國 立 交 通 大 學
資 訊 工 程 學 系
碩 士 論 文

P2P 軟體對網路影響之研究

A Study on the Impact of P2P Software

指導教授：蔡文能　教授
研 究 生：賴俊廷

中 華 民 國 九 十 四 年 十 月

P2P 軟體對網路影響之研究

A Study on the Impact of P2P Software

指導教授：蔡文能　　Advisors：Wen-Nung Tsai

研 究 生：賴俊廷　　Student：Chung-Ting Lai

國 立 交 通 大 學

資 訊 工 程 研 究 所

碩 士 論 文

A Thesis Submitted to
Institute of Computer Science and Information Engineering
College of Electrical Engineering and Computer Science
National Chiao Tung University
in Partial Fulfillment of the Requirements
for the Degree of Master
in
Computer Science and Information Engineering

July 2004

Hsinchu, Taiwan, Republic of China

中 華 民 國 九 十 四 年 十 月

P2P 軟體對網路影響之研究

學生：賴俊廷　　指導教授：蔡文能 教授

國立交通大學資訊工程學系（研究所）碩士班

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。

A Study on the Impact of P2P Software

student：Chung-Ting Lai　　Advisors：Dr. Wen-Nung Tsai

Institute of Computer Science and Information Engineering

National Chiao Tung University

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.

致　　謝

在兩年的努力之下，終於完成了我的碩士論文，中間經過了許多的困難與問題，也受到許多人的幫助，在此一一感謝。首先要感謝的是我的指導教授-蔡文能教授，在碩士班的兩年，他給我許多方面的指導，讓我受益良多，也成長了不少。再者，感謝父母給我一個平穩的環境，讓我能夠專心學習，還有心靈上的支持。接著要謝謝的是實驗室的學長姐、同學們，大家一起工作、學習，互相幫忙，讓我有個難忘的碩士回憶。

目　　錄

表　目　錄

圖　目　錄

第一章 緒論

1990 年代網際網路興起，網路使用者數量急遽上升，越來越多應用都是透過網路來進行。到了 1990 末期，隨著用戶端設備功能日漸增強，以及網路頻寬加大，點對點應用儼然成為最熱門的網路傳輸方式。雖然目前已有許多不同種類點對點網路架構，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。由於 P2P 技術的成熟與進步，P2P 軟體隨手可得，目前的 P2P 軟體大部分都應用於檔案的分享，而造成了網路頻寬的消耗，版權的保護也成為 P2P 技術令人爭議的問題。

1.1 動機與目的

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。

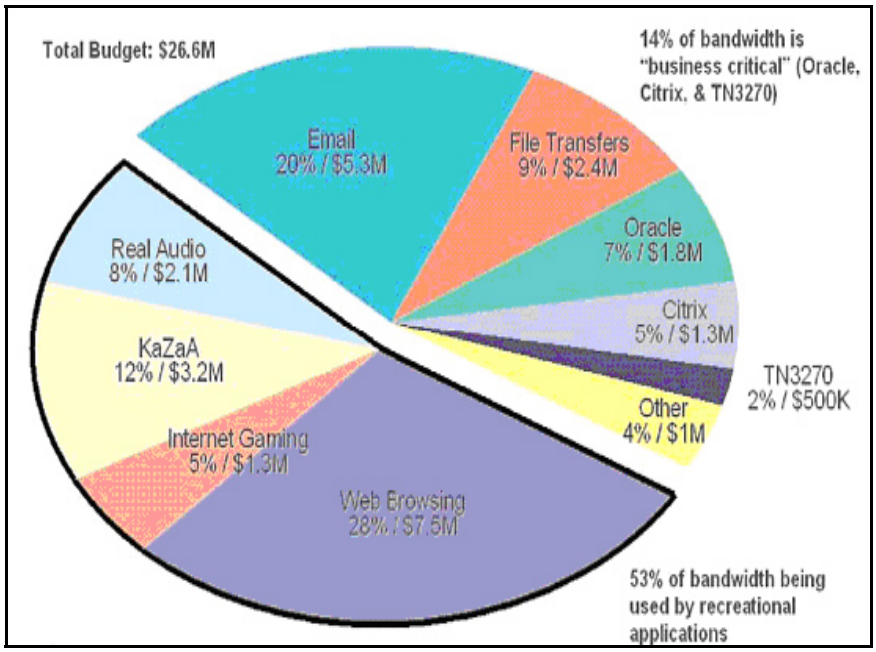

圖 1 各種網路應用比例圖（資料來源[33]）

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。

P2P 是 Peer-to-Peer 之簡稱，是一種分散式的網路技術，在 P2P 網路中每台電腦同時身兼兩種角色：Client 和 Server。任兩台電腦間可以直接連線而不用透過伺服器，可以對等的互相分享資源，有別於過去個人電腦，一定要連結上某個網站或公司的伺服器，才可以下載取得檔案資料的主從式(Client/Server)架構作業模式。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，除此之外，萬一該用戶故障，則和該用戶相關的資訊不僅全部斷除，更需要花上許多回復成本及浪費頻寬。點對點應用上所傳輸的內容及流量對整體網路的壅塞，是否也是影響了網路效能的原因。

在 P2P 的發展過程中，產生「集中式（centralized）」與「分散式（decentralized）」兩類不同的架構，雖然都採用分散式檔案共用的設計，但集中式 P2P 為爭取搜尋效率，在中央伺服器中建有目錄，提供檔案名稱或索引之管理；分散式 P2P 則未利用到中央伺服器，所有搜尋、傳輸及重製均發動及完成於使用者之間，在這個搜尋以及下載檔案的過程中，所消耗的頻寬，對於網路的效能造成了很大的影響。

根據 Yankee Group research firm 在 2002 年對美國 P2P 軟體使用情形所做的調查結果顯示，當時在美國約有 5 千 7 百萬的人使用 P2P 軟體，而透過 P2P 軟體分享的檔案數量超過 50 億，此數據持續增加。雖然在網路中流通的檔案數量極為龐大，但是著作權人(影音業者)卻很難從中獲取利潤，甚至因為 P2P 軟體的普及而招致損失。RIAA 便表示消費者利用 P2P 軟體非法下載 MP3 造成唱片銷售量的減少，損失金額至少超過數百萬美元。

在本篇論文中，我們希望能在真實的環境之下進行封包的搜集，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，以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。

1.3 論文架構

論文將以點對點傳輸架構演進的順序，說明其改進的緣由以及優缺點，並介紹目前熱門的點對點檔案分享軟體，同時講述相關知識。另外，再藉由分析網路流量來探討各種 P2P 軟體使用的比例以及 P2P 軟體使用情形對網路的影響。

本論文的整體架構如下，第二章介紹與本論文相關的背景知識，包括 P2P 技術的應用、P2P 架構的演進以及目前熱門的 P2P 熱門應用軟體。第三章探討目前提出分析網路流量的方法，如協定基礎分析方法（Protocol-Based Traffic Analysis）和資料基礎（Packet Payload Based Analysis）分析方法。第四章說明本論文用來分析網路流量方法，首先說明針對網路各種應用如 HTTP、FTP、POP3 等所佔網路流量比例的分析方法，然後再講述分析 P2P 流量中各種軟體所佔的比例以及 P2P 檔案分享內容所佔比例的方法。第五章列出我們分析的結果包括網路應用比例、P2P 軟體使用比例、P2P 檔案分享比例。第六章對本論文研究進行討論、提出結論，並說明未來可能的研究方向。

第二章 背景知識

在此章節將一一介紹點對點應用的實例、分類以及相關知識。2.1 介紹實際點對點的應用包含檔案下載以及最熱門的即時通訊軟體。2.2 將點對點架構依照其發展的時間與特性作分類說明。2.3 介紹目前常見的 P2P 軟體。2.4 介紹如何截取網路傳輸的封包。

2.1 點對點應用(Peer to Peer Application)

點對點應用主要概念是在現有實體網路上做邏輯佈設，而非重新建構新實體網路，用戶端不管選用哪種邏輯架構，最終資料傳輸與訊息傳送，還是藉由實體網路的 IP 傳輸。兩個用戶端之間各種傳輸，不管是直接傳輸或透過其他用戶端的間接傳輸，基本上都是兩個用戶端間互相溝通傳遞。然而這些邏輯架構的優劣比較就在於佈設上是否穩固、有效率及所提供服務的多寡。在此小節，我們將介紹目前 P2P 技術最熱門的應用：檔案下載和即時通訊軟體。

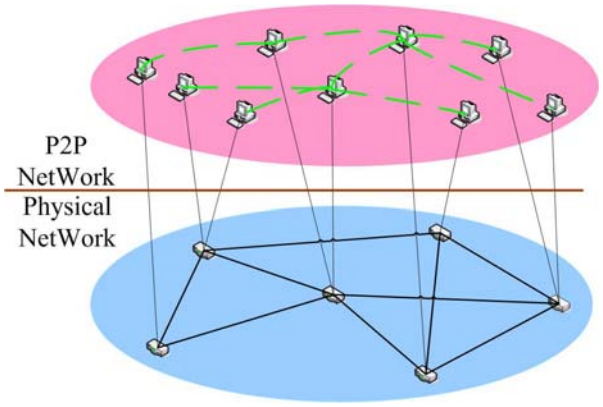

圖 2 邏輯架構與實體網路的對應

2.1.1 檔案分享（File Sharing）

由於網路頻寬的改善、電腦硬體設備的提升、P2P 軟體技術的進步，加上現代人使用科技的能力提升，越來越多人將 P2P 檔案分享軟體作為檔案取得的來源，根據調查報告指出，2003 年九月分台灣共有 335 萬人曾造訪檔案分享網站，200 萬人曾經使用過 Kuro[20]下載音樂，由此可見 P2P 技術已成為最熱門的檔案分享技術。

在 P2P 檔案分享的領域中，可以分成三個階段的應用軟體。首先是最早的 P2P 檔案分享軟體 Napster[15]，其功能很單純，就是純 MP3 音樂檔案的分享，此軟體具有一個集中式伺服器，做為已分享檔案的索引，因此整個網路的資源消耗較低。檔案在傳輸過程也沒有加解密，明顯的優點就是簡單使用和延展性較高。第二階段的代表軟體是 Gnutella[16]，其改成純分散式的架構，沒有集中式的伺服器，相對地網路的延展性變低，容易造成壅塞。Gnutella 支援的檔案格式比較多，幾乎所有的多媒體格式皆可，使用者在下載過程也無法匿名或資料加密。最後是 Freenet[32]，如其名，給予使用者完全自由的網路空間，提供不同資源的分享，不侷限在檔案種類。在 Freenet 中，檔案本身會依據熱門程度，適當的移動或複製到需求量高的位置，相對地，一旦一個檔案在 Freenet 上乏人問津，時間一久即會自然消失。

P2P 技術蓬勃發展，目前約有上百種的 P2P 檔案分享軟體，由於軟體的發展大都採用開放原始碼的方式（Open Source），因此種類仍然繼續增加中，雖然各種軟體的使用介面和功能大不相同，但底層所使用的技術原理皆相同，就是整合伺服器、客戶端和路由功能於單一的軟體中，讓個人電腦不需透過特定的伺服器進行連線及分享資源，而能分享的檔案類型包括音樂、影片、軟體、檔等。

eMule、eDonkey、BitTorrent 是目前台灣最熱門的點對點檔案分享軟體，WinMax、Kuro、Ezpeer 可算是目前台灣最熱門的點對點音樂分享軟體。這些軟體的主要目的就是做到資源共用。

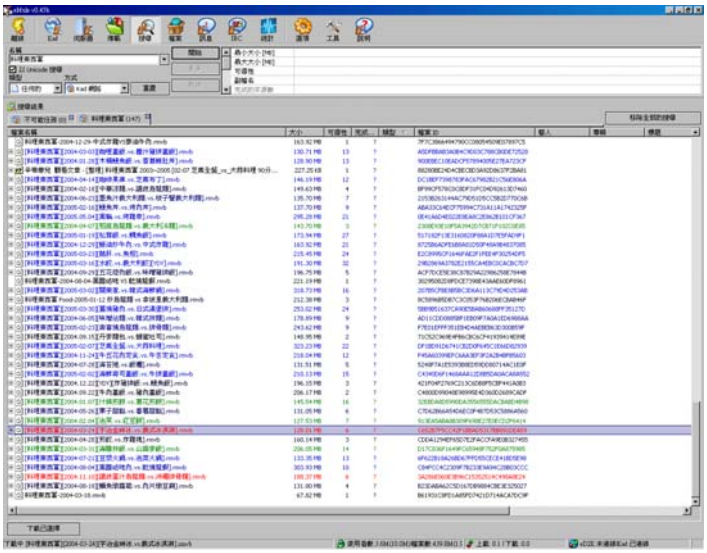

圖 3 搜尋介面(eMule 搜尋介面)

搜尋可謂此類軟體的主要服務之一，除了 BitTorrent 必須由使用者自行尋找所求檔案的種子(SEED)之外，其他軟體皆提供搜尋功能與介面(如圖 3)，使用者可輕易找出所要檔案的來源用戶端。然而由於 P2P 網路是屬於分散式的架構，因此一個良好的搜尋方式影響了軟體對使用者的接受度。因此搜尋廣度、搜尋速度、檔案可取得性、搜尋頻寬等都影響了使用選擇該軟體的意願。

	eMule	eDonkey	BitTorrent	KaZaA	Kuro	Ezpeer
架構世代	1/3	1	N	2	1	1
搜尋功能	有	有	無-SEED	有	有	有
通過 NAT	不可	不可	不可	可	不可	不可

