建構在 Web 上之網路流量統計分析系統
The Analytic System of Network Flows over Web Architecture

黃圳柏* 羅啟文* 黃文祥* 邱基峰**
*國立高雄應用科技大學電機系
**國立成功大學電機工程研究所
E-mail:adrian@wshlab2.ee.kuas.edu.tw, ken@wshlab2.ee.kuas.edu.tw
wshwang@mail.ee.kuas.edu.tw, gary@hpds.ee.ncku.edu.tw

摘 要

網路的迅速發展於近年來持續的成長，為了有效掌握網路頻寬使用情形，讓現有的頻寬做最好的應用乃為現今網管人員重要課題之一，然而欲達到網路流量透明化及有效掌握目前網路使用情況，流量統計與分析為不可或缺之重要項目。本文中，我們研製一個建構在 Web 上之流量統計分析系統，並利用三層式架構建構出本系統。於本系統中，我們以 WWW 模式顯示出流量分析圖，如此網路管理人員將可方便、迅速的了解此網域中頻寬使用情況，以及各 IP 使用狀況，並加以統計作為日後分析查詢使用。具備此相關資訊，網管人員即能利用 Internet 立即得知網域內現階段網路使用情況，並且需要此網域頻寬使用情況時亦可作為依據。我們將在本論文中對此系統作詳細的描述與討論。

關鍵詞:流量分析、流量繪圖、網路管理、WWW

1. 前 言

近年來由於 Internet 的蓬勃發展，人們使用網路之情況亦迅速成長，透過網際網路我們可方便、省時處理生活中的一些問題。例如透過網際網路可以與全世界任何地方溝通，達到知識交流，也因此造成網路使用之頻繁，身為一位網管人員，掌握網域上頻寬使用情況一直是網管人員非常關切的問題，欲了解網路使用情況可由網域流量分析得以了解目前該網域流量分布情形，網管人員即可很快地藉由流量負載，來判斷網路或設備發生問題的可能原因。MRTG(Multi Router Traffic Grapher)乃作為流量分析之一工具，但由於 MRTG 所提供的資訊仍然有限，為了提供更多的流量資訊便於網管人員方便掌握及分配資源，我們研製一更多功能之流量統計分析系統，提供給網管人員更詳細的資訊。系統中，我們於乙太網路上研製一個以 Linux 為平台之封包擷取系統，此系統將負責擷取此網域之封包並加以解析、儲存於資料庫中，同時我們將透過 JAVA Applet 於 WWW 上顯示出流量分析圖。利用 JAVA Applet 方式可使得 Server 繪出流量分析圖之負載分散於 Client 上，對此 Server 將更有效率的工作。

2. 背 景

2.1 PCAP

Packet Capture Library(PCAP，封包擷取函式庫)，是位於 Linux 作業系統底下所提供的函式庫之一。利用此函式庫中所提供之函式[6]，即可於乙太網路上擷取所有被廣播出來的封包。圖 1 為乙太網路封包框架圖。利用此技術所擷取到的封包格式將如圖 1 所示，只要我們加以解析，即能成為我們所迫切需要得知資訊。圖 2 為乙太網路標頭格式，於圖中第三欄位指明其所收到封包為那一類型之封包(IP、ARP、RARP 等等)。因此，利用此欄位資料將可便於分辨封包種類。圖 3 顯示 IP 標頭格式，於 IP 標頭格式中，我們需要此封包來源位址與目的位址欄位的資訊得知此封包屬於那一 IP 所送出及接收。

