於階層式疊層網路下具延遲上限應用層群播之 P2P 會議服務

蘇暉凱*[a]、潘建廷[b]、林哲价[b]、陳景章[b]、楊明達[c]
國立虎尾科技大學電機工程系[a]
國立中正大學通訊工程研究所[b]
工業技術研究院[c]

摘要 — P2P-SIP 多媒體會議是透過網路中的使用者彼此分享會議資源，改善傳統集中式的會議模型架構下負載集中、單點失效和昂貴的基礎建設成本等問題。有別於傳統集中式的架構，Peer-to-Peer 的環境容易造成疊層路徑與實體路徑的落差問題，因此如何取得底層資訊並有效地建立會議的應用層群播樹就成了 P2P-SIP 多媒體會議的關鍵。[1]

一、 簡介

在傳統集中式的多媒體會議模型架構下，有硬體建置成本高、擴展性低和單點失敗等問題。隨著多媒體會議的需求日益增長，上述問題將會對服務業者造成龐大的基礎設施建置成本。

因此近期很多研究紛紛朝向 Peer-to-Peer 的環境發展，在分散式環境中，使用者可以分享彼此的會議資源以及相關訊息，進而降低伺服器的建置成本。但在此環境中，由於 Peer-to-Peer 形成的疊層網路與實際的 IP 網路路由之間有落差，這會造成查詢或會議進行時的不必要路由；另外，相比於傳統集中式的會議服務，分散式環境中的會議系統因為缺乏網路底層的資訊，因此無法建立適用於即時多媒體會議的應用層群播樹。

二、 背景介紹

由於多媒體會議相關研究越來越被重視，IETF 因此成立了許多 Working Group，進行多媒體會議標準發展以及相關議題討論，目前也制定出許多 RFC 文件如中 RFC 4353[1]、RFC 4579[2]、RFC 5850[3]。

P2P-SIP 則是融合 Peer-to-Peer 分散式結構與傳統集中式 SIP 的 Client-Server 結構，利用疊層網路為 SIP 提供完全分散式的資源定位與訊息傳輸服務，具有去集中化、低成本、高容錯性與高可擴充性等優點，改善集中 SIP 系統所需的龐大維護管理開銷，避免單點故障與效能瓶頸等問題。目前 P2P-SIP 主要採用 SIP-over-P2P 的架構，即利用 Peer-to-Peer 協定為 SIP 協定提供位置服務，實現 SIP 使用者註冊、定址等行為，並擴展到其他應用。目前 IETF 將[4]列為標準草案，積極發展相關標準協定。

另一方面，由於目前網路環境普遍不支援 IP 層的群播功能，因此應用層群播就成了實現會議串流繞送的主流方法，但在分散式環境中，經過雜湊函數運算後的節點要如何規劃出合適的應用層群播樹，目前的研究對於此類的討論尚不多。

[1] 本研究由國科會贊助，計畫編號 NSC100-2221-E-150-077 與 NSC 101-2221-E-150-001。

三、 P2PSIP Conference System

3.1 階層式疊層網路

P2PSIP 會議系統是由所有加入疊層網路的 Peer 所形成的環境，每個 Peer 會和疊層網路中的其他 Peer 互連，並透過疊層網路的路由演算法維護路由資訊；同時疊層網路中的 Peer 也提供儲存空間，儲存保管分配到的資料。在先前研究[5]中，提出以 IP prefix 劃分區域疊層網路，並透過共同疊層網路來為各個區域疊層網路做跨疊層的路由或訊息轉送，藉此可以加強各個區域疊層網路的地域性，解決疊層網路與實體網路間的落差，如圖 1。Peer 可以視其能力分享資源給其它 Peer-to-Peer 中的使用者，擔任 P2P Focus 或 P2P Mixer 等多媒體會議服務的角色。P2P Focus 負責會議初始化、管理、與會議資源協調管理。P2P Mixer 負責會議語音與視訊資料混合重新編碼，提供串流給所有會議參與者。

