以網路流量分析網路使用者之行為
——以淡江大學為例

黃明達、陳正宏
淡江大學資訊管理學系

摘要

本研究是以實際網路流量數據，來分析使用者的網路行為。目前大部分流量分析主要是以 Netflow 資料為主，但此種方式會造成「無法分類」(Others)類別流量過大的情況。因此本研究設計並實作一套資料收集方法及軟體，不但採用 Netflow 資料且結合監聽封包的方式，來改善其統計結果。本研究是以淡江大學校園網路作為個案研究。

本研究發現：佔用頻寬最大比例者為 P2P 軟體之檔案資料傳輸，約佔總頻寬 34%，其中最大項目為 eDonkey，佔總頻寬使用量約 24%；群組的流量，大部分是群組內少數使用者所使用；寬頻及窄頻使用者行為有顯著不同，其中，寬頻使用者是以下載檔案為主，而窄頻使用者是以收信為主；在檔案的傳輸中，以副檔名為 dat、mpg、avi 等影片檔及副檔名為 iso、mdf 、img 等光碟映射檔為主要傳輸類別。

關鍵字：使用者行為、Netflow、網路流量

A Study of Network User Behavior Based on the Network Traffic Flow –A Case Study of TamKang University

Ming-Dar Hwang、Chen Cheng Hung
Department of Information Management, TamKang University

Abstract

Based on the network traffic flow data, this study will analyze the network user's behaviors. Currently, most of the traffic analysis reports are mainly based on the Cisco Netflow data only, which will categorize a lot of services belonging to 'others'. Therefore, we use a different approach and design software to collect data. Our approach, combining Netflow data and packet-sniffing, will reduce the percentage of 'others'. TamKang University campus network environment is our study case.

Our study has the following findings: most of the network bandwidth is used by P2P software to transfer the files, approximately 34%. Among the P2P software, the largest service item is the eDonkey, approximately 24%. Most of the traffic in each group is used by only a few users. The behaviors of the broadband users are significantly different than the narrowband users. To download the files is the major purpose of the broadband users, but the narrowband users are used to get their Emails. The files with the video files extensions of dat、mpg、avi and with image file extensions of iso、mdf 、img are the primary file types transferred by campus users.

Keywords: User Behavior, Netflow, Network Traffic

壹、緒論

一、研究背景與動機

隨著網際網路(Internet)的興起以及寬頻的普及，各式各樣的服務，以網路作為媒介，傳遞給使用者。而隨著越來越多樣的服務，以及越來越多的使用者，對於頻寬的使用量也日益增大。因此對於網路使用者行為的研究有其必要性。

目前一般的網路使用者行為研究多以問卷的方式來進行，藉以研究使用者在網路上所從事的行為。而這樣的方式雖然可以瞭解到使用者的使用目的，但是對於以網路管理者的觀點來說，卻無法得到較準確的數據。因為使用者最常從事的行為，不一定是網路頻寬最大的使用來源。(蕃薯藤數位科技，民 92；教育部統計處，民 92)

除了使用問卷方式之外，也有直接使用統計分析網路流量的方式(杜家麟，民 89；翁瑞鋒，民 90；張仕達，民 88)。

目前台灣學術網路(Taiwan Academic Network，TANet)(教育部資訊網，民 92)上最常使用的網路流量分析方法為使用 Cisco 公司路由器(Router)所丟出來的 Netflow 資料來做分析。Netflow 為 Cisco 公司所制定的一種流量統計的技術(Cisco IOS Netflow)。透過 Netflow 技術，路由器會不斷地將所經過的封包資料經過初步整理後以 UDP(User Datagram Protocol)的形式丟給負責收集資料的伺服器。而以 Netflow 資料為主的分析方式乃是以埠號(Port Number)作為判斷所使用服務(Service)種類的依據。但此種方式容易造成服務種類辨別上的誤判或無法辨識，尤其一些地下 FTP(File Transfer Protocol)站台，常常將埠號改變成其他服務的埠號，企圖魚目混珠，避免被發現。或是像 FTP 這類的傳輸協定(Protocol)，使用一個以上的連結(Connection)來運作，其中傳輸 FTP 命令所使用連結是採用固定埠號，但資料傳輸的連結則使用隨機埠號。由於資料傳輸所使用的連結是主要的流量來源且日漸廣泛使用的 P2P（Peer-to-Peer）軟體，許多都提供可以更改埠號的功能，因而對於流量統計更是一大挑戰。因此，僅使用埠號來作為服務種類的判斷依據，較容易造成流量統計上的誤判及 Others 流量過高的情形。

另一種的流量分析方式為使用封包監聽軟體，監聽所有流經路由器的封包，並分析每個連結的服務種類，統計流量大小等(簡榮成，民 88；楊明宮，民 87)。但是此種方法的困難點在於，若單只使用此方法，在一些網路頻寬較大的骨幹上，封包流量太大，很有可能會因機器處理速度不夠快而掉封包，導致統計分析準確性的降低。

目前 TANet 各區網中心的流量統計分析，大部分皆以 Netflow 資料分析及 MRTG(Multi Router Traffic Grapher) 為主。MRTG 所產生的流量資訊非常的少，只能看出整個單位或是整個學校依時間排序之流量統計。但到底是哪些使用者，以及其所使用的服務種類，則完全無從得知。而單以 Netflow 所產生的流量統計，在目前大量使用 FTP 及 P2P 軟體的環境下，往往造成 Others 流量超過 30~40%，甚至超過 50%以上亦常常可見(教育部電算中心，民 92)。如此一來，對於該區網或校內的網路頻寬使

用分析上，等於有可能超過一半以上的頻寬使用情況是未知。

網路管理者無法得到足夠的資訊，對於該區網、學校內使用者網路行為亦無法得知甚詳，造成管理上的漏洞。

因此，如何有效的分析使用者網路行為，並瞭解目前大學生所從事的網路行為，且提供給決策者作為參考，是相當重要的議題。

二、研究目的

- 瞭解大學內校園網路之行為與狀況
- 瞭解在不同系所群組間，其使用者網路行為差異性
- 作為網路管理者決策之參考。

三、研究對象

淡江大學目前約有 3 萬名師生，為全國人數最多之大專院校之一(教育部統計處，民 92)。眾多師生的大學校園網路流量行為，應有相當的代表性。

貳、文獻探討

一、網路流量收集方法

流量收集為使用者行為分析的第一步，根據張仕達(民 88) 對網路流量資料的收集方式可為三種。

1.伺服器導向(Server Oriented)