表 1 台灣常見點對點軟體比較

2.1.2 網路即時通訊（Instant Message）

P2P 技術的另一項應用即時通訊，可謂現今最熱門的網路應用軟體，從早期的 ICQ 到 MSN、Yahoo Messenger、Skype。其使用的技術基本上都是基於點對點傳輸之上。尤其後來紛紛加入語音通訊（VoIP）的功能，使得點對點傳輸特性更為明顯。即時通訊的概念相當容易，藉由設備將用戶端發出的訊息包裝成封包，經由媒體傳輸到對方接收設備，可處理這類工作的設備不再限於電腦，目前已有商業販售的其他硬體設備。概括分類這些 VoIP 應用，依演進過程可分為 PC to PC、PC to Phone、Phone to Phone 等三種，以圖 4 說明這三種系統的整體架構。

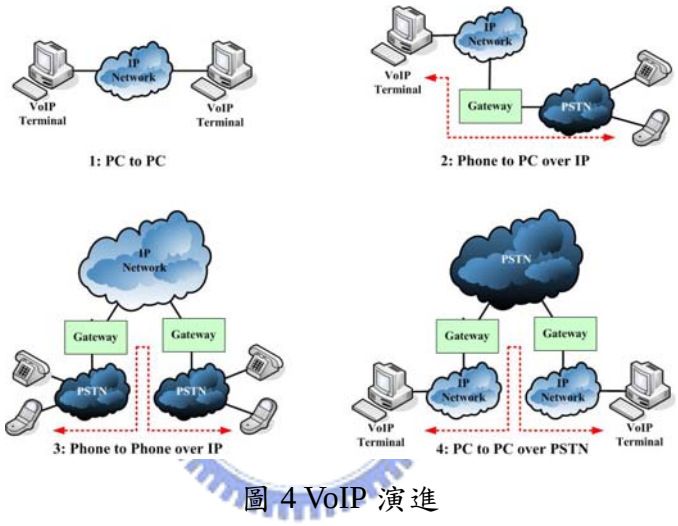

圖 4 VoIP 演進

這些 VoIP 應用中最引人注目的應屬 Skype，因為其音質最佳。除了採用適當的編碼 GIPS iSAC and iLBC codecs[31]之外，其封包繞送也有別於一般 VoIP 軟體的直接點對點傳輸。雖然其架構到目前為止尚未正式公開，但是分析[7]指出，其採用了特殊的應用層路由(Application layer routing)來輔助封包傳送。藉充分利用用戶端網路設備來增加傳輸的穩定性以求提供更好的通話品質，然而這類作法在實際產品 KaZaA -Lite 中證實是可行的[29]。表 2 為針對目前最熱門的即時通訊軟體所做的簡單優劣比較。

	ICQ	Yahoo Messenger	MSN	Skype
通話品質	可	可	可	優
視訊	有	有	有	無
通過 NAT	不可	不可	不可	可

表 2 台灣常見及時通訊軟體比較表

2.2 點對點傳輸架構世代

本節以演進時間作為對點對點邏輯架構的分類：主從架構(Client and Server)、第一代、第二代、第三代。此外各小節也將講述各個架構的基本概念和限制，並提出優缺點。

2.2.1 主從式傳輸（Client and Server）架構

主從架構是最早透過網路傳輸的方式。由用戶端(Client)和伺服器(Server)組成。雙方必須以兩方皆認同的通訊埠(Port)和通訊協定(Protocol)來溝通。常見應用首推檔案傳輸協定(FTP)。加入主從式傳輸架構用戶端只需要知道伺服器網路位置、通訊埠與通訊協定即可。用戶端搜尋資料時，直接向伺服器發出詢問即可。

此一架構的優點歸納說明如下：

- 通訊協定簡單。
- 一般擁有較穩定的傳輸速度。
- 容易　　。
- 伺服器可有效控管檔案的讀取權限。

此一架構的缺點歸納說明如下：

- 有固定的伺服器，容易成為攻擊的目標，一旦伺服器被攻破或者故障，則資訊的來源隨即停止。

- 分享服務只能由伺服器端提供，用戶端要分享資源時，必須經由伺服器同意，然後上傳到伺服器。造成檔案分享的限制。
- 用戶端必須知道伺服器相關的連線資訊。
- 伺服器的負載重，除了提供搜尋還要負責傳輸資料。

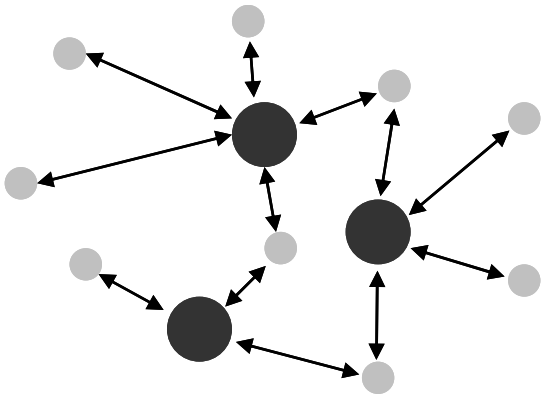

圖 5 主從傳輸架構

2.2.2 第一代點對點傳輸架構

第一代的 P2P 網路架構簡單地說，就是具有索引伺服器(Index Server)輔助檔案分享架構。主要是由索引伺服器和一般的點(Peer)組成，點即為用戶端。索引伺服器顧名思義，就是伺服器上擁有大型集中式索引，其中記錄相連點有哪些資源可供分享。此架構的運作方式，如同主從式架構，點只需知道索引伺服器網路資訊即可。當點搜尋資料時，直接向索引伺服器詢問，而索引伺服器則回傳擁有所需資料的用戶端資訊。所需資料的實體來源是回傳中的部分用戶端，並非索引伺服器。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構的代表。

此一架構的優點歸納說明如下：

- 不需處理檔案傳輸，伺服器負載減輕許多。
- 一般狀況下，因為來源多所以傳輸速度提升許多。
- 通訊協定簡單，用戶端不需維護其他用戶端資訊。

此一架構的缺點條歸納說明下：

- 已經被認定為非法的檔案分享架構。美國唱片工業協會（Recording Industry Association of America，RIAA）於一九九九年十二月六日對 Napster 提起訴訟，經過一年多的纏訟，舊金山聯邦上訴法院在二〇〇一年二月十二日，裁定 Napster 構成對於錄音及音樂著作之著作權人重製權與散佈權之「輔助侵害（contributory infringement）」與「代理侵害（vicarious infringement）」並下令要求 Napster 內部必須建置篩選機制，封鎖受版權保護的歌曲經過其 server。[35]
- 因為有固定的索引伺服器，容易成為攻擊目標，一旦索引伺服器被攻破或者故障，則搜尋功能完全喪失。

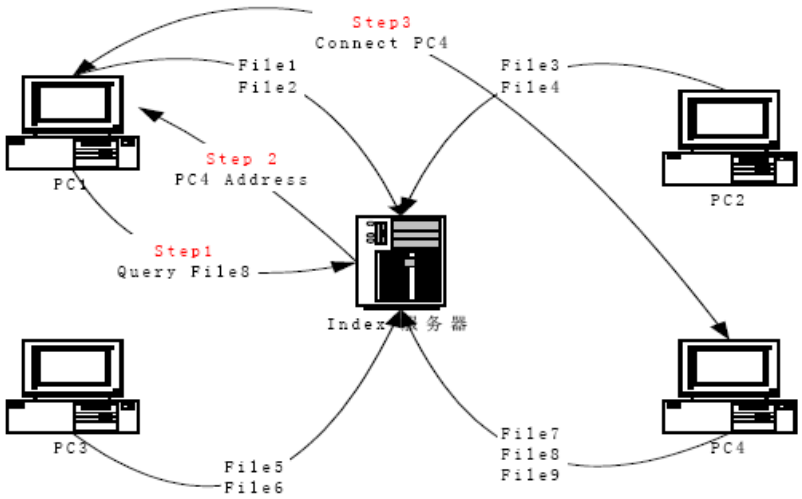

圖 6 第一代點對點傳輸架構

2.2.3 第二代點對點傳輸架構

第二代的點對點傳輸架構主要是為解決第一代架構中索引伺服器所引發的各種問題，尤其是違法問題。在第二代的架構下整個邏輯網路皆由點（Peer）所組成。所以架構中的服務必須由各點互相合作，因此出現鄰居(Neighbor)的概念，因為網路資源分散各處，所以此架構又被稱為分散式服務。每點除了為接收端之外，亦會建立本地端可分享資源的索引表，提供其他點搜尋。加入此類架構，必須藉由已在架構下的點輔助資訊交換認識其他鄰居。當點搜尋資料時，只需將搜尋封包傳送給所認識的鄰居。

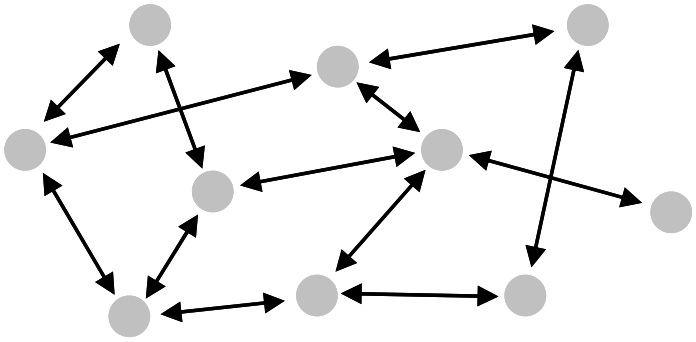

圖 7 第二代點對點網路傳輸架構

此世代的代表性軟體為 Gnutella，其搜尋時將搜尋封包傳送給所有鄰居，然後鄰居各自回應原始點，並轉送搜尋封包給第二層鄰居，如此遞迴下去。此種作法，造成相當嚴重的搜尋洪流(Query Flooding)。被後來 FastTrack[24] (即 KaZaA 的底層技術)迅速掘起取代其地位，改為搜尋封包在點間傳送，直到找到檔案實體才將搜尋結果回傳給原始點，並非每一點都做回應。而這種技術仍相當不便，容易產生無用的搜尋迴圈，當大量搜尋同時執行將造成網路效能大幅下降。此問題最簡單的處理方式就是限定搜尋深度(Gnutella 限制往下搜尋七層)，但是這個作法讓搜尋只限定於邏輯網路的某些區域[15]。

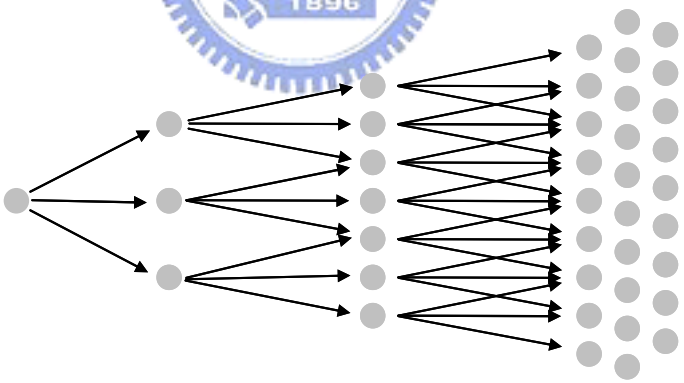

圖 8 Gnutella 搜尋邏輯

但是限制搜尋深度仍然消耗很大的網路成本。以 Gnutella 來說，T 為搜尋深度，N 為每個用戶端相連的個數，則每次搜尋所影響的用戶端個數如圖 9。

	T=1	*T=2*	*T=3*	*T=4*	*T=5*	*T=6*	*T=7*
N=2	2	4	6	8	10	12	14
N=3	3	9	21	45	93	189	381
N=4	4	16	52	160	484	1,456	4,372
N=5	5	25	105	425	1,705	6,825	**27,305**
N=6	6	36	186	936	4,686	**23,436**	**117,186**
N=7	7	49	301	1,813	**10,885**	**65,317**	**391,909**
N=8	8	64	456	3,200	**22,408**	**156,864**	**1,098,056**

圖 9 Gnutella 的搜尋影響

此一架構的優點歸納如下：

- 消除了中央伺服器帶來的不利因素，由於沒有中央控制點，不會因為一點故障就導致癱瘓，是真正的分散式網路

此一架構的缺點歸納如下：

- 雖然利用限定搜尋深度，但是依然造成網路壅塞，而搜尋封包在網路中來回傳送，造成搜尋時間過長
- 只能做到區域性搜尋(Local Search)。
- 較多人知道的點往往成為熱點(Hot Point)。造成該點負載遠高於其他點。
- 如果預知點剛好都不在網路架構下，則等於無法加入該網路架構。

2.2.4 第三代點對點傳輸架構

第三代的點對點傳輸架構是目前 P2P 軟體主流的技術，主要是因為加入分散式雜湊表(DHT，Distribute Hash Table)的概念，加快了檔案搜尋的速度，搜尋不再像第二代中利用點和點間封包轉送來做搜尋，取而代之的是有系統、漸進式的尋找。所謂分散式雜湊表概念，即利用雜湊概念來定位用戶端及資料的邏輯位置。在此網路架構下，用戶端利用 IP、Port 等網路屬性做為其特徵值，而資料通常以資料的檔案名稱、大小、類型等做為特徵值，再藉由這些特徵值經由 Hash Function 計算出用戶及資料的編號。

當用戶端的用戶端編號算出後，經由一個已在該邏輯網路架構下的用戶端協助，進入該邏輯網路架構。進入之後，用戶端將取得鄰居資訊，並藉由一連串的動作觸發相關鄰居，更新相關鄰居資訊。而當資料的資料編號算出後，該資料的仲介資料(Meta Data)將被應對到某用戶端上。所以搜尋時，只要算出資料編號即可找到該資料的仲介資料，進而取得資料實際位置資訊。另外，資料總數往往遠多餘用戶端總數，所以在此架構的資料編號空間都會大於用戶端編號空間，一般皆採取餘數的概念來解決此問題，以圖 10 為例，編號為 0122 的用戶端，會負責所有資料編號餘數為 0122 的資料。