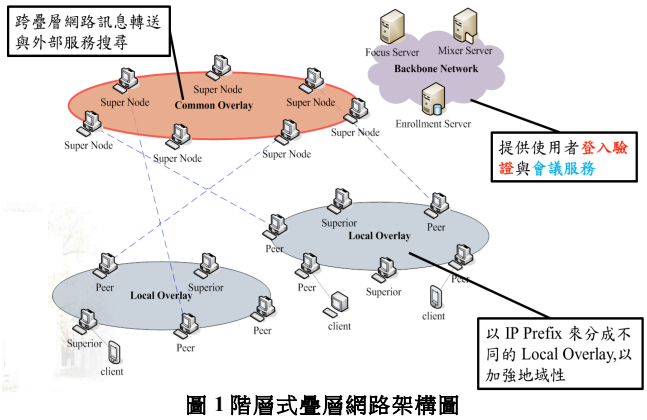

圖 1 階層式疊層網路架構圖

如果在 P2P 網路中頻繁的測量網路延遲，將會對網路造成很大的負擔；如果在會議發起時才測量，則會使會議建立時間增加，因此在本論文參考[6]中的 Landmark 估測機制來估測出區域疊層網路間的網路延遲。本論文中，區域疊層網路裡的節點彼此間的網路延遲是極小的，甚至可以忽略不計。 每個區域疊層網路之 SN 將會對所有的 Landmark 週期性測量 RTT ，並得到相對應的 landmark vectors，如圖 2。

圖 3則是以兩個不同的區域疊層網路對三個 Landmark 測量 RTT 後所得到的 landmark vectors 示意圖。

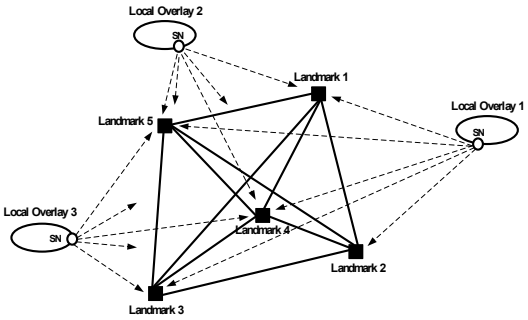

圖 2 各區域疊層網路之 SN 對 Landmarks 測量 RTT

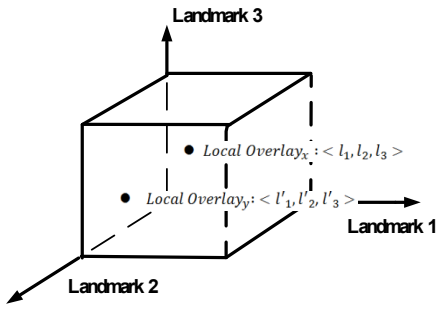

圖 3 landmark vectors 示意圖

3.2 Priority Weight

3.2.1 Delay Latency

透過上節介紹的 landmark vectors，並以[7]提出之(1)，可以估測出兩區域疊層網路間的網路延遲。其中 x 和 y 代表兩個不同的區域疊層網路，而 d_i^x 則是區域疊層網路 x 對 Landmark i 所測量到的網路延遲，而 n 為 Landmark 的總個數，如此一來就可以估算出兩個區域疊層網路間的網路延遲。

$$d(x,y)=\sqrt{\sum_{i=1}^{n}(d_i^x-d_i^y)^2}$$

由於算出來的 d(x ,y) 將與下節測出來的網路頻寬去算出 Priority Weight ，因此我們必需將 d(x ,y) 做標準化的動作，其公式如下：

$$d_n(x,y)=\frac{300ms}{d(x,y)}\begin{cases}>1, \text{愈大代表該路徑延遲愈低}\\ \le 1, \text{該路徑不能用}\end{cases}$$

由於 ITU 定義 VoIP 中多媒體的延遲不得超過單向上限 150ms，而我們測量延遲是以雙向的 RTT 為單位，在忽略雙向路徑不一致的情況下，這邊正規化的單位以兩倍的單向上限延遲來定義。

3.2.2 Peer Out-degree

在本論文中，每個節點會週期性地估測或量測其可用頻寬，但此處不討論可用頻寬的估計方法。在考慮到 Priority Weight 的標準化，公式如下：

$$c_i=\left[\frac{\text{目前可用頻寬}}{\text{傳串流連線所需的頻寬}}\right]=\begin{cases}>1, \text{愈大代表連線數愈多}\\ \le 1, \text{無法建立連線}\end{cases}$$

本論文以每一條串流所需要的頻寬來做為正規化的單位，而 i 代表該 Peer 在此多媒體會議中屬於某區域疊層網路的會議成員，c_i 即代表該 Peer 可以支援的連線數。

