬性

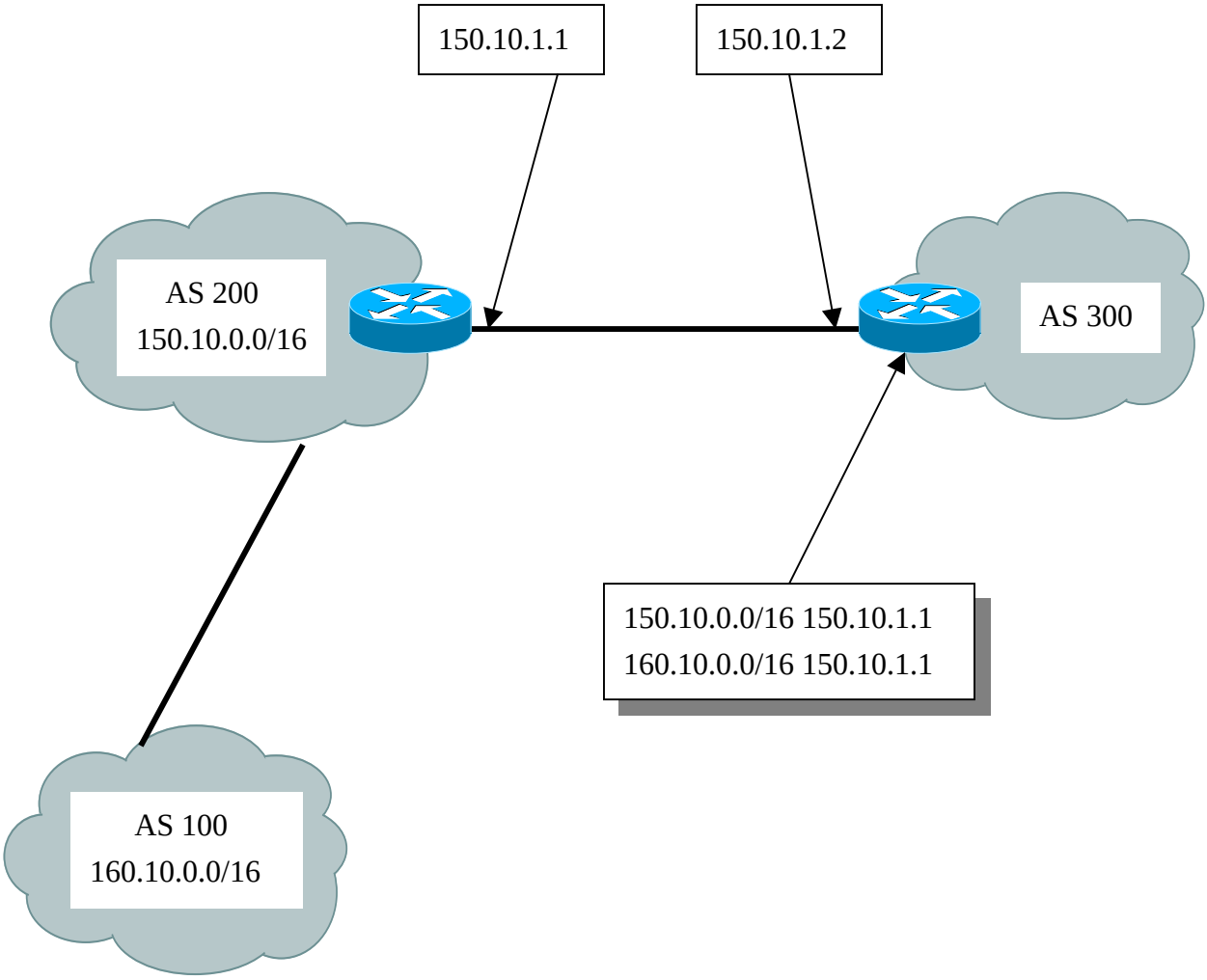
BGP 屬性是描述一個 prefix 特性的一組參數，BGP 依據這些屬性經由決定程序步驟(詳 2.5.7)來選擇它的最佳路徑，下面將一一檢視這些屬性及如何設定它們以控制路由行為。