伺服器導向的資料收集方式大多以分析各種伺服器的紀錄檔(Log File)為主。以這種方式的使用者行為分析通常是針對單一特定網站或是針對網頁瀏覽行為作分析。而不管是針對特定網站或是對使用者網頁瀏覽行為作分析，其都只佔整個網路使用量的一部分，無法代表整個網路的使用狀況。例如無法分析到 FTP 檔案傳輸或是其他服務類型，且這些項目反而佔用網路頻寬相當大的比例。

2.使用者導向(Client Oriented)

以使用者導向的行為分析，需要在每一個 Client 端上裝置代理人(Agent)來紀錄使用者使用網路的相關資訊。但若欲對於大範圍的樣本做紀錄，需要找到足夠的使用者安裝此軟體，使用者對於其隱私權也會有其顧慮，因此使用此種方式並不容易。

3.通道導向(Channel Oriented)

通道導向以路由器或監聽封包的方式，收集所有經過該節點的封包。目前大部分的路由器都提供流量統計的功能，可以將所有經過路由器的流量，以特定的形式遞送

出來。以 Cisco 公司的 Netflow 技術為例，因其所提供的資料都已經過路由器初步整理，方便後續的處理，所以目前大部分的流量分析方式都以此方式來達成。而監聽封包的方式為收集所有流經過的封包，需自行做處理及統計，因此若所監聽的網路流量太大時，效能將是一大考驗。

二、相關文獻整理

如表 1。

參、研究方法及設計

一、研究方法

本研究是採用個案研究法。主要是以實際網路流量，來分析網路使用者之行為。由於流量資料的收集需耗費較長之時間，資料量亦非常龐大，且網路流量資料的擷取會牽涉到校園網路安全及網路設備配置，因作者之一為淡江大學資訊中心主任，有其運作方便性，故本研究僅以淡江大學作為個案研究。

表1：相關文獻

篇名	作者/年份 出版單位或期刊	與本研究相關之處
資料挖掘在網際網路流量之分析研究	杜家麟/2000 靜宜大學資訊管理研究所碩士論文	單以 Netflow 資料為主，有資料不完整的問題，並著重於使用者分群。
以流量為基礎之網路分析系統	林育生/2001 國防大學中正理工學院電子工程研究所碩士論文	單以 Netflow 資料為主，但其主要在使用者行為的分群上，而不在使用者行為分析。
廣域網路流量監測與分析工具之製作	簡榮成/1999 國立中正大學電機工程研究所碩士論文	單以封包過濾法，於大流量的骨幹上分析會有資料不完整的問題。
以 Flow 為基礎之網際網路流量量測分析	張傑生/1997 國立交通大學資訊管理研究所碩士論文	提出 Flow 概念。並比較 Flow 與傳統解封包的優缺點。
網路流量監視器內封包過濾器之實現	楊明宮/1998 大同工學院資訊工程研究所碩士論文	單以封包過濾法，於大流量的骨幹上分析會有效能上的問題。
從網路流量中發掘網際網路使用者之使用模式	張仕達/1999 國立中山大學資訊管理研究所碩士論文	以 Netflow 資料為主，其主要目的在於偵測異常網路狀況。
NetFlow: Information loss or win?	Robin Sommer, Anja Feldmann/ IMW2002	本篇論文在比較 3 種不同的流量資料來源所造成的影響。

二、資料收集方式

本研究在資料收集上，是以 Netflow 所產生資料為主，而封包分析為輔。

由於目前並沒有適合的現成軟體可以達到我們的需求，因此本研究自行撰寫有關封包收集及分析的軟體。以下簡稱該軟體為 PacketAnalyzer。

PacketAnalyzer 全部以 C 語言撰寫，程式碼總共約 2,000 行，使用 OS 為 FreeBSD 4.7，資料庫為 MySQL。

三、資料收集硬體架構

本研究資料收集硬體架構如圖 1。使用兩台 PC，皆連接到校內 Cisco 6509 交換器 (Switch)。其中一台(為一般 PC，Intel P3 等級)，負責接收由 Cisco 公司 7507 路由器所送出來的 Netflow 封包。另一台機器(Intel P4 2.4G 雙 CPU、1 GB 記憶體、Intel Gigabit 網路卡，OS 為 FreeBSD 4.7)執行本研究所撰寫之封包收集分析程式 PacketAnalyzer，透過 Mirror Port 的方式來監聽 6509 與 7507 連結介面上所流經過的封包。

