網路架構自動偵測之研究

計畫編號： NSC 90-2623-7-009-020

執行期限： 90/01/01-90/12/31

指導教授 ： 交通大學資工所

劉振漢教授

學　生　 ： 交通大學資工所

蔡弘晉 ， 黃聖懿

摘要

網際網路在設計的基本精神上提供很大發展空間，一個機構只要對外符合TCP/IP 的介面，內部的組織架構可有很大的彈性。機構內的各部門也可以獨立自主的發展。一個機構的網路主管對各部門內的網路配置也無法完全掌控。

對一般的使用者而言，網際網路是一個無窮大的世界，可以任意遨遊。但是「網路之間是如何相連，有些什麼電腦，提供什麼服務」等等，是對網際網路的瞭解的重要一環。以網際網路的重要，不管是系統主管，網路技術員乃至一般使用者都應該對此加以探索。

本計畫的目標在提供一個工具，透過它，可以針對某一網址（IP address）範圍內，及/或一個網域名稱（domain name）下，探索機器數目，機器間的連線，以及每部機器的網址、名稱、作業系統、服務項目等等。不同的人可能會將這個工具用在不同的用途，例如尋求類似下面問題的答案：（1）我任職的機構內共有幾部電腦，它們如何連結，使用什麼網址、名稱。（2）我國分配到的網址範圍，國內的網址如何分配。

Abstract

Internet is designed to allow independent growth. An organization, from a country to a small company, can freely allocate its quota of IP addresses to various sectors. It can create (sub)domain names, provide all sorts of services, such as FTP, e-mail, Web sites, etc.

For an ordinary user, Internet is a boundless world to surf, he has very little idea as to where a server is located, how machines are connected, what ip addresses are assigned. But such information is an important part of one's knowledge about Internet, for network administrators as well as ordinary clients of Internet.

In this project, we try to design software tools. With them one can concentrate on certain scope of IP addresses, or domain names, investigate number of machines, their operating systems, their topology, services provided, etc. One can answer questions, such as (1) How are computers and domain names used in my university; (2) In Taiwan, how are IP addresses and domain names allocated and how do they correspond to each other.

一. 研究動機

在現在有線的網路環境之中，常常是相同的一個機構或組織會自成一個網路的群組，如公司機關，校園網路等等，而在這一個群組當中，有著一定程度的關係性或依賴性，如封包傳遞的方式，網路實體布線的連結等等，而這些資訊是我們所感興趣的，例如有時候我們會想要知道一個機關或校園中有哪些機器? 它們所能提供的服務是什麼? 要傳送資料到群組中的電腦時中間是經過哪些 router? 或是想要統計校園中所提供的網路服務的數量情形等等，因此這個計劃的目的便是希望能夠提供使用者收集以上資訊以利研究及發展

二. 研究重點

(1). 偵測網路上提供服務的主機，以及提供哪些服務和使用的作業系統
(2). 偵測網路上 Router 分布架構，並追蹤網路封包經過的所有 Router
(3). 偵測出網路上整體架構，並可針對某一網路封包來源的區域網路做網路整體架構的分析與偵測

以下程式均以 **Visual C++ 6.0** 撰寫,執行環境爲 **Windows 2000**,以 multi-thread 的方式來作偵測.統計數據則以 **Microsoft Excel** 作分析與統計.

三、流程一：找出一個單位裡裡所有的 IP 和 domain name

目標

我們以「交通大學」爲例。透過搜尋引擎，我們找到

www.nctu.edu.tw

這個網址。用

ping www.nctu.edu.tw

命令，得到

Pinging www.nctu.edu.tw [140.113.250.5] with 32 bytes of data:
Reply from 140.113.250.5: bytes=32 time=45ms TTL=246
Reply from 140.113.250.5: bytes=32 time=41ms TTL=246
.....

得知此網頁的IP是140.113.250.5。

我們再進入交通大學各單位的網頁去看，可以看到他們的domain name都是

?.?.nctu.edu.tw

而他們的IP都是

140.113.?.?