Next_hop 屬性

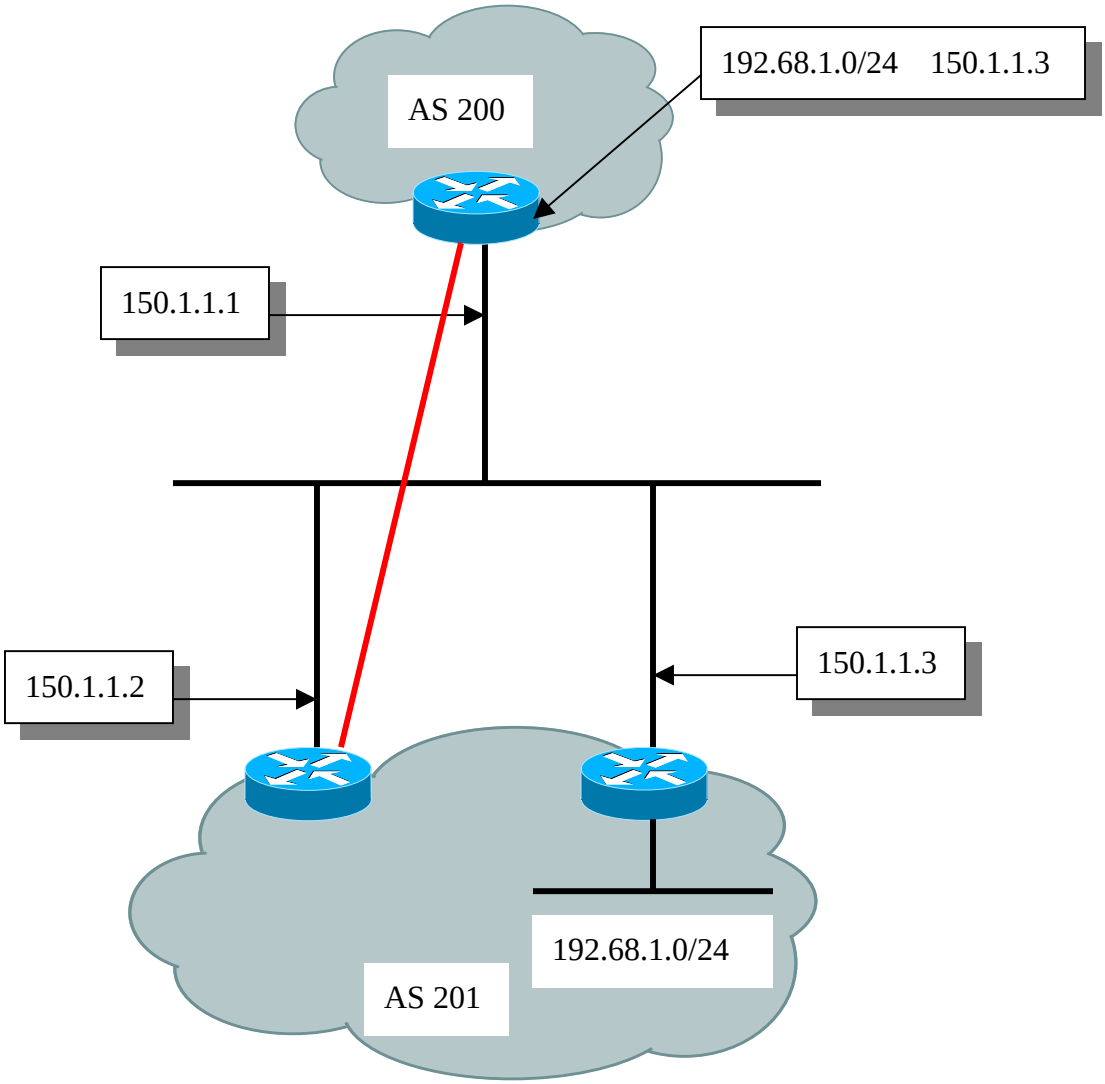
在 IGP 中到某個路徑的下一站(next hop) 指的就是這條路徑的路由器上相連介面的 IP 位址。

BGP 的下一站有下列三種形態:

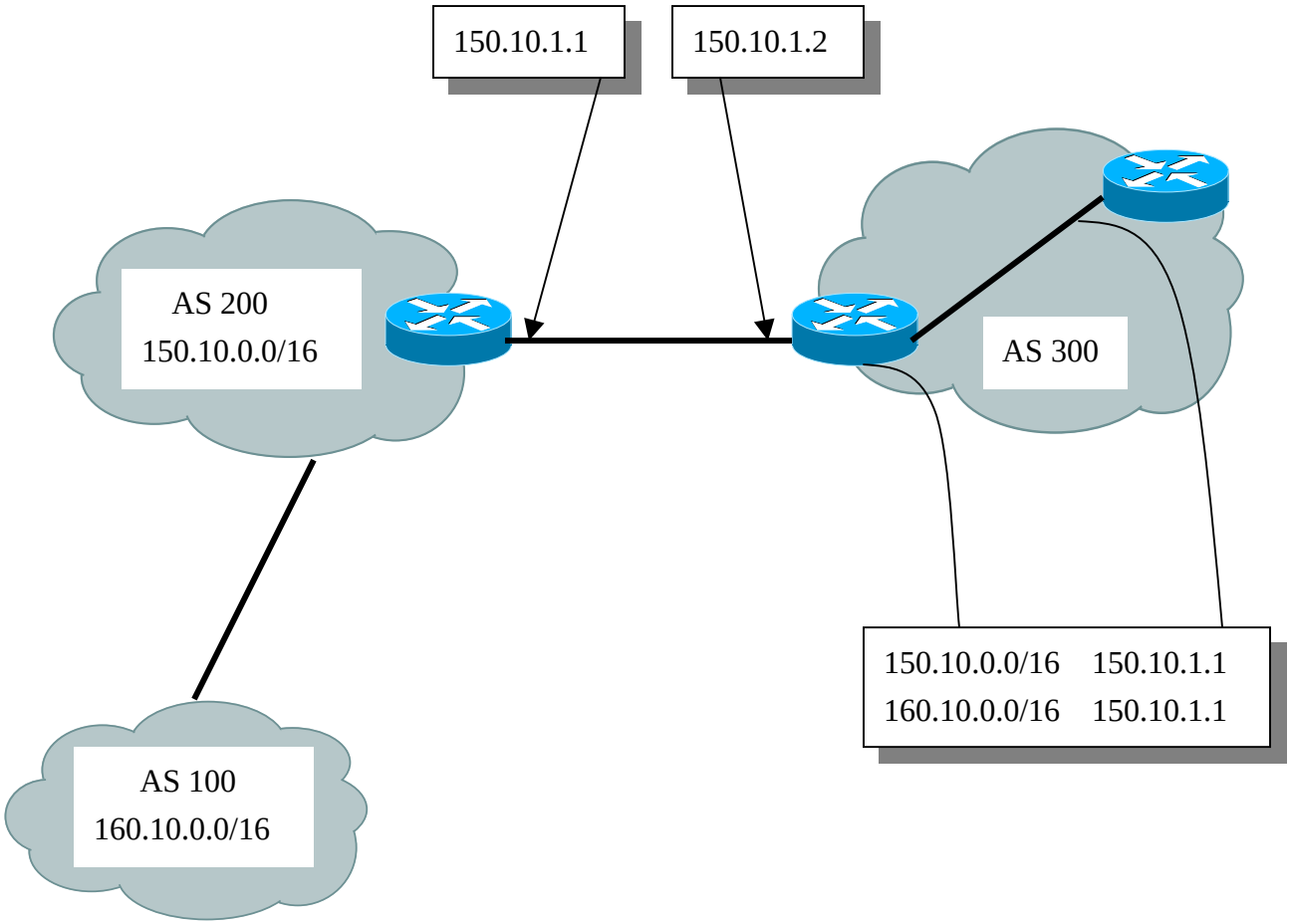
1. 對 EBGP 連線而言:下一站就是宣告這條路徑的鄰居的 IP 位址。以圖六說明該型態之 Next_hop。
2. 對 IBGP 連線而言:對從 AS 內部產生的路徑，下一站就是宣告這條路徑的鄰居的 IP 位址，以圖七說明該型態之 Next_hop。
3. 如果路徑是在一個多重存取媒介(像是Ethornet, Frame Relay)上宣告，下一站通常就是連接到這個媒介，並產生該路徑的路由器介面的 IP 位址，以圖八說明該型態之 Next_hop。



