

結合 CDN 之階層式 P2P 疊層網路隨選視訊服務架構與機制設計

蘇暉凱^a、黃敬庭^b、張榮佐^b、蕭紋旭^c、陳景章^d

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

吳鳳科技大學應用數位媒體系^c

國立中正大學電機工程學系^d

摘要 — 隨著網路技術蓬勃發展，人們已經可透過日益增快的網路速度獲得高品質的視聽娛樂，隨選視訊服務就是其中之一。但集中式的隨選視訊架構在面對大規模的服務數量時，可能在設備與網路頻寬方面會產生系統瓶頸，甚至服務會遭受阻斷。因此本文提出結合 CDN 的 P2P 隨選視訊服務架構，希望對於服務提供者而言，能夠以適當的成本提供盡可能多服務數量；而從使用者的角度來看，也確保串流服務能夠順暢不中斷。

一、簡介

IPTV 為目前網路的熱門應用之一，根據服務類型的不同，IPTV 可分為 Push-Based 的 Live TV 與 Pull-Based 的隨選視訊(Video on Demand, VoD)兩種形式。

相較於 Live TV，VoD 對於網路延遲的要求較低，因此有部分的 VoD 供應商[1]為了兼具服務數量與成本控制，摒棄了傳統伺服器的形式，採用 P2P 的架構來提供串流服務。P2P VoD 不僅能讓使用者獲得服務，也可提供服務給其他人，如此一來伺服器的功能與負載便能分散到使用者上，進而減少建設成本並增加服務規模。但這類的做法若不採用目錄伺服器的方式來協助使用者搜尋串流資源，則容易造成 P2P 疊層網路(Overlay Network)與實際網路差異而產生較大的延遲。此外 P2P 系統可能會由於節點頻繁進出，造成部分資源不可用的狀況。

為避免上述問題可能造成的服務中斷，本文提出了結合 CDN 的階層式 P2P 網路架構來進行改善。此服務架構可透過階層式 P2P 降低網路差異所產生的延遲，並透過 CDN 服務對遺失的串流資源進行快速的填補。

在本文的第二章會對此架構使用的相關技術及背景進行探討；在第三、四章會對系統架構及其運作機制做介紹；第五章採用模擬的方式對本系統的設計進行驗證；最後在第六章進行本文的總結。

二、相關背景

2.1 P2P 系統與 DHT

Peer-to-Peer 系統為一種分散式的資料分享系統，此系統不存在特定的檔案伺服器，而是透過使用者互相交換訊息達成檔案資源的定位與共享，因此比起集中式的 Client/Server 模式，P2P 可提供更高的擴展性與服務規

模。早期的 P2P 系統如：Napster[2]，採用目錄伺服器 Tracker 來幫助使用者進行節點與資源定位，但 Tracker 仍會面臨傳統集中式架構的問題。因此純 P2P 架構的系統[3]應運而生，使用者節點會相互連結型成網狀拓樸的邏輯網路，即疊層網路。在疊層網路中，使用者節點透過泛洪(Flooding)的方式向鄰居節點廣播搜尋訊息，但此種方式必須透過設置 TTL(Time to Live)來避免廣播風暴(Broadcast Storm)，所以可能會造成資源存在卻因為 TTL 的限制而無法搜尋到。

因此有人提出了利用分散式雜湊列表(Distributed Hash Table, DHT)[4][5]來組織使用者節點，透過 Hash 的方式為每個節點產生唯一的辨識 ID，讓每個節點各自維護一個辨識 ID 與 IP 的對應表，並經由此表互相通訊以一步步的定位資源或節點。

2.2 P2P VoD 系統

P2P VoD 系統即是透過 P2P 的系統機制達成 VoD 的影片串流服務。典型的 P2P VoD 服務[10]，大多採用入口網站(Portal)的形式提供影片列表，使用者在選定影片後會加入該影片的 P2P 群組(Swarm)，因此使用者可能會根據觀看影片的多寡，同時加入多個 P2P 群組。加入群組的同時入口網站會告知使用者此群組的 Tracker 位置，讓使用者節點透過此 Tracker 來獲得提供串流資源的 Seeder 位置。Seeder 可能為系統提供供應商提供的影片伺服器，或者是分享影片資源的使用者節點。

2.3 內容分發網路(Content Delivery Network, CDN)

