

具有動態差別群播服務等級之集中式控制管理系統架構探討

文志超 古文彬 陳景章

國立中正大學電機工程研究所

E-mail: {ccwen, ieejkc}@cn.ee.ccu.edu.tw

摘要

此篇論文提出一個在 DiffServ / MPLS 網路環境中具有動態差別群播路由服務架構，有效控制管理群播網路，並提供訊息同時傳送接收給多個群組接收端。現今的群播系統大多屬於分散式群組控制演算法，需同時參與封包傳送。群播控制機制包括路由計算、資源保留、還有群組管理。一般 IP 群播路由協定會隨著群組加入數量增加與使用者 QoS 需求不一致，造成資源分配不足與加入失敗的問題。本文的目標在於將群播封包經過 DiffServ 分類，加上動態差別等級，並採用集中式群播控管機制，將封包傳送功能從群播機制中分離。MPLS 群播封包傳送機制能夠建立點對點以及點對多點標籤傳送路徑(LSP)，而集中式管理不僅可保證 QoS，更可以對於群組和網路資源做有效管理。我們將探討此架構之運作，並提出分析評估。

關鍵詞：群播路由服務、DiffServ、MPLS、集中式管理控制

Abstract

In this paper we propose a Differentiated Service (DS) class supported Multicast architecture with Centralized Control and Management (DS-MCCM). With consideration of the concept of centralized multicasting, this architecture can enhance interactive multicast control and management capabilities for receiver-initiated heterogeneous QoS requirements, multicast routing and resource reservation, and dynamic join/leave of group management. Since current IP multicast suffers from scalability and QoS guarantee as the number of multicast groups increases, a router requires maintaining multicast forwarding state of multicast tree for all group members. Our proposed architecture DS-MCCM based on IP PIM-SM multicast routing tree can support flexible group management and dynamic resource management in DiffServ/MPLS network. Finally, the proposed system architecture with centralized control and management operations are evaluated.

Keywords: Multicast routing service、DiffServ、MPLS、Centralized control and management.

1. 簡介

近年來，由於網路的越來越普遍以及網路技術的越來越成熟，以至越來越多的群組應用 (Group Applications) 得以在網路上實現並且流行。其中視訊會議、遠距教學等即是常見的點對多點 (p2mp) 或是多點對多點 (mp2mp) 的群播 (Multicast) 應用。這些點或多點對多點的群播應用必須要有適合的群播路由協定來支援，再加上有效的網路資源管理與網路傳輸品質的協助；因此使用符合網路資源有效運用目的的群播路由協定，對於提供點對點或是多點對多點的群播應用服務或是新世代網路基礎建設，都是發展的重要議題[1]。

群播路由是以群組 (Group) 為傳送的目標，一旦加入同一群組的使用者逐漸增多，負責傳送資料給此群組的路由器所需記錄的狀態便會增多，如此便容易增加路由器的負擔，因此會造成 Scalability 的問題。此外，因為群播路由的群組為開放性群組 (Open Group)，使用者可任意做狀態上的變動，使得整個網路上充斥著加入與離開群播樹需求的控制訊息，而造成網路資源的浪費和壅塞。本文使用了集中式控管系統，利用 PIM-SM 群播路由計算 [2]，產生的分享樹來收集群播需求和來源樹來傳送資料的互動機制，使得信令 (Signaling) 控制和資料傳送 (Data Forwarding) 的訊息獨立運作，並且利用集中式控管機制來平衡整個網路資源的使用情形[3][4]。

本文的內容組織，可區分如下：第 2 節先介紹集中式品質策略之管控模型；第 3 節提出動態差別等級群播服務架構，說明其機制及運作；第 4 節探討這個整合架構上下層協定互動關係；最後，提出分析與評估。