我們可以推斷所有交通大學下的單位的domain name 都以.nctu.edu.tw結束。但是不能確定所有140.113.之下的網址都是分配給交大使用。

我們設計了一個程式，輸入IP網址的開頭數字，以及domain name 的結尾名稱，程式就把所有符合條件的網址和domain name存入檔案中。

程式名稱

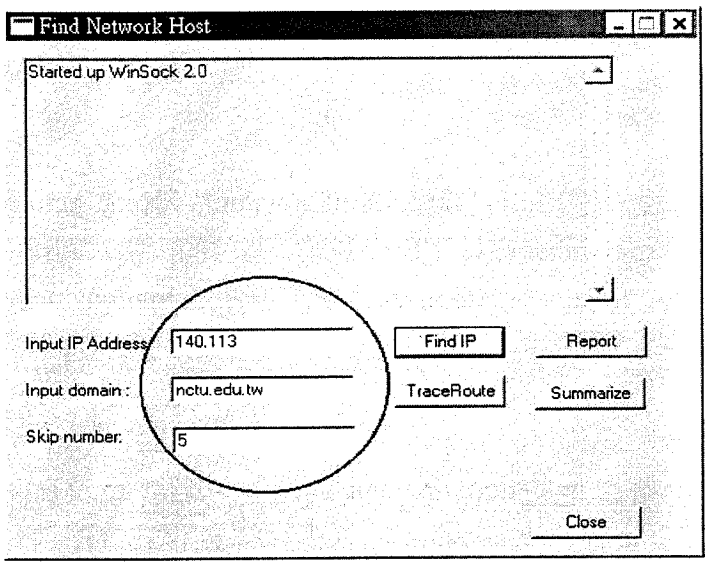
host.exe的"Find IP"

輸入資料

IP開頭片段(ex.140.113),domain name結尾片段(ex. nctu.edu.tw)

程式執行

執行程式後出現下面的對話盒：



在這個使用者介面中,使用者將下面三個欄位填好：

Input IP Address: 填入IP網址的開頭數字

Nput domain: 填入domain name的結尾部分

Skip number: 填入一個數字，表示跳號

有些單位的網址很多，要清查每一個網址，機器要跑很久的時間。Skip number大於1表示不清查每個網址。在上圖的設定中，會嘗試找出

140.113.1.1

140.113.1.6

140.113.1.11

……..

140.113.251.251

的網址是否有分配給某台電腦使用。有的話，它的domain name是什麼。

按下"Find IP"後開始執行.

輸出資料

在執行完"Find IP"後,我們將符合的資料儲存在檔案"a.txt"中.

下面是程式輸出的資料的樣品：

140.113. 1. 10 : MOEsun.NCTU.edu.tw

140.113. 1.245 : news2.cc.NCTU.edu.tw

140.113. 2. 5 : sophi5.adm.nctu.edu.tw

……

140.113.251. 5 : EIC-fddi.nctu.edu.tw

每一列均爲符合要求host的ip與domain name.以交大爲例,即IP前端爲"140.113",domain name片段爲"nctu.edu.tw"的所有電腦均會列在檔案"a.txt"中.

程式所用方法

將所欲查詢網域的ip片段(e.g. 140.113)以及domain name片段(e.g. nctu.edu.tw)輸入其中,並給定一個skip number.在按下Find IP後,程式將利用gethostbyaddr,尋找一個b class範圍所包含ip的domain name,並將找出的domain name與使用者輸入的domain name片段做比較,紀錄包含該domain name片段的ip與對應的domain name.

當我們以 gethostbyaddr 尋找 ip 的 domain name 的過程中,若該 ip 沒有使用或在 dns server 中並無該筆 ip 與 domain name 的對應資料,則程式會停頓一段時間來等帶回應.爲了加快搜尋速度,我們利用 multiple thread 一次對多個 ip 作詢問.

四、流程二: 找出到達每台電腦所經過的路徑

目標

經過步驟一後,我們可以找到某一單位下其電腦的ip與domain name分布.爲了整理出該單位所有電腦的網路架構.我們有必要先蒐集由本地端電腦到該單位各電腦所需經過的路徑(router).
利用本程式,可以將到達該單位每台電腦所經過路徑紀錄在一個檔案中,以供接下來的步驟處理.