圖六 ：Next_Hop 說明範例



圖七 : Third-Party Next_Hop 說明範例



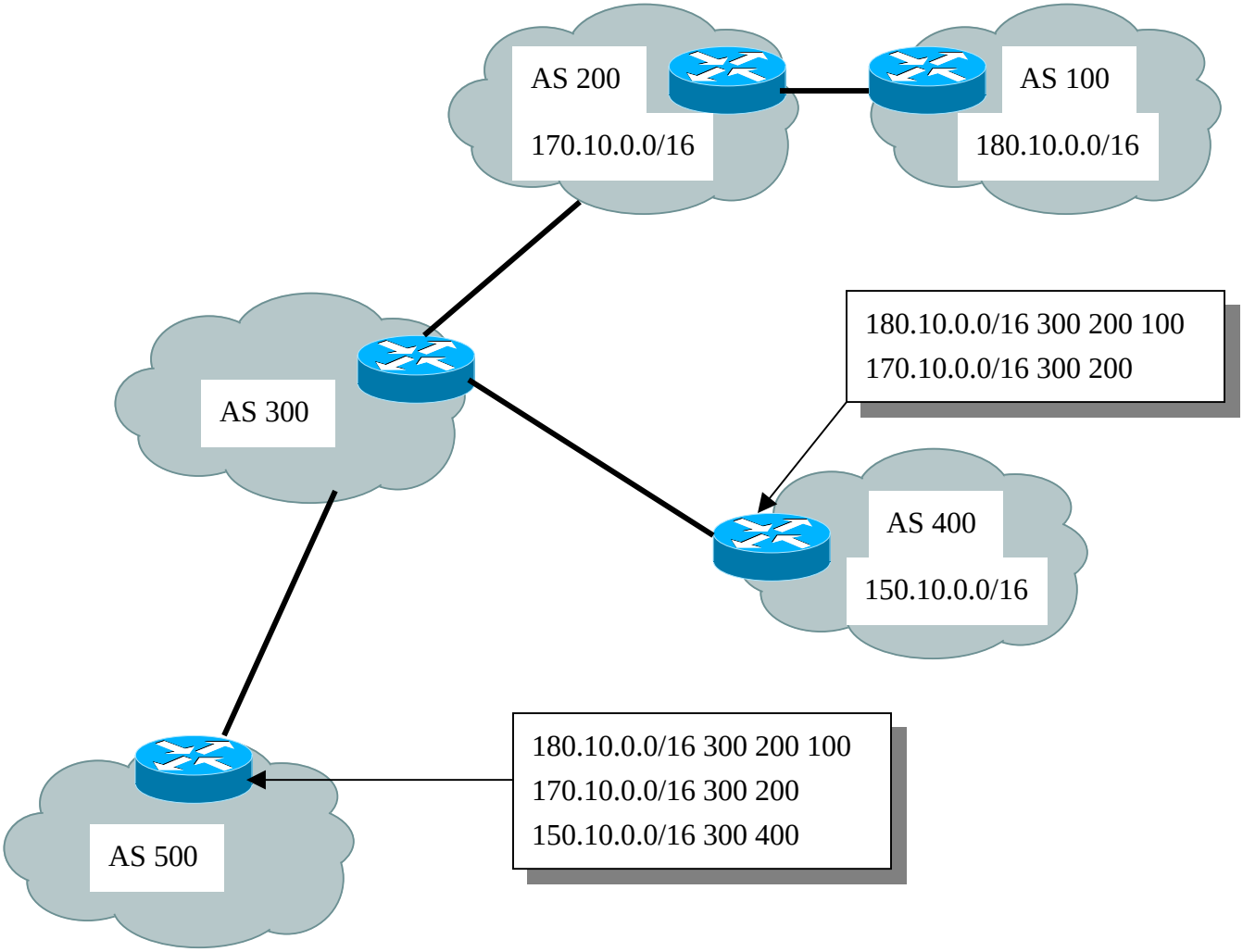
圖八 : IBGP Next_Hop 說明範例

AS_Path 屬性

AS_Path 屬性是一連串的自治區系統編號(AS Number)，也就是一條路徑到達目的地途中所經過的 AS，經過路徑的 AS 在把路徑傳給其他 BGP 同伴時，會把自己 AS 編號加在名單上，而最後這名單就表示路徑經過的所有 AS 編號，而最先發出這條路徑的 AS，它的 AS 編號排在名單最後，此名單稱為 AS-Sequence，因為所有的 AS 編號是依前後順序排列的。

BGP 利用 AS-path 屬性以確保 Internet 上的無迴圈(Loop detection)架構，另 BGP 也可依據 AS-path 屬性以決定要到目的地應該選取的最佳路徑。