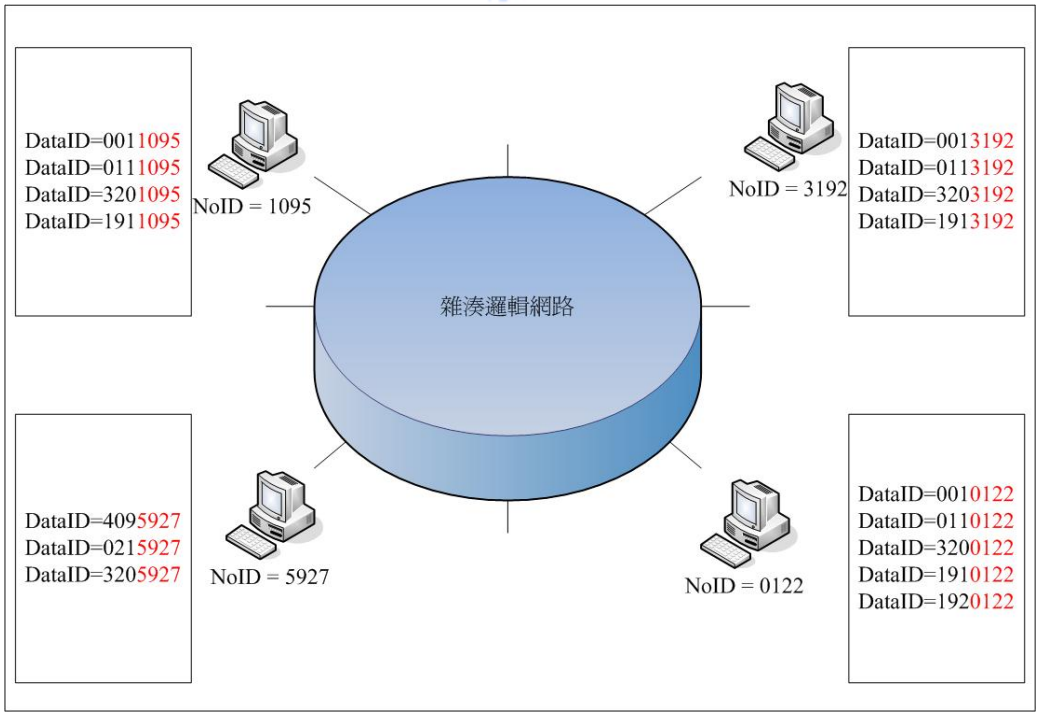

圖 10 資料與用戶對應示意圖

搜尋時，點會計算出要搜尋資料相對應的資料編號，然後藉由該資料編號計算出屬於哪一個用戶端編號，然後從所有鄰居中找出最接近該用戶端編號的一個，送出查詢封包。基本上，此法解決了搜尋洪流，同時也不需限制搜尋長度，相對地，可做到廣域搜尋(Global Search)。再者良好的鄰居選擇策略可以降低搜尋跳躍數目。

第三代點對點傳輸架構則是以 EDonkey-Kademlia 及 Morpheus[26]為代表，此外還有一些較小的獨立軟體開發商，配合特殊的改善方式，使這些工具比以前更有效率。根據網路監視公司 BayTSP[30]的 2004 年 9 月使用者數量評比，eDonkey[18]超越 KaZaA[25]成為全世界最受歡迎的點對點傳輸軟體。

此一架構的優點歸納如下：

- 解決搜尋洪流。
- 可做到廣域搜尋。

此一架構的缺點歸納如下：

- 每點的責任相對變大，點離開將會造成搜尋失敗率變高。而且必須花額外的頻寬來維護點與資料的關係。
- 部分架構加入其他技巧如樹狀搜尋、快取等方法加速搜尋，但同時也增加許多維護訊息。

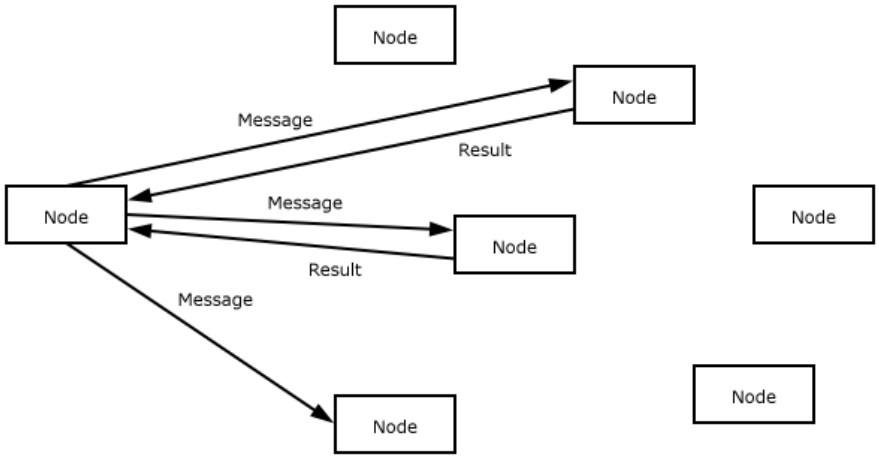

圖 11 第三代 P2P 網路架構

2.3 常見分享軟體簡介

目前比較常見的 P2P 軟體有 eDonkey[18]、eMule[17]、ezPeer[21]、Kuro[20]、BitTorrent（BT）[27]、KaZaA[25]、Gnutella[16]、FastTrack[24]等，以下各小節分別簡介其運作方式。

2.3.1 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，此因於 Merkur 不滿意原始 eDonkey2000 客戶端使用介面，並且堅信能提升程式的效能，建立更穩定的網路架構，因此他聚集了周圍的開發人員，開始 eMule 專案。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，由於 Merkur 等人的努力，eMule 成為了最受歡迎、最可靠的 P2P 檔案分享軟體。

因為 eMule 是採取開放原始碼的方式進行開發，eMule 更新速度快，其頻率約在 1～3 週，便有新版本的出現，每個新版的釋出使這網路更為有效率。eMule 具有排隊和額度系統，有助於確保每個人經由上載回饋的方式，可以順利的在網路取他想要的檔案。eMule 還具有自動檢查下載的檔案是否損壞的功能，以確保檔案的正確性，eMule 智慧損壞控制有助於快速矯正損壞的部分。其自動優先權及來源管理允許您一次下載許多檔案而不須監視他們。除此之外，預覽功能允許您在下載完成之前查看您的影像或檔案。在搜尋檔案時，eMule 提供了一個大範圍可能的搜尋，包含了：伺服器(本地和全球)、Web 基礎(Jigle 和 Filedonkey)以及 Kad 網路，eMule 也允許您使用非常複雜的布林搜尋使搜尋更為的靈活。eMule 還具有即時通訊的功能，可以傳送訊息到其他的客戶端並可將他們加入成為好友，以內建的 IRC 客戶端，能在全世界和其他的下載者聊天。

2.3.2 FastTrack

FastTrack 和 Kazaa 是由荷蘭人 Scandinavians、Niklas Zennstrom 和 Janus Friis 所設計的，2001 年 3 月，在一家荷蘭公司 Consumer Empowerment 的產品中最先使用。那時候正是第一代 P2P 網路的末期。FastTrack 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，此軟體屬於第二代的 P2P 架構，使用者要交換檔案並不需要經過任何廠商作仲介。

FastTrack 在 P2P 的網路上做了一些加強，包含 Super nodes 技術，如果一個用戶端電腦性能足夠，而且網路連接速度快，那它就自動成為 Supernode。Supernode 為那些速度較慢的節點提供索引服務，這樣的改變大幅加快了搜尋以及下載速度。一開始的時候，用戶端中保存了一個 supernode 的 IP 位址的列表，它會試圖連接這些 IP 直至找到一個在網路中的 supernode，它會向這個 supernode 索取當前活動的 supernode 的列表，並更新自己保存的列表。用戶端也會把自己分享的檔列表告訴 supernode，也從 supernode 查詢自己想要的檔案，當查詢到檔案的位置時，就會直接與檔的所有者連接，並透過 HTTP 協議進行下載。

2.3.3 Kazaa

Kazaa(採用 FastTrack 技術)，到目前為止尚未公開其架構，現有的文獻[13]是經由分析該軟體的行為，反推而得的結果。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。

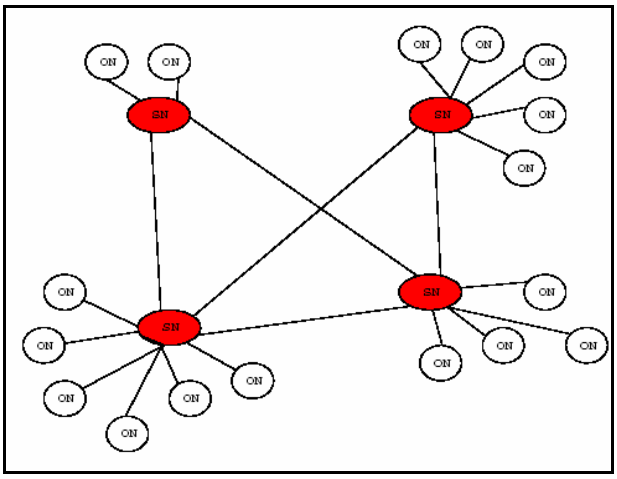

圖 12 Kazaa 網路架構

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)拓樸，如圖 12 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。

2.3.4 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。

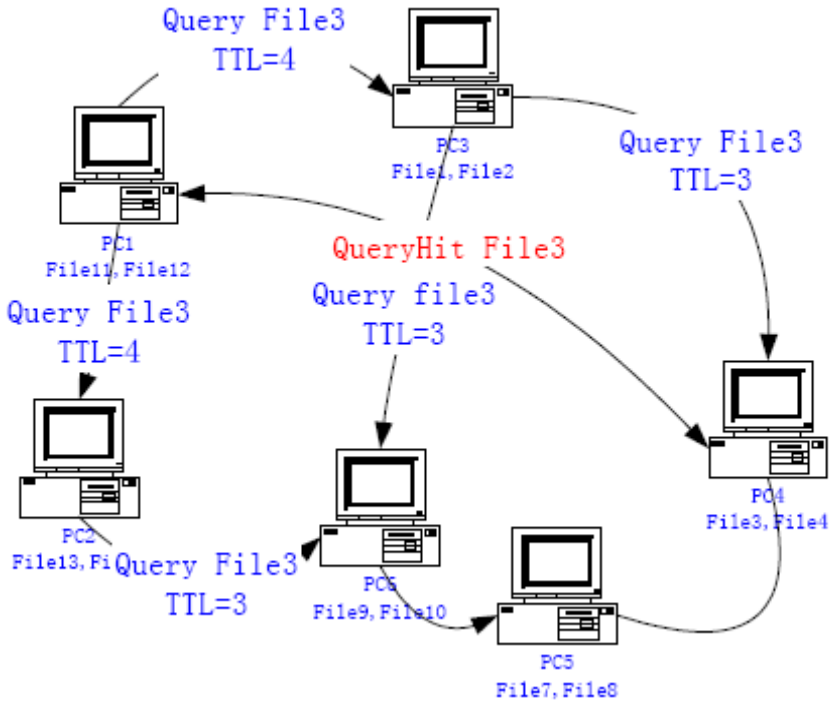

圖 13 Gnutella 網路架構

在最初的 Gnutella 協定中，使用的 Flooding 方法，來進行檔案的搜尋，此種搜尋方式被稱為盲目搜索（Blind Search），意思是在網路中，每個節點都不知道其他節點擁有那些資源。當某個節點要尋找某一特定資料時，此節點會把這個查詢要求傳發送給它的相鄰節點，如果相鄰節點含有這個資源，就返回一個 QueryHit 的資訊給 Requester。如果它相鄰的節點都沒有命中（QueryHit）這個被查詢檔案，就會把這條搜尋要求轉發給自己的相鄰節點。這種方式訊息就像洪水在網路中各個節點流動一樣，所以叫做 Flooding 搜索。

Flooding 搜尋方式，使得控制訊息大量的被發送，消耗了大量的網路頻寬，並很快造成網路擁塞甚至網路的不穩定。同時，局部性能較差的節點可能會導致 Gnutella 網路被分片（fragment），從而導致整個網路的可用性較差，另外這類系統更容易受到垃圾資訊，甚至被惡意的使用者利用而造成阻斷式攻擊（DoS attack）。

2.3.5 ezPeer 與 Kuro

ezPeer 與 Kuro 是國內本土公司採用會員收費機制的 P2P 檔案分享軟體。ezPeer 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 沒有索引服務，它採用類似 Gnutella 的運作方式，ezPeer 的特色在於透過驗證主機控管並且對會員收費。它的搜尋功能不需付費，只有下載需要付費。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，所支援的分享檔案類型除 MP3 音樂檔案外，ezPeer 也可搜尋或傳輸文字檔、圖片檔、影片檔等檔案格式

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他使用者，藉著其他使用者連上 Kuro 的 P2P 網路。

Kuro 並沒有集中的檔案索引 Server，供使用者直接查詢存放音樂之檔案資訊，以進行搜尋。在分享音樂及搜尋音樂方面，是向其他使用者發出請求，若有找到此音樂，便會向擁有音樂的使用者要求下載。Kuro 的 Server 並不會存放音樂檔案的索引檔，而是僅存放 Kuro 使用者的資訊。（目前在 Server 存放使用者資訊，還在司法程式中認定是否違法。）