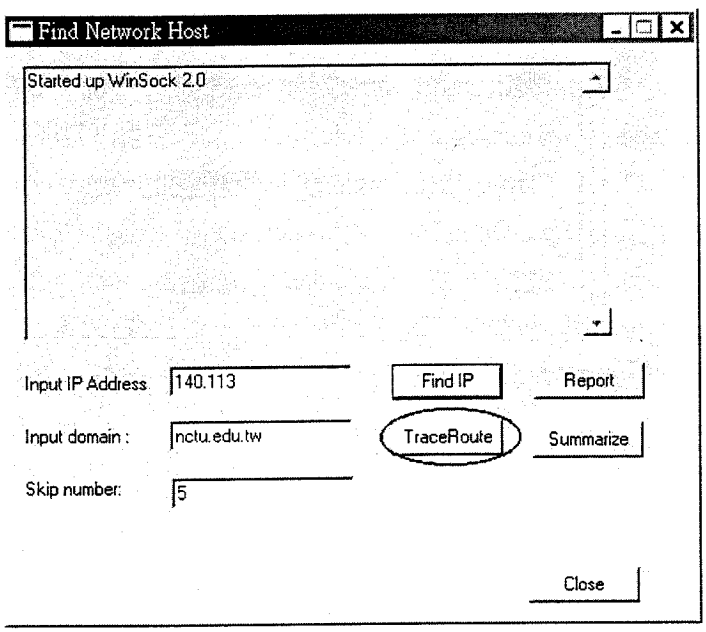
程式名稱

host.exe的"TraceRoute"

輸入資料

執行完"Find IP"後所得到的檔案"a.txt"

程式執行



將先前收集的ip與domain name資料---"a.txt"當成輸入資料,按下"TraceRoute"後,程式會針對a.txt中所有列出的ip進行traceroute的動作,找出本地端到該ip所經過的router並記錄下來.另外,對於不完整或有錯誤的route flow則不予紀錄.

在這個過程中,根據所查詢單位的電腦數量,會需要相當長的時間.

輸出資料

在對所有的ip進行過traceroute後,會將所有結果紀錄在TR.txt中.
下面是程式輸出的資料的樣品：

140.113. 1. 10 =140.113.216.254+140.113. 53. 10+OK+

140.113. 1.245 =140.113.216.254+140.113. 53. 10+OK+

140.113. 2.140 =140.113.216.254+140.113. 53. 28+OK+

.....

140.113.250.200 =140.113.216.254+140.113. 53.254+OK+

每一列均爲某個IP的route flow.其中,"="左邊的IP爲執行traceroute的destination IP."="右邊的IP串爲所經過的router IP,每個router均由"+"作分隔,以"OK+"作爲該route flow的結尾.
以140.113.1.10爲例,由140.113.216.123對該destination作routing時會經過router

140.113.216.254

140.113. 53. 10

後到達該destination.

程式所用方法

在收集完該網域中所有存在的ip與domain name對應後,針對這些符合的ip,利用traceroute找出由本地端到該ip所需經過的router並予以紀錄.我們利用ICMP封包,依次改變封包的TTL(Time To Live)來找尋所經過的router.同樣的,爲了加快搜尋速度,我們用multiple thread來執行traceroute.但是由於ICMP沒有對每個socket指定port,因此當電腦收到一個packet後,每個thread均會收到該packet.因此對與每個thread所發送的封包必須在加入一個id欄位,用來辨別收到的reply packet是否屬於該thread.

五.流程三: 找出該單位所有電腦的作業系統與提供的服務

目標

對於一個區域的電腦,我們有興趣的不只是整體的網路架構,同時也想知道該區域電腦所使用的作業系統與可能提供的服務.因此本程式提供了一個簡略的分析來得知這些訊息.