由於計算時是以區域疊層網路為單位，因此 C_j 代表會議成員在某區域疊層網路的總支援連線數， j 表示多媒體會議中的某個區域網路， n 為該會議中所有的區域疊層網路總數。

$$C_j=\sum_{i=1}^{n}c_i$$

3.2.3 Stability Prediction

ALM 應用中， Peer Stability 是重要的討論重點，因為節點的穩定度太低會造成 Churn 的現象。此現象指的是因為 P2P 網路中的節點頻繁的加入或離開而造成整體 P2P 行為的崩壞或是負擔，因此在 Live Streaming 的環境中考慮應用層群播樹的建立時，一般都會將 Stability 比較高的節點擺在群播樹的上層，以期降低因為 Churn 而被影響到的節點可以降到最少，因為如果應用層群播樹中間的節點離開的話，就意味著該節點以下的所有節點都無法再獲得多媒體串流。

綜合以上三節，在建立應用層群播樹時，如何在 Peer-to-Peer 間的網路延遲和 Peer 頻寬之間取捨就成很重要的議題，本論文中我們提出以 Priority Weight 的方式來解決問題，該連線的 Priority Weight 判別依據如下：

$$PW=\alpha\times d(\mathrm{x},\mathrm{y})+(1-\alpha)\times \mathrm{C_y}, 0\le\alpha\le 1$$

其中 $d(x,y)$ 為兩點之間的網路延遲， C_y 為在應用層群播樹中，被挑選節點的可用頻寬，即該節點能支援的連線數。另外，α 為系統參數，可依應用需求在網路延遲和頻寬之間做權值。

在多媒體會議應用中，由於該會議的所有成員知道自己正在開會，因此相對於 Live Streaming 的應用，會議的應用層群播樹中 Churn 的發生機率會比較低，因此我們只對低於 Stability Threshold 做處理，判斷該節點是否為群播樹中的葉節點，讓該節點的底下不再接其他的節點，以降低 Churn 發生時的影響程度。

3.3 ALM tree Policy

除了以 PW 作為應用層群播樹的挑選策略之外，由於本論文的環境是階層式的疊層網路，因此我們特別以此環境的特性為考量，另外為區域疊層網路內的節點設計節點挑選機制。

本論文提出以區域疊層網路為單位建立出一棵 Global View 的應用層群播樹，這邊的 Global View 指的是以共同疊層網路的角度來看的應用層群播樹。在 Global View 中的每一個應用層群播樹節點都代表一個區域疊層網路。而在每一個區域疊層網路中，可能會有多個會議的成員，因此必須為他們制定一個 Local View 的串流傳輸策略。

在 Local View 的應用層群播樹中，我們希望 Local View 內的節點可以為 Global View 節點提供更大的 Out-degree ，透過降低整體應用層群播樹的深度，進而減少整體會議成員間的網路延遲，機制說明如下表與下圖 4。假設每個節點的可用頻寬都可以支援兩條串流連

線，Caller 會以成員數最多的 No.1 來搜尋 Focus 和 Mixer ，亦即 No.1 為此會議的應用層群播樹的串流來源區域疊層網路。假設此例中的 Mixer 之 Out-degree 為 4，因此它最多可以支援四個子節點，假如我們僅以區域疊層網路為單位並以 PW 值來規劃此群播樹，那麼便會形成如下圖左下的 Global View 樣子的應用層群播樹，其深度在此例中為三層。由於在多媒體會議中，Churn 的發生機率遠比 Live Streaming 來的小，因此我們對於 Stability Threshold 的門檻可以降低許多，在此前提之下，高於門檻值的節點都可以任意的擺放在 Local View 的應用層群播樹中。以下圖右為例，如果我們將所有高於門檻值的節點都放到 Mixer 底下，那麼以 Global View 的角度來看的話，此應用層群播樹的來源端 Out-degree 將會增加到 8 個，達到增加 Global View 節點 Out-degree 並降低整體網路延遲的效果。