圖 14 ezPeer 和 Kuro 軟體

2.3.6 BT(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。

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。

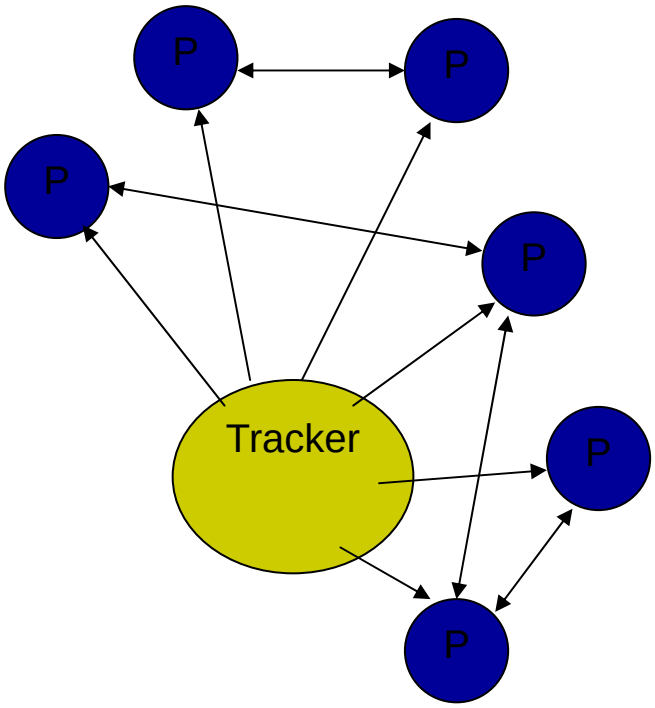

圖 15 BitTorrent 傳輸架構

2.4 封包攔截技術

任兩台電腦間，要透過網路互相傳送資料時，會先將資料交由網路卡封裝打包，然後透過網路線或電話線傳送至對方，對方電腦再藉由網路卡接收資料。在正常的情況下，我們傳送給對方的資料只有對方可以收到，但由於資料在傳送過程中是完全開放的（Broadcast），也就是每個人都可以透過某些技術就可以收到原本不屬於自己的網路封包，以下將介紹這項技術原理。

2.4.1 封包攔截原理

網路模型架構（OSI，Open System Interconnection）將網路通訊功能劃分為七層模型，各司其職，互相依存、合作。這七層分別為：應用層(Application Layer)、表示層(Presentation Layer)、會話層(Session Layer)、傳輸層(Transport Layer)、網路層(Network Layer)、資料連結層(Data Link Layer)及物理層(Physical Layer)。而 TCP/IP 體系也同樣遵循這七層標準，只不過在某些 OSI 功能上進行了壓縮，所以實際上我們打交道的 TCP/IP 僅僅有四層而已，網路上的分層結構決定了在各層上的協議分佈及功能實現，從而決定

了各層上網路設備的使用。

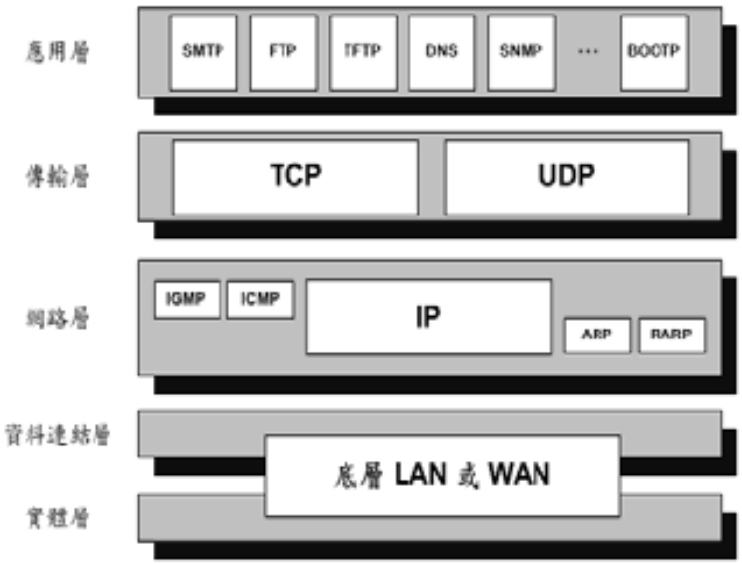

圖 16TCP/IP 層級模型

乙太網路（Ethernet）和 TCP/IP 可以說是相互相成的，兩者的關係幾乎是密不可分，乙太網路在網路存取層提供物理上的網路連線功能，而 TCP/IP 工作在上層，使用 32 位元的 IP 位址，乙太網則使用 48 位元的 MAC 位址，兩者間使用 ARP 和 RARP 協議進行相互轉換。從我們圖 16 TCP/IP 的模型圖中可以清楚的看到兩者的關係。

載波監聽/衝突檢測(CSMA/CD)技術被普遍的使用在乙太網路中，所謂載波監聽是指在乙太網路中的每台站點（Host）都具有同等的權利，在傳輸自己的資料時，首先監聽傳輸通道是否空閒，如果空閒，就傳輸自己的資料，如果通道被佔用，就等待通道空閒。而衝突檢測則是為了防止發生兩個站點同時監測到網路沒有被使用時而產生衝突。乙太網路採用廣播機制，所有與網路連接的工作站都可以看到網路上傳遞的資料。

因為在乙太網路中，所有的通訊都是廣播的，也就是說通常在同一個網段的所有網路介面都可以訪問在物理媒體上傳輸的所有資料，而每一個網路介面都有一個唯一的硬體位址，這個硬體位址也就是網卡的 MAC 位址，大多數系統使用 48 bit 的位址，這個位址用來表示網路中的每一個設備，一般來說每一塊網卡上的 MAC 位址都是不同的，每家網卡製造公司會得到一段位元址，然後用這段位址分配給其生產的每個網卡一個位址。在硬體位址和 IP 位址間使用 ARP 和 RARP 協議進行相互轉換。 在正常的情況下，一個網路介面應該只回應這樣的兩種封包資料：

1.與自己硬體位址相匹配的封包資料。

2.發向所有機器的廣播封包資料。

在一個實際的系統中，資料的收發是由網路卡來完成工作的，當網路卡接收到傳輸來的資料時，電腦上的網卡驅動程式會根據設置的接收模式判斷該不該接收，認為該接收就接收，接收後交上層處理，認為不該接收就丟掉不管，所以不該接收的資料就會被網路卡忽略，電腦根本就不知道。而對於網卡來說一般有四種接收模式，如表 3 所列。

廣播方式：	該模式下的網卡能夠接收網路中的廣播資訊。
組播方式：	設置在該模式下 的網卡能夠接收組播資料。
直接方式：	在這種模式下，只有目的網卡才能接收該資料。
混雜模式：	在這種模式下的網卡能夠接收一切通過它的資料，而不管該資料是否是只傳給它的。

表 3 網路卡接收模式

根據上面所述，我們知道在乙太網路中是基於廣播方式傳送資料的，也就是說，所有的實體信號都要經過我的機器，再者，網路卡可以置於一種模式叫混雜模式（promiscuous mode），在這種模式下工作的網卡能夠接收到一切通過它的資料，而不管實際上資料的目的地址是不是他。這實際上就是我們封包攔截的基本原理，讓網卡接收一切他所能接收的資料。

2.4.2 封包抓取函式庫（Libpcap）

Libpcap 全名是 Packet Capture Library，為 Unix/Linux 平臺下的網路封包攔截函式庫，由 Berkeley 大學 Lawrence Berkeley National Laboratory 研究院的 Van Jacobson、Craig Leres 和 Steven McCanne 編寫，目前的最新版本為 0.9。大多數網路監控軟體都以它為基礎。Libpcap 可以在絕大多數類 Unix 平臺下工作，此函數庫提供的 C 函數介面可用於需要截取經過網路介面封包的系統開發上。該函數庫支援 Linux、Solaris 和 BSD 系統平臺。利用 Libpcap 函數庫來完成一個網路封包截取程式主要可以分成五個步驟：

（1） 首先要決定監聽網路介面卡名稱，在 Linux 系統上通常是 eth0，在 BSD 系統中就是 x11，在這個步驟中我們也可以利用 Libpcap 所提供的函數來取得預設的網路卡名稱。

（2） 在這個步驟中主要是初始化一個封包監聽會話（sniffing session），藉由 Libpcap 所提供的函數，我們可以設定我們所要監聽的網路介面卡。

（3） 在第三步驟中，定訂封包過濾規則，例如我們只對 TCP 且通訊埠為 23 的封包感興趣，則我們可以設定如下規則字串。

〝 tcp and port 23 〞

（4） 在步驟 3 所設定的過濾規則，必須透過 BPF compiler 編譯，然後利用 Libpcap 提供的函數套用在特定的封包監聽會話中。

（5） 當封包流經我們監聽的網路卡時，若符合我們步驟 3 所設定的過濾條件，則會被截取交由後端程式進行分析，如不符則會乎忽略捨棄。

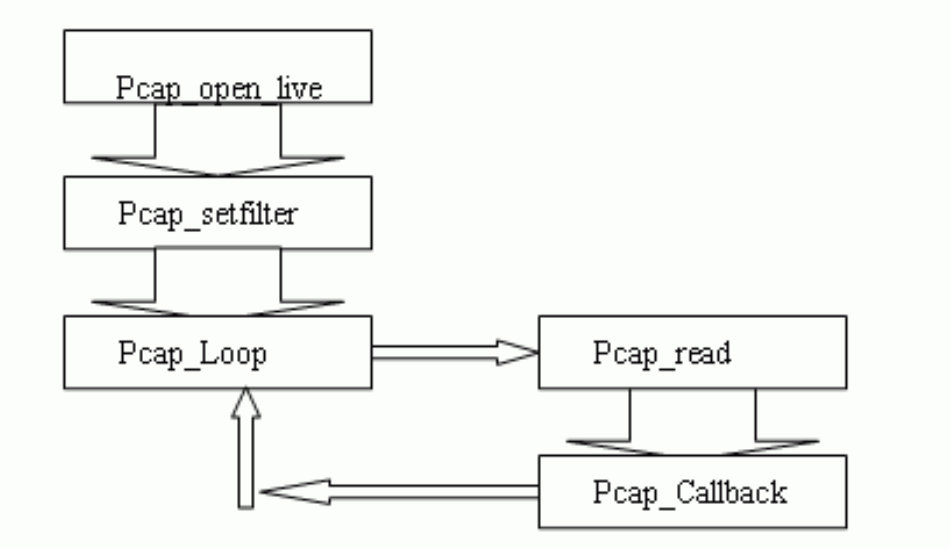

圖 17 網路資料截獲流程

第三章 相關研究

本章節將介紹與本論文相關的一些研究。3.1 是研究如何在傳輸層（Transport Layer）上，藉由網路連線行為來分辨出屬於 P2P 的傳輸。3.2 介紹利用軟體特徵（Application Signature）來進行 P2P 軟體的辨識。3.3 是利用社會學上的一些理論來探討 P2P 軟體對使用者行為及對社會上的影響。

3.1 傳輸層中辨別P2P傳輸（Transport Layer Identification of P2P Traffic）

在第一代 P2P 技術中，由於每種 P2P 軟體皆使用固定的通訊埠來進行溝通，所以我們可以容易的辨別出何者是屬於 P2P 的傳輸。然而隨著 P2P 技術的進步以及法律上的考量，在第二代及第三代的 P2P 軟體已經可以讓使用者自行設定通訊埠來進行傳輸甚至隨機產生，因此也造成了無法再依照固定的通訊埠來辨別 P2P 流量的困難。

Thomas Karagiannis 等人[11]在 2004 年的 Internet Measurement Conference 中提出一套在網路傳輸層中來辨別 P2P 流量的方法，藉由觀察網路連結行為樣式（connection pattern），來判定 P2P 流量。此方法是藉由大量的網路連線行為觀察中，歸納出兩個辨別原則，進而設計出 PTP 演算法（P2P Traffic Profiling algorithm）。這兩條歸納原則是 TCP/UDP IP pairs 原則和{IP, Port} pairs 原則，分別說明如下

- TCP/UDP IP pairs 原則：

  在目前熱門的 P2P 軟體中，例如 eDonkey、FastTrack、Gnutella 等，連線雙方都會同使用 TCP 和 UDP 傳輸協定。一般而言，都是利用 UDP 來傳送控制訊息，利用 TCP 來傳送資料。反觀，非 P2P 軟體會同時使用 TCP 及 UDP 做為傳

輸協定的軟體非常少，如圖 18 列出了此類的軟體，因此我們可以利用此原則來辨別屬於 P2P 的網路流量。

Ports	Application
135,137,139,445	NETBIOS
53	DNS
123	NTP
500	ISAKMP
554,7070,1755,6970,5000,5001	streaming
7000, 7514, 6667	IRC
6112, 6868, 6899	gaming

圖 18TCP/UDP IP pairs 原則例外通訊埠

◆ {IP, Port} pairs 原則：

在這個原則中，是利用{IP, Port} pairs 連線行為樣式做為分析的依據，因為自從集中式的 P2P 網路架構被認定為違法後，P2P 網路朝向分散式（distributed network）或者是混合式（hybird network）網路模式發展，不管是分散式或混合式的 P2P 網路中，當一個新節點要加入這個網路時，這個新節點必須知道一些已經在網路中的節點列表，這些節點稱為超節點（superpeer），這樣可以簡化建立連線的過程。

當新節點與超節點建立連線後，新節點會藉由超節點將自己的連線資訊也就是 IP 位址及設定通訊埠號碼（{IP, port} pair）廣播至整個網路中，以便讓網路中其他的節點可以進行通訊。通常在 P2P 網路中，任兩個節點只會存在一條連線。因此，考慮當網路中有二十個節點要與新節點連線時，這些節點會打開一個暫時的連接埠和新節點進行連線，此時，新節點的連線資訊（{IP, Port}）就會同時與二十個不同的{IP, Port}有關連，整個連線過程如圖 19。藉由上述的觀察結果，當發現到同一個{IP, Port} pair 同時與多個不同的{IP, Port} pair 進行連線時，我們就可以將此傳輸判定為 P2P 傳輸。