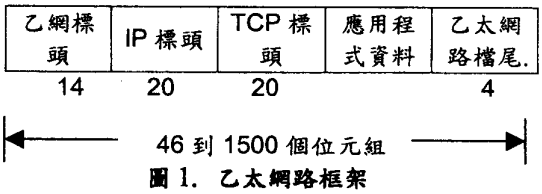

圖 1. 乙太網路框架

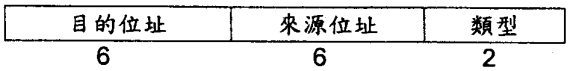

圖 2. 乙太網路標頭格式

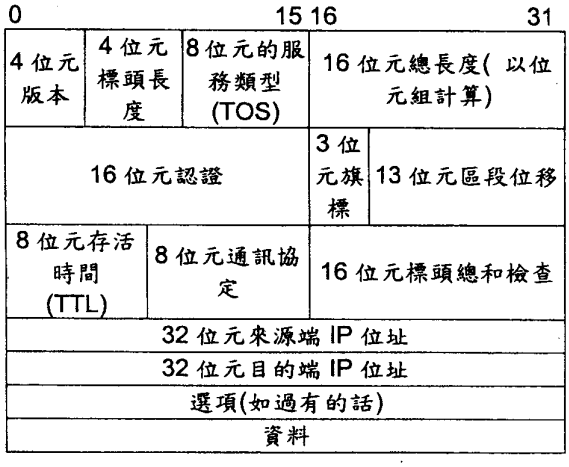

圖 3. IP 標頭各欄位

2.2 JAVA Applet

JAVA Applet 是一種以視窗為基礎(window-based)的應用程式。它的架構與一般以控制台為基礎(console-based)的應用程程式有所不同。Applet 以事件驅動(event driven)為導向，一個 Applet 就像是一組中斷服務程式(Interrupt Service Routine)，Applet 將一直處於等待狀態，直到某一個事件被觸發為止。當被觸發時，AWT(Abstract Window Toolkit)[7]將通知 Applet，同時經由 Applet 所屬的事件處理器(event handler)所呼叫的事件產生。一但事件被觸發後，Applet 必須選擇適當的動作然後迅速地把控制權交還給 AWT 系統，因為 AWT 必須一直聆聽事件是否被觸發，其處理過程如圖 4 所示。有鑑於此，我們便藉助 JAVA Applet 能力，並使用 Applet 本身所提供的繪圖功能，實作流量分析繪圖系統。同時 WWW 模式之 JAVA Applet 執行過程中，利用 Server 端所編譯之 byte code 傳送至 Client 端，並於 Client 端進而將此 byte code 編譯為可執行之應用程式，如此將負載分攤於 Client 端，Server 端之處理效能亦能大大提昇。

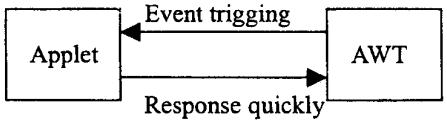

圖 4. Applet 與 AWT 之間控制權交換圖

2.3 三層式架構

資料庫具有儲存、管理以及查詢功能，也由於網際網路的快速成長使得資料庫由單機操作之作業環境轉移到三層式架構-分成客戶端(Client)、web 伺服器(web server)、資料庫伺服器(Database Server)三層，如圖 5 所示 Web 伺服器負責 Web Server 與 Client 端之間連線建立與取得資料庫伺服器之資料，資料庫伺服器則負責儲存呈現給客戶端的資料，而客戶端便利用瀏覽器與 Web Server 連接，透過「首頁」的呈現方式即可讓 Client 與 Web Server 做雙向溝通。在面對龐大、複雜需耗費大量人力的資料，如果藉由三層式架構來輔助，把工作分散於各主機上，自然可降低各主機的負載。

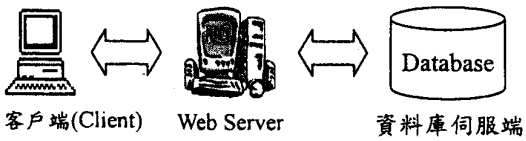

圖 5. 三層式架構

3. 系統架構

本研究中，為了實作出三層式架構(其目的是乃為希望分攤 Linux 平台上之負載)，故我們將實驗平台分成二部分。一部分屬於 Linux 平台(後端)，其主要的工作是負責擷取封包並解析之，進而將其送至資料庫；而另一部份為 Windows(Web server)部分，其主要的工作為連至 Linux 平台取得資料庫之資料，並將結果顯示於網頁上，供網管人員參考。底下將詳細討論每個平台之實作細節。

3.1 Linux 平台上實作

於 Linux 平台下，我們利用 PCAP 函式庫所提供之函式，即可擷取乙太網路上所有被廣播出來的封包，圖 6 為一例子，圖中顯示擷取到的封包並將其解析後的結果。第一行指出擷取到的封包屬於那一種類型之封包(可分為 IP、ARP、RARP 等等)，而 source port 與 dest(destination) port 指的是來源端與目的端之 port number(埠號 23 為 telnet 使用)，protocol 是指在 IP 標頭中 protocol field(協定欄位)之值，意思為其屬於何種協定，例如：協定欄位之值為 1，即表示為 ICMP，2 表示為 IGMP，6 表示 TCP ，17 表示 UDP。

至此，我們將所接收到的封包經過解析，並將其存至資料庫，同時使用 MySQL 提供於 C 的相關 API[5]，將解析後所得之資料送至資料庫，以等待 Client 端要求 Java Servlet 連至資料庫取得所需的資料並顯示於 Client 端。

圖 7 為擷取封包流程圖，系統啟動後將一直持續擷取封包，並且有二個 thread[1]執行著，一個為負責加總封包流量大小，另一個則負責計時，直到經過五分鐘後再將其所接收之資料送至資料庫，如此反覆執行。