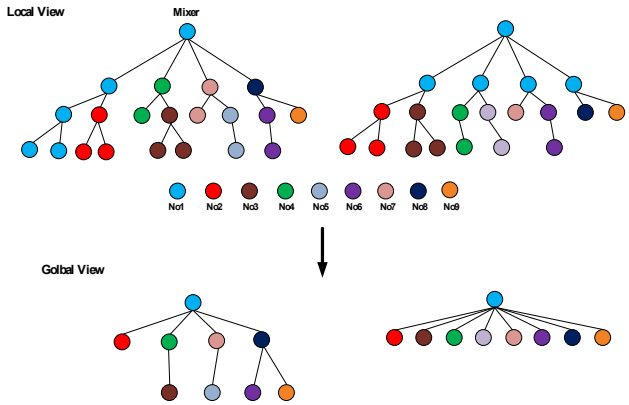

圖 4 群播樹節點挑選策略

四、系統運作

當一個使用者作為會議發起者要建立一個會議時，它的流程如下：

(1) 首先必須設定這個會議的 Conference Profile，包含會議的總人數、會議成員的 SIPURI、等機制。

(2) Caller 會先以 SIPURI 過濾出所有成員的分布，並選定最多成員數的區域疊層網路，透過服務搜尋的機制，到該區域疊層網路中找到可用的 Focus。由於此階段 Caller 能夠掌握的訊息僅有所有會議成員的 SIPURI，因此如果出現有兩個以上的區域疊層網路都可以有搜尋 Focus 的狀況時，則從這裡面亂數決定一個區域疊層網路來作為 Focus 的所在位置。

(3) 當找可用的 Focus 之後，會議發起者要透過 SIP 的 Invite 訊息，向此 Focus 提出建立會議的請求，同時將這個會議的成員名單 (INVITE-Contained Lists) 透過 SDP 夾帶於 SIP Invite 的訊息中送給 Focus 。

(4) 當 Focus 收到會議請求之後，要根據會議成員名單到各個區域疊層網路中尋找會議的參與者，中間利用 Super Node 進行跨疊層的繞送，透過疊層網路找到這些會議成員的 IP 位址。此階段 Focus 在跨疊層搜尋會議成員時，會向各個區域疊層網路的 Super Node 詢問該區域疊層網路的 landmark vectors 。landmark vectors 的詳細說明在 3.2.1 節。

(5) 會議成員回應其節點的基本資訊，包含 3.2.3 節中討論的節點可用頻寬和節點 Stability value。

(6) 此外， Focus 還要依照會議的設定，再次以服務搜尋的機制到疊層網路上尋找可用的 Mixer 。

(7) 找到可用的 Mixer 之後， Focus 則向 Mixer 送出 SIP 的 Invite 訊息。

(8) Mixer 接受請求後將回應 200 OK 給 Focus ，並於其中夾帶 SDP 提供 Media Mixing Level 的資訊。

(9) Focus 再送出 SIP 的 Invite 訊息，告知會議成員名單中的參與者，並同時 Re-Invite 會議發起者，告訴它們 Mixer 的資訊。此階段發送的訊息內容包含由 Focus 依 3.2 節所介紹的演算法計算出來的應用層群播樹訊息，以此告知各個節點與 Mixer 在應用層群播樹中的位置。

(10) 當會議參與者收到 Invite 後，即根據上面的資訊加入會議，並以它在該會議的應用層群播樹中的位置，接收多媒體串流以及做為媒體中繼節點來轉送多媒體串流給它的子節點。

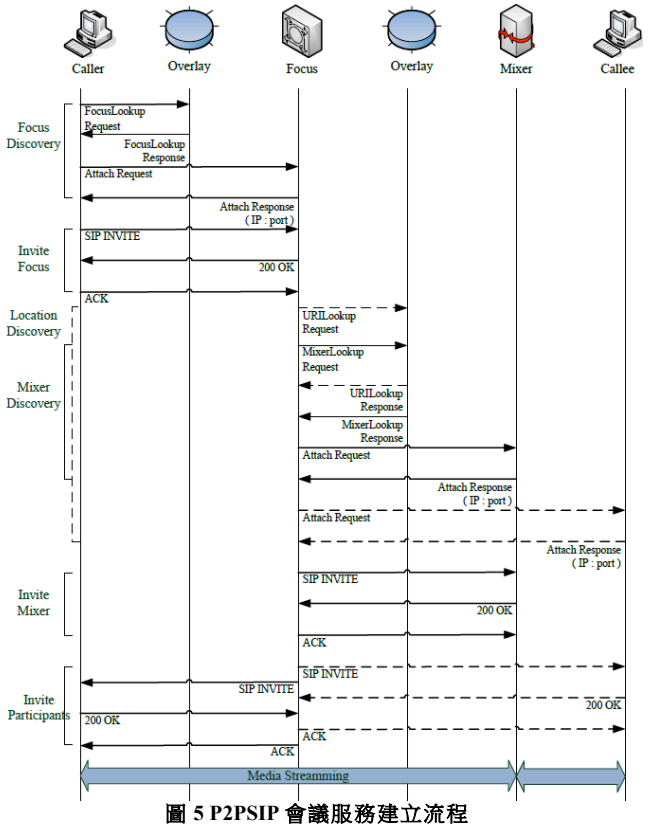

圖 5 P2PSIP 會議服務建立流程

五、系統模擬分析

本論文將以[8]中的測量結果作為我們的網路拓樸，進而驗證效能。在[8]中，實際測量的日本的 ISP 業者所佈署總計共 13 個的地區網路與地區網路之間的網路延遲 (以 RTT 表示) 如下表 1