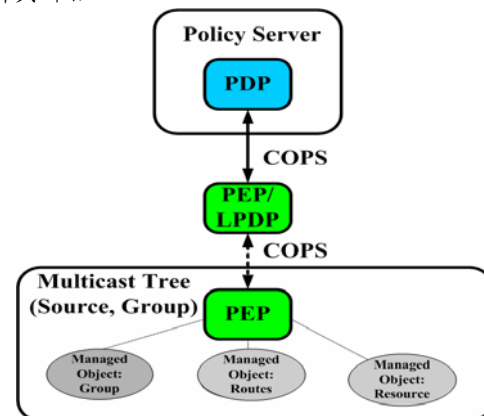


圖 1: COPS 群播策略管理控制模型

2. 集中式品質策略與管理控制模型

網路品質策略模型[5][9][10]與管理控制架構的相關研究[4][6][11]，一直是網路品質服務的重要議題。而品質策略制定，則為網路業者提供不同服務品質保證的參考依據(service profile)。IETF(Internet Engineering Task Force)組織針對 IP 網路服務品質管理，制定許多相關協定標準[7][8][9][10]。其中 Common Open Policy Service (COPS)，主要是利用伺服器 (PDP: Policy Decision Point) 和其客戶端 (PEP: Policy Enforcement Point) 之間的要求 (Request)/回覆 (Decision) 的溝通機制來對於網路設備與服務作統籌性控管，藉著設計與使用運作策略(Policy)於網際網路，使得網管系統知道該如何進行網路元件的管控，被管理網路設備該如何處理不同的封包。Policy 大致可分為傳輸品質(QoS, Quality of Service)與加密(Security)等兩大類，負責傳送 Policy 可藉由 SNMP、COPS、HTTP 等任一通信協定來完成之。但本文的需求為傳送 Policy 所專門設計的 COPS，以提供中央管理者運作管道。圖 1 所示在受控管網域內只需有一個 Policy Server，藉由 NIS (Network Information System) 控管所有路由器，而受控管群播路由器 (匯集節點:RP /傳送節點:TPs) 也可使用 COPS 將網路狀況資料回報給管理者知道，以利檢視 Policy 執行效果，如此一來就可以此作為改進 Policies、Resource、Multicast routing 的依據，使網路運作更合理順暢。再者 COPS 為 Stateful 通訊協定，所以 PDP 和 PEP 均會瞭解要求/回覆的決策指令的狀態，避免若決策指令有時間先後關係時執行發生順序錯亂而造成執行失敗的情形發生。

在 COPS 中 PDP 負責下決策指令和存取一些需設定的網路參數，而 PEP 負責執行決策指令和回報指令執行狀況及網路狀況。而由圖 2 所示，PDP 通常是扮演管理者的角色透過通訊協定和相關網管伺服器做訊息交換，隨時掌握整體網路狀況；PEP 則是一般受控管的網路設備負責執行和回報。PEP 和 PDP 兩者間互動就靠 COPS 進行溝通，PEP 中的 LPDP (Local Policy Decision Point) 為在 PEP 中存取 PDP 所下過的決策，以避免 COPS 連線突然中斷造成服務中斷。而 PDP 和 PEP 之間 COPS 連線的互動關係，PEP 先對 PDP 提出需求 (Request)，接著 PDP 發決策指令 (Decision) 給 PEP，之後 PDP 會將決策指令執行結果回傳給 PDP。COPS 協定管控群播連線，主要有下列幾個特點：

1. 在 TCP 連線上，建立可靠的 PDP 決策指令傳輸。
2. PEP 和 LPDP 可利用 COPS 中的 Keep-Alive 訊息來確保連線以及對方的存在。
3. 當 PEP 和 PDP 在中斷後重新連線時，可利用 COPS 中的同步 (Synchronization) 訊息來同步彼此的狀態。
4. PDP 可以對已執行之決策，透過 COPS 網路狀況回報要求重新調整修改。
5. PEP 可以對 PDP 報告決策執行情況。

