

P2P 軟體對網路影響之研究

劉大川 蔡文能
交通大學計算機網路中心
{ltc,tsaiwn}@faculty.nctu.edu
u.tw

蔡宗易 黃威力 宋嘉銘 賴俊廷
交通大學資訊工程學系
{tytsai,weili,chiaming,ctlai}@csie.nctu.edu.tw

摘要

隨著電腦進步與頻寬成長，網路使用型態已有所改變。根據本論文之分析，P2P 應用漸漸成為網路流量的主要來源。藉由 P2P 軟體，用戶可以有效地互相分享各種檔案，但是卻造成智慧財產權的侵犯以及網路壅塞的問題。本篇論文分析網路流量，配合 OSI 第七層內容過濾方式取得 P2P 流量佔總流量之比例，以及各種 P2P 軟體之比例。接著以隨機取樣方式，觀察 P2P 之間傳輸檔案的類型比例。

關鍵詞：P2P，智慧財產權。

Abstract

Network usage has been entirely different now with the improvement in computer technology. According to the research of this paper, P2P applications have become one of the majority sources of network traffics. Based on P2P applications, people can share resource with each other in efficient and convenient manner. However, it may cause the Intellectual Property Rights and network traffic congestion problems. In this paper, we propose several methods to analyze network traffics and conclude with statistics results, including P2P traffics ratio with respect to total traffics and various P2P software ratio. Then, we observe the distribution of various data type.

Keywords: P2P, Intellectual Property Rights.

1. 前言

最近十幾年來，網際網路使用者數量呈曲線指數上升，越來越多生活應用都是透過網路來進行，如商業行為、生活、娛樂等。到了 1990 年代末期，隨著用戶端設備不斷更新，以及網路頻寬日漸加大，點對點應用儼然成為最熱門的網路傳輸方式。雖然目前已有各式各樣的點對點網路架構，但是主要基礎架構多出於集中索引(Index Server)、分散式服務(Distribute Service)、分散式雜湊表(DHT, Distribute Hash Table)等三種，從技術及法律的種種層面來看，各有其優缺點，但尚未出現一個最佳的架構。

雖然網路頻寬變得越來越大，傳輸的效能也越來越好，但是由於現在網路應用程式的普及化與多樣化，使得我們並沒有從網路的發達中得到同樣多

的好處，更糟的是，我們碰上了更多的效能問題。

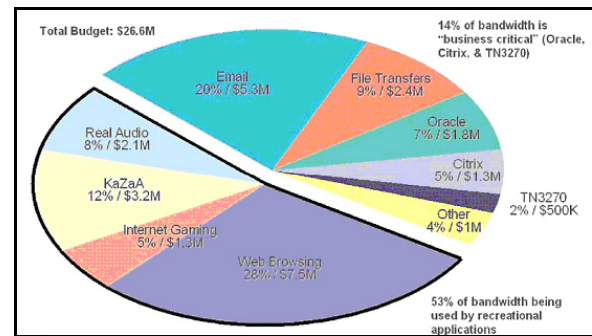


圖 1 各種網路應用比例圖(資料來源[13])

舉個例子來說(如圖 1)，在 2003 年，IDC 曾經做出研究報告指出，某家典型的美國大型企業每年花費約兩千六百萬美金在網際網路上面，卻只有 14% 的頻寬用在必要的工作上，其餘的都是網頁瀏覽、電子郵件傳送、P2P 程式等等，甚至還有網路遊戲。

現行的點對點應用架構，多著重於資料查詢或者與其他用戶端之間的溝通維護。但是往往沒有考慮到部分用戶因被選為熱點(via point, 中繼點)，不僅僅造成該用戶端網路壅塞及效能低落，更嚴重地，影響整體網路的效能，造成網路癱瘓。除了點對點本身架構改善能解決部分網路壅塞問題之外，點對點應用上所傳輸的內容及流量對整體網路的壅塞，是否也是一大主因之一。