程式名稱

host.exe的"TraceRoute"

輸入資料

執行完"Find IP"後所得到的檔案"a.txt"

程式執行

在執行"TraceRoute"的同時,我們也一併收集了該電腦的資料.當該電腦處在開機狀態時,最後會收到一個destination回應的封包.這時可以順便偵測該電腦的作業系統與提供的服務.

輸出資料

在偵測完成後,程式會將每台電腦的相關資訊存放在"summary.txt"中.
下面是程式輸出的資料的樣品:

140.113.1.10(MOEsun.NCTU.edu.tw) : solaris 2.x : ftp,telnet,SMTP,

140.113.1.245(news2.cc.NCTU.edu.tw) : FreeBSD 3.x/linux 2.2.x/solaris 8 : telnet,SMTP,web,

140.113.2.140(sec2140.adm.nctu.edu.tw) : windows 9x/NT/2000 : unknown or no service

.....

140.113.250.200(sppmgr.NCTU.edu.tw) : solaris 2.x : unknown or no service

每一列均紀錄該電腦的ip, domain name, 所使用的作業系統,與可能提供的服務.每個欄位均以":"作間隔.

程式所用方法

在偵測os部分,我們利用ICMP protocol,對每個ip發送ICMP封包.在收到reply封包後,判斷該封包的TTL.若TTL介於60到64,我們判斷該主機所使用的os爲FreeBSD 2.x/linux2.2.x/solaris 8.若TTL介於110到128,該主機所使用的os可能爲windows 9x/NT/2000.若TTL介於240到254,該主機所使用的os可能爲solaris 2.x.

在偵測主機所提供的服務方面,我們嘗試對該ip的幾個well known port作TCP connect的動作.如果connect成功,我們便假設該主機有提供該項服務.這裡我們針對FTP(port = 21),telnet(port=23),smtp(port=25),dns(port=53),和http(port=80)作測試.

六.流程四: 分析所有電腦的網路架構

目標

收集完每個ip的route flow後,可以針對這些route flow作分析,整理出該區域電腦的網路分布狀態.進而了解封包在該區域routing的可能路徑.

程式名稱

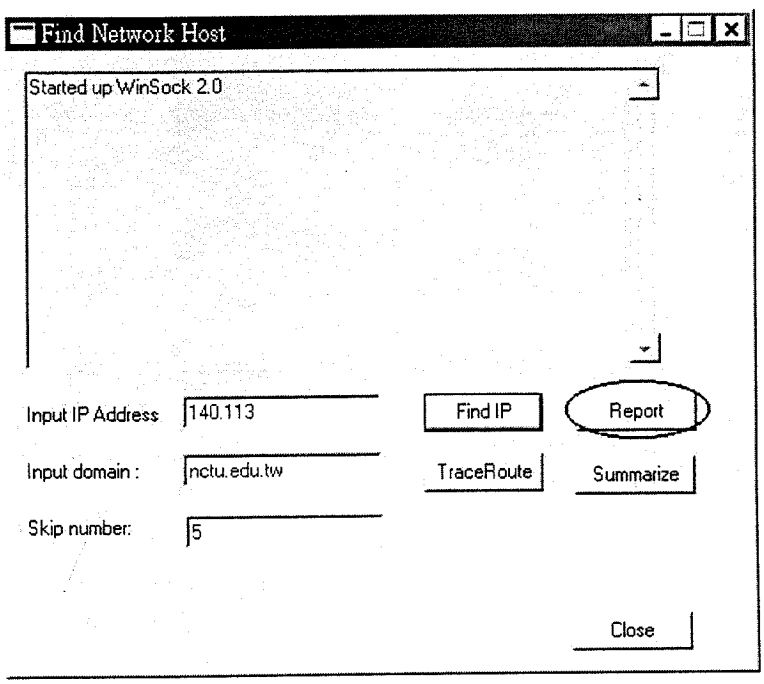
host.exe的"report"

輸入資料

執行完"TraceRoute"後所得到的檔案"Tr.txt"

程式執行

以先前所收集的所有ip的route flow作爲輸入資料,執行"report".程式會自動的分析每個node(包含router)的關係.最後在以樹狀結構的方式將所有節點紀錄在檔案"report.txt"中.
這份資料是一個暫存的文字資料.是爲了之後建立更詳細的圖形資料,以方便使用者查詢.