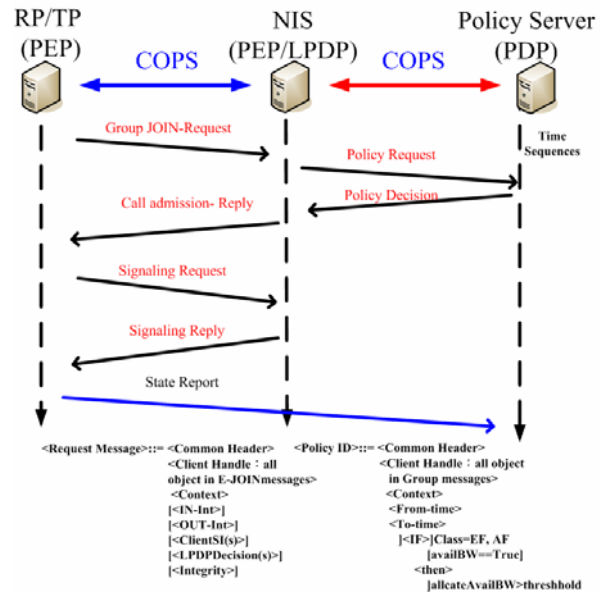


圖 2: COPS 與 PDP/PEP 協定語法

3. 動態差別等級群播服務架構

隨著群組加入增加與使用者 QoS 需求不一致，造成資源分配不足與加入失敗率提高的問題。我們提出具有動態差別群播路由服務架構，目標在於將群播封包分類，加上動態差別等級，並採用集中式群播控管機制，將封包傳送功能從群播機制中分離。針對整體控管功能與流程，在以下小節說明。

3.1 可調式差別等級群播服務控制機制

圖 3 是我們提出具有差別等級群播服務之網路架構，其中 COPS 的應用主要是在 NIS 和 RP 間作溝通群播樹、路由訊息和監控網路狀態的介面。在此服務架構中，COPS 協定中 PDP 由 Policy Server 扮演決策者的角色，PEP 由 RP/TP 扮演執行者的角色，而 NIS 則做為兩者中介協調的角色。在上述小節中提到 COPS 主要會影響群播樹建置和群組管理兩個層面：

I. 群播樹建置：

由圖 3 可知 COPS 在群組建置時所扮演著 NIS 和群播網路節點 RP/TP 間溝通者的角色，接著將說明 NIS 和 RP/TPs 利用 COPS 的互動情形。

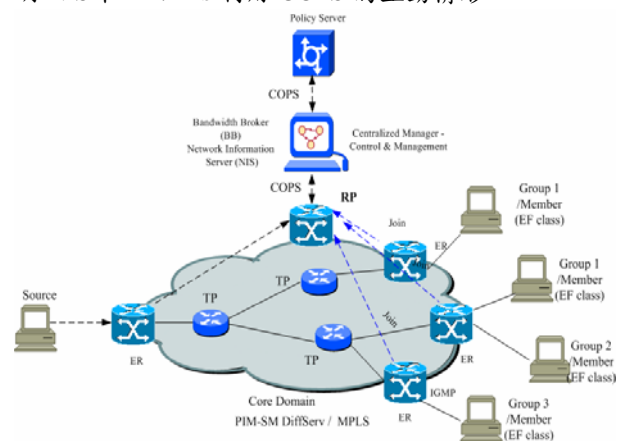


圖 3: 動態差別群播服務等級之網路架構

1. 當 RP 收到 TPs 的群播需求時，RP 會透過 COPS 對 NIS 發出請求訊息並將 TPs 群播需求包在其中，以方便 NIS 和 BB 作群播樹計算。
2. 當 NIS 和 BB 計算完群播樹後，會透過 COPS 對 RP 發允入 (Admission) 訊息並將群播樹和 LSP 訊息包在其中，以方便 RP 透過共享樹通知 TPs 建置來源樹。
3. 當 RP 接收到允入接受訊息後，會透過 COPS 對 NIS 發信令要求 (Signaling Request) 訊息，NIS 將路由訊息傳給 RP，TPs 則透過回報 (Report) 訊息將網路使用情形和群組狀態回報給 NIS，以方便 NIS 作群組管理。
4. 而當 RP 收到 TPs 解除群組需求時，RP 會透過 COPS 對 NIS 發出解除需求 (Termination Request) 訊息並將 TPs 解除群組需求包裝在其中，以方便 NIS 和 BB 作群組解散和網路資源的重組。
5. 當 NIS 和 BB 完成群組狀態消除和網路資源重組時，會透過 COPS 對 RP 發解除訊息並將需解散的群組和路徑訊息包裝在其中，以方便 RP 透過共享樹通知 TPs 來解散群組和釋放網路資源。

II. 群組管理：

COPS 在群組建置時所扮演著 NIS 和 RP 間溝通者的角色，群播服務因為網路環境變動和群組變動而造成改變，在網路變動時往往必須重建群播樹；而在群組變動時有可能需重新建置群播樹（使用者加入/離開時）。但新群組加入也有可能發生資源競爭與分配保留衝突的問題，服務等級需動態調整，正是本文提出集中式管控架構針對資源管理的解決方案[10]。