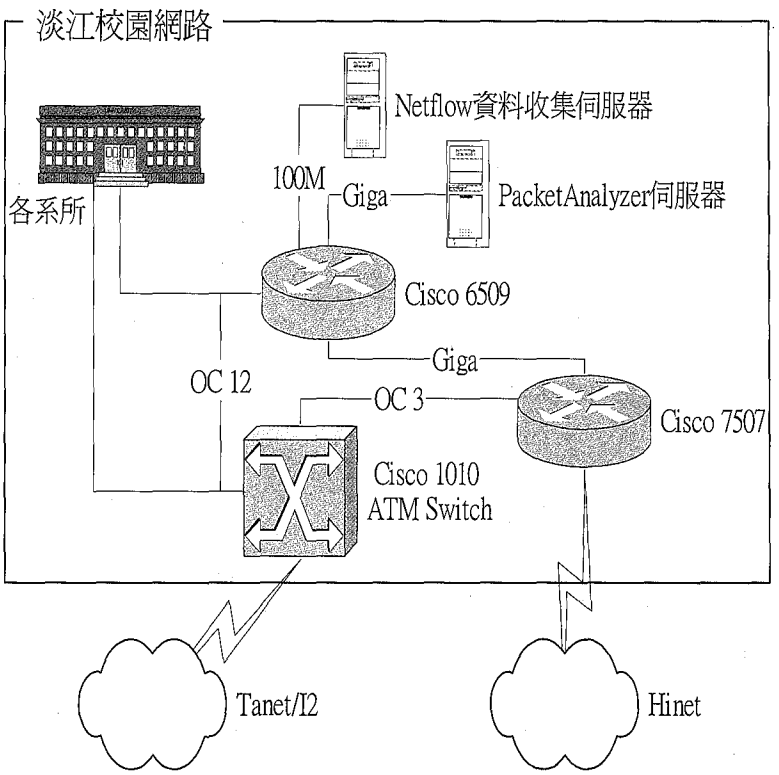

圖1：流量收集硬體架構圖

四、軟體架構

如圖 2 所示，本研究資料來源以 Netflow 資料為主，PacketAnalyzer 所收集到的資料為輔。路由器所送出的 Netflow 資料乃為一段時間內，所有經過該路由器的封包統計資料，而這些資料僅為每個連結(Connection)的一部分。因此我們需將這些資料先彙整成以連結為單位的流量資訊，然後再匯入資料庫內。而 PacketAnalyzer 是以即時(Real time)的方式，辨認每一個連結封包內容的特定字串，然後將所辨認出來的通訊協定種類，與該次連結的相關資訊(如時間、來源 IP 位址、目的 IP 位址、來源埠號、目的埠號、通訊協定種類等)，記錄於 Log file 中。之後則依照 PacketAnalyzer 所辨認出來的連結，對由 Netflow 資料所建立的資料庫做更新的動作。PacketAnalyzer 最主要的目的在於調整 Netflow 以埠號為判斷依據所無法辨認以及錯誤辨認的部分。

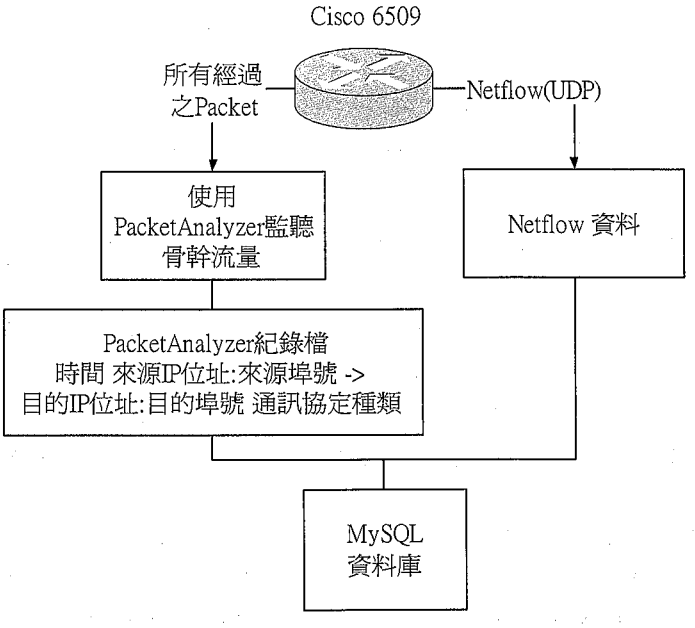

圖2：流量資料架構圖

五、PacketAnalyzer 程式資料處理流程

本研究所撰寫 PacketAnalyzer 程式之處理流程如圖 3。在底層封包資料的收集上使用 Libnids (Rafal Wojtczuk，2002)函式庫來處理。Libnids 為 Rafal Wojtczuk 所設計，原先設計是作為網路入侵偵測系統(Network Intrusion Detection System)的一個模組。其發展至今已有三年，並經過穩定度的測試。

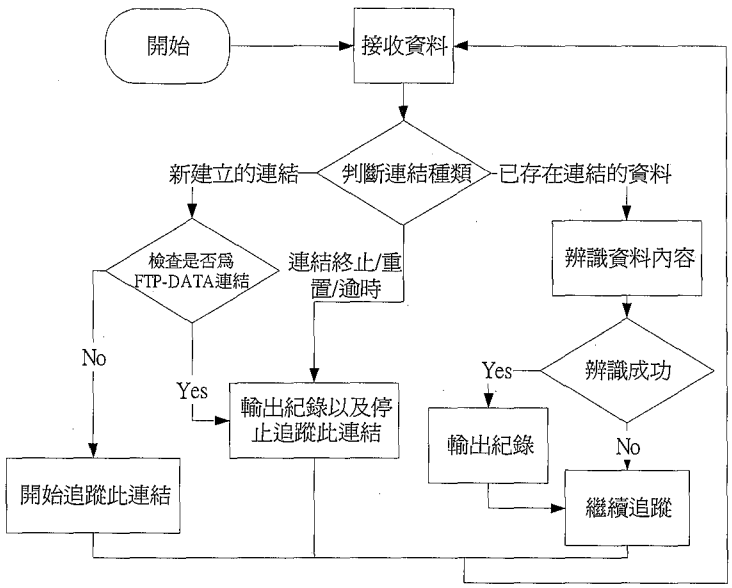

圖3： PacketAnalyzer 對封包資料的處理流程圖

本 PacketAnalyzer 程式在收到由 Libnids 所處理好的應用層資料後，可由 FTP command 中特定的幾個字串來辨識該連結是否為 FTP 連結。這些字串包括： PASV、PORT、SYST、TYPE A、TYPE I、CWD。若 PacketAnalyzer 發現到傳輸資料的開頭有以上這 6 個字串之一，就會將此連結視為一個 FTP 連結，並開始監控此連結。當其偵測到 PASV 或是 PORT 指令時，會將其後面所接之網路 IP 位址及埠號紀錄下來，因為這組網路 IP 位址及埠號，即為此 FTP 連結稍後所要建立的另一個 FTP-DATA 連結的關鍵。

在 PASV 或是 PORT 字串之後，通常都是 GET 或是 PUT 指令，分別表示下載與上傳檔案。本程式在發現此兩指令時，會將其後所接之檔名，配對到剛剛所紀錄的網路 IP 位址及埠號。這時因為 FTP 程式已經溝通好 FTP-DATA 連結所建立的方式以及所要傳輸的內容並開始傳輸。因此 PacketAnalyzer 程式將會偵測到一個新的連結建立，而根據之前所記錄到的網路 IP 位址及埠號，我們將可以知道此連結是否為 FTP-DATA 的連結，若是的話，我們還可以知道此連結所傳輸的內容以及確切的連結資料。這時 PacketAnalyzer 程式將會把這些連結資料包括時間、雙方的網路 IP 位址、埠號、及傳輸的檔名等紀錄下來，並繼續監聽。