內容分發網路為一種改善傳統 Client/Server 缺點，提供更多服務流量的網路架構。傳統的集中式伺服器隨著使用者規模的成長，大量的存取行為會開始對設備產生負擔，最終可能會造成單點失效(Single Point Of Failure)等問題。因此必須購買更高等級的伺服器設備或者建立伺服器叢集來提供服務，但若使用者規模繼續成長，此時除了設備成本之外，必須要花費更大在增加網路頻寬上，因此 CDN 採取了分散的方式來解決此項問題。

CDN 提供商會在各地的架設具 Cache 功能的伺服器—Edge Server，這些 Edge Server 會根據提供商策略的不同，使用 Pull-Based 或 Push-Based 的方式從原始內容伺服器—Origin Server 獲取服務內容，最後再透過特殊的

CDN DNS 將使用者的存取要求引導至適當的 Edge Server。如此一來便可增加系統頻寬同時分散系統負載。

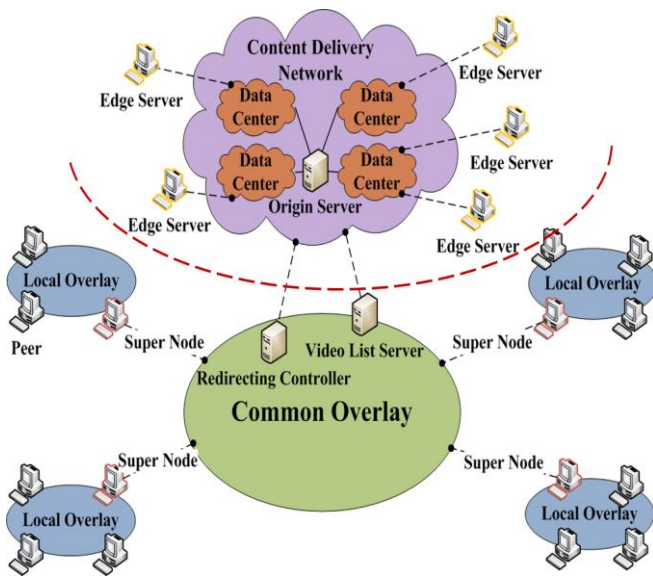
除了平衡負載的方式，CDN 也可採取地理感知的技術，判斷使用者來自何處，再讓 CDN DNS 指派離使用者最近的 Edge Server 提供服務，如此便能降低延遲，讓 CDN 服務不僅能使用於檔案及網頁，更能滿足於對延遲要求較為嚴格的 IPTV 等服務。

三、 系統架構

本系統採取 P2P 系統做為共享影片資源的方式，而為了改善 P2P 疊層網路與實際網路差異導致的延遲，將距離相近的 Peer 使用地理感知的技術進行分區，再透過 Chord 的 DHT 演算法[4]組織，形成階層式的架構。

此外，本系統在 P2P 網路之上結合能分散負載且能快速提供服務的 CDN，以進一步改善傳統 P2P VoD 系統中 Seeder 的負載、延遲與資源分配等問題，整體系統架構如圖一所示，分為 CDN 與階層式疊層網路兩個層級，在本章會由系統的底層開始依次介紹各層部件的功能。

在本文的架構中，階層式疊層網路可分為兩個子層級，分別是區域疊層網路與共同疊層網路。



圖一：結合 CDN 之階層式 P2P VoD 系統架構

3.1 階層式疊層網路 - 區域疊層網路(Local Overlay Network)

區域疊層網路是由地理與網路狀態相近的使用者們組織而成，因此可能會存在多個不同的區域疊層網路。使用者在此可透過 P2P 的資源搜尋機制以及 CDN 服務獲得所需的影片資源。

Peer：使用者在系統中的角色，同時也是構成區域疊層網路的主要部件，除了發起搜尋以獲得資源之外，還需要儲存區域疊層網路的資訊與維護路由。此外，也要保存觀看過的串流 Segments 做為副本註冊至區域疊層網路，以提供給其他 Peer 搜尋使用。

Super Node：在區域疊層網路中，透過[11]發覺穩定

性比較高的節點，功能為與共同疊層網路連結的橋梁，並非唯一。當目標資源無法在區域疊層網路獲得時，須經由 Super Node 來做進一步的搜尋以獲得需要的資源。

3.2 階層式疊層網路 - 共同疊層網路(Common Overlay Network)

此層級主要的功用是對在區域疊層網路上無法獲得的資源去做更進一步的搜尋，因此在這裡加入了 CDN 的介面，讓 Super Node 不僅能搜尋其他區域疊層網路，也能向 CDN 發送資源的請求。