輸出資料

下面是程式輸出的資料的樣品：

140.113.216.254 (tw.edu.nctu.csie.e3rtn-216)

140.113. 53. 10 (not-a-legal-address)

140.113. 1. 10 □ (tw.edu.NCTU. MOEsun)

.245 □ (tw.edu.NCTU.cc. news2)

140.113. 40. 20 □ (tw.edu.nctu.adm. acct4020)

140.113.200.100 □ (tw.edu.NCTU.cc. liao-100)

……

這是一份階層式的資料.每個要到達底層ip的封包均需要經過上層的router才可到達.以140.113.1.10爲例,封包需經過

140.113.216.254

140.113.53.10

才可到達.

程式所用方法

收集完每個 ip 的 route flow 後,我們對所有找出的 ip(包含 router)作整理,紀錄每一個 node 的 parent 爲何.之後,以本地端 ip 爲 root,根據每個 node 的紀錄,建立一個 route tree.同時將每個 node 的資料,如 domain name, os,以及 server 均附在在這個 route tree 上.這些資料均暫存在 report.txt 檔案中,在之後的資料整理會使用到.

七.流程五：將所有電腦的網路架構以 Tree View 方式顯現

目標

針對我們所蒐集到的資料，我們以程式來做以下的分析：

- → 分析網路上電腦分布的 **Topology**，之間的連結關係
- → 了解特定網域下的電腦數量，**service** 分布狀況
- → 搜尋與觀察
- → 以 **Tree View** 的方式來顯現

程式名稱

DataSum

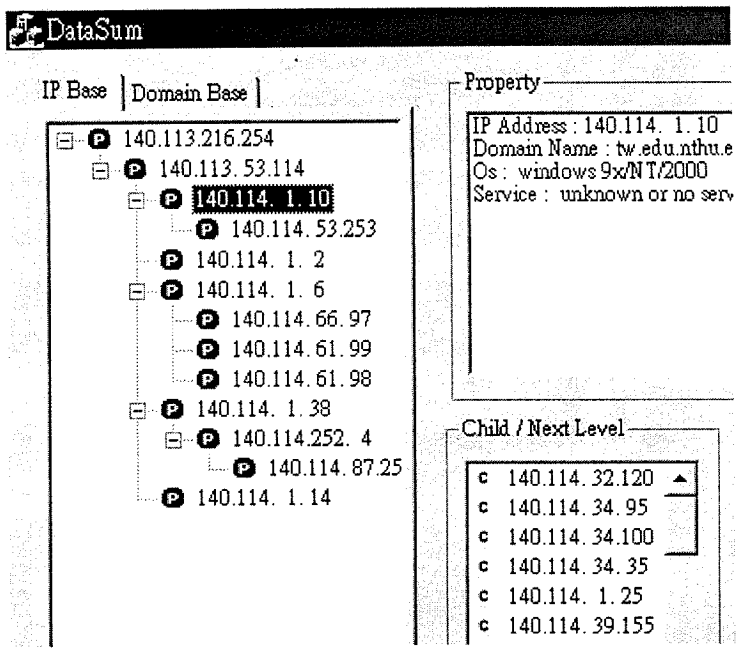
輸入資料

執行完"TraceRoute"後所得到的檔案"Tr.txt"，"report.txt"，及 "summary.txt"

程式執行

我們的程式中表示資料的方法都是使用 Tree View 的方式來表示，即類似檔案總管的方式，將全部的資料做階層式的整理，分出一些所屬關係，而所用來分類的關係我們使用了兩種模式，IP Base Mode 及 Domain Base Mode：

Data Summary (IP Base)： 我們在做分析的時候使用了兩種不同的模式，首先是 IP Base 的模式，這個模式下是以 IP Address 作爲分類的基準，而由於整個網路的架構及 Protocol 的設計，以 IP Address 作爲分類的基準的話，事實上就是以電腦在網路上實際的連結狀況來作爲分類，也是電腦在網路上分布的 Topology，下面是 IP Base 時的分析圖