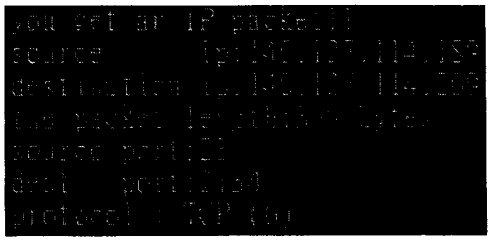

圖 6. Capture packet with PCAP

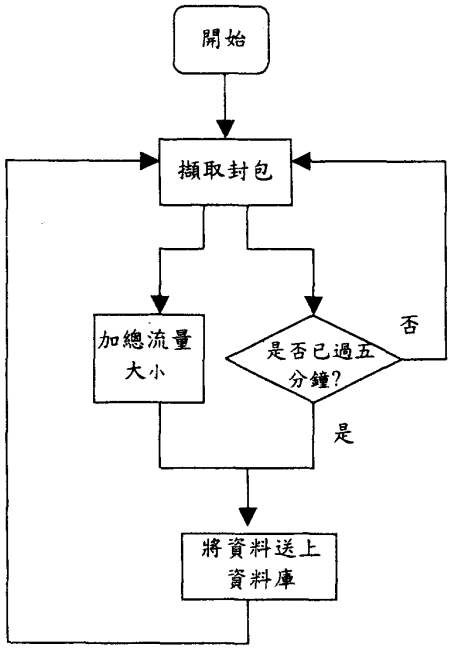

圖 7. Linux 平台上擷取封包流程圖

3.2 Windows 平台上實作

Windows 平台上，我們撰寫一支 JAVA Applet 程式，以取得 Linux 平台上資料庫資料，並將其繪成流量分析圖。對於一般 Applet 而言，Applet 自動被定義為不受信任(untrusted)的 Applet，所謂"不信任"意指這個 Applet 只能由下載此 Applet 的機器讀取資料(或者只能連線到自己本台機器之資料庫)，若欲連往其他的 server，則 Applet 會產生 Security Exception 例外事件，表示無法連線。因此，我們必須使用 Applet 與 Servlet 之間的通訊來完成工作。透過 Applet 連線到 Servlet，此時 Servlet 將連至 Linux 上 MySQL 資料庫取得所需資料，間接地再將資料送回至 Client 端，圖 8 為整個系統動作原理示意圖。其上半部屬於 Linux 平台部分，下半部則屬於 windows 平台部分。

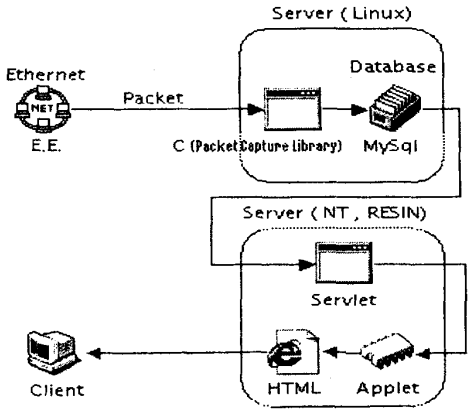

圖 8. 流量統計分析系統動作原理圖

4.實作結果

圖 9 為 Client 端呈現之流量主分析圖，圖中橫軸為時間(小時)，紀錄著最近 24 小時內所測得之數據(以每五分鐘計算一次)，縱軸為流量大小(KBytes)，並於網頁底下顯示於過去 24 小時當中最大流量時間點、最小流量時間點以及平均流量值，同時利用滑鼠移至流量分析圖上，它將自動顯示此點的流量大小，例如圖中顯示之值為 2315.476(KB)。

除了主流量分析外，圖 10 所示為各個 IP 於某一時間點上(只要於圖九上針對某一點按滑鼠左鍵)送出與接收之流量圖。如此即可讓網管人員了解於某一時間點上，那個 IP 正在使用網路頻寬。由此網管人員即能隨時掌握網路上頻寬使用之情況。在此由於 140.127.114.255 之 IP address 乃屬於廣播位址，故我們並沒有將它顯示於網頁上。同時各 IP 每日的流量排名也顯示於系統中，如圖 11 所示，其主要顯示出過去 24 小時內 IP address 的總流量排名(以 Send_bytes 與 Receive_bytes 相加後排名)，於各個 IP 詳情中，其結果顯示於圖 12，主要把所測得數據繪成流量分析圖。與主流量分析圖不同的是此功能顯示出某一 IP 過去 24 小時內之輸出與輸入流量圖，此網頁底下分別指出其最大輸入與最大輸出之大小、時間點，讓網管人員能更進一步了解各 IP 輸入與輸出流量分布情況，因此不正常使用頻寬之 IP address 即一一曝光於此。