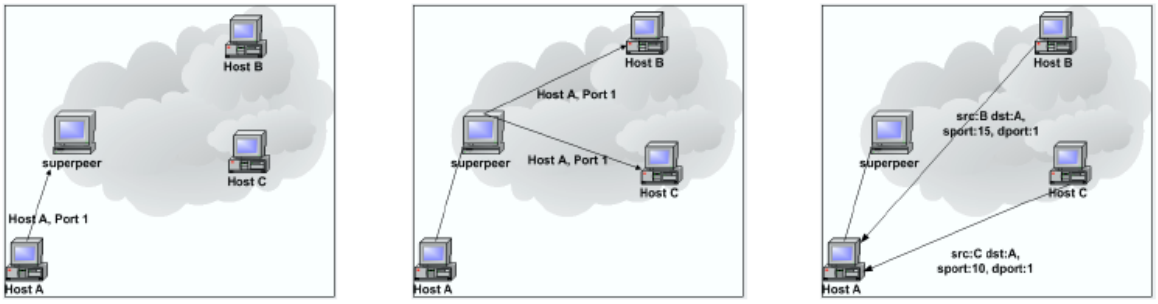

圖 19{IP, Port} pairs 原則示意圖

Algorithm 1 Nonpayload algorithm for P2P flow identification

```
1: procedure PTP                                   ▷ P2P Traffic Profiliing
2:     FT ← Flow Table
3:     for every src-dst IP pair in FT do
4:         if TCP/UDP pair then
5:             P2PIP.insert(srcIP)    ▷ TCP/UDP heuristic
6:             P2PIP.insert(dstIP)
7:     for all flows in FT do
8:         if src IP or dst IP in P2PIP then
9:             print flow          ▷ found by TCP/UDP pairs
10:            P2PIP.insert(srcIP)  ▷ put both IPs in P2P
                   list
11:            P2PIP.insert(dstIP)
12:        else if DNS heuristic is true then
13:            RejectedPairs.insert(src Pair)              ▷
                   pair=={IP,port}
14:            RejectedPairs.insert(dst Pair)
15:        else if src and dst IP not in MailServers then
16:            for src and dst IP-port pair do
17:                if pair in P2PPairs then
18:                    print flow ▷ found in previous interval
19:                    P2PPairs.insert(src pair)  ▷ put both
                           pairs in P2PPairs list
20:                    P2PIP.insert(src pair)
21:                else if pair not in Rejected then
22:                    Udpate sets for pair
23:                    IPPort.insert(pair)
24:                else if pair in Rejected then
25:                    Rejected.insert(src pair)
26:                    Rejected.insert(dst pair)
27:    for pairs in IPPort do
           ▷ examine pairs that were added during
           ▷ previous intervals and have not been yet classi-
           fied
28:        if IP not in MailServers and pair not in Rejected
               then
29:            if IP in P2PIP or pair in P2PPairs then
30:                P2PPairs.insert(pair)
31:                print all flows of pair
32:            else
33:                diff ← |pair.IPSet.len−pair.PortSet.len|
34:                if diff < 2 or (diff < 10 and port in
                       KnownP2PPorts) then
35:                    if Check_if_Mailserver == true then
36:                        MailServer.insert(IP)
37:                    else if Check_if_Malware == true
                           then
38:                        Rejected.insert(pair)
39:                    else if Check_if_scan == true then
40:                        Rejected.insert(pair)
41:                    else if Port_History heuristic=true
                           then
42:                        Rejected.insert(pair)
43:                    else
44:                        P2PPairs.insert(pair)
45:                        print all flows of pair
46:                else if diff> 10 then
                       Rejected.insert(pair)
```

圖 20 PTP 分析演算法

依據前述的兩項原則所提出的分析演算法如圖 20 所示，此演算最大的特點是能夠辨別未知的 P2P 通訊協定，將此演算法用於實際的環境進行分析，所得的結果如圖 21 所示，結果指出 PTP 演算法成功的辨識出 90%的 P2P Bytes 和超過 99%的 P2P flows。

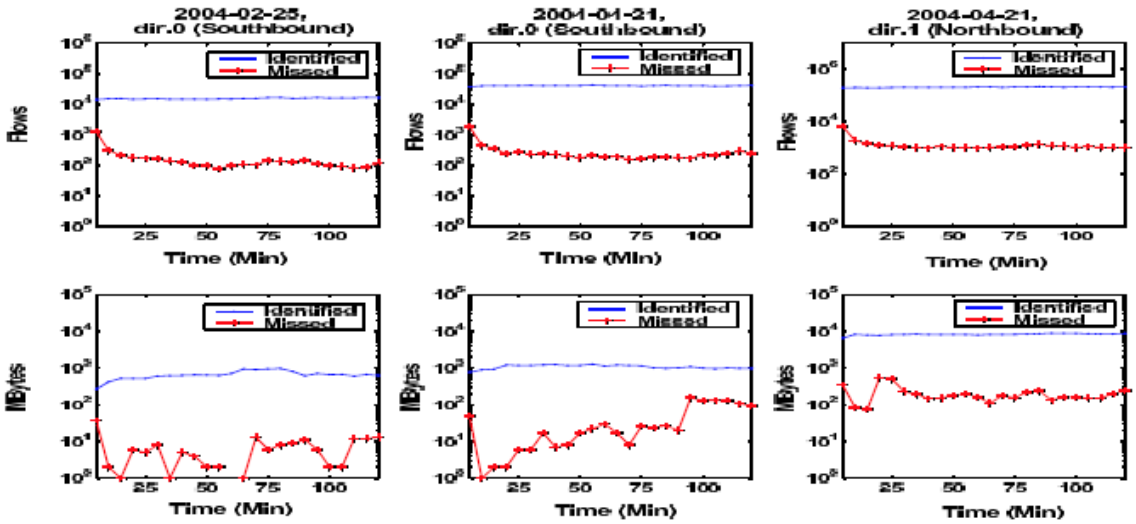

圖 21 PTP 演算法實驗結果

由於這個演算法是透過P2P 軟體在網路中連線的行為模式來判定是否屬於P2P 的流量，但是網路的應用中存在其連線行為模式和此方法中所提到的第二個原則相同，卻不是 P2P 的應用，例如 web server 的連線行為就和{ip, port} pairs 原則一樣，就是 server 的連線資訊可能同時和多個相異的{ip, port} pair 有關連。雖然這類的連線可以透過 well-known port 而過濾，但仍有可能影響到我們對 P2P 流量的判定。

3.2 利用軟體特徵辨別P2P流量（Accurate, Scalable In-network Identification of P2P Traffic Using Application Signatures）

P2P檔案分享軟體越來越盛行，佔用的網路頻寬越來越大，對於網路管理人員或是ISP業者想要對P2P流量進行管理，首先要能準確的辨別P2P流量，因為P2P技術的進步，以往藉由預設的通訊埠來判斷P2P流量已經不是那麼準確。

封包內容分析技術已經廣泛的應用於網路安全中，例如網路入侵偵測系統。Subhabrata Sen、Oliver Spatscheck、Dongmei Wang[6] 將此概念用於P2P流量的辨別，就是利用分析每個封包的內容來判斷此封包是否是由P2P軟體所產生的，針對目前比較熱門的P2P軟體，如Gnutella、eDonkey、DirectConnect、BitTorrent、Kazza，從分析這些軟體所使用的通訊協定及相關的文件，粹取每種應用程式的特徵（Application Signature），將這此特徵用於網路流量的辨識中。

針對Gnutella通訊協定分析所獲得的軟體特徵為：

- 封包開頭字串為“GNUTELLA”、“GET”、“HTTP”
- 若封包開頭字串為“GET”、“HTTP”，則其內容必包函含下面其中一個字串，“User-Agent: <name>”、“UserAgent: <name>”、“Server: <name>”

針對eDonkey通訊協定分析所獲得的軟體特徵為：

- 封包開頭的第一個Byte為eDonkey marker，其值固定為0xE3
- 封包的第二到第四個表示封包內容的長度

針對DirectConnect通訊協定分析所獲得的軟體特徵為：

- 封包開頭的第一個字元為“$”，而最後一個字元為“|”
- 在“$”緊接一個以空白字元結束的字串，此字串為DirectConnect命令。

針對BitTorrent通訊協定分析所獲得的軟體特徵為：

- 封包的第一個Byte，其值為0x19
- 接下來的19個Bytes，為一個長度19的“BitTorrent protocol”字串

針對Kazaa通訊協定分析所獲得的軟體特徵為：

- 封包開頭字串為“GET”、“HTTP”，且封包中包含”X-Kazaa”字串

因為在封包中搜尋字串相符的字串需要大量的計算時間，為了能夠在高速的網路中快速的進行辨識，將所定義的P2P軟體特徵分成兩個部分：固定位移比對（Fixed Offset Match）、變動位移比對（Variable Offset Match）。固定位移比對，主要是將P2P軟體特徵和TCP封包內容中特定的Byte進行比對，這樣比對方式所需要的計算相當的低。而變動位移比對則是透過正規表示式（Regular expression）來比對。

將這些定義的P2P軟體特徵實作在Gigascope[1]中進行驗證，Gigascope為一高速網路流量監測器，所能監測的網路流量最高可到達OC-48（2.4Gbps），實驗的結果證明定義的P2P軟體特徵可以精確的辨別出各種P2P軟體所產的網路流量，且精確度高於一般以通訊埠分辨的方法，實驗結果如圖22，對於Kazaa的分辨，Signature-Based的辨別率比Port-Based的辨別率比高出3倍。

Protocol	All Connections	
	Port-based (MB)	Signature-based (%)
Gnutella	487.12	145
Kazaa	548.41	347.38
DirectConnect	2000.75	163.45
BitTorrent	54444.67	90.97
eDonkey	2149.84	102.37

圖 22 Port-based 和 Signature-based 的精確度比較

利用軟體封包的特徵來判別 P2P 的流量，除了可以提高辨識率，也能對各種不同 P2P 軟體所產的流量進行統計，提供相關的資訊給研究人員，針對特別的 P2P 軟體來進行效能上的改進。在本研究中也將利用相同的分式，來探討目前網路上各種熱門的 P2P 軟體的使用情形。

3.3 點對點檔案分享軟體使用行為之研究

由於 P2P 技術的蓬勃發展，有許多文獻從技術以及法律觀點來討論點對點檔案分享軟體，卻僅有少數研究從使用者的觀點出發。然而想要了解點對點技術對使用者所造成的影響，唯有從使者的動機和行為才能真正了解點對點檔案分享軟體所造成的影響。再者 P2P 技術改變了傳統傳播模式，賦予閱聽人更高的媒體選擇權，閱聽人又是如何看待再此一新興的媒體技術呢？

中山大學傳播管理研究所，劉怡玟在其論文[36]中，從使用者的角度出發，依據使用與滿足理論，建立如圖 23 的研究架構，利用模式中的動機、個人特質、使用行為，來探討點對點檔案分享軟體使用者的動機和行為架構。在此研究架構中包含了三個變數，個人特質（Characteristics of Individual）、動機（Motivations）和點對點檔案分享軟體使用行為（P2P Uses），分別定義如下：

一、個人特質：

個人特質所代表的因素是考慮使用者的社會身分、角色而不討論內在人格特質。

二、動機：

動機等同於預期媒體所帶來的效用(Utility)或者使用媒體的理由，在此包括娛樂(Entertainment)、習慣(Habit)、資訊(Information)、社會互動(Social interaction)、打發時間(Pass time)、放鬆(Relaxation)、刺激(Arousal)、便利性(Convenience)、經濟(Economy)、自我認同(Personal identity)等。

三、點對點檔案分享軟體使用行為：

點對點檔案分享軟體允許使用者主動散佈訊息，與過去媒體傳播方式使用者僅能被動的接受資訊不同，因此所指的使用行為不僅包括傳統衡量媒體使用行為，也包括分享行為在內。

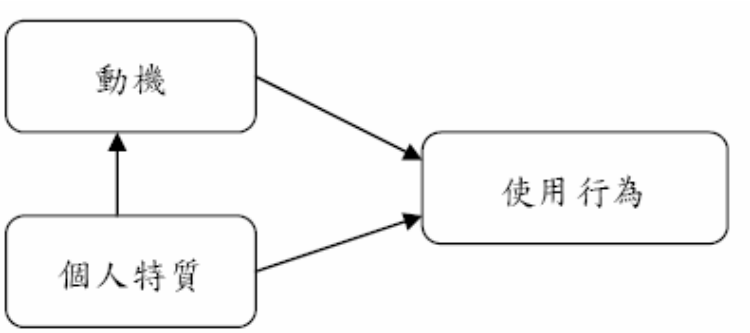

圖 23 使用與滿足研究模型架構

根據研究架構，所設計的問卷包含三部分，第一部份為點對點檔案分享軟體使用者行為和特徵，第二部份為點對點檔案分享軟體的使用動機，第三部份則是個人基本資料。

在第一分問卷回收的結果為：(一) 使用者人口特徵：填寫問卷樣本上有89.2%為男性，而女性僅占了10.8%，年齡分佈上以21～25歲居多，佔37.8%，其次是16～20 歲佔23.7%，而26～30 歲亦佔21.3%，總計16～30歲共佔全部樣本82.8%，從以上數字顯示，點對點檔案分享軟體使用者以男性居多，年輕族群為主要使用者。另外，教育程度大學佔51.3%，職業方面以學生最多，佔56.9%。(二) 點對點檔案分享軟體使用情形：，大部分的人使用點對點檔案分享軟體經驗都在二年以下，佔百分之七十左右，其餘百分之三十使用點對點檔案交換軟體資歷為兩年以上，其中四年以上僅佔百分之六。可見點對點檔案分享軟體是近兩年才被大量採用。在調查樣本中有86.7%的人使用Bit Torrent，45.9%的人使用eMule/eDonkey，另外有18.8%的人使用Kuro/ezPeer 收費軟體(見圖24)。