就上述的情形而言，動態調整差別服務等級，可以將 COPS 在群組管理的應用上分為 NIS 被動應變調整和 NIS 主動應變調整兩類。

(1) NIS 被動應變等級調整 (Passive class remarking)：

NIS 被動應變主要是針對網路環境變動和群組變動時所造成的群播樹變動而採取的應變模式。首先 TPs 會定期回報群播樹狀態和網路使用情形給 RP，RP 會透過 COPS 回報訊息將收集來的群播樹狀態和網路使用情形送給 NIS。一旦 TPs 回報訊息中帶有網路環境變動和群組變動等異常活動時，NIS 會重新計算/調整群播樹，BB 調整新加入群組的等級並透過 COPS 將新的群播訊息傳送給 TPs。

(2) NIS 主動應變 (Active class remarking)：

NIS 主動應變主要是針對新群組 (S,G) 加入和 Policy 要求改變時，所造成的不同群播樹之間資源分配衝突問題而採取的應變模式。當新群組增加時，NIS 和 BB 在評估過新使用者需求後會透過 COPS 主動發出要求 Policy 變更，來調整群播樹和群組服務等級，並透過 RP 傳送訊息給 TPs。

3.2 系統運作流程

群播路由為點對多點或多點對多點的傳輸服

務，因此群播路由的首要步驟就是先將具有相同需求的使用者組成群組，而在本文架構中是以共享樹做為蒐集群播需求的動作，系統整個處理流程如圖 4 所示。傳送端路由器和接收端路由器會先將其群播需求傳送給 RP，但和一般共享樹群播路由需求有所不同，傳送端路由器和接收端路會將 QoS 的需求包裝在群播需求訊息中一併傳送給 RP，如此可以完全滿足每一個使用者需求而非只是單純以傳送端需求為考量點，之後 RP 再將其收到之使用者訊息傳送給 NIS 和 BB 做計算具 QoS 群播樹和允入控制使用。

在 NIS 和 BB 收到使用者需求後，會根據當時網路使用狀況和使用者需求來算群播樹。在 NIS 計算完來源樹後會再將來源樹訊息回傳給 RP，之後 RP 再將來源樹訊息傳送給需轉換之 TPs。在計算過程中，NIS 負責計算來源樹並且也會將底層對應之 MPLS LSP 路徑計算完成，並且會對以計算好的路徑作 TE (Traffic Engineering)；BB 負責對使用者作允入控制，並協助 NIS 作 TE 的工作；RP 負責當作 NIS 和 RPs 之間的溝通媒介。圖 4 執行的程序，如下說明。

1. TPs 會將其群組和 QoS 需求傳送給 RP。
2. RP 再將其收集到的群播訊息傳送給 NIS 和 BB。
3. NIS 會根據使用者需求、OSPF-TE Database 和 Multicast Database 來計算群播樹和執行 Traffic Engineering。
 - I. OSPF-TE Database 中存取的為經由 OSPF-TE 所蒐集而來的網路資源使用資訊。
 - II. Multicast Database 中存取的為使用中的群播樹使用資訊。
4. NIS 會和 BB 合作對使用者做允入控 (Admission Control) 以避免網路資源消耗殆盡和違規使用者佔據網路資源。
5. NIS 會將計算好的群播樹訊息回傳給 RP。
6. RP 會將群播樹訊息傳送給需建置來源樹 Transit Points。此訊息包含了 L3 群播樹上下游路由傳送關係和建立 L2 P2P/P2MP LSP 所需傳送訊息。