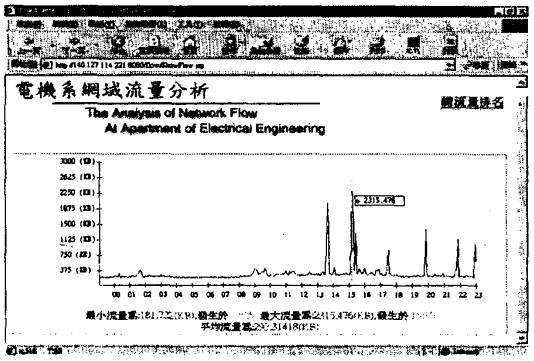

圖 9. 流量主分析圖

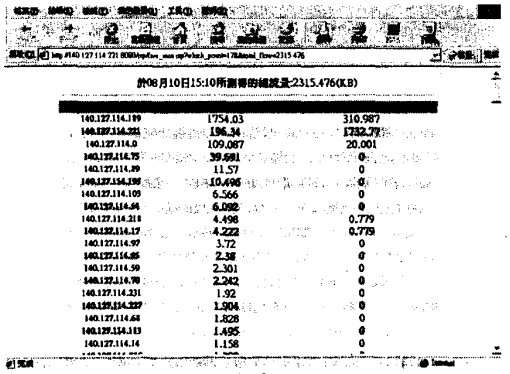

圖 10. 各個 IP 於某時間點上之流量排名

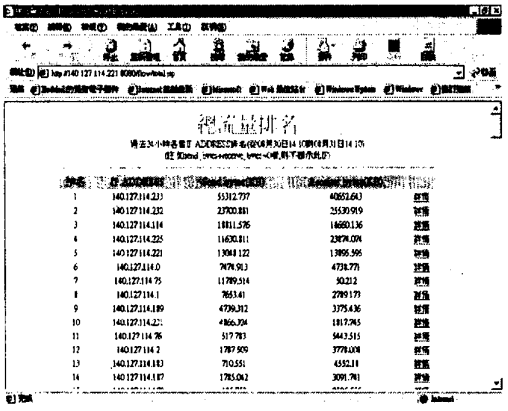

圖 11. 過去 24 小時總流量排名

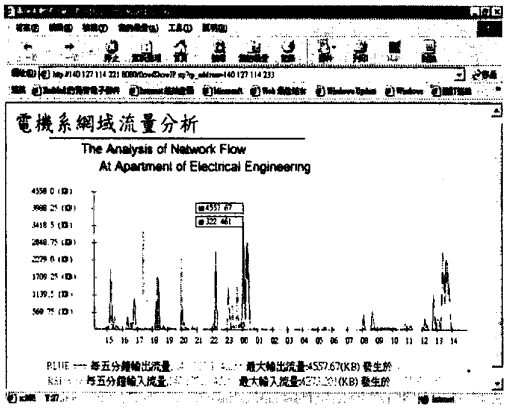

圖 12. 於過去 24 小時，某一 IP 流量分析圖

5.結論與未來工作

本文中，我們研製一個建構在 Web 上之流量分析統計系統，系統中可方便、迅速提供網管人員得知網路頻寬使用狀況，以利於網管人員對網路資源做最佳的分配，同時網管人員可了解每一時間點之流量大小、各 IP 流量分布之情形以及某一 IP 之流量分布圖，提供一個參考之依據。

未來我們將繼續完成更完善的功能，例如：觀察 TCP、UDP 之個別流量分析圖。並且增加一即時流量顯示功能，加強整個系統之完整性，以提供網管人員更多的資訊。

6.參考文獻

[1] Richard Stones and Neil Matthew "Beginning Linux Programming", WROX , October 1999.

[2] Danny Ayers, Hans Bergsten, Michael Bogovich...etc, "Java Server Programming" , WROX, 1999

[3] W. Richard Stevens, "Unix Network Programming, Volume 1, 2/e", October 1997

[4] Richard Stevens 原著, 資策會中文化部門編譯, "TCP/IP Illustrated, Volume 1 國際中文版", 學貫行銷股份有限公司 , September 2000

[5] Paul DuBois, Michael Widenius 原著, 劉春成等編譯, "MySQL 徹底研究", 博碩文化, October 2000

[6] Packet Capture Library 說明文件

[7] Patrick Naughton and Herbert Schildt 原著, 朱光宇 斐宜編譯, "Java 2 設計實務", McGraw Hill, April 2000

[8] George Reese 原著, 陳建勳 編譯, "JDBC 與 JAVA 資料庫程式設計", O'REILLY, February 2001

[9] Linux 超級手冊第五版 , Tackett & Burnett , 碁峰資訊有限公司 , May 2000

[10] Jason Hunter and William Crawford 原著, 李國熙 編譯, "Java Servlet 設計", O'REILLY, July 1999

[11] http://java.sun.com/applets/

[12] http://www.linux.org