在本篇論文中，我們針對交通大學校園網路環境中各種應用所佔比例，進行測量，並嘗試對目前點對點應用對網路生態造成的影響進行研究，及點對點應用對目前法律及生活所帶來的新層面衝擊進行探討。

2. 背景知識

1999 年 1 月，一位 18 歲的美國 Northeastern University 大學生 Shawn Fanning 寫了第一個 P2P 程式 Napster，容許 Web 使用者開放硬碟與別人交換 MP3 檔案，短短一年中創造 3800 萬會員。從此，在網際網路上已開啟 P2P 之門，P2P 應用如雨後春筍般湧現，知名軟體公司亦參與其中。如微軟公司正進行代號 Avalanche 檔案分享應用程式的開發 [15]。昇陽公司亦主導一項對等式架構技術開發計畫，Project JXTA [14]。

以下將對 P2P 傳輸架構及熱門的應用程式作簡單的介紹。

2.1 P2P 架構演進

點對點應用，如 Napster、Gnutella、Emule、Edonkey、Kazaa、ezPeer、Kuro 等，主要概念是在現有實體的網路上做新邏輯佈設，並非重新建構新的實體網路，用戶端不管選用哪種點對點的邏輯架構，最終資料傳輸與訊息傳送，還是依靠實體網路的 IP 來進行傳輸。每個用戶端之間在使用點對點的應用時，不管是直接傳輸或透過其他用戶端的間接傳輸，基本上都是用戶端兩兩互相溝通傳遞，如圖 2 所示。透過不同的點對點架構，所建構出來的點對點網路，是否穩固、有效率及其所能提供服務之多寡，便成了影響實體網路重要的因素。

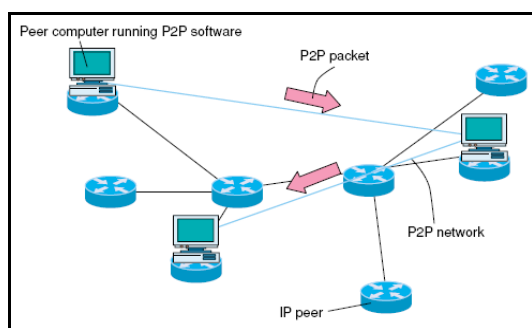


圖 2 點對點網路

目前點對點傳輸架構以演進時間做為依據，分類如下：主從式架構(Client and Server)、第一代、第二代、第三代點對點傳輸架構。

2.1.1 主從式架構

主從式傳輸架構是由用戶端 (Client) 和伺服器 (Server) 組成。雙方必須以協定好的通訊埠 (Port) 和通訊協定 (Protocol) 來溝通，如圖 3 所示。

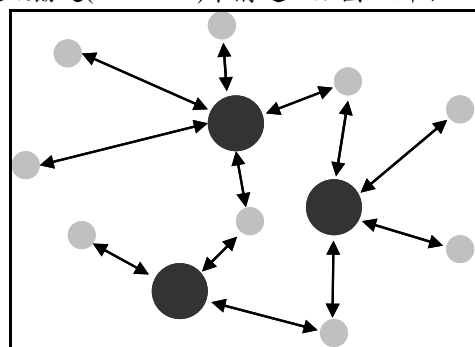


圖 3 主從式架構

用戶端只需要知道伺服器的 IP 位置、通訊埠與通訊協定即可。此架構最廣泛的應用為檔案傳輸協定 (FTP)。

2.1.2 第一代點對點傳輸架構

主要是由索引伺服器 (Index server) 和一般用戶端所組成，如圖 4 所示。索引伺服器，就是在伺服器上存放大量的集中式索引，用來記錄分享資源的連結。而傳輸運作模式則如同主從式架構，一般用戶端只需知道索引伺服器網路位置即可。當用戶端搜尋資料時，直接向索引伺服器詢問，索引伺服器則回傳擁有所需資料的用戶端資訊。用戶端取得相關資訊後，便向擁有此資料的用戶端要求檔案。這也是此架構與主從式架構最大的差異。Napster 為第一代點對點傳輸架的代表。