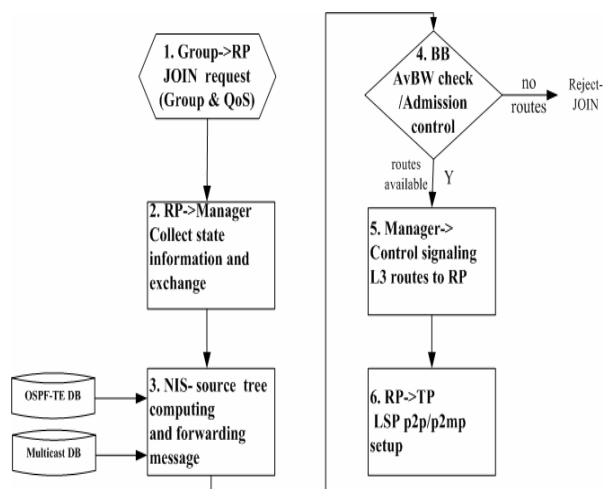


圖 4：系統運作流程

4. 上下層協定互動運作

本文所提架構與運作流程，可以依照 OSI (Open System Interconnection) 模型[11][12]，將圖 3 架構分為三個上下層面：L4 群播服務管理層面、L3 群播路由控制層面和 L2 群組資料傳輸層面，如圖 5 所示。L4 管理層面主要是利用 COPS 來作為溝通介面，而在管理層面中需要根據蒐集的資訊做為決策的判斷，因此在管理層面中需要聯結相關網管伺服器或資料庫以取得完整資訊，而在本文中設定了決策資料庫 (Policy Database)、認證資料伺服器 (LDAP: Authentication server) 和 SNMP (Simple Network Management Protocol) 伺服器。決策資料庫存取著關於系統管理者對於網路使用策略的設定參數，認證資料伺服器主要是對使用者作認證服務而 SNMP 伺服器主要是存取著關於整體網路的運作資訊。L3 群播路由層面主要是利用共享樹和群播樹的互動來完成，而本文架構中群播樹均是由 NIS 計算，因此 NIS 需要對整體網路上群播樹的使用情形和網路資源狀態有完整的瞭解，因此在群播路由層面需要群播樹資料庫 (Multicast Tree Database) 和 OSPF-TE 網路狀態資料庫 (OSPF-TE Database) 來輔助 NIS 計算群播樹。群播樹資料庫主要是存取著還在使用中群播樹的狀態，而 OSPF-TE 資料庫主要是存取著關於網路資源使用狀態資訊。此外，NIS 也會和 BB 合作對使用者作允入控制。L2 資料傳輸層面主要是利用 P2P 或 P2MP LSP 來完成。根據圖 5 中的 COPS 運用，使得上下層間路由協定之互動分為兩種互動模式，其一為 NIS 和 RP 兩者間 COPS 應用之上下層間路由協定之互動模式；而另一為 NIS、RP 和 TPs 三者間 COPS 應用之上下層間路由協定之互動模式。其中，在 NIS、RP 和 Transit Points 間 COPS 應用之中 COPS 可以直接傳送 LSP 相關資訊給 TPs 來建立底層 LSP 傳輸路徑，並不需要 RSVP-TE 來輔助 MPLS 建立 P2P/P2MP LSP，節省處理與傳遞延遲的時間。

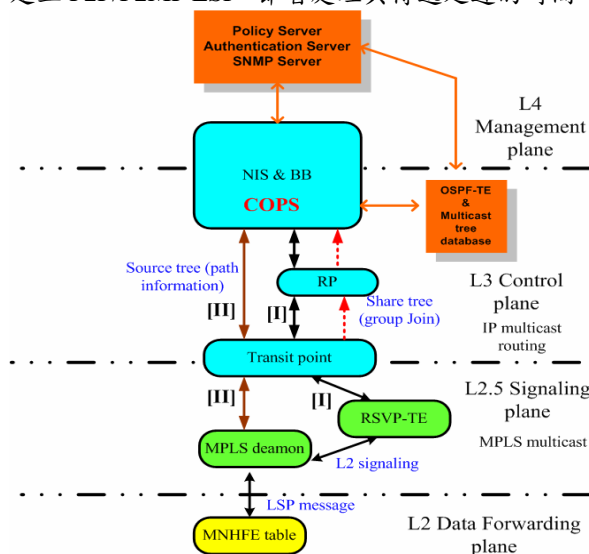


圖 5: 群播控管上下層協定互動

5. 集中式系統評估與探討

本文架構屬於集中式管理模式而使得 NIS 的工作量過大而影響系統運作，因此根據群播路由動作來平衡 NIS 負載，並且評估上述兩種群播路由服務模式，探討交互運用及效能。如圖 5，將以 [I] 代表原先 NIS、RP 和 TP 間 COPS 應用服務模式，[II] 代表改進的 NIS 和 TPs 間 COPS 群播服務模式做比較和評估。