至於 eDonkey 的部分則跟 FTP 辨識方法類似，惟辨識之關鍵字串不同。

六、研究過程

本研究從 91 年 9 月開始進行，前期以相關研究及程式撰寫為主，後期以資料分析為主，其中較為重要之時程如表 2所示。

表2：研究過程的重要日誌表

時間	事項
91/9/15	開始收集文獻
91/10/2	開始研究 Cisco 公司 Netflow 技術，並參考 TANet 上各大學所用之流量統計系統
91/10/14	著手開始寫程式，將 Netflow 資料收集彙整且匯入資料庫
91/11/3	研究 TCP/IP 以及各種常見通訊協定
91/11/18	參考 dsniff 程式0，並使用 Libnids 函式庫來撰寫本研究所需之封包分析程式
91/12/1	於校園骨幹上試跑
91/12/7	適逢校園骨幹變動，原先為 PacketAnalyzer 主機直接接上 Cisco 6509 switch，因 switch port 不夠改用分光器轉接，發現使用 Libnids 函式庫於分光器轉接下無法收集到資料。後來因為學校有借測的 Foundry router 有多餘的 port 可以接，才又收得到資料。中間暫停約 2 個星期。
91/12/23	發現封包分析程式有資料遺失的問題，後來發現應為當初封包分析程式參考 dsniff 所寫，dsniff 使用 hash table 對於 PacketAnalyzer 來說並不適合，造成程式不斷的在做排序的動作，以及部分程式邏輯上的問題，而造成部分連結資料沒被紀錄到，後改用 avl tree 取代 hash table 解決此問題。
92/1/3	撰寫將 PROXY log、PacketAnalyzer log、Netflow 整合程式
92/1/15	重新抓取資料
92/2/1	決定分析 eDonkey 資料，開始研究 eDonkey
92/2/24	新版程式上線，開始收集 eDonkey 資料
92/3/20	重新抓取兩個星期網路流量資料，作為本研究使用

肆、資料分析

一、分析架構

圖 4。其一是依所使用的通訊協定來分析，其二則從 FTP、HTTP、eDonkey 等通訊協定所存取的副檔名來分析。

通訊協定的分析以 Netflow 的資料再加上 PacketAnalyzer 所辨識出來的資料作為分析來源。

根據本研究發現，HTTP、FTP 及 eDonkey 約佔了本校頻寬使用量 60%，因此副檔名的分析採用以 PROXY 的 log 再加上 PacketAnalyzer 所擷取到的檔案名稱為來源，應具有相當的代表性。

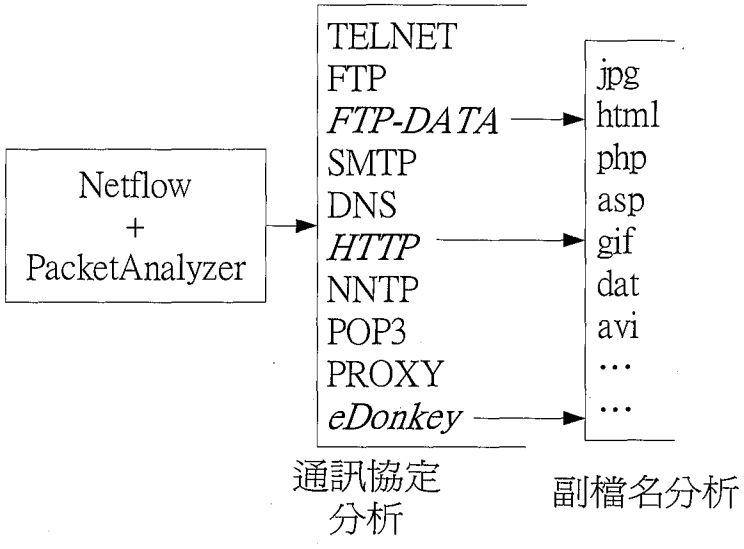

圖4：資料分析架構圖

二、分析說明

在流量資料分析上，我們採用 IN/OUT(流入/流出)流量加總計算的方式。

在通訊協定使用狀況分析上，我們主要分成以下 21 個項目，括號內為其通訊協定所採用之埠號：

1. FTP-DATA(20)—使用 FTP 軟體時，用於傳輸資料的連結。
2. FTP(21)—使用檔案傳輸軟體時，用於傳輸 FTP 命令的連結。
3. TELNET(23)—遠端登入，常用於 UNIX、Linux 等系統及 BBS。
4. SMTP(25)—通常用於 Email 軟體寄信。
5. DNS(53)—通常用於網域名稱解譯。
6. HTTP(80)—即目前最廣泛使用的 WWW。
7. POP3(110)—通常用於 Outlook 等 Email 軟體收信。
8. NNTP(119)—用於傳輸新聞群組資料。
9. NetBios-ns(137)—微軟 NetBios 名稱查詢。
10. NetBios-ssn(139)—微軟 NetBios 連結控制。
11. HTTPS(443)—經過 SSL 加密的 WWW。
12. KaZaA(1214)—為一知名 P2P 軟體。
13. MediaPlayer(1755) —微軟 MediaPlayer 之串流多媒體。
14. Lineage(2000) —為一知名線上網路遊戲，名稱為天堂。
15. Proxy(3128) —代理伺服器，此處泛指校內使用者使用 Proxy 伺服器的流量。

16.eDonkey(4661、4662、4665)—eDonkey 為目前熱門之 P2P 軟體。每個 eDonkey 的使用者連接伺服器後，可先向伺服器查詢特定的關鍵字或檔案類型，然後找出且向其他有提供檔案分享的使用者來抓取所要檔案。其最主要的特點是使用 MFTP(Multisource File Transfer Protocol)技術，取代了傳統只向單一來源抓取檔案的方式。於 MFTP 技術下，會將每個檔案切割成細小的區塊，然後同時向不同的來源，也就是其他使用 eDonkey 分享該檔案的使用者來抓取。第二個特點在於對檔案的辨認方式不再是使用檔名辨識，而是根據檔案內容所計算出來的 MD4(R. Rivest, 1992)值。第三個特點是它使用非集中式的伺服器。在查詢檔案時，不僅可以向目前所連結的伺服器查詢，還可以對其他名單上的伺服器作延伸查詢。換言之，eDonkey 將所有的使用者都連結在一起。當使用者連上伺服器之後，可以看到目前所有在運作中的伺服器數目、連接人數、及分享的檔案數目。一般使用者都是超過 50 萬人，且分享超過 4,000 萬個檔案(MetaMachine，2003)。