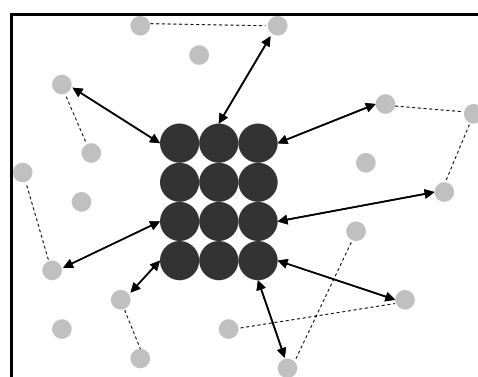


圖 4 第一代點對點傳輸架構

2.1.3 第二代點對點傳輸架構

第二代點對點傳輸架構主要是為了解決第一代中索引伺服器違法問題。因此第二代架構移除了索引伺服器，讓架構裡只剩下一般用戶端，如圖 5 所示。所以架構中的服務必須由每個用戶端來維護。因此，除了當接收端之外，每個端點亦會建立分享資源的索引表，提供其他點搜尋。傳輸運作模式，首先必需知到此架構中的某個點，再借由此點加入此架構，再進行分享及要求檔案的動作。Gnutella 為第二代點對點傳輸架構的龍頭。

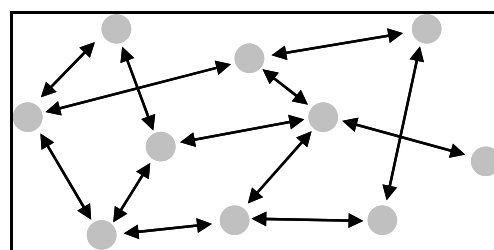


圖 5 第二代點對點傳輸架構

2.1.4 第三代點對點傳輸架構

第三代是目前點對點應用的主流，主要原因是因為加入分散式雜湊表 (DHT, Distribute Hash Table) 的概念，不再是第二代中雜亂無章的封包轉

送方式，而造成網路資源浪費。取而代之的是有系統、漸進式的尋找。整體架構與第二代相同，只剩下點。但由於加入了 DHT 的方法，使得每個點擁有一個唯一的用戶端編號。而所有的分享檔案也都有自己的資料編號，並且會對應到特定用戶端編號的點上。用戶端要加入此架構進行搜尋或傳輸時，與第二代相同，必須知道至少一個以上已在架構下的點，藉由此點來交換資訊，找到必須負責的所有資料編號，進而將來能供其他點詢問及分享檔案。eDonkey、eMule 為第三代點對點傳輸架構的代表。

2.2 常見分享軟體簡介

目前比較常見的 P2P 軟體有 eDonkey[7]、eMule[8]、ezPeer[9]、Kuro[12]、BitTorrent (BT) [6]、Kazaa[11]、Gnutella[10]、FastTrack[16]等，以下各小節分別簡介其運作方式。

2.2.1 eDonkey 與 eMule

eDonkey 為一個不需要依賴中心伺服器就可以搜尋網路上分享的檔案。搜尋的速度快且具有全域搜尋的功能。它允許客戶端傳輸任何類型的檔案，它能自動的交換來源繼續傳輸檔案，使用者將能從多重來源下載一個相同的檔案以確保傳輸是快速的可能性。

eMule project 開始於 2002 年 5 月，Merkur 不滿意原始 eDonkey2000 客戶端並且堅信能做得更好。因此他開始製作，聚集了周圍的開發人員，eMule 專案就此誕生。此專案的目標是將 eDonkey 好的部分保留，並增加新功能與改善使用者圖形介面。