圖九說明 AS_Path 屬性，RTA 收到之 180.10.0.0/16 會有 300 200 100 之 AS Sequence list, 而 RTA 收到之 150.10.0.0/16 則有 300 400 之 AS sequence list。

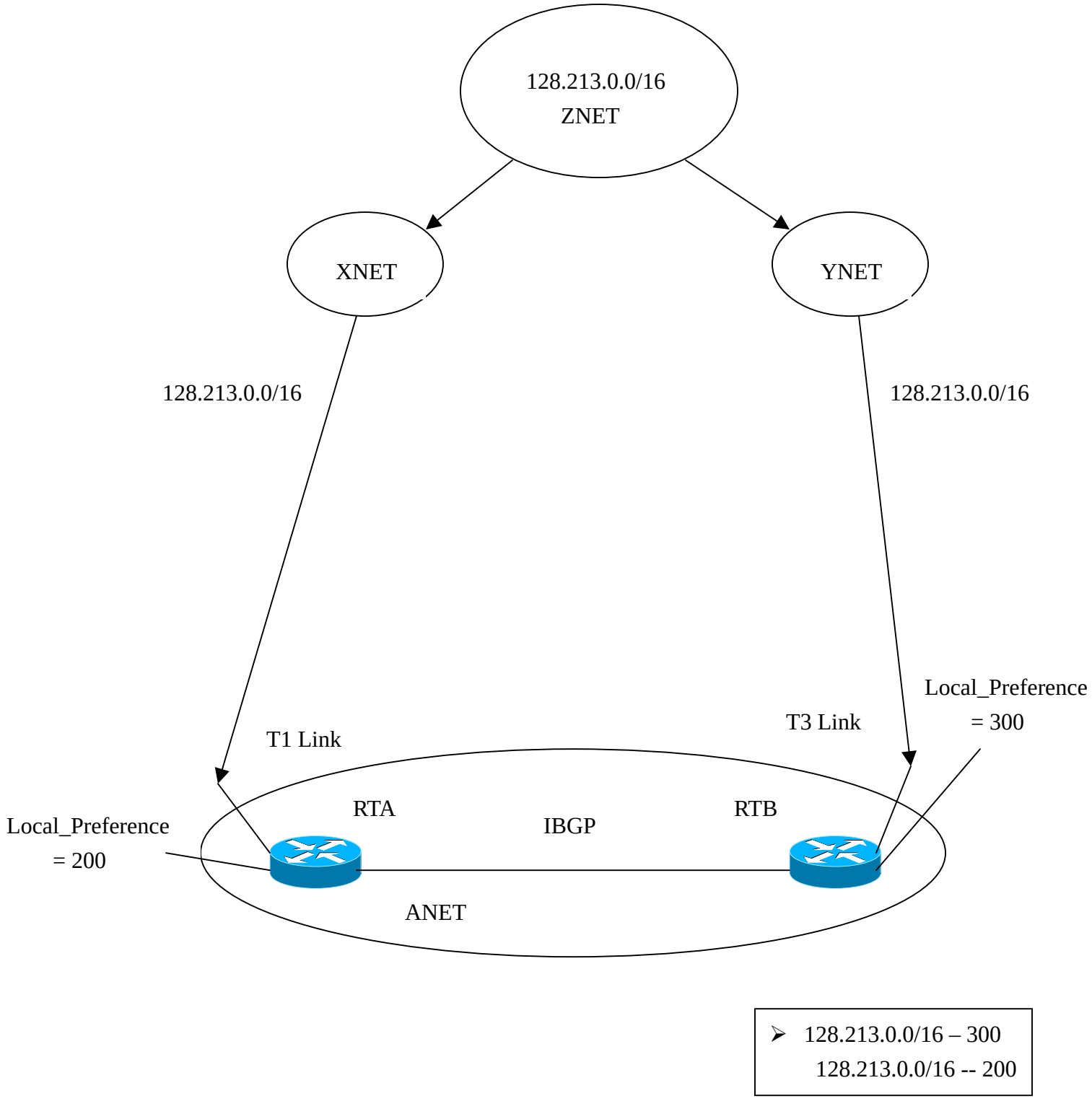


圖九 : AS Path 說明範例

Local_preference 屬性

Local_preference 是對通往相同目的地的路徑，提供偏好程度的比較，Local_proference 值較高表示比較偏好這條路徑，Local 的意思對於 AS 來說就是只在該 AS 之 IBGP peer 之間交換而不會傳給 EBGP peer。