Super Node：組成共同疊層網路的主要部件，因此必須儲存共同疊層網路的資訊與維護路由。除此之外，Super Node 還需要接受自身區域疊層網路 Peer 的委託，進行跨疊層網路的 P2P 資源搜尋，所以需要透過額外的搜尋機制與空間去儲存「影片名稱」來幫助我們定位串流資源。這裡的「影片名稱」指的是「在這個區域疊層網路曾經播放過的影片名稱」。除了協助 Peer 進行跨疊層網路的資源搜尋外，Super Node 還需要在本層搜尋影片列表檔。影片列表檔指的是目前本系統可提供的影片列表以及串流的資源 ID 列表所集成的檔案。Peer 可透過影片列表檔得知目前有甚麼樣的影片是可以觀看。而需要注意的一點是，Super Node 不會在共同疊層網路註冊影片串流的 Segments，因此在本層不會存在任何的影片資源。

Video List Server：CDN 層在共同疊層網路加入的節點，並非唯一，其用途為定時提供最新的影片列表檔，以供 Super Node 及其區域疊層網路進行影片列表的更新，而不必透過集中式的伺服器來得知目前可觀賞的影片以及 Segments 的資源 ID。

Redirecting Controller：CDN 在共同疊層網路提供的介面，並非唯一。當無法在各個區域疊層網路搜尋到串流資源時，Super Node 可以透過此節點來要求 CDN 提供服務，以確保服務不中斷。因此除了做為節點註冊至共同疊層網路之外，還需要將「Edge Server」此字串做為資源註冊至共同疊層網路，像是 Super Node 註冊影片名稱一樣，讓 Super Node 可透過搜尋「Edge Server」，在共同疊層網路找到 Redirecting Controller。

3.3 CDN 層

本層主要功能為儲存原始多媒體資源及提供存取介面讓使用者距離最近或延遲最低的資料中心主機獲得服務。

Origin Server：儲存原始影片串流的伺服器，所有影片的 Segments 均存放於此。其資源僅提供 CDN 內部的設備存取，不會對使用者提供服務。

Edge Server：CDN 中主要提供使用者服務的伺服器。根據使用者要求，Edge Server 向 Origin Server 抓取影片資源提供給使用者，並同時進行資源的暫存。當有其他使用者要求相同資源時，Edge Server 就能更快速的提供服務。相同國家或地域的 Edge Server 會以資料中心叢集的方式呈現。

Video List Server：定時向 Origin Server 獲取目前可用的串流 Segments 並 Hash 獲得資源 ID，再與影片名稱包裝成影片列表檔註冊至共同疊層網路。

Redirecting Controller：根據 Peer 的地理資訊與網路狀態提供適合的資料中心與 Edge Server 位置，讓其為使用者提供快速的服務，因此 Redirecting Controller 必須維護一張地區與 Edge Server 對應的表格。此外，在新使用者加入的時候，Redirecting Controller 也要為使用者尋找合適的區域疊層網路，並提供該區域疊層網路的引導節點(Bootstrap Node) IP，以供新使用者以 Peer 的身分加入系統。

四、系統機制

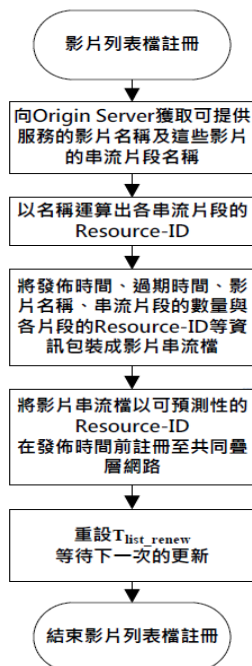
本系統架構為結合 CDN 的 P2P 影片串流系統，因此在本節將會對於如何透過 P2P 系統獲得 CDN 上的資源，以及 CDN 如何提供距離使用者較近、延遲較低的服務等機制進行介紹。

4.1 區域疊層網路的加入

新加入的使用者透過 IP 定位的功能獲得到的資訊(所在之國家與 ISP 業者)來算出自己的 Special-ID，並傳送給 Redirecting Controller，利用 Redirecting Controller 引導新使用者加入適合的區域疊層網路以及為使用者分配適合的 Edge Server。