您目前所使用的 P2P 軟體	次數	次數佔總次數的比例	次數佔總樣本比例
KURO/ezPeer	183	11.2	18.8
eDonkey/eMule	447	27.5	45.9
KaZaa	74	4.5	7.6
BT	844	51.8	86.7
Apia	29	1.8	3.0
其他	51	3.1	5.2
總計	1628	100.0	167.3

圖 24 使用點對點檔案分享軟體種類之次數分配與百分比

在下載檔案類型方面(見圖25)，目前大多下載電影，佔全部樣本人數77.2%，軟體是大家第二常下載的檔案佔64.2%，音樂則是第三，佔54.1%，另外色情影片佔48%，遊戲也占了46.2%，過去點對點檔案分享軟體發展初期僅能下載mp3，音樂一直被認為是最常被下載的檔案，然而隨著點對點檔案分享軟體功能提升、配合網路頻寬的改善，下載電影、軟體的次數反而比音樂多。

經常下載哪些類型檔案	次數	次數佔總次數的比例(%)	次數佔總樣本比例(%)
音樂	526	15.1	54.1
電影	751	21.6	77.2
軟體	625	18.0	64.2
遊戲	450	12.9	46.2
電子書	112	3.2	11.5
動畫	370	10.6	38.0
電視	170	4.9	17.5
A片	467	13.4	48.0
其他	9	.3	.9
總計	3480	100.0	357.7

圖 25 分享檔案種類與下戴次數百分比

第二分問卷主要是針對點對點檔案分享軟體使用動機進行調查研究，利用主成分分析法(Principle component analysis)對問卷調查結果進行分析，發現使用點對點檔案分享軟體使用動機主要有五個，按照高低順序為：資訊動機、便利動機、收藏動機、人際互動和自我成就。

根據問調查結果所獲得的以下三點的結論：

（1）點對點檔案分享軟體使用者仍以 16～30 歲的年輕族群為主，使用者大多為男性，職業以學生居多，其中大學生仍為主要使用族群。此結果可以和點對點檔案分享軟體的興起連結，Napster 作者 Shawn Fanning 在大學設計此軟體，Napster 自大學校園被創造、盛行，點對點檔案分享軟體能發展到今日，一部份也仰賴大學生的支持，大學生的特色為：沒有收入來源、經常遨遊網路、樂於接受新事物、對於資訊娛樂需求量大，點對點檔案分享軟體的出現對學生是一大福音，使得點對點檔案分享軟體在大學校園中廣為流傳。

（2）對於點對點檔案分享軟體使用行為，使用者有半數每天都使用點對點檔案分享軟體，而有60%的人開啟程式後會持續使用十個小時以上，使用點對點檔案分享軟體的次數頻繁且使用時間長，下載檔案是主要活動，點對點檔案分享軟體使用仍以”享”為主。在分享行為上，結果顯示有90%的人曾將檔案移入分享夾與他人共享，顯示出主動分享有增加的趨勢。

（3）在使用點對點檔案分享軟體種類，調查結果顯示最多人使用的點對點檔案分享軟體為BT，BT盛行也暗示著使用者所在意的是檔案下載的速度，使用者重視軟體的穩定、快速和可靠，由於BT下載速度快，使得下載檔案失敗的風險降低，使用者可以快速地下載到檔案，即使在操作上不如其他軟體便利，大家仍然願意使用，另外一項發現是，有許多人有同時使用多種點對點檔案交換軟體的習慣，有超過半數的人使用兩種或以上的軟體，這可能和每個軟體所能提供的利益不同有關。至於檔案分享種類，目前以電影下載量最多，其次依序為軟體、音樂、A 片、遊戲、動畫、電視劇、電子書最少。

在該篇論文中，透過問卷調查的方式，對目前社會大眾使用P2P軟體的情形進行研究，其研究結果顯示，P2P軟體已經成為一種被大眾接受的新興傳播媒體，藉由P2P來分享資原、取得資訊，由此可見P2P軟體對大眾而言已是生活中的一個重要要資訊來源。然而由於研究者在進行資料收集時皆採用匿名方式，所以收集到的資料可能會有所誤差，或者是受訪者會提供不實的資料，因此在我們的研究中，將藉由收集網路上封包的方式來對P2P軟體使用情形進行一個比較客觀的分析。

第四章 研究方法

隨著電腦進步與頻寬成長，網路使用型態已有所改變。本章節將說明如何藉由分析網路流量來取得網路各種應用所產生的比例。在 4.1 中藉由對通訊協定封包檔頭的瞭解，分析各種網路應用。在 4.2 中配合 OSI 第七層內容過濾方式取得各種 P2P 軟體之比例。在 4.3 中藉由分析內容的副檔名，來統計 P2P 之間傳輸檔案的類型比例。接著在 4.4 以隨機取樣方式，來了解目前分享內容的合法性。

4.1 網路應用比例研究方法

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. 影音資料的網路共用漸漸廣泛。

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. 即時的影像、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。

在表 4 中列出目前常見的網路應用，我們依照其提供的服務以及所使用的通訊協定加以分類。首先是傳統的網路應用程式如 http、ftp、telnet、smtp 等，這類的網路應用是採取主從式架構進行通訊，客戶端透過由一個固定的通訊與伺服端進行溝通，伺服端可以同時與多個客戶端進行通訊，而客戶端與客戶端間不會任何的通訊，所傳遞的任何資料皆需藉由伺服端傳送。這類網路應用程式所產生的流量可以藉由所用的通訊埠來進行辨識。

第二類是網路多媒體串流，這類的應用程式利用應用層通訊協定如 RTSP、SIP、Q.931 和 H.245 來做為其通訊協定。常見的軟體如 QuickTime 和 RealTime 便是利用 RTSP 和 RTP/RTCP 來建立串流和傳送資料。

第三類是目前最熱門的 P2P 檔案分享軟體，利用對等式的架構來做資源分享，打破了傳統主從式架構，允許網路上任兩個節點可以直接連線進行傳輸，常見的軟體有 eDonkey、eMule、BT、Kazaa 等。P2P 的另外一項熱門應用就是即時通訊，如 MSN Messenger。第四類就是目前很流行的網路遊戲，透過網路可以讓玩家一同進行遊戲與溝通，目前熱門的遊戲有 Starcraft、Warcraft、Diablo。

種類	應用程式 / 應用層通訊協定
主從式架構	http，https，ftp，telnet，ssh，nntp，dns，smtp，pop3 etc.
多媒體串流	rtsp，sip，mms，rtp/rtcp，rdt，q.931，h.245，etc.
P2P	ezPeer，Kuro，eMule，BT，MSN Messenger, etc.
遊戲	Starcraft，Warcraft，Diablo，Counter Strike, etc.

表 4 常見網路應用程式

想要分析網路上的各種應用使用的情形，最簡單的方式就是透過監聽。監聽到封包後，檢查他的封包檔頭，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。

在分辨各種網路流量上，我們採取以通訊埠為辨別的基礎（Port-base），也就是利用各種應用程式的預設通訊埠來辨別流量是由那一種應用程式產生的，對於上述各種網路應用，我們藉由實際操作觀察以及各種文件列出每種應用程式預設的通訊埠，如表5。

Application name	Representative port	TCP well-known ports
WWW	80	80, 8080, 443
FTP	21	20, 21
POP3 / POP3S	110,995	110, 995
SMTP	25	25
MSN Messenger	1863	1863, 6981-6990, 14594
Windows Media	1755	1755
KaZaA	1214	1214
eDonkey	4661	4661~4665
eMule	6688	6600-6802
BT	6881	6881-6889

表 5 常見網路應用軟體預設通訊埠

在分析步驟上，可分成三個步驟，首先我們對收集到的封包資料進行資料流（flow data）分類，就是將相同來源IP/port到相同目的IP/port的單向封包序列（unidirectional packet sequence）收集到同一個資料流裡，再對每組資料流找到其重要通訊埠（Critical Port Number），所謂的重要通訊埠主要是用來辨識該組資料屬於何種網路應用的通訊埠，最後根據網路應用通訊埠表將每個資料流對應到各個應用程式，進行分類和統計，流程如圖26所示。

對於每組資料流的重要通訊埠決定，是依據 TCP 三向交握（Three-way handshake）的觀念來做選擇。在每條 TCP 連線中，伺服端會開啟一個通訊埠等待客服端的連線要求，而伺服端所開啟的通訊埠就是我們判斷流量屬於何種應用程式的依據，也就是我們所定義的重要通訊埠。在每條 TCP 網路連線建立時，客戶端會先發送一個 SYN 的封包給伺服端，而伺服端在收到 SYN 封包後會回覆一個 SYN-ACK 封包給客戶端，利用這兩個封包就可以決定每條 flow 的重要通訊埠，也就是 SYN 封包的目地通訊埠，SYN-ACK 封包的來源埠。

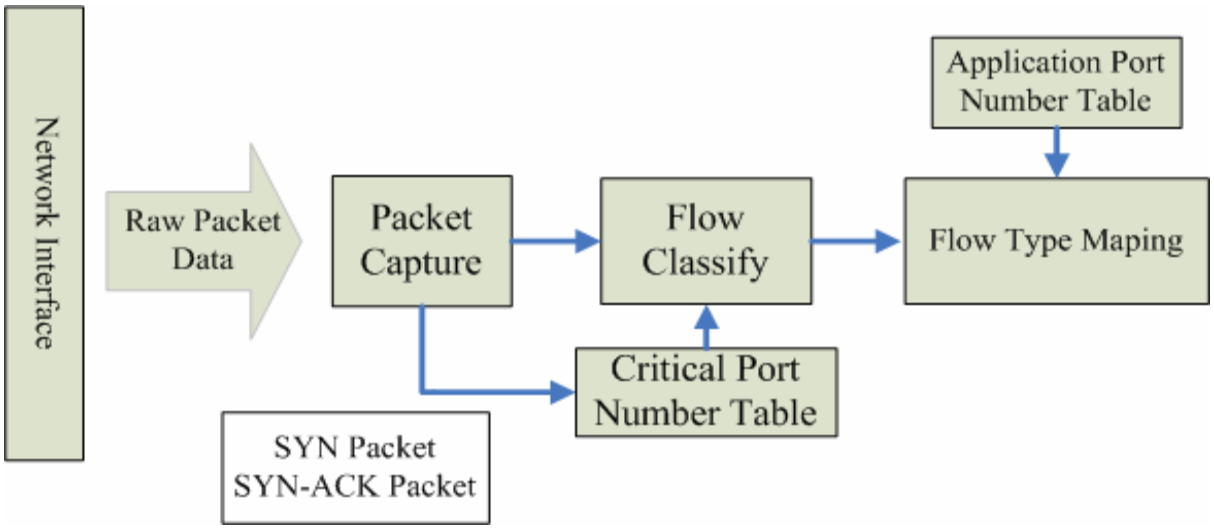
Network Interface
Raw Packet Data
Packet Capture
Flow Classify
Application Port Number Table
Flow Type Maping
Critical Port Number Table
SYN Packet
SYN-ACK Packet

圖 26 流量分析流程

4.2 P2P 軟體比例研究方法

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。由於 P2P 技術的進步，目前大部分的 P2P 軟體都允許使用者自行設通訊埠或者隨機產生，造成利用 well-know port 來對 P2P 軟體進行分類，會使原本屬於 P2P 的流量沒有被辨識出來，為了提高判斷的精確性，我們採用封包 payload 的分析方式，針對每種常見的 P2P 軟體所使用的協定進行分析，經由參考各種通訊協定的文件以及對各種軟體使用的觀察，歸納出每種 P2P 軟體所傳輸的封包特徵，用此特徵來進行 P2P 軟體使用比率的分析。

在 P2P 網路中所傳送的封包主要可以分成兩類，一種是控制訊號封包（signaling packet）另一種是資料封包（data packet）。控制訊號封包主要是點和點之間用來做溝通的，如搜尋特定檔案、要求檔案下載等，資料封包是用來傳輸下載的檔案，也是 P2P 流量的主要來源。下面將介紹我們對各種 P2P 軟體的這兩類封包特徵的分析。

eDonkey / eMule：

eDonkey 和 eMule 這兩種軟體使用同一種通訊協定，使用的預設通訊埠有 4661～4665/tcp 和 4665/udp，資料的傳輸是透過 TCP，而控制訊息可以透過 TCP 也可以經由 UDP 傳送。通訊協定的資料封包和控制封包開頭的第一個位元組的值是固定的，eDonkey 封包開頭第一個位元組的值為“0xe3”，而 eMule 是“0xc5”。在接下來的 4 個位元組表示整個封包的長度，我們可以利用這 5 個位組來判定屬於 eDonkey / eMule 封包。

Kazaa：

Kazaa 底層所使用的通訊協定與 Fasttrack 相同，預設的通訊埠為 1214/tcp 和 1214/udp，但允許使用者自行設定軟體所使用的通訊埠。Kazaa 封包格式類似於 HTTP 封包格式，在辨別上比較困難。Kazaa 在檔案的傳輸上是利用 HTTP 協定中的 GET 方法，但在 URL 的部分所代表的是要傳輸檔案的 hash 值，開頭字串為“/.hash=”，因此我們可以將字串“GET /.hash=”做為其檔案傳輸封包的特徵。控制封包的特徵比較明確，可以分成兩大類，第一種是控制封包內容大小為 5Bytes，則其封包開頭字串為“0xc028”、“ 0x290000002901”、“0x280000002900”、“0x270000002980”、“0xc1”，另外一種是大小為 2Bytes 的其封包開頭字串為