17.MSN(6257)—微軟公司的即時傳訊軟體 MSN messenger。

18.Napster(6677、6699)—Napster 為另一知名的 P2P 軟體，以分享 MP3 為主。

19.RO(6900、4000、5000) —為一知名線上網路遊戲，名稱為仙境傳說。

20.CS_game(27015)—為一知名的連線對戰遊戲，名稱為 Counter-Strike。

21.Others—若不屬於以上任何一種通訊協定即歸類為 Others。

因為時間上的關係，本研究只擷取 2003/3/20~4/2 這兩週的資料。以下所有的表格，皆為這兩週資料取日平均值所分析出來的結果，表上不再註明。

淡江大學校園網路有被鄰近 7 所的國中及國小所連接。本研究只以淡江的網路 IP 位址(163.13.*.*)為處理範圍，非此範圍的資料將不採用。

於研究期間，淡江大學每秒的資料傳輸量平均為 290Mbits(流入 116Mbits/流出 174Mbits)，尖峰時可達 380Mbits(流入 152Mbits/流出 228Mbits)，平均一天的資料量為 3.1TB(Terabytes)。Netflow 一天的資料約為 2.5GB。PacketAnalyzer Log 一天的資料約為 2GB。Netflow 和 PacketAnalyzer 資料經過彙整後匯入 MySQL，一天約為 4,500 萬筆連結資料，容量大小約為 6GB。

資料彙整、分析程式、及 MySQL 資料庫，皆執行於一台 P4 2.4G 雙 CPU、記憶體 1GB 的 PC 上。一天的 Netflow 資料整理及匯入 MySQL 約要 10 個小時，接著再將 PacketAnalyzer Log 的資料更新至 MySQL 約需 8 小時，總計處理一天資料約需 18 小時。

資料彙整的工作，是使用自行撰寫之軟體。主要分成兩大部分。一為將 Netflow 原始資料整理後，匯入 MySQL 資料庫，此部分為使用 C 語言，程式碼約 1,000 行。第二部分是將 PacketAnalyzer 資料更新至 MySQL 資料庫內，此部分為使用 PHP 語言來完成，程式碼約 500 行。

三、分析結果

由表 3 全校網路通訊協定之使用狀況，我們可以發現本研究所提出的方法，的確可有效降低 Others 的流量百分比。我們也發現除了 Others 外，eDonkey 是頻寬使用的主要來源，約佔了全部頻寬的 24%，而全部 P2P 檔案交換軟體(包含 eDonkey、Napster、KaZaA)共佔了總頻寬約 34%。在連結次數排行上，eDonkey 因採用 MFTP 架構，使得連結次數遠高於第二名 Others。而在平均流量大小上，排行第一的為 NNTP，乃因新聞群組於連結建立後，會保持連線一段時間，故其每次平均大小遠高於其他通訊協定；排行第二為 Napster，主要內容為 MP3 分享，而一般普通 MP3 約為 3~4MB，因此其平均大小也遠高於其他項目。

在平均連結時間上，較為特殊的是 FTP 項目（01:03:55），較其他項目高，應為 CuteFTP 等知名 FTP 軟體提供持續連線功能所導致的結果。而 DNS 一般連結時間應該都很短，這邊卻很高，實屬不合理。由原始資料檢查後，發現應為其他服務也使用 53 埠號所導致。HTTP 項目偏高（00:44:10），其原因應為常於網頁下載檔案。

表3：各通訊協定流量大小及連結次數統計表(全校總和)

通訊協定	流量大小			連結次數			平均大小		連結時間		平均連結時間 (時:分:秒)
	單位: MB	%	序位	單位: 千次	%	序位	單位: KB	序位	單位: 千天	%	
Others	769,339	24.52	1	11,664	24.58	2	66	10	296.55	38.99	00:36:20
eDonkey	743,029	23.68	2	24,471	51.58	1	30	16	158.12	20.79	00:09:18
FTP-DATA	630,932	20.11	3	408	0.86	7	1,546	3	1.8	0.24	00:06:20
HTTP	481,631	15.35	4	6,415	13.52	3	75	9	196.82	25.88	00:44:10
Napster	323,405	10.31	5	175	0.37	13	1,844	2	0.54	0.07	00:04:25
Proxy	39,472	1.26	6	744	1.57	5	53	14	25.24	3.32	00:48:49
SMTP	25,424	0.81	7	388	0.82	8	66	11	0.48	0.06	00:01:46
CS_game	21,366	0.68	8	357	0.75	10	60	12	0.49	0.06	00:01:58
FTP	18,644	0.59	9	1,001	2.11	4	19	17	44.42	5.84	01:03:55
RO	17,836	0.57	10	209	0.44	12	85	7	2.36	0.31	00:16:13
POP3	15,289	0.49	11	276	0.58	11	55	13	0.17	0.02	00:00:52
Telnet	14,123	0.45	12	370	0.78	9	38	15	4.01	0.53	00:15:35
NNTP	12,887	0.41	13	6	0.01	21	2,046	1	0.11	0.01	00:24:46
KaZaA	6,950	0.22	14	87	0.18	15	80	8	0.54	0.07	00:04:25
MediaPlayer	6,633	0.21	15	25	0.05	18	266	4	0.35	0.05	00:20:16
NetBios-ssn	4,714	0.15	16	44	0.09	17	107	6	0.02	0	00:00:35
DNS	2,360	0.08	17	536	1.13	6	4	19	26.32	3.46	01:10:41
Lineage	2,240	0.07	18	10	0.02	20	214	5	0.11	0.01	00:14:38
Https	752	0.02	19	81	0.17	16	9	18	0.17	0.02	00:03:02
MSN	204	0.01	20	155	0.33	14	1	20	2.29	0.3	00:21:19
NetBios-ns	21	0	21	23	0.05	19	1	21	0.29	0.04	00:18:25

表 4 中，因淡江所有的對外 HTTP(80)通訊協定，都要經過代理伺服器(Proxy Server)，再加上淡江的 FTP 站台使用率很高，因此使得校級主機的流量非常高。在 ADSL 及 Cable 這兩個群組的使用量上，雖然 Cable 提供較高的傳輸速度，但淡江 ADSL 的使用人數約是 Cable 的 2.5 倍，因此 ADSL 群組的使用量排行第一。比較特殊的是 ADSL 群組的連結數目佔了全校約 40%，為所有群組之冠。可能是 eDonkey 使用人數眾多以及 eDonkey 本身採分散式傳輸所造成的結果。

表4：群組流量大小及連結次數統計表(流量大小排行前 10 名)