以圖十來說明 Local preference 屬性，路由器 RTA 對來自 XNET 的路徑給予 Local_preference=200，路由器 RTB 對來自 YNET 的路徑給予 Local_preference=300，由於 RTA 與 RTB 會經由 IBGP 交換路由更新，因此會選擇使用 YNET 作為往目的地 128.213.0.0/16 的出口，因為 Local_prefence=300 比較高。

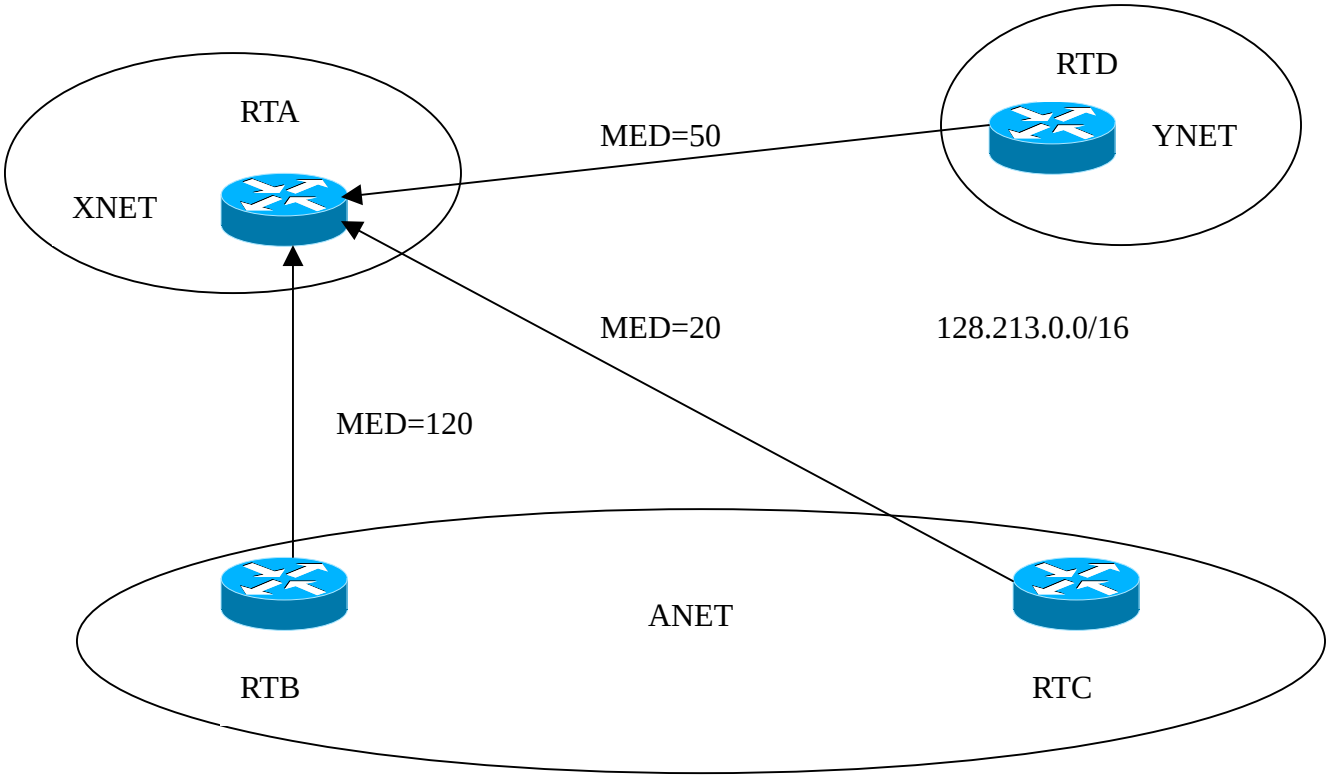


圖十：Local_Preference 屬性範例

Multi_Exit_Discriminator(MED)屬性

有多個連接點的兩 AS 間，可利用 MED 向外部 peer 暗示其偏好的路徑，MED 也就是一條路徑的外部路徑值(metric)，較低的 MED 值比較高的 MED 值受到偏好。

圖十一說明 MED 之作用，RTA 從三個不同來源收到有關 128.213.0.0/16 的路由更新:RTB(MED=120)，RTC(MED=200)與 RTD(MED=50)，RTA 會比較來自 ANET 的兩個 MED 值，並選擇 RTB，因為其宣告之 MED 值較小，如果 RTA 路由器使用了 bgp always-compare-med 設定指令，那麼它就會比較來自 RTD 的 MED，因此選擇路由 RTD 到達 128.213.0.0/16