“0x2a”。

BitTorrent：

BitTorrent 是迅速增長的一個新興 P2P 軟體，所使用的通訊協定預設通訊埠為 6881～6999，當 BitTorrent 被執行時，會從最小的預設通訊埠開始嘗試開啟，若最小的預設通訊埠無法取得，則會往上嘗試，直到最大的通訊埠。由於 BitTorrent 不具搜尋的功能，檔案的定位及搜尋皆透過使用者分享出來的種子（seed）來完成，因此並沒有搜尋檔案的控制封包，所以我們在辨別時主要是針對其檔案傳輸的封包。BitTorrent 資料封包的開頭有固定的格式，第一個 Byte 為一固定值：“0x13”，接下來的 19 個 Bytes 代表一個固定的字串：“BitTorrent protocol”，我們將這 20 個 Bytes 的值當作 BitTorrent 的封包特徵。

Kuro / ezPeer：

Kuro 和 ezPeer 為付費的 P2P 軟體，其所使用的通訊協定並沒有公開，所以我們僅能以藉由觀察軟體時際在運做時所產生的封包，來定義封包特徵。

Kuro 所使用的預設通訊埠範圍介於 6600 和 6802，在使用 Kuro 時必須先付費，取得帳號及密碼後才能進入 Kuro 網路進行檔案的下戴和分享。圖 27 為 Kruo 軟體整個運作過程，在過程中所傳遞的封包開頭，每個步驟皆不同，但具有共同的格式“?0B0”其中“?”會隨每個步驟而改變，我們可以此為 Kuro 的封包特徵。

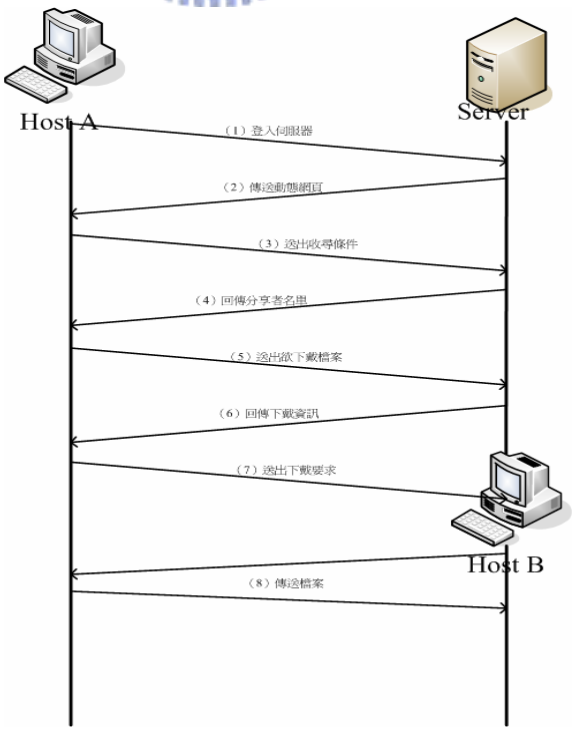

圖 27 Kuro 運作流程

ezPeer 所使用的通訊埠範圍可以區分成三段，分別是 6800～6890、7880～7890、8860～8890，ezPeer 也是必須經過認證才能登入 ezPeer 網路，進行檔案的搜尋和下戴。由於 ezPeer 使用者間，所傳輸的封包經過特殊的編碼處理，所以我們無法區分所傳送的是控制封包或者是資料封包。經由觀察封包的原始資料，發現每個由 ezPeer 所產生的封包，其開頭的 4 個位元組，都具有相同的值：“0x474e”，所以我們可以此為封包特徵，做為辨識的依據。

表 6 是我們從相關研究[6][11]以及藉由分析各個 P2P 軟體使用文件所歸納出來的結果，我們以此表做為我們統計的依據。

軟體種類	封包開頭字串	傳輸協定	預設通訊埠
eDonkey /eMule	0xe3 0xc5	TCP/UDP	4661-4665
Kazaa	“GET /hash 0xc028, 0x290000002901, 0x280000002900, 0x270000002980,0xc1, 0x2a	TCP	1214
BitTorrent	0x13BitTorrent	TCP	6881-6889
Kuro	?0B0	TCP	6600-6802
ezPeer	0x474e	TCP/UDP	6800~6890 7880~7890 8860~8890

表 6 常見 P2P 軟體封包的開頭字串及預設通訊埠

4.3 P2P 分享內容比例研究方法

依據檔案的副檔名，我們可以將 P2P 分享內容分成好幾類，如音樂檔、影像檔、圖檔、壓縮檔等，透過各類檔案所佔的比例，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，來對目前 P2P 網路所分享的檔案進行統計。

我們利用各種 P2P 軟體的搜尋功能來進行資料的收集，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。

為了避免我們收集到的檔案是損毀的，也就是說搜尋到的分享內容是無法開啟的錯誤檔案，所以，我們對收集到的檔案先進行過濾，再加以統計。過濾的方式依照每種檔案類型的檔案格式來進行過濾，當我們收集到的檔案，其檔案格式與其副檔名不和，我們將會此檔案認定為損毀，不列入統計中。

接著我們在三個不同的時段來進行資料的收集，我們隨機將上述 4,017 個常用字及常用詞，輸入給模擬程式（robot），此程式會模擬人對各種 P2P 軟體的操作行為，依序將常用字及常用詞作為搜尋字串，當 P2P 軟體傳回搜尋結果時，模擬程式會將結果依照檔案進行分類，然後過濾每損毀的檔案，再進行統計。統計時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，整個流程如圖 28

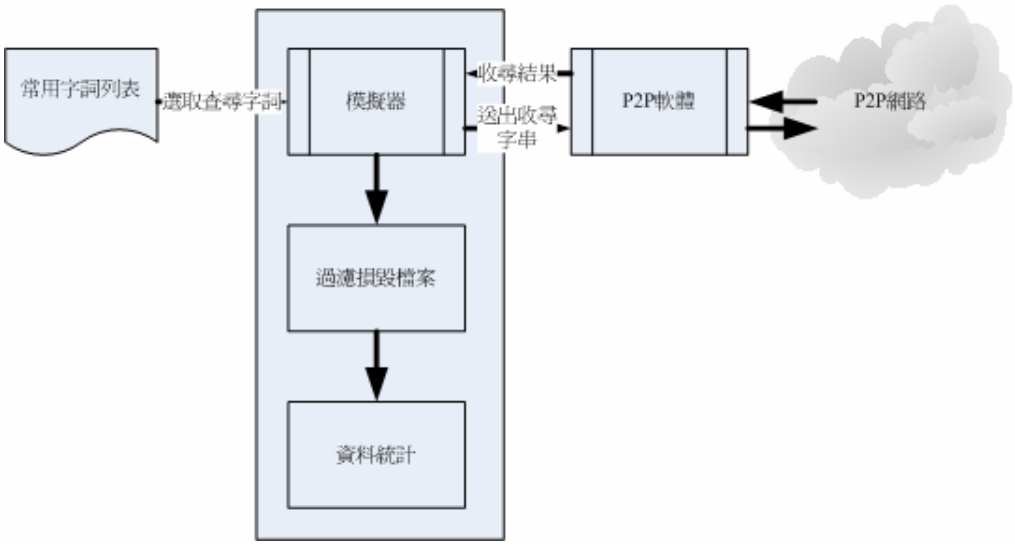

圖 28 分享內容分析流程圖

4.4 P2P 分享內容合法性研究方法

藉由 P2P 的架構，使用者可以任意分享其檔案，並且在分散式共享的推波助瀾下，快速搜尋與散播資料。藉由我們在 4.3 中收集到的資料中，我們知道目前傳輸的檔案大多屬多媒體資訊，如音樂與影片。因此在本節中，我們將對收集到的分享內容進行合法性的分析，觀察究竟有多少比例是屬於未經合法授權而進行的搜尋與下載。

在我們收集到的分享內容中，我們針對音樂檔，也就是副檔名為 mp3 或是 wma 的檔案來進行分析。從我們收集的檔案中，過濾出 20,519 首有效的音樂檔案，再從這些檔案中隨機挑選 1,500 首歌曲，我們將此名單交由 IFPI Taiwan 人員，由 IFPI 人員將此名單與該會錄音著作資料庫進行比對，統計該 1,500 個歌曲檔案中有多少檔案未是經授權而在網路上分享的。

第五章 研究結果

這在一章節中，將說明我們的實驗結果。5.1 說明目前網路的使用情況，5.2 列出我們所測得的個種 P2P 軟體的傳輸速度以及使用的比例，5.3 說明目前 P2P 檔案分享軟體的分享內容趨勢及各種內容所佔的比例。

5.1 網路應用比例研究結果

藉由在 4.1 中所提出的研究方法，我們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如圖 29、圖 30 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。

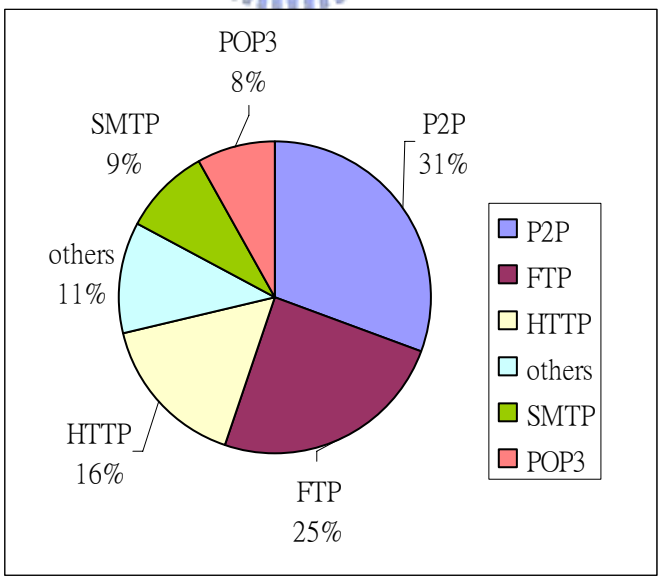

圖 29 交大資工流量分析結果

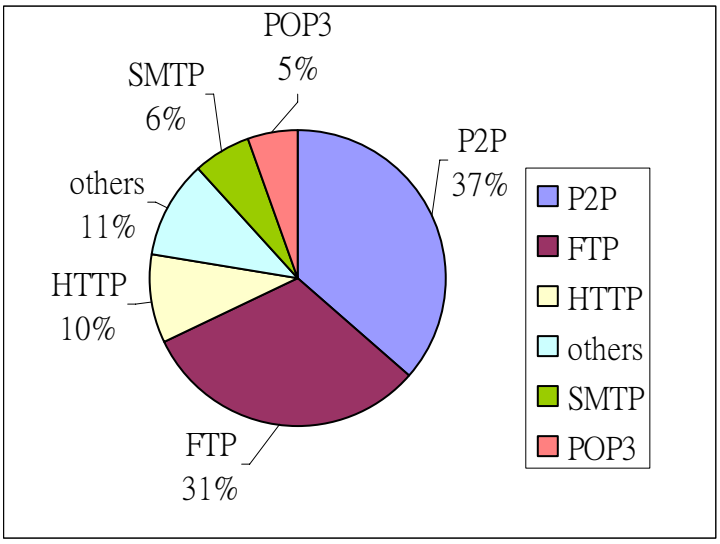

圖 30 交大第十三學生宿舍分析結果

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。

由以上的統計資料可以知道，P2P 軟體的盛行所造成的現象，例如網路效能的影響、甚至包括合法與非法的探討，早已經是必須重視與解決的課題。因此，接下來我們將說明我們對各種 P2P 軟體在網路上的使用的情形的研究結果，並且說明 P2P 的傳輸資料，的內容與類型，徹底分析出 P2P 在網路上的全部細節，以讓那些想要更進一步研究的人員或者執法單位，能當作一個有用的參考依據。

5.2 P2P 軟體比例研究結果

利用我們自行設計的 P2P 軟體流量分析程式針對交大資工與交大學生第十三宿舍的網路流量進行分析，統計結果如圖 31 為各種軟體在 P2P 網路流量中所佔的比例，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%、其他 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[36]的結果類似，也是 BT 與 eDonkey / eMule 佔絕大多數。

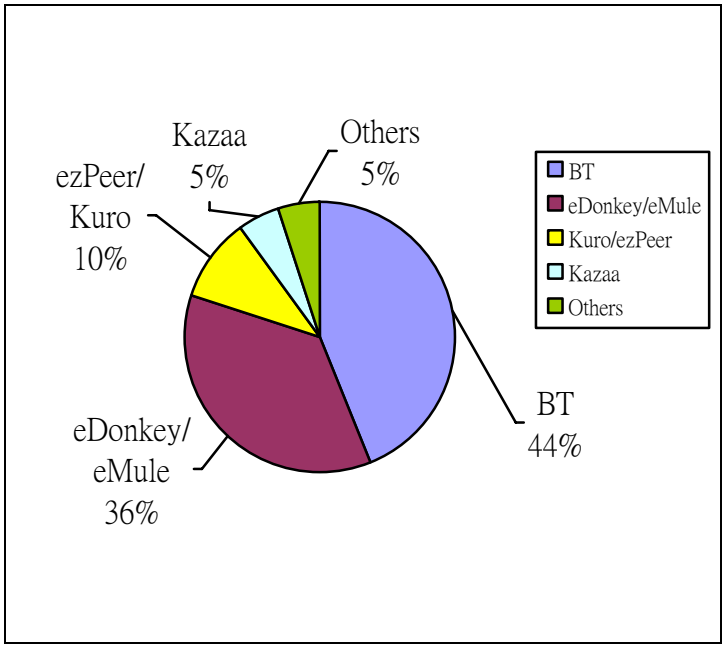

圖 31 各 P2P 分享軟體使用比例

由上面的結果中我們可以看出 eDonkey、eMule 及 BT 為目前最熱門的 P2P 檔案分享軟體，其中比較特別的是，對於 P2P 軟體而言最重要的功能之一，就是對檔案的搜尋，然而 BT 卻不具此功能，但在分析結果中，佔有率卻遠遠大於具有搜尋功能的 P2P 軟體，可見對使用者而言，具有良好的傳輸速率對使用者是比較有吸引力的。