表 1 Delay latency of ISP (ms)

	No.1	No.2	No.3	No.4	No.5	No.6	No.7	No.8	No.9	No.10	No.11	No.12	No.13
No.1	4.46	72.50	55.04	46.42	70.72	63.06	55.24	59.28	73.56	69.02	75.38	78.08	105.70
No.2	72.74	5.26	42.16	33.84	56.78	49.00	42.80	46.38	61.46	56.68	62.72	68.76	91.40
No.3	54.88	41.88	2.90	19.52	41.90	32.32	28.50	32.10	40.70	42.46	48.46	64.50	75.58
No.4	46.42	34.46	19.70	3.16	33.98	26.28	19.76	23.34	37.56	33.26	39.48	41.62	68.64
No.5	69.80	55.66	41.66	33.44	1.46	45.78	42.02	48.20	54.44	55.86	62.04	78.28	89.16
No.6	62.86	49.24	32.06	25.16	45.74	1.54	35.86	41.94	48.10	49.40	52.08	72.02	90.82
No.7	55.40	42.84	28.26	19.32	42.36	35.86	2.18	32.40	46.46	41.84	48.46	50.74	78.84
No.8	58.96	46.60	32.00	23.24	48.16	41.14	32.26	1.88	50.60	46.22	52.42	55.00	83.78
No.9	73.26	61.06	40.36	37.34	54.54	48.18	46.40	50.74	1.90	60.40	60.86	69.22	91.54
No.10	69.14	56.54	42.30	33.04	55.84	49.50	41.86	46.34	60.38	2.50	62.14	64.64	92.72
No.11	75.26	62.94	48.62	39.22	62.08	52.18	48.10	52.54	60.88	62.10	1.50	74.72	95.36
No.12	77.98	68.74	64.74	41.90	78.24	72.16	50.72	55.24	69.30	64.74	75.04	1.96	106.20
No.13	105.54	91.24	75.26	67.78	89.42	90.82	78.76	84.14	91.50	92.78	95.56	106.20	1.36

針對本論文的效能驗證，我們將每一個地區網路看作是一個 Local Overlay ，在估測網路延遲方面，我們需要挑選合適的 Default SNs ，這部分不管是在集中式或分散式的架構都尚有很大研究討論空間。在這裡挑選 No.1、No.4 、No.7、No.10 和 No.13 作為 Default SNs 來做為 Landmarks。接著每個 SN 將會分別對以這五個 SNs 測量出 landmark vectors，以供 Focus 在建立應用層群播樹時運算出 PW 值。

表 2模擬時之節點分佈組態檔

區域疊層網路編號	區域疊層網路成員數
No.3	1
No.6	2
No.10	5
No.11	7
No.12	4
No.13	1
會議總成員數	20
會議總區域疊層網路數	6

另外，[9]提出以 RELOAD 中的 Finger Table 來作為其應用層群播樹的建樹依據，以最簡單的條件來建出應用層群播樹，該機制稱作 Adjacency Matrix。另一方面，在 P2P 即時串流的相關應用中，因為害怕 Churn 會造成應用層群播樹的崩壞，因此通常都會在建樹的機制中考慮節點穩定度 (Stability) 的問題，[10]為近年來較為代表性的機制。本論文之模擬將以這兩者作為比較組並在 Worst Case 的狀況下進行模擬。本論文定義的 Worst Case 為每個節點的頻寬僅能支援兩條輸出串流連線 (out-degree)，以此模擬最大會議成員數時 (20 位會議成員) ，在不同的 Local Overlay 分布時的網路延遲情形。