群組名稱	流量大小			連結次數			平均大小		連結時間		平均連結時間(時:分:秒)
	單位:MB	%	序位	單位:千次	%	序位	單位:KB	序位	單位:千天	%	
ADSL	735,039	24.45	1	14,755	39.68	1	50	35	2.26	11.12	00:00:13
校級主機	619,047	20.59	2	6,429	17.29	2	96	24	15.12	74.43	00:03:23
CABLE	361,059	12.01	3	5,255	14.13	3	69	27	1.32	6.51	00:00:21
電機系	151,443	5.04	4	1,408	3.79	5	108	22	0.16	0.77	00:00:09
航空系	138,496	4.61	5	214	0.58	18	646	2	0.02	0.12	00:00:09
資管系	138,406	4.6	6	306	0.82	16	452	3	0.04	0.18	00:00:10
機械系	127,525	4.24	7	434	1.17	14	294	9	0.04	0.21	00:00:08
資工系	117,664	3.91	8	1,846	4.96	4	64	29	0.24	1.18	00:00:11
化學系	99,172	3.3	9	640	1.72	8	155	16	0.07	0.32	00:00:08
松濤三館	66,490	2.21	10	479	1.29	11	139	18	0.31	1.51	00:00:55

表5：電話撥接流量大小及連結次數統計表前 3 名

群組名稱	通訊協定	流量大小			連結次數			平均大小		連結時間		平均連結時間(時:分:秒)
		單位:MB	%	序位	單位:千次	%	序位	單位:KB	序位	單位:千天	%	
電話撥接	SMTP	15	0	340	0	0	395	49	268	0	0	00:00:14
電話撥接	FTP	13	0	348	2	0	256	8	460	0	0	00:00:07
電話撥接	POP3	13	0	349	0	0	463	103	186	0	0	00:00:06

各群組的通訊協定使用量排行前 3 名，如表 6，大部分含 eDonkey 及 FTP-DATA。

從表 5 及表 6 可知，窄頻(電話撥接)使用者，其網路行為與寬頻(ADSL、Cable)有顯著不同。寬頻使用者以檔案傳輸為主，而窄頻使用者以收發 Email 為主。

表6：總流量前 10 大群組內通訊協定流量大小排行前 3 名

群組名稱	通訊協定	流量大小			連結次數			平均大小		連結時間		平均連結時間 (時:分:秒)
		單位: MB	%	序位	單位: 千次	%	序位	單位: KB	序位	單位: 千天	%	
ADSL	eDonkey	315373	10.49	2	12,260	32.97	1	26	339	1.13	5.56	00:00:07
	Others	142370	4.74	3	1,284	3.45	5	111	179	0.34	1.66	00:00:22
	Napster	136349	4.54	4	113	0.31	39	1,201	66	0.02	0.11	00:00:17
校級主機	HTTP	436935	14.53	1	4,645	12.49	2	94	196	11.79	58.04	00:03:39
	FTP-DATA	108183	3.6	7	123	0.33	37	878	73	0.1	0.48	00:01:07
	SMTP	20758	0.69	27	248	0.67	21	84	205	0.03	0.13	00:00:09
CABLE	eDonkey	132008	4.39	5	4,246	11.42	3	31	319	0.42	2.05	00:00:08
	Others	82143	2.73	8	454	1.22	10	181	141	0.14	0.69	00:00:26
	FTP-DATA	68308	2.27	12	40	0.11	69	1,709	54	0.01	0.03	00:00:14
電機系	eDonkey	80150	2.67	10	944	2.54	6	85	204	0.11	0.57	00:00:10
	Others	40556	1.35	21	170	0.46	29	239	118	0.01	0.06	00:00:05
	FTP-DATA	15130	0.5	33	11	0.03	105	1,369	64	0	0.01	00:00:17
航空系	Others	81102	2.7	9	70	0.19	52	1,163	67	0.01	0.03	00:00:08
	FTP-DATA	44420	1.48	17	7	0.02	125	6,466	21	0	0.01	00:00:14
	FTP	6400	0.21	64	3	0.01	184	1,871	52	0	0.01	00:00:29
資管系	Others	77527	2.58	11	41	0.11	68	1,882	51	0	0.02	00:00:09
	FTP-DATA	43463	1.45	18	5	0.01	150	8,583	14	0	0	00:00:10
	eDonkey	6027	0.2	67	48	0.13	65	127	167	0.01	0.03	00:00:09
機械系	Others	67663	2.25	13	87	0.23	43	778	79	0.01	0.03	00:00:06
	FTP-DATA	44551	1.48	16	6	0.01	143	7,991	15	0	0.01	00:00:24
	eDonkey	12577	0.42	41	321	0.86	15	39	295	0.03	0.16	00:00:08
資工系	Others	47208	1.57	15	219	0.59	23	216	123	0	0.02	00:00:01
	FTP-DATA	40981	1.36	20	19	0.05	87	2,152	47	0	0.01	00:00:08
	eDonkey	19554	0.65	30	1,495	4.02	4	13	422	0.1	0.52	00:00:06
化學系	Others	42488	1.41	19	85	0.23	44	498	94	0.01	0.03	00:00:06
	FTP-DATA	19742	0.66	29	3	0.01	215	7,809	16	0.01	0	00:00:18
	eDonkey	13632	0.45	35	466	1.25	9	29	324	0.05	0.24	00:00:09
松濤三館	Napster	24709	0.82	24	6	0.02	138	4,156	29	0	0.01	00:00:39
	Others	22527	0.75	25	296	0.8	17	76	217	0.27	1.31	00:01:17
	FTP-DATA	11104	0.37	47	7	0.02	123	1,554	59	0	0.01	00:00:14

表 7 為各群組內個別 IP 位址流量的各項統計數據，我們可以發現大部分的群組流量都是由少部分的 IP 所使用的。而由變異係數(=標準差/平均流量)知，幾乎所有的變異係數都是大於 1，也就是說，不管在任何群組內，每個 IP 位址之間的使用量，都有非常大差異。

表7：各群組內個別 IP 位址流量最大/平均/標準差/變異係數統計表(變異係數排行前 10 名)