2.2.2 ezPeer 與 Kuro

ezPeer 是一款國人自製的檔案共享軟體，為全球數碼科技公司之產品。它以點對點的網路架構技術，讓每一個使用者的電腦可以直接與另一個使用者交換 MP3 音樂或多媒體檔案(包括電影或小遊戲)。早期的 ezPeer 類似 Napster，為第一代點對點架構，有一個主 Index Server 紀錄歌曲來源。改版後的 ezPeer 似無 Index Server。

ezPeer 的特色在於透過驗證主機控管並且對會員收費。它也提供線上立即娛樂、歌曲邊載邊聽、內建 MP3 音樂播放器、即時訊息傳遞、線上音樂聊天室等等功能。

Kuro 為飛行網公司所開發的 p2p 軟體，其推出的主要功能是音樂搜尋與音樂分享，目前為付費軟體，必須申請帳號及付費才能正式使用。會員需透過 Kuro 主機通過付費驗證後，才能進行搜尋及下載。當安裝完 Kuro 的軟體之後，第一次使用此軟體，Kuro 會自動連向固定的 Server 來做溝通，並取得其他使用者的相關資訊，往後在登入 Kuro 時，

不會再連向固定 Server，而是連到第一次記住的其它使用者，藉著其它使用者連上 Kuro 的 P2P 網路。

Kuro 並設有集中索引的 Server，供使用者直接查詢存放音樂之檔案資訊，以進行搜尋。若有找到此音樂，便會向擁有音樂的使用者要求下載。

2.2.3 FastTrack

FastTrack 是由荷蘭程式設計師所研發的技術。FastTrack 可允許使用者交換任何形式的數位檔案，包含音樂、軟體、視訊、電影等，此技術屬分散式架構，使用者要交換檔案並不需要經過任何廠商作中介。

FastTrack 網路在 P2P 的網路上做了一些加強，包含 Supernodes 技術，用戶端會使用更多的頻寬以及 CPU 來儲存快取檔案(Cache Files)以及提供多使用者的連線。這樣的改變大幅加快了搜尋以及下載速度。

這種 P2P 網路串聯了終端使用者本身電腦的處理能力，而當終端使用者人數成長，也就是新的使用者加入之後，則整個網路的處理能力同時跟著本身使用人數成長。

2.2.4 Kazaa

Kazaa(採用 FastTrack 技術)，到目前為止尚未公開其架構，現有的文獻是經由分析該軟體的行為，反推而得的結果[3]。但是推測的架構的確是可行的，而且由非官方單位依其架構開發的軟體 Kazaa-Lite 現在也廣為流行。

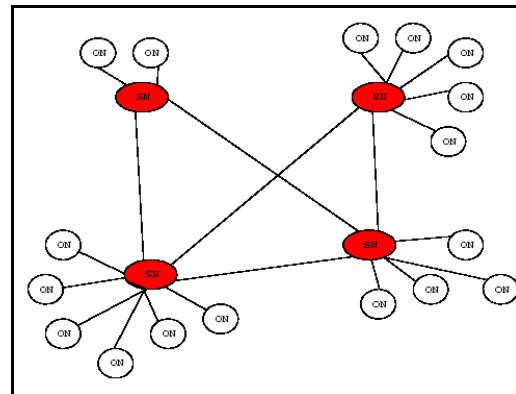


圖 6 Kazaa 網路架構

此架構下也是只有一般用戶端，但是會因各用戶端傳輸能力而被區分為 ON(Ordinary Node)、SN(Super Node)拓撲，如圖 6所示。ON 為一般用戶端，SN 為網路資源大的用戶端(如頻寬大)。不過使用者無法直接得知本身為 ON 還是 SN，因為 ON 可能因為被整個系統判定為高資源者，進而轉變成 SN 來處理較多的事件。由於在這樣的架構下可以提升傳輸速，使得 Kazaa 成為了目前熱門的點對點軟體。

2.2.5 BitTorrent

Bittorrent(簡稱 BT)沒有提供搜尋功能,使用者需自行尋找所需種子(SEED, .torrent 檔)。種子中包含追蹤者(Tracker)網路位置、最初來源網路位置、檔案雜湊值等檔案相關資訊,利用該種子才能下載所需檔案。