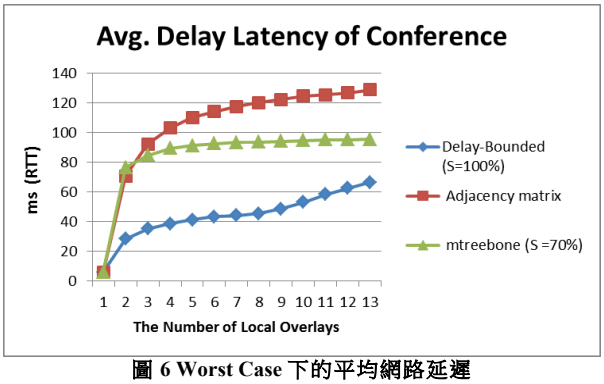

圖 6 Worst Case 下的平均網路延遲

由圖 6可以看出本論文所提出的機制效能明顯比另外兩個方法來的好。因為本論文的方法是將所有的成員依區域疊層網路來分群組建應用層群播樹，以 Global View 的角度來看的話，依成員分布不同以及疊層網路總量不同兩個關係的影響，所形成的應用層群播樹將可能會有 1~3 不同的層數，因此本論文的機制在平均延遲的效能曲線上，將有三個不同階段以階層式的上升。而由於我們提出的機制有將會議成員分組，並計算出 PW 值來建應用層群播樹，因此即使在更多疊層網路的情況下，整體的平均網路延遲仍會優於其他兩種方法。

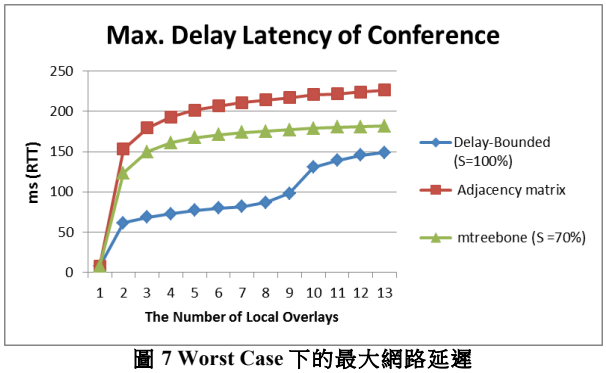

圖 7 Worst Case 下的最大網路延遲

圖 7中的效能曲線一樣有階層式的提升，而本論文提出的機制所呈現出來的最大網路延遲，在最差的情況下大約在 150ms 左右，僅有 ITU 的上限 300ms 一半左右。

參考文獻

[1] Rosenberg, J., "A Framework for Conferencing with the Session Initiation Protocol (SIP)", RFC 4353, February 2006.

[2] Johnston, A. and O. Levin, "Session Initiation Protocol (SIP) Call Control - Conferencing for User Agents", BCP 119, RFC 4579, August 2006.

[3] Mahy, R., Sparks, R., Rosenberg, J., Petrie, D. and Johnston, A., "A Call Control and Multi-Party Usage Framework for the Session Initiation Protocol (SIP)", RFC 5850, May 2010.

[4] Jennings, C., Lowekamp, B., Rescorla, E., Baset, S., and H. Schulzrinne, "REsource LOcation And Discovery (RELOAD) Base Protocol", *draft-ietf-p2psip-base-22*, July 2012.

[5] Hui-Kai Su, Chien-Min Wu and Wang-Hsai Yang, "Design of Location-Based Hierarchical Overlay Network Supporting P2PSIP Conferencing Service", The 7th International Conference on Autonomic and Trusted Computing (ATC 2010), Xi'an, China, 26-29 October, 2010.

[6] Z. Xu, C. Tang, and Z. Zhang, “Building Topology-Aware Overlays Using Global Soft-State,” Proc. Int’l Conf. Distributed Computing Systems, May 2003.

[7] Z. Li, G. Xie, K. Hwang, and Z. Li, “Churn-Resilient Protocol for Massive Data Dissemination in P2P Networks,” In IEEE Transactions on Parallel and Distributed Systems, Vol. 22, No. 8, Aug. 2011.

[8] K. YOSHIDA, Y. KIKUCHI, M. YAMAMOTO, Y. FUJII, K. Nagami, I. NAKAGAWA and H. ESAK, "Inferring POP-level ISP Topology through End-to-End Delay Measurement", 10th Passive and Active Measurement Conference, April, 2009.

[9] M. AMAD, Z. HADDAD, and L. KHENOUS, “A Scalable based Multicast Model for P2P Conferencing Applications,” International Conference on Ultra Modern Telecommunications & Workshops, 12-14 Oct. 2009.

[10] F. Wang, Y. Xiong, and J. Liu, “mTreebone: A Collaborative Tree-Mesh Overlay Network for Multicast Video Streaming,” In IEEE Transactions on Parallel and Distributed Systems, Vol. 21, No. 3, Mar.2010.