群組名稱	總流量 (單位:MB)	群組內單一 IP 位址流量最大值 (單位:MB)	最大單一 IP 位址流量佔總流量%	平均流量 (單位:MB)	群組內單一 IP 位址流量最小值 (單位:MB)	IP 位址總數	標準差	變異係數
資訊中心	143,447	75,807	52.85	560	84	256	5,149	9.19
圖書館	158,555	51,962	32.77	363	100	437	3,069	8.46
外語學院	15,235	13,139	86.24	203	92	75	1,519	7.48
文學院、會文館	21,559	18,423	85.45	295	100	73	2,171	7.35
大傳系	45,134	39,492	87.50	765	100	59	5,159	6.74
統計系	42,620	38,577	90.51	836	176	51	5,397	6.46
國際學院	1,720	1,655	96.22	400	100	43	252	6.30
工學院	110,318	59,298	53.75	927	100	119	5,728	6.18
松濤一館	88,068	80,929	91.89	1,957	100	45	12,056	6.16
數學系	20,479	15,248	74.46	330	100	62	1,949	5.90

在 FTP 傳輸檔案的副檔名統計上(表 8)，dat、mpg、avi 等多媒體格式包辦了前 5 名中的 3 個，佔了大約全部流量的 63%。由此可見，電影、MTV、動畫等多媒體檔案在網路上互傳的程度非常高。而 iso、mdf、img 等皆為光碟映射檔，排行也在前 10 名內，這與寬頻及燒錄機的普及，應有很大的關聯性。在傳輸次數上，前兩名分別為 zip 及 exe，而且傳輸的次數遠比其他的高很多，可見應用程式的下載，亦為一般使用者的 FTP 項目。

表8：使用 FTP 傳輸流量統計前 10 名(依副檔名排序)

副檔名	傳輸大小			傳輸次數			平均大小/次	
	單位: MB	佔全部%	序位	單位: 次	佔全部%	序位	單位: KB	序位
dat	189,096	31.42	1	3,315	2.78	9	57,039	7
mpg	145,084	24.1	2	3,959	3.32	6	36,647	9
iso	58,233	9.67	3	2,775	2.32	10	20,985	11
exe	56,975	9.47	4	23,229	19.46	2	2,453	67
avi	42,525	7.06	5	1,347	1.13	14	31,575	10
mdf	18,990	3.15	6	89	0.07	50	213,846	2
img	15,426	2.56	7	126	0.11	42	122,620	4
bin	8,450	1.4	8	111	0.09	46	76,124	5
mp3	8,212	1.36	9	4,307	3.61	5	1,907	71
zip	8,068	1.34	10	31,191	26.13	1	259	110

在 HTTP 傳輸流量排行前 10 名副檔名中(表 9)，我們可以發現和 FTP 傳輸前 10 名(表 8)有很大的不同點。HTTP 傳輸流量最多的項目為 jpg 圖檔類，其次是 zip、exe、rar 等應用程式類、再其次才是多媒體檔案如 rm、wmv 及 mp3，而 FTP 傳輸量最多的項目，則為多媒體檔案或是光碟檔。

在傳輸次數上，很明顯的，FTP 傳輸的次數比 HTTP 少很多。而 HTTP 流量排行較高的副檔名，其平均大小大都小於 FTP。由此可見，FTP 傳輸副檔名流量較高的項目應是由少部分的人所創造出來的，而 HTTP 副檔名流量較高的項目，則是由大量的使用者存取所造成的。

表 9 中，副檔名 rar 的平均大小為 4KB，屬偏低，較不合於經驗值，下次重新統計時，可注意之。

表9：使用 HTTP 傳輸流量統計前 10 名(依副檔名排序)

副檔名	傳輸大小			傳輸次數			平均大小/次	
	單位:MB	佔全部%	序位	單位:次	佔全部%	序位	單位:KB	序位
jpg	40,839	18.9	1	2,899,947	15.65	2	14	74
zip	25,170	11.65	2	104,387	0.56	13	241	49
exe	24,392	11.29	3	175,875	0.95	12	139	60
gif	18,042	8.35	4	8,376,234	45.21	1	2	90
rm	11,964	5.54	5	68,677	0.37	14	174	54
wmv	11,962	5.54	6	18,111	0.1	21	661	7
rar	10,773	4.99	7	2,604,801	14.06	3	4	86
mp3	9,031	4.18	8	27,446	0.15	17	329	39
swf	8,722	4.04	9	269,004	1.45	9	32	68
mpg	8,341	3.86	10	12,636	0.07	25	660	8

在使用 eDonkey 傳輸檔案副檔名統計上(表 10)，mpg、dat、avi 等多媒體格式包辦了前 3 名，佔了大約全部 eDonkey 流量的 72%。比較表 8及表 10，我們可以發現 eDonkey 主要傳輸項目和 FTP 可說是非常類似。而在傳輸次數上，其次數都遠大於 FTP，應跟 eDonkey 採用 MFTP 有很大的關聯。

表10：使用 eDonkey 傳輸流量統計前 10 名(依副檔名排序)

副檔名	傳輸大小			傳輸次數			平均大小/次	
	單位:MB	佔全部%	序位	單位:次	佔全部%	序位	單位:KB	序位
mpg	210,220	29.84	1	5,924,865	29.48	1	35	37
dat	189,921	26.96	2	2,895,822	14.41	3	66	26
avi	107,590	15.27	3	5,675,391	28.24	2	19	51
bin	42,084	5.97	4	2,057,362	10.24	4	20	49
iso	29,051	4.12	5	591,008	2.94	6	49	30
rar	25,330	3.6	6	861,931	4.29	5	29	41
img	22,124	3.14	7	551,403	2.74	7	40	33
mdf	15,677	2.23	8	119,666	0.6	12	131	14
nrg	14,954	2.12	9	171,976	0.86	9	87	22
zip	14,498	2.06	10	362,211	1.8	8	40	34

由 FTP、HTTP 及 eDonkey 總合傳輸檔案大小副檔名排行來看(表 11)，不外乎圖檔，程式，光碟映射檔等。且前 10 排行的副檔名就已佔了所有傳輸檔案大小的 85%。

表11：副檔名傳輸大小排行前 10 名(HTTP + FTP+eDonkey)