5.1 群播路由服務評估

因群播樹之路由由計算大部分均應用於大量且及時性的資料，使得信令時間變為需要考量的要素。而基於系統時限和系統成本的考量，系統複雜度 (Complexity) 也是必須列入系統評估的考量。因此在群播樹計算層面，本文主要是在信令時間和系統複雜度來比較 [I]、[II] 兩種服務模式。

— 信令時間 (Tree signaling time)

因本文採取集中式管理控制，所以系統在建置群播樹時之信令時間就比一般分散式群播路由協定長。由圖 5 可看出在 L3/IP 與 L2.5/MPLS 建置群播樹時，模式 [I] 多出 T_{RP} 與 $T_{RSVP-TE}$ 的訊息處理時間；而 [II] 在建置群播樹時 NIS 以 COPS 直接程序交換訊息至 TP 與 MPLS，因此可以推估出 [II] 的信令時間 $T_{Sig[II]}$ 較 [I] $T_{Sig[I]}$ 短， $T_{Sig[I]} > T_{Sig[II]}$ 。

— 群組管理維護 (Reaction vs. Overhead)

群播路由中使用者可以自由進出任何群組，使得群組的狀態隨時在改變；群播服務因為網路環境變動造成改變。再加上群播路由大多應用於傳輸大量且即時性資料，所以在群組管理中對於群組變動的反應時間 (Reaction Time) 變得相對的重要。在模式 [I] 或 [II] 服務模式中，對於群組管理皆是以定期回報的方式做為掌握群組資訊的狀態。不論是網路狀態回報和群組管理訊息從 Transit Points 傳送至 NIS 過程中，模式 [I] 必須經過 share tree 與 RP，反應時間變慢，但回報訊息 (*,G) 可在 RP 整合與過濾後送出，可降低回報訊息量 (overhead)。模式 [II] 之 source tree 網路狀態 (S,G) 回報則不需經過 RP，反應時間快但每個群播樹都需回報的訊息量會增大。由上述群組管理訊息的傳遞過程中，可知群組管理的反應時間與回報訊息量存在 trade-off 的關係。

— 資料傳輸控制效能: (Control Efficiency)

不論是在 [I] 或 [II] 服務架構中，資料的傳輸均是以 P2P 或 P2MP LSP 做為底層群播樹的資料傳輸 (Data forwarding)。單純從具備 L2 MPLS 傳輸能力比較，[I] 或 [II] 服務模式並無直接關係。然而在 L2.5 的 P2P 和 P2MP LSP 傳輸控制中，[I] 採用 RSVP-TE 計算路徑標籤，[II] 直接利用 NIS 同時計算路由與相對應的路徑標籤。就效能而言，可知 [II] 較優於 [I]。但 [I] 的傳輸路徑具有 TE 的考量，較能滿足實際 QoS。

— 系統複雜度:

在本文集中式架構中 [I] 和 [II] 兩種服務模式，支

接之管理與控制協定 COPS 及 PIM/SM 隨運作範圍的不同，對於時間與空間的複雜度影響而不同。假設在一個網路中，節點總數為 N ，來源端數目為 S ，群組數目為 M 且 $M \leq N$ 。在[I]和[II]中 NIS、RP 和 Transit Points 所支援的管控通訊協定不同。由表 1 中可得知[I]和[II]中 RP 和 Transit Points 都支援的 PIM/SM，因此其時間複雜度皆相同。其中[I]的 Transit Points 不支援 COPS，RP 與 NIS 將共同分擔保留系統狀態訊息。而[II]的 Transit Points 支援 COPS，故可直接傳送狀態訊息給 NIS，則需較多記憶儲存系統狀態訊息，空間複雜度較高。

表 1 [I]、[II]服務模式需支援通訊協定處理比較

Network Nodes		Supported Protocol	Multicast: PIM/SM *	Time complexity	COPS **	State
服務模式 [I]	Transit points		*			
	RP		*	$O(M)+O(N)$	**	$S+M$
	NIS			$O(SM)$	**	$S+M$
服務模式 [II]	Transit points		*		**	
	RP		*	$O(M)+O(N)$	**	
	NIS			$O(SM)$	**	SN