圖十一：MED 屬性的作用範例

Community 屬性

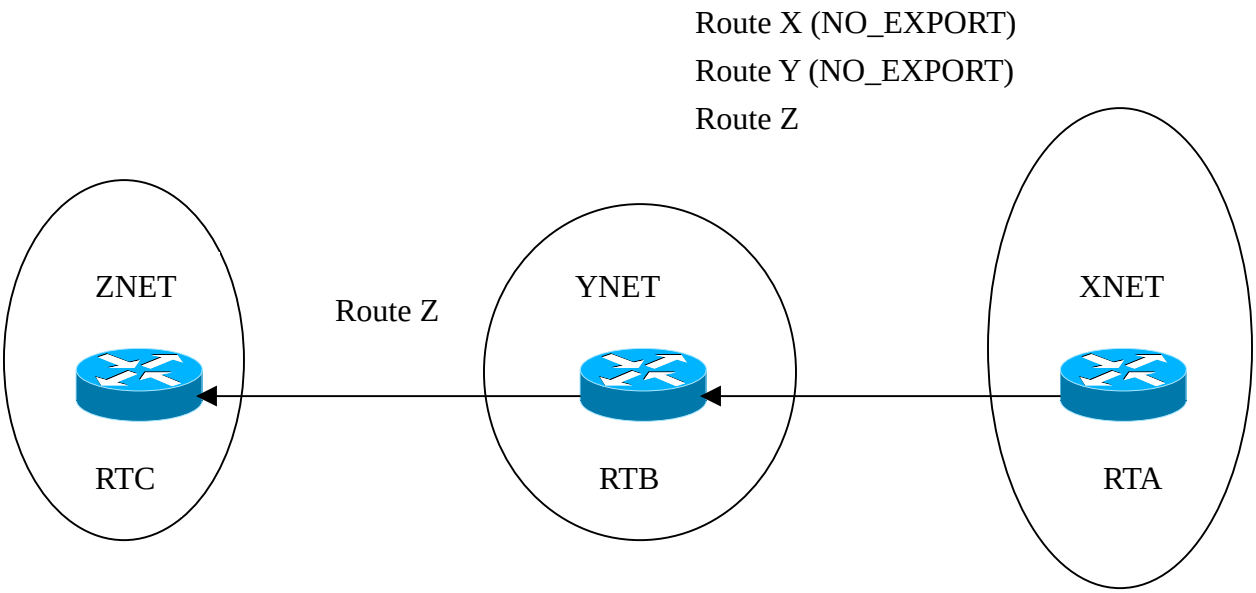
以 BGP 而言 Community 是一群分享某些共同特性的目的地。Community 可用來簡化 Route Policy，其方法是根據邏輯特性而非只有 IP prefix 或 AS 編號來辨識路徑。

Community 屬性值在 0X00000000 到 0X0000FFFF 以及 0XFFFF0000 到 0XFFFFFFFF 範圍內的 Community 屬性值是保留的，也就是說它們已定義成具有全球性共同意義的。

例如:

- NO-EXPORT(0XFFFFFF01):攜帶此 Community 值的路徑不可向 Confederation 或是同一 AS 以外的 BGP peer 宣告。
- NO-ADVERTISE(0XFFFFFF02):收到了攜帶此 Community 值的路徑不可向任何的 BGP peer 宣告。

圖十二顯示使用 Community 屬性一個範例，XNET 向 YNET 傳送 X、Y 與 Z 路徑，X 和 Y 中攜帶 Community 屬性 No-EXPORT，路徑 Z 則沒有，路由器 RTB 只會將路徑 Z 傳播給路由器 RTC，不會傳播路徑 X 與 Y，因其攜有 NO-EXPORT 屬性。