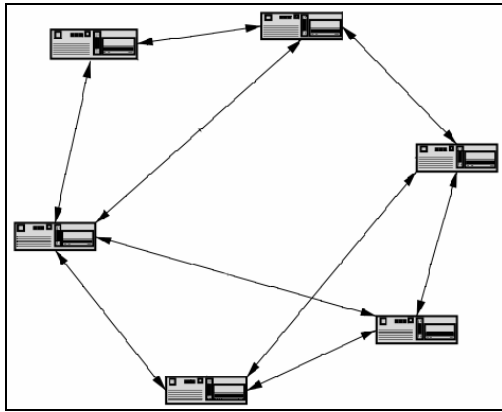


圖 7 BitTorrent 傳輸架構

BitTorrent 的架構由追蹤者、種子來源、用戶端組成。追蹤者在 BitTorrent 佔最重要的角色,追蹤者會輔助尋找檔案,因為追蹤者會紀錄有哪些用戶端擁有所需資料的片段可供下載。種子來源提供取得種子的角色,通常是網站論壇或使用者之間互相交換。點則是每個用戶端,不僅是接收端,同時也是來源端。用戶端加入 BitTorrent 的方式,不需要透過複雜溝通,只要得到種子即可。然而搜尋則相當於是尋找種子的動作。取得種子後,BitTorrent 軟體會解析種子的內容,取得追蹤者的網路位置,並與追蹤者溝通,然後詢問追蹤者是否有擁有資源的來源點,獲得來源點的網路位置後,點就開始向這些點要求傳輸。

2.2.6 Gnutella

Gnutella 係由 AOL 的子公司 NullSoft 開發出來的程式。Gnutella 為一分散式的 P2P 軟體,其主要的功能是提供一個溝通搜尋管道,使用者只要下載 Gnutella 程式,不需透過中央伺服器,即可以匿名的方式,和所有已安裝 Gnutella 的同好交換儲存在其電腦內的音樂和其他檔案(如電影)。Gnutella 之運作是一種類似電話連線網路,呈現一種多重樹形,各節點提供轉接的服務。當登入網路後,首先它會盡可能告訴你,網路現在有多少你的平行同儕(horizontal peers)。不過,你僅認識鄰近同儕,而其中一位會認識其本身之同儕,依此類推而擴展至整個網路。雖然 Gnutella 利用幾個伺服器來暫存同儕的連線資訊,方便新使用者加入,但運作時並沒有集中管理,是完全分散式地,任何一個節點(Peer)擁有完全相等的能力。

3. 各種網路應用之比例分析

網路已經成為生活上與工作上必備的工具,於是網路流量和效能就成為迫切需要重視的問題。而面對這個問題,第一件要研究探討的主題就是:誰在用網路?針對網路流量的增長,不外乎以下幾個主要的原因:

1. 越來越多的網路應用程式產生。
2. 影音資料的網路共享漸漸廣泛。
3. web-based application 的開發已經成為趨勢。
4. 分散式系統的發展越來越普及。
5. 及時的影像、聲音傳遞也越來越發達,例如網路電話等等。

接下來我們要研究分析的重點將放在第一項,也就是網路上所傳送的資料到底都是屬於哪些應用類型。

3.1 網路應用比例研究方法

想要分析網路上的資料屬於何種類型,最簡單的方式就是透過監聽。監聽到封包後,檢查他的封包表頭,透過封包資訊即可判斷這筆資料的類型。判斷完類型,即可計算接下來傳輸的流量,直到連線結束為止。

我們根據網路上較為熱門的項目,選擇以下分類方式,針對交大資工與交大學生第十三宿舍所測得之資料來統計結果:

包括 HTTP, FTP, SMTP, POP3, P2P (ezPeer, Kuro, eMule, BT)以及其他等。

3.2 網路應用比例研究結果

統計資料結果如圖 8、所示。在這兩個結果中我們發現,交大資工系館 P2P 軟體佔據約 30% 的頻寬!而交大宿舍的分析則可以更明顯看出, P2P 和 FTP 的傳輸量佔據絕大部分的頻寬。