4.2 影片資訊的發布

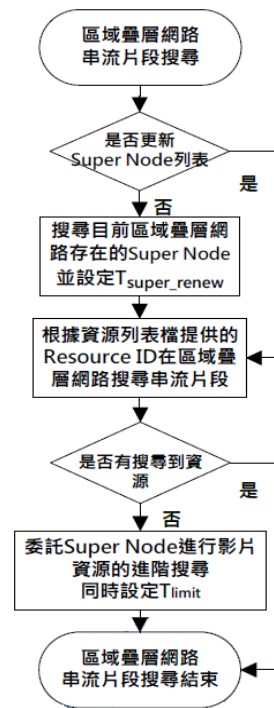
本系統的影片列表採用 P2P 機制而非 Portal 的方式提供給使用者，這樣除了可避免影片列表伺服器遭到服務阻斷(DoS)，也能降低服務提供者的系統負載。



圖二：Video List Server 的影片列表檔案發佈流程

4.3 串流資源的搜尋

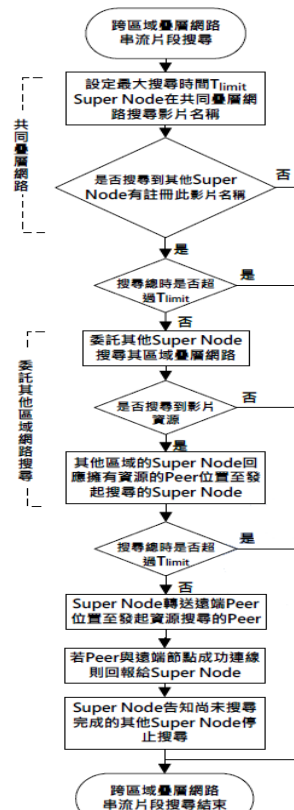
本系統採用階層式的 P2P 結構，組織位置相近的 Peer 以利提供低延遲的串流服務，同時也會根據搜尋條件與狀態的不同，可能會讓遠端節點或 CDN 提供服務。



圖三：Peer 的區域疊層網路搜尋機制

4.3.1 進階搜尋

進階搜尋事實上為 Super Node 在共同疊層網路上兩個行為的並稱：跨區域疊層網路搜尋與 CDN 連線的初始化，讓 Peer 能夠透過剩餘的播放長度來靈活判斷要採用何種存取方式來獲得串流片段。



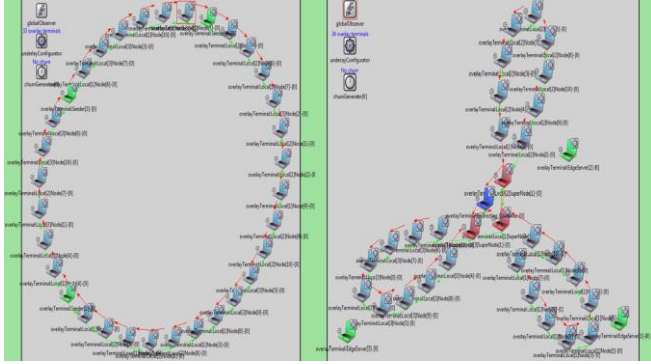
圖四：跨區域疊層網路的資源搜尋

五、 模擬與分析

本文採用 OverSim 對系統機制與效能進行模擬。OverSim 是以 OMNet++ 基礎開發的 P2P 模擬框架。OverSim 除了內建如 Chord、Pastry 等 P2P 模組外，使用者可根據需求使用 C++ 撰寫自定義的模組，並透過圖型化介面觀察網路及節點行為。