副檔名	傳輸大小			傳輸次數			平均大小/次	
	單位:MB	佔全部%	序位	單位:次	佔全部%	序位	單位:KB	序位
dat	380,222	25.15	1	2,901,037	7.55	6	131	68
mpg	363,645	24.05	2	5,941,460	15.47	2	61	80
avi	152,674	10.1	3	5,688,834	14.81	3	27	86
iso	87,285	5.77	4	594,117	1.55	11	147	66
exe	82,947	5.49	5	256,108	0.67	18	324	45
bin	50,534	3.34	6	2,057,473	5.36	7	25	87
zip	47,736	3.16	7	497,789	1.3	14	96	74
rar	42,170	2.79	8	3,467,946	9.03	4	12	95
jpg	41,532	2.75	9	2,901,864	7.56	5	14	93
img	37,550	2.48	10	551,529	1.44	12	68	79

結論與建議

一、結論

1. 根據教育部(民 92)所做之大學生時間運用調查，大學生上網目的有 73.37%為收發 Email 及 66.50%為找資料(複選題)。這些統計資料是從最常使用的功能來統計，但這些功能項目於實際頻寬使用尚未達 20%（表 3，HTTP+SMTP+POP3）。
2. 本研究所設計之收集方法的確可有效改善一般流量統計 Others 比例過高的問題。實際以淡江大學網路環境測試結果，可有效將 Others 流量降低至 25%左右。雖然仍佔總頻寬一定比例，但比起傳統單用 Netflow 為主動輒超過 50%改善許多。
3. 網路頻寬使用上，以 eDonkey 資料傳輸所佔比例最大。以本校為例， eDonkey 傳輸流量佔總頻寬約 24%，為頻寬使用之最大項目。
4. 大部分的群組，其流量為少部分的 IP 所創造出來。大部分群組，其單一 IP 位址流量最高者，皆超過整個群組總流量的 30%，更有近 70%群組，其單一 IP 位址最大流量超過整個群組的 70%(表 7)，推測可能有許多群組皆有架設伺服器，因此佔掉大部分的頻寬使用量。
5. HTTP 和 FTP 傳輸型態有顯著不同。以 HTTP 傳輸方式，其副檔名佔總頻寬較大者，是因由較多的傳輸次數所造成。而以 FTP 傳輸方式，其副檔名佔總頻寬較大者，是因為傳輸的檔案通常屬較大所造成 (表 8、表 9) 。
6. 寬頻使用者與窄頻使用者網路行為上有顯著不同。我們可以發現窄頻使用者以收發

信(POP3/SMTP)佔頻寬使用量最高。而寬頻使用者以 eDonkey 所佔頻寬最高(表 5、表 6)。

7. 大學校園網路使用 P2P 軟體分享檔案佔了頻寬使用量約為 34%左右。而隨著越來越多的 P2P 軟體的出現，該比例勢必將會提高 () 。
8. 校園中，多媒體檔案及光碟映射檔傳輸量高。副檔名以 dat、mpg、avi 等影像檔為 FTP 及 eDonkey 傳輸的最大項目。而除了影片之外，寬頻及燒錄機的普及，也使得光碟映射檔為另一個頻寬使用項目。可見校園網路使用於非學術領域的比例並不低。

二、建議

1. 本研究調查對象為淡江大學，不同學校之間應有其差異性，其研究結果應也會有不同。未來可針對不同學校，比較不同學校間的網路使用者行為差異。
2. 本研究所設計之資料收集方法雖已將 Others 降低至 25%左右，但仍有改進空間，後續研究者可以朝此方向進行。
3. 本研究雖已記錄下使用者所傳輸之檔名，但僅對副檔名作分析，未來可針對整個檔名作分析，對於瞭解網路使用者行為應該更有幫助。
4. 本研究著重在整體網路使用者行為之探討，後續研究者可以針對特定通訊協定的行為做更深入研究。

參考文獻

1. 吳琮璠，"資訊管理個案研究方法"，資訊管理學報，第四卷第一期，7-11 頁，民 86。
2. 杜家麟，"資料挖掘在網際網路流量之分析研究"，靜宜大學資訊管理研究所碩士論文，民國 89 年 6 月。
3. 周鍾驥，"網路訊務量量測與特性研究"，中央員警大學資訊管理研究所碩士論文，民國 89 年 6 月。
4. 林育生，"以流量為基礎之網路分析系統"，國防大學中正理工學院電子工程研究所碩士論文，民國 90 年 6 月。
5. 翁瑞鋒，"網頁瀏覽者行為之泛化分群分析"，國立交通大學資訊科學研究所碩士論文，民國 90 年 6 月。
6. 張仕達，"從網路流量中發掘網際網路使用者之使用模式"，國立中山大學資訊管理研究所碩士論文，民國 88 年 6 月。
7. 張傑生，"以 flow 為基礎之網際網路流量量測分析"，國立交通大學資訊管理研究所碩士論文，民國 86 年 6 月。
8. 教育部統計處，"大專院校校別學生數"，http://www.edu.tw/statistics/service/s210190.xls，民國 92 年。

9. 教育部統計處，"大學生時間運用調查結果摘要分析報告"，http://www.edu.tw/statistics/publish/time- arrange.doc，民國 92 年。
10. 教育部資訊網，"台灣學術網路", http://140.111.1.22/ tanet/，民國 92 年。
11. 教育部電算中心，"國內網路流量統計分析"，http://nnst1.moe.edu.tw/，民國 92 年。
12. 楊明宮，"網路流量監視器內封包過濾器之實現"，大同工學院資訊工程研究所碩士論文，民國 87 年 6 月。
13. 蕃薯藤數位科技・開拓文教基金會，" 2002 年台灣網路使用調查"，http://survey.yam.com/survey2002/ index.html，民 92。
14. 簡榮成，"廣域網路流量監測與分析工具之製作"，國立中正大學電機工程研究所碩士論文，民國 88 年 6 月。
15. Cisco, "Cisco IOS NetFlow," http://www.cisco.com/warp/public/732/Tech /nmp/ netflow/ index.shtml, 2003.
16. Dug Song, "dsniff," http://naughty.monkey.org/ ~dugsong/dsniff/, 2003.
17. MetaMachine , "eDonkey," http://www.edonkey2000.com, 2003.
18. MySQL.com，"MySQL"，http://www.mysql.com/, 2003.
19. R. Rivest, "The MD4 Message Digest Algorithm," ftp://ftp.rfc-editor.org/in-notes/rfc1320.txt, April 1992.
20. Rafal Wojtczuk, "Libnids," http://www.packetfactory.net/projects/libnids/ , August 2002.
21. Robin Sommer and Anja Feldmann, "NetFlow: Information Loss or Win?," Proceedings of ACM SIGCOMM Internet Measurement Workshop (IMW'2002), November 2002.