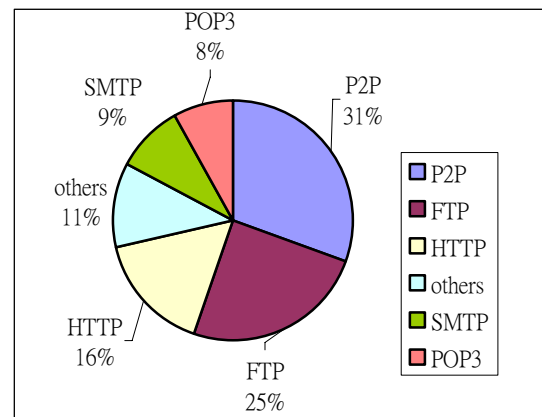


圖 8 交大資工流量分析結果

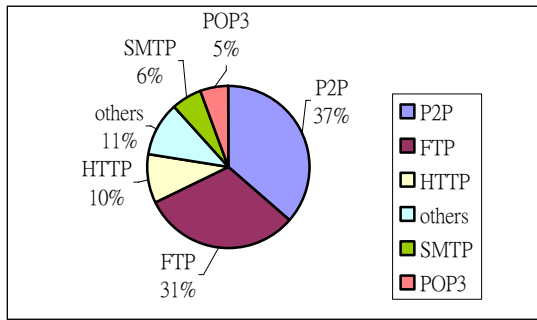


圖 9 交大第十三學生宿舍分析結果

美國版權工業，例如唱片業、動畫業協會的一個代表機構曾經做過研究指出，P2P 幾乎佔用美國大學網路 75% 的頻寬。而我們的研究結果只有 30%~40%，主要是因為交大系館或宿舍中，學生目前仍然以 FTP 傳輸為主要檔案交流方式。但 P2P 的頻寬已經凌駕在 FTP 之上，可見得 P2P 已經慢慢取代傳統 FTP 而成為主流。

4. 各種 P2P 軟體之比例分析

接著，我們針對以下幾個 P2P 應用軟體：eDonkey、eMule、BitTorrent (BT)、Kazaa、Kuro、ezPeer 等，藉由攔截並分析網路上之封包，來分析這些 P2P 軟體在網路上使用的比率。

4.1 P2P 軟體比例研究方法

針對 P2P 軟體種類流量的分析，我們藉由監聽網路封包的方式，配合各個 P2P 軟體的通訊協定、封包開頭特徵字串，以及所使用的通訊埠[5] (如表 1)，來區分不同的 P2P 軟體，以此統計各種 P2P 軟體在 P2P 流量中所佔的比例。

表 1 常見 P2P 軟體封包的開頭字串及預設通訊埠

軟體種類	封包開頭字串	傳輸協定	預設通訊埠
eDonkey	0xe319010000	TCP/UDP	4661-4665
	0xc53f010000		
Kazaa	0x270000002980	TCP	1214
BitTorrent	0x13Bit	TCP	6881-6889
Kuro	"?0B0"	TCP	6600-6802
ezPeer	0x474e5554454c	TCP/UDP	6800~6890
			7880~7890 8860~8890

4.2 P2P 軟體比例研究結果

如圖 10 所示，在我們所分析的 P2P 網路流量中，以 BT 所佔的比例最多，佔 P2P 總流量的 44%。第二名為使用相同協定的 eDonkey 及 eMule，佔總流量的 36%。而以分享音樂為主的 ezPeer 及 Kuro

則佔 10%。相較之下，在台灣使用較不普遍的 Kazaa 則是佔 5%、其它 P2P 軟體約佔 5%。我們藉由實際分析網路流量所得之軟體比例表，與另外一項使用問卷調查之相關研究[17]的結果類似，也是 BT 與 eDonkey (eMule) 佔絕大多數。