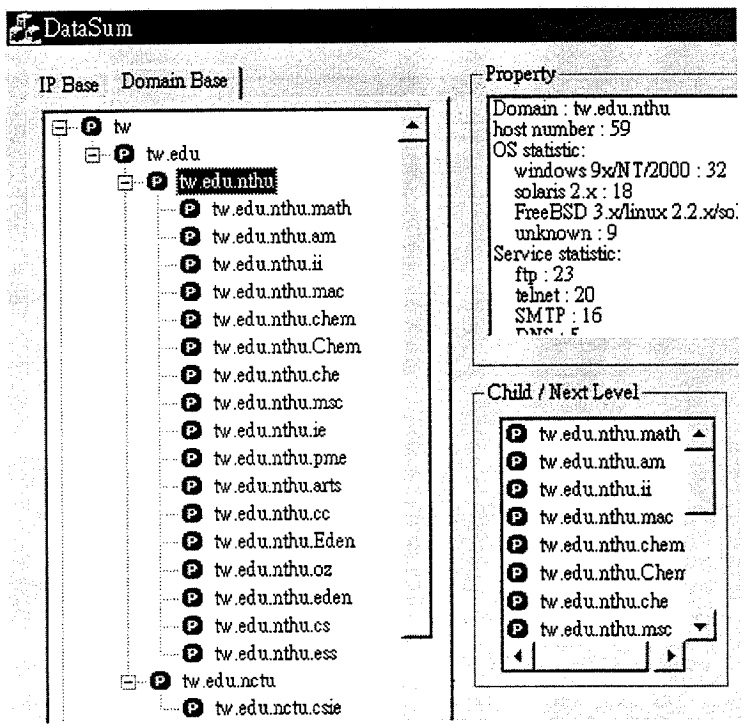


就圖上來看的話，我們可以清楚的知道這些 IP 的連結情形，如 140.113.53.114 下面連結的電腦有 140.114.1.10 、 140.114.1.6 、 140.114.1.38 等等，而.1.6 下面又連接了 .66.97 、 .61.99 、 .61.98 三台電腦，而在 tree 中所表示出來的 IP，在網路的架構上來看的話都是 Router，而單就 tree 的結構賴看的話就是都扮演著某些點的 Parent，IP 前面的 P 就是這個意思，而對網路上來說，一定有一些電腦是最底層的，大部分的使用者都是屬於最底層的，不負責網路的 routing，這些電腦的話我們就另外來處理，當我們選定了某一台 Router 時，在旁邊的 Child/ Next Level 視窗我們便把連結到它的所有電腦 IP 列出，在這個視窗中不管它是不是 Router 都會列出，不是 Router 的話，我們用黃色的 C 來分別，這樣做的原因是，通常最底層的電腦會很多，如果全部列出來的話看起來有點累贅，而且，我們要看一個網路的 Topology 架構時，通常我們所關心的就是 Router 之間的連結情形，網路繞送也是看這些可以明白清楚的了解，因此我們才 default 只列出 router 級的點出來，另外，右上方的 property 視窗是顯示我們所選取的電腦（140.114.1.10）的一些相關資訊

(Note)： 由於我們的程式結果會根據當時所執行時本機的網路位置有所不同，不過也只有最上層的幾個結果不同，例如，我們由交大跑這個程式去測清大時，最上層的電腦一定是經由交大，再連到清大，不過若我們放到台大跑的時候，最上層的電腦就會是台大的電腦，所以當我們要看這個資料的時候，基本上是種相對的資料，重要的是我們要了解我們重視的是什麼，像圖上是由交大去測清大的資料，因此

最上層的 140.113.216.254 便不是那麼的重要

Data Summary (Domain Base) : 跟 IP Base 不同的是 , Domain Base 在
tree view 時使用來分所屬關係的是直接使用 Domain Name 的關係性 , 將 Domain Name 類似的當成同