圖十二：Community 屬性的簡單應用範例

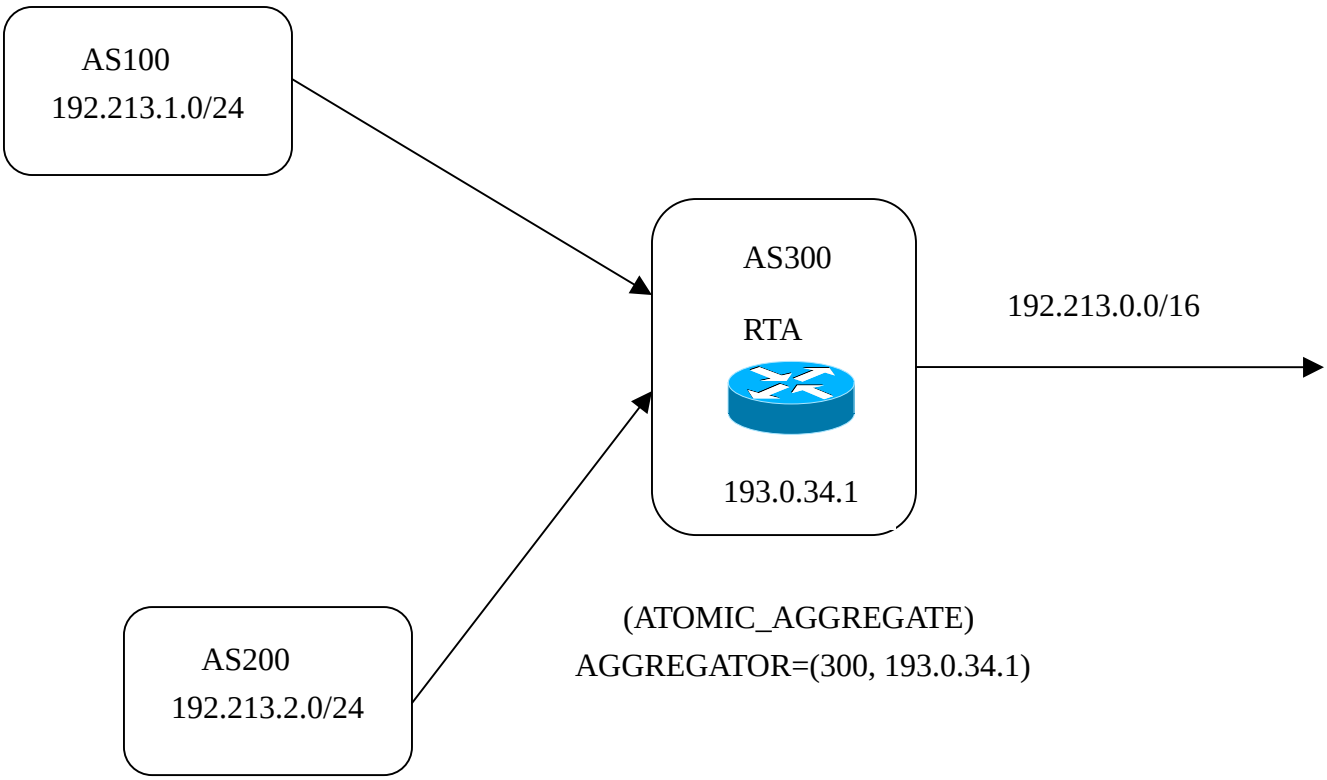
ATOMIC_AGGREGATE

由於路集合來自不同的來源，而各來源可能有不同屬性，因此可能造成資訊的漏失，因此如果系統傳播一個集合時造成了資料的漏失，就需要在此路徑裡附加 ATOMIC_AGGREGATE 屬性。

AGGREGATOR 屬性

AGGREGATOR 指出產生集合的 AS 與路由器，執行路徑整合的 BGP 發言者(Speaker)可能會加上 AGGREGATOR 屬性，包含這個發言者的 AS 編號與 IP 位址(大部份就是這個路由器的 RID).

圖十三說明 AGGREGATOR 屬性，AS300 從 AS100 及 AS200 分別收到路徑 192.213.1.0/24 及 192.213.2.0/24，當 RTA 產生整合 192.213.0.0/16 時可選擇括入 AGGREGATOR 屬性，其中包含 AS300 與產生此整合路由器 RTA 之 RID193.0.34.1。