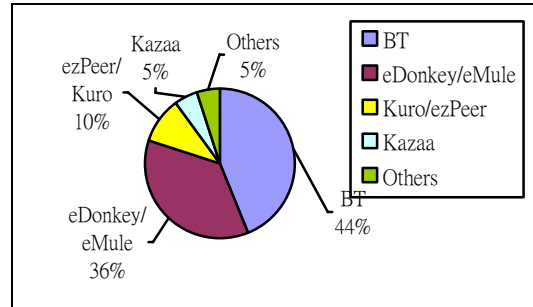


圖 10 各 P2P 分享軟體使用比例

5. P2P 分享內容比例

P2P 分享內容依其副檔名可分成好幾類，如音樂檔、影像檔、圖檔、壓縮檔等，透過其檔案副檔名所佔的比例，可得知目前 P2P 傳送最多的資料及佔用網路頻寬最大的為何種檔案類型。我們藉由隨機取樣之方式，觀察得出網路上分享的內容主要以音樂與影片為主，其次是應用軟體。

在接下來的兩個章節，我們假設在 P2P 分享內容中並沒有偽造檔案發生[1]，也就是說使用者搜尋到的分享內容並非無法開啟的錯誤檔案，亦不是檔名與檔案內容不相符的分享內容。

5.1 P2P 分享內容比例研究方法

為確保取樣過程之公正性，在選取採樣資料字詞來源時，特別注意應選取全體字詞中具有普遍代表性者，而避免刻意選定特定檔案、個人等傾向慣用之字詞。因此，我們使用微軟相關字詞編輯工具共 4,017 個常用字，以及微軟相關字詞編輯工具中，每個常用字配對的第一個最常用詞共 4,017 個常用詞，作為搜尋字、詞之來源。

接著我們隨機將上述 4,017 個常用字及常用詞分配給數名研究人員登入各種 P2P 分享軟體，依序將常用字及常用詞作為搜尋字串，當搜尋到每個字詞的檔案時，將檔案名稱依照不同的副檔名進行排序，使擁有相同副檔名的檔案在 P2P 軟體的搜尋表單內串連在一起，以方便統計人員進行統計，然後將每種副檔名的檔案個數記錄下來，得到出每個 P2P 軟體不同副檔名檔案的個數，並將相同副檔名的檔案個數加總後取平均，最後統計出各種 P2P 軟體分享內容比例的平均值。

5.2 P2P 分享內容比例研究結果

統計的結果如圖 11 所示。其中以副檔名為 mp3 數量最多，佔總比例六成以上，副檔名為 avi、rmvb 次之，再來是副檔名為 wmv、mpg，其他像副檔名為 jpg、zip、rar 等檔案所佔的比例就明顯比其它三者小很多，此外還有接近兩成的封包為其他類型檔案，並不在上述的幾種副檔名之內。

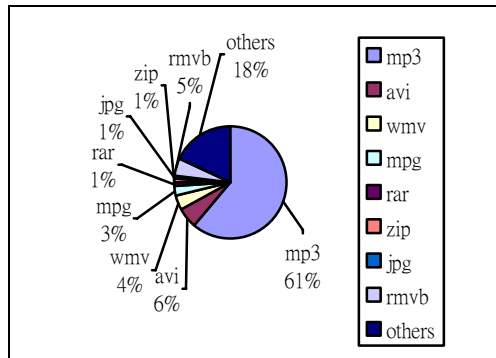


圖 11 網路分享内容比例

從封包所佔比例的角度來看，副檔名為 mp3 (即音樂檔) 的數量最多，有超過六成比例的 p2p 使用者在搜尋或是傳輸音樂檔。再來是副檔名為 avi 及 wmv 的檔案(即影像檔)，而影像檔根據實際搜尋結果，大部分影像檔為電影檔；再來是副檔名為 mpg 的檔，也是同樣為電影檔。不過 p2p 分享及傳輸的音樂檔、影像檔，大部分都是不合法的檔案，其中又以音樂檔居多，所以對目前的音樂產業會有一定的衝擊。