5.2 系統負載平衡評估

集中式管理最大的缺點為中央管理者的負荷過大。而在本文架構中，RP 利用共享樹負責蒐集群播訊息，MPLS 來源樹負責資料傳送以及 NIS 負責群播樹計算和群組管理。因此 NIS 的負荷會造成明顯的瓶頸(bottleneck)，而 RP 只是擔任群組資訊交換的代理(Agent)。若是根據 MOSPF 劃分群播路由區域(Domain)與事件屬性，來重新分配區域內部(Intra-domain) NIS 和 RP 的工作性質，區域外部(Inter-domain) 由 Policy Manager 協調跨區群播連線，則可達成系統負載的平衡。此外，由上述的比較中可以看出[I]和[II]服務模式各有優缺點，所以有效利用[I]和[II]服務模式交互運用來減輕系統負擔和將可增進系統效率。

6. 結論

這篇論文最主要的目的是在提出一個在 MPLS 網路環境之具有 DiffServ 群播路由服務的控制管理架構，以因應未來網路應用於多方互動的視訊會議和 QoS 的需求的多媒體群播服務。集中式的架構主要以 QoS 考量，來源樹和共享樹互動架構主要是來自 PIM-SM 的切換模式，不但可以擷取來源樹和共享樹的優點，並且也可以保持 multicast 原有的基礎來改良。而以集中式管理來計算和管理群播樹，IP Multicast 機制可以在 DiffServ/MPLS 環境上實現，使用 P2P/P2MP LSP 來負責傳送群播資料。主要是解決底層網路環境的改變和不同 QoS 的需求。

本文提出集中式群播管理架構的優點，包括將控制與資料傳輸分開處理，可減輕群播路由器的負

擔。在群組管理方面則利用 RP，透過分享樹掌握所有群組的動態訊息。定期回報的方式，NIS 根據狀況判斷運作的應變機制，以針對網路環境改變和使用者變動等不同狀況做調整服務等級的應變措施。未來可以繼續探討將 COPS 做更進階的應用在階層式 NIS 間的互動，達到跨區整合式的群播服務。

參考文獻

- [1] A. Striegel and G. Manimaran, "A survey of QoS Multicasting Issues," IEEE communications Magazine, June 2002.
- [2] D. Estrin, D. Farinacci, A. Helmy, D. Thaler, S. Deering, and so on, "Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification," RFC 2362, June 1998.
- [3] R. Neilson et al., "A Discussion of Bandwidth Broker Requirements for Internet2 Qbone Deployment," Internet2 Qbone Advisory Council, Aug. 1999.
- [4] S. Keshav and S. Paul, "Centralized Milticast," IEEE/ACM Trans. Net., Apr. 1998.
- [5] M. Sloman, "Policy Driven Management For Distributed Services," J. Net. and Sys. Mgmt., vol. 2, no. 4, Dec 1994.
- [6] A. Boudani and B. Cousin, "A New Approach To Construct Multicast Trees in MPLS Networks," IEEE ISCC'02.
- [7] D. Durham, J. Boyle, R. Cohen, S. Herzog, R. Rajan, and A. Sastry, "The COPS (Common Open Policy Service) Protocol," RFC2748, January 2000.
- [8] S. Biswas, R. Izmailov, B. Rajagopalan, "A QoS-Aware Routing Framework for PIM-SM Based IP-Multicast," <draft-biswas-pim-sm-qos-01.txt>, January, 2005.
- [9] S. Blake et al., "An Architecture for Differentiated Services," IETF RFC-2475, Dec. 1998.
- [10] R. Bless, K. Wehrle, "IP Multicast in Differentiated Service (DS) Networks," RFC 3754, April 2004.
- [11] P. Trimintzios et al., "A Management and Control Architecture for Providing IP Differentiated Services in MPLS-based Networks," IEEE Commun., Special Issue in IP Operations and Management, vol. 39, no. 5, IEEE, May 2001.
- [12] D. Ooms, B. Sales, W. Livens, A. Acharya, and so on, "Overview of IP Multicast in a Multi-Protocol Label Switching (MPLS) Environment," RFC 3353, August 2002.