5.3 P2P 分享內容比例研究結果

藉由 4.3 的研究方法，我們針對交大資工與交大學生第十三宿舍的網路流量進行分析。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 所分享最多的檔案類型。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。

整體而言，統計的結果如圖 32 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他三者小很多，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。

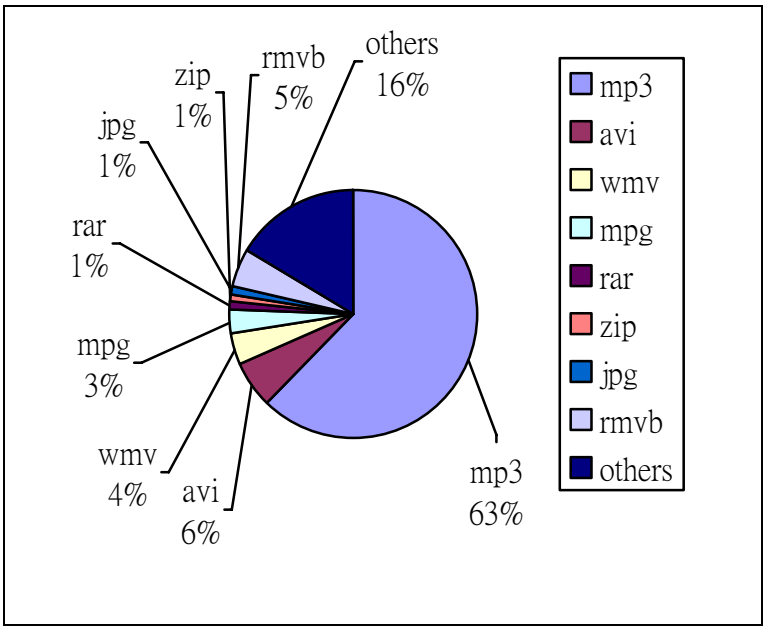

圖 32 網路分享內容比例

從檔案個數所佔比例的角度來看如表 7，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，這也表示音樂檔目前是最受歡迎的檔案。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，表示電影檔也是非常受 P2P 使用者所歡迎，再來是副檔名為 mpg 的檔，也是同樣為電影檔。

音樂類型	
Type	%file number
mp3	61.54
wma	1.69
wav	1.43
m4a	1.33
Subtotal	65.99

影像類型	
Type	%file number
wmv	4.10
mpg	2.86
avi	6.49
asf	0.14
mov	0.07-0.08
rmvb	5.06
Subtotal	18.72

表 7 主要音樂檔案和影像檔案類型

以使用者的角度來看，最受歡迎的就是這些能夠提供使用者娛樂的檔案，不過大多數的 P2P 分享及傳輸的音樂檔、影像檔，大部分都是不合法的檔案居多，所以對目前的音樂產業及電影產業會有一定的衝擊。且大多數使用者都是傳輸音樂檔居多，因此專門傳輸音樂檔的 P2P 軟體也因應需求而開發出來，甚至有公司專門為此而營運，如 Kuro、ezPeer 等著名 P2P 分享音樂軟體。

NTT technical review[4]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，且在非法檔案及色情檔案的阻檔方面，更難有所發揮。

5.4 P2P 分享內容合法性研究結果

IFPI Taiwan 人員根據我們所提供的 1,500 個歌曲檔案列表與該會錄音著作資料庫一一比對，所得統計分析結論如下(如圖 33 所示)：

（1）其中 1,079 檔案，即 71.93%之採樣檔案係由 IFPI Taiwan 會員公司享有著作權之歌曲，該等歌曲均未經其授權而於 P2P 網路中、傳輸、下載。

（2）另外約有 300 首歌（約 20%之採樣檔案），係屬中英文流行歌曲之音樂檔案。根據 IFPI Taiwan 提供之初步分析，非屬其會員之其他獨立唱片公司等唱片市場之佔有率約有二至三成，據此可合理估計 1500 個檔案中至少仍有 20%為他人享有著作權之歌曲。該等歌曲均未經其授權而於 P2P 網路中供搜尋、傳輸、下載。

（3）以上兩者合計 1,379 檔案，共約 90%之採樣檔案確屬享有著作權之歌曲，或合理估計極可能為享有著作權之歌曲。換言之，未經合法授權而於 P2P 網路中供搜尋、傳輸、下載等歌曲檔案約佔九成，P2P 網路中交換之檔案絕大多數未經合法授權。

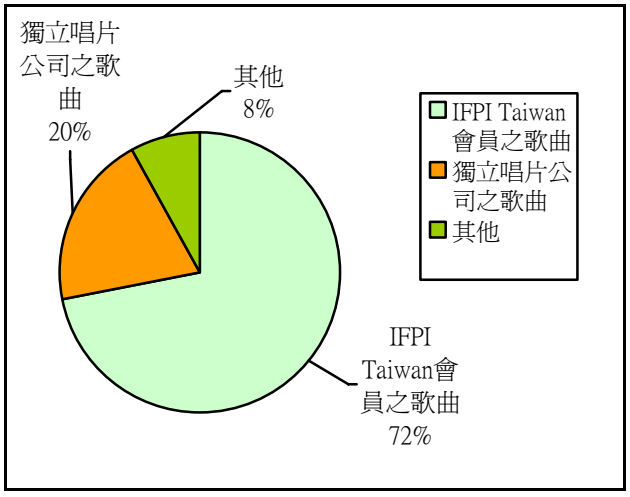

圖 33IFPI Taiwan 歌曲統計分析比例

第六章 結論

6.1 討論與結論

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。

目前，最熱門的網路應用程式莫過於 P2P 檔案分享軟體，P2P 技術的出現，打破了傳統過去網路運作的模式，讓使用者能夠對等的進行通訊，不用透過中央伺服器的幫助。P2P 技術被廣泛應用於檔案的分享，藉由一群點（Peer）使用相同的通訊協下，建立一個邏輯的網路，使用者在這個邏輯網路內互相分享自己所擁有的資源。由於 P2P 網路的運作的方式，產生了大量的封包，消耗大量的網路頻寬，造成網路效能的低落，影響了網路的其他應用。如何有效管理 P2P 軟體對網路造成的影響，成為網路服務業者（ISP）以及網路管理人員的重大問題。

P2P 技術支援使用者可以利用 Internet 和其他使用者自由的交換各種檔案資料，在本篇論文的研究結果指出，有將近六成的分享內容屬於 mp3 音樂檔案，並且這些檔案中超過九成是屬於未經合法授權就在網路上散播的檔案。這種利用網路交換音樂和電影檔案的行為被認為是侵犯業者的智慧財產權，所以國內外各主要音樂業者都傾全力的圍堵與抵制提供 P2P 軟體的製造業者。

P2P 侵犯智慧財產權的議題，最早始於 2001 年 Napster 事件，當時一位十九歲的美國高中生 Shawn Fanning 發明這項網路檔案交換技術，由於使用人數快速增加而引起電影和唱片業者的圍剿，最後美國法院裁決 Shawn Fanning 的 Napster 敗訴，主要是因為 Napster 的 P2P 技術需經過中央伺服器來進行分享，隨後的 P2P 軟體改採取分散式的架構，來躲避法律問題，然而仍然受到多方的爭議。

雖然現在許多 P2P 軟體被和侵犯智慧財產權劃上等號，但 P2P 技術的本身是無罪的，其問題在於使用者的心態可議與其行為有罪。P2P 技術出現是為了解決目前網路遭遇的問題，它有效的利用網際網路上眾多的電腦資源，讓複雜的工作能分散平行的處理。網際網路最大的意義在於資源的分享，而 P2P 的技術正是達到資源分享的新模式。從過去主從式架構的計算模式，由伺服器提供資源，改變成現在的對等式架構，由網路上的各個電腦來提供與交換各自的資源。P2P 的技術本身沒有善惡之分，全看開發者及使用者如何的應用這項技術。

6.2 未來工作

在本研究中利用攔截封包的方式，針對交大資工與交大學生第十三宿舍的網路流量進行分析，所獲得的結果與目前的相關研究結果大致符合。由於現在大多數的網路路由器（router）都具有網路管理功能，也就是支援簡單網路管理協定（Simple Network Management Protocol, SNMP），於硬體直接提供 MIB 資料庫，未來我們可以藉由 SNMP，直接從各大骨幹網路的路由器上取得流量資訊，利用我們所提的研究方式來進行分析，以對目前 P2P 軟體使用情形進行更詳細的研究，例如 P2P 軟體用戶分佈的情形、P2P 網路動態以及分享內容。

再者，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，以保障企業與學校的網路使用品質。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，如何限制互相溝通的內容可能才是未來需要努力的課題，例如分享內容只能分享受到 DRM 管制許可之檔案。

參考文獻

[1] CD Cranor, T Johnson, O Spatscheck, V Shkapenyuk .Gigascope: A Stream Database for Network Applications. In Proceedings of the 2003 ACM SIGMOD international conference, pages: 647 – 651, San Diego, California 2003

[2] Jian Liang, Rakesh Kumar, and Keith W. Ross. The Kazaa Overlay: A Measurement Study. In Proceedings of the 19th IEEE Annual Computer Communications Workshop,pages 45-70 Oct. 2004

[3] J. Liang, R. Kumar, Y. Xi, and K. Ross. Pollution in P2P file sharing systems. In IEEE Infocom, Miami, FL, USA, March 2005

[4] 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

[5] RI：A. Crespo and H. Garcia-Molina. Routing Indices For Peer-to-Peer Systems. In Proceedings of the 22nd International Conference on Distributed Systems, pages 23-32, Vienna, Austria, 2002.

[6] S Sen, O Spatscheck, D Wang. Accurate, Scalable In-Network Identification of P2P Traffic Using Application Signatures. Proceedings of the 13th international conference on World Wide Web

[7] 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

[8] Richard Muntz, Miodrag, Potkonjak, Sasha Slijepcevic, Vincent Busam. Analysis of Internet music content distribution. Dimi Program Final Report D99-10

[9] 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

[10] T. Karagiannis, A. Broiodo, N. Brownlee, kc claffy, and M. Faloutsos. Is P2P dying or just hiding? In Globecom, Dallas, TX, USA, November 2004.

[11] Thomas Karagiannis, UC Riverside; Andre Broido, CAIDA, SDSC; Michalis Faloutsos, UC Riverside; Kc claffy, CAIDA, SDSC." Transport Layer Identification of P2P Traffic," IMC2004

[12] Thomas Karagiannis, Andre Broido, Nevil Brownlee, K. Claffy, and Michalis Faloutsos.

Filesharing in the Internet: A characterization of P2P traffic in the backbone. Technical report, University of California, Riverside, November 2003.

[13] J. Liang, R. Kumar and K.W. Ross, "Understanding KaZaA," submitted, 2004

[14] M. Jovanovic, F.S. Annexstein, and K.A. Berman. Scalability issues in large peer-to-peer networks - a case study of gnutella. Technical Report, University of Cincinnati, 2001.

[15] Napster. [Online]. Available : http://www.napster.com/

[16] Gnutella. [Online]. Available: http://www.gnutella.com/

[17] Emule [Online]. Available: http://www.emule-project.net/

[18] Edonkey. [Online]. Available:http://www.edonkey2000.com/

[19] WinMax. [Online]. Available: http://www.agry.purdue.edu/max/

[20] Kuro. [Online]. Available: www.kuro.com.tw/

[21] Ezpeer. [Online]. Available: http://www.ezpeer.com/index.html

[22] Skype. [Online]. Available: http://www.skype.com/

[23] MSN. [Online]. Available: http://www.skype.com/

[24] FastTrack [Online]. Available:http://www.slyck.com/ft.php?page=1

[25] KaZaA [Online]. Available : http://www.kazaa.com/us/index.htm

[26] Morpheus [Online]. Available : http://www.morpheussoftware.net/

[27] Bittorrent [Online]. Available : http://www.bittorrent.com/

[28] Yahoo Messenger [Online]. Available : http://www.yahoo.com/

[29] KaZaA-Lite [Online]. Available:http://c2p.6x.to/

[30] Baytsp http://www.baytsp.com/

[31] SkypeOut and SkypeIn Gateways http://463west.blogspot.com/2005/05/skypeout-and-skypein-gateways.html

[32] The Freenet Project http://freenet.sourceforge.net/

[33] Packeteer, Inc., Strategies for Managing Application Traffic. [Online] available: http://www.packeteer.com/resources/prod-sol/MngAppTraf.pdf

[34] Pastry Routing Table http://www.scs.cs.nyu.edu/V22.0480-005/notes/l24.pdf

[35] U.S. Court of Appeals for the Ninth Circuit http://www.ce9.uscourts.gov/web/newopinions.nsf/0/c4f204f69c2538f6882569f100616b06?OpenDocument

[36] 劉怡玟：點對點檔案分享軟體使用行爲之研究，2004；國立中山大學傳播管理研究所碩士論文