NTT technical review[2]也在 2004 年發表類似的研究，該研究指出的結果跟目前所研究的結果大致上相符合，顯示出 P2P 軟體對網路使用生態所帶來的衝擊性。除此之外，P2P 軟體的快速發展，短期間內的變動是相當大的，從第三代的 P2P 軟體 BitTorrent 崛起開始，比起 emule 等不特定分享檔案的軟體，更容易的傳輸分享大型檔案，比起一首 mp3 音樂更容易佔用大量的網路頻寬。

6. 結論

根據本篇論文的研究，P2P 軟體已經逐漸與 FTP 軟體平起平坐，成為新一代分享檔案的利器。然而因為 P2P 軟體的分散式架構與搜尋後傳輸的特性，使得網路使用環境有相當大的改變。不僅是網路頻寬受到擠壓，分享檔案的智慧財產權問題更是造成創作人本身權益受損。

根據我們的研究結果，有將近六成的分享内容屬於 mp3 音樂檔案，並且這些檔案中超過八成是屬於未經合法授權就在網路上散播的檔案。正因如此，有許多學者專家致力於研究數位智慧財產權管理(DRM-Digital Right Management)。

針對 P2P 分散式架構，有心人士可以稍加利用，轉而成為攻擊的跳板。不論是大量發送廣播搜

尋封包來癱瘓整個網路，還是假造資料來源使得某個受害主機成為熱點 (Hot Spot)，都是變相的分散式網路阻絕攻擊(DDoS)。因此，目前已經有許多防火牆加入阻斷這類 P2P 封包的功能，然而網路第七層的內容過濾往往使得防火牆負載過高，反而導致影響程度更為加重。在 P2P 軟體成為新一代網路應用且快速發展時，這樣的影響仍會持續加深。因此，P2P 應用成為必然的趨勢，如何限制互相溝通的內容可能才是未來需要努力的課題，例如分享軟體只能分享受到 DRM 管制許可之檔案。

參考文獻

- [1] J. Liang, R. Kumar, Y. Xi, and K. Ross. Pollution in P2P file sharing systems. In IEEE Infocom, Miami, FL, USA, March 2005.
- [2] Keita Ooi, Satoshi Kamei, and Tatsuya Mori, "Survey of the State of P2P File Sharing Application," Ntt Technical Review, May 2004 Vol.2 No.5.
- [3] Nathaniel Leibowitz, Matei Ripeanu, and Adam Wierzbicki. Deconstructing the Kazaa network. In WIAPP '03: Proceedings of the The Third IEEE Workshop on Internet Applications, page 112, Washington, DC, USA, 2003. IEEE Computer Society. ISBN 0-7695-1972-5
- [4] Stefan Saroiu, P. Krishna Gummadi, and Steven D. Gribble. A measurement study of peer-to-peer file sharing systems. In Proceedings of Multimedia Computing and Networking, 2002, January 2002.
- [5] T. Karagiannis, A. Broiodo, N. Brownlee, kc claffy, and M. Faloutsos. Is P2P dying or just hiding? In Globecom, Dallas, TX, USA, November 2004.
- [6] Bittorrent, Available: <http://www.bittorrent.com/>
- [7] Edonkey, Available: <http://www.edonkey2000.com>
- [8] Emule, Available: <http://www.emule-project.net/>
- [9] ezPeer, Available: <http://www.ezpeer.com/>
- [10] Gnutella., Available: <http://www.gnutella.com/>
- [11] Kazaa, Available: <http://www.kazaa.com/us/>
- [12] Kuro, Available: <http://www.kuro.com.tw/>
- [13] Packeteer, Inc., Strategies for Managing Application Traffic. [Online] available: <http://www.packeteer.com/resources/prod-sol/MngAppTraf.pdf>
- [14] Project JXTA. <http://www.jxta.org/>
- [15] <http://taiwan.cnet.com/news/software/0,2000064574,20099823,00.htm>
- [16] <http://en.wikipedia.org/wiki/FastTrack>
- [17] 劉怡玟。2004。點對點檔案分享軟體使用行為之研究。國立中山大學傳播管理研究所。