在系統模組會針對不同節點數量的串流服務取得時間與傳統 VoD 服務架構進行比較。

下面左圖為結構化 P2P 的 VoD 服務架構，右圖為本論文所提出的架構。



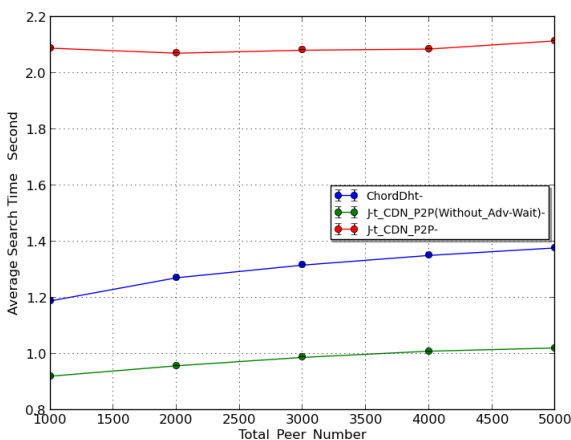
圖五：Oversim 圖形化介面

5.1 模擬環境

在這裡的模擬情境為系統已運行一段時間，P2P 網路上已存在著部份串流片段(70%)，系統中共 10 部影片，每部長度約為 15 分鐘，每個串流片段長度為 10 秒；而使用者會在看完影片後 10 分鐘離開系統，並於離開後 20 分鐘重新上線後選取觀賞下一部影片。為顯示節點遠近差異，近端節點間的單向傳輸延遲設定為 5ms、遠端節點為 30 ms。

5.2 串流搜尋時間

本模擬將使用者分為 10 個不同的區域疊層網路，並以 100 個節點為單位去逐次增加區域疊層網路的節點數量，同時 Edge Server 與 Media Server 均僅能提供 300 個串流片段同時傳送。最後由圖六對本論文的系統的架構(J-t_CDN_P2P)與一般結構化 P2P 的系統架構(ChordDht)進行平均搜尋時間的比較。



圖六：平均搜尋時間的模擬比較

從圖六可發現，本論文提出的系統架構效能遠低於一般的結構化 P2P 系統，其原因在於本系統的搜尋情境

中，有一項是執行跨疊層網路搜尋，直到 Tlimit 到期後才與鄰近的 Edge Server 取得連線，這表示即使階層式疊層網路中不存在目標資源，也必須等待至搜尋時限才會向 CDN 獲取資源，因而造成大量的搜尋時長。若將此搜尋情境排除(J-t_CDN_P2P(Without_Adv-Wait))，則平均搜尋時間優於普通的結構化 P2P 系統。

六、 結論與未來展望

本服務架構設計分散式的影片列表提供機制，以避免傳統VoD影片列表伺服器所可能遭受的單點失效，增加系統的穩定性及強韌性。

透過模擬數據，本論文提出的結合CDN之階層式P2P疊層網路隨選視訊系統雖然可大幅改善因網路差異的延遲所帶來的各種缺點，但仍有許多待討論改進的部份，例如跨疊層網路搜尋時，若定位到遠端的資源，如何評估與選擇遠端節點是否適合傳輸串流；進階搜尋中，最大搜尋時間限制Tlimit若設定太小，則容易造成凍結次數的上升，設太大則會把流量集中到CDN去，因此Tlimit的最佳化也是未來可探討的研究主題；此外，CDN層的能力是否能透過用雲端建構的方式更有效的利用也是能進一步研究的議題。

致謝

感謝科技部研究計畫NSC102-2221-E-150-001與103-2221-E-150-020支持。

參考文獻

- [1] Yan Huang, Tom Z. J. Fu, Dah-Ming Chiu, John C. S. Lui and Cheng Huang, "Challenges, Design and Analysis of A Large-Scale P2P-VoD System," ACM SIGCOMM 2008 Conference on Data Communication, Aug 17-22 2008, pp.375-388.
- [2] Napster Message, <http://opennap.sourceforge.net/napster.txt>.
- [3] Gnutella Protocol Development, <http://rfc-gnutella.sourceforge.net/>
- [4] Ion Stoica, Robert Morris, David Karger, M. Frans Kaashoek and Hari Balakrishnan, "Chord: A Scalable Peertopeer Lookup Service for Internet Applications," 2001 ACM SIGCOMM Conference, Aug 27-31 2001, pp.149-160.
- [5] Sylvia Ratnasamy, Paul Francis, Mark Handley, Richard Karp and Scott Shenker, "A Scalable Content-Addressable Network," 2001 ACM SIGCOMM Conference, Aug 27-31 2001, pp.161-172.
- [6] Rui S. Cruz, Mario S. Nunes, Yingjie Gu, Jinwei Xia, Joao P. Taveira, and Deng Lingli, "PPSP Tracker Protocol-Base Protocol (PPSP-TP/1.0)", IETF Draft, draft-ietf-ppsp-base-tracker-protocol-03.txt, Dec 31, 2013.
- [7] Lo V, Dayi Zhou, Yuhong Liu, GauthierDickey C. and Jun Li, "Scalable supernode selection in peer-to-peer overlay networks," Hot Topics in Peer-to-Peer Systems, 2005. HOT-P2P 2005. Second International Workshop on, Jul 21 2005, pp.18-25.