圖十三 : AGGREGATOR 屬性範例

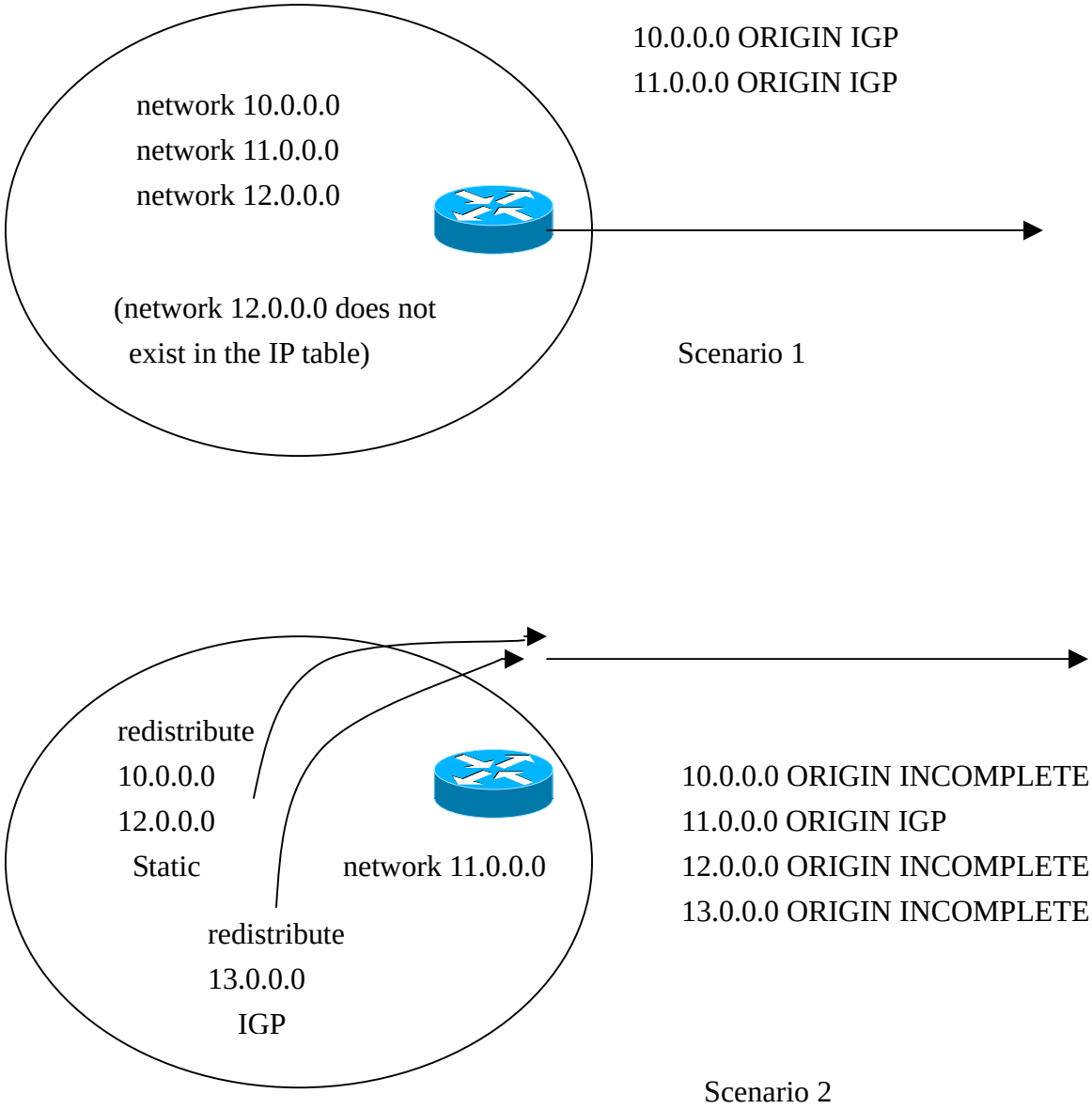
ORIGIN 屬性

藉由 network 指令或整合方式宣告的網路，BGP 認為是屬於 AS 內部的，會在每條路徑資料中加註 ORIGIN 屬性為 IGP，相反地，當某一路徑經由注入(inject)進 BGP 時(不論是靜態的或動態的)，路徑的 ORIGIN 屬性都會註記成 INCOMPLETE，因為注入的路徑可能來自任何的地方，BGP 共有三種 ORIGIN 型態:

- IGP
- EGP
- INCOMPLETE

BGP 會在路由選擇決定程序中考慮 ORIGIN 屬性，IGP 優先於 EGP，EGP 又優先於 INCOMPLETE。

圖十四說明上述情形，在狀況一中所有網路都經由 network 指令列在 BGP 程序下，BGP 將 10.0.0.0 及 12.0.0.0 是經由靜態路徑注入，而 13.0.0.0 是由 IGP 動態注入學到的，所以 11.0.0.0 會以屬性 IGP, 10.0.0.0 12.0.0.0 及 13.0.0.0 以屬性 INCOMPLETE 送出去。



圖十四 ：ORIGIN 屬性使用方式比較範例

2.5.7 BGP 路由選擇決定步驟

當有多條路徑通往相同的目的地時，BGP 會根據屬性值來作決定選擇一條最佳路徑，下列步驟列出 BGP 如何依序選擇最佳路徑：

1. 如果無法存取到 Next_hop，就忽略掉這條路徑，因此一定要有 IGP route 可到達 Next-hop 的原因。
2. 選擇 Weight 值較的路徑(Weight 是 Cisco 專有的屬性值)。
3. 如果 Weight 相同，選擇 Local_preference 屬性值最大的路徑。
4. 如果 Local_preference 屬性值相同，選擇從本地這個路由器所產生的路徑。
5. 如果 Local_preference 屬性值相同，選擇 AS_path 名單長度最短的路徑。
6. 如果 AS_path 的長度相同，選擇 ORIGIN 型態最小的路徑(IGP<EGP<INCOMPLETE)。
7. 如果 ORIGIN 型態相同，選擇 MED 屬性值最小的路徑。
8. 如果路徑的 MED 相同，依下列順序選擇路徑:EBGP 優於 Confederation，Confederaion 又優於 IBGP。
9. 如果上列條件都相同，則選擇可經由最近的 IGP Next_hop 到達的路徑，也就是說在 AS 內選擇最短的內部路徑到達目的地。
10. 如果內部路徑都相同，BGP 路由器的 RID 就是最後的決定者，選擇來自路由器 RID 較小的 BGP 路由器(路由器 RID 通常是路由器上最高的 IP 位址或是邏輯的 loopback 位址)。

三、 觀感與建議

Internet 的興起不僅創造了 ISP，ICP 等新興行業，也衍生了許許多多嶄新的商業經營模式(Business Model)，更開創了數據通信發展重要里程碑，同時世界各國電信自由化及市場開放，新業者使用 IP 新技術建構之網路更具彈性及可提供多樣化服務加入競爭，做得傳統電信業者感受極大的威脅。

IP 封包化技術演變至今已為所有通信業者，設備製造商公認為下一代電信網路的標準，我們正面臨這整個產業技術變革的時代，極需有所適切因應措施，以免落伍於世界潮流之後更有甚者形成競爭劣勢。

以下就 IP 網路之特性分析，茲建議數點以供參考：

維運集中化—IP 網路設備由於具有共通的底層通信標準，可提供所有相關設備整合網管之良好基礎，因此維運體系應配合此項特點作單純化之配置，以期有效降低維運成本，提昇市場競爭優勢。

服務全球化—IP 網路使用世界共通的 TCP/IP 通信標準，所以具有服務全球化特性，也就是說世界各國的新、舊業者均可輕易地利用其作全球性之服務，因此任何服務提供之市場觀點均需以全球性作考量。

競爭國際化—使用 IP 之新一代電信服務，無疑地使電信服務競爭國際化，本土市場之絕對優勢需加以發揮形成區域競爭優勢，更需思考如何運用策略聯盟，併購等手法增加國際競爭力，其中培育或招募熟悉 IP 技術，涉外及法務人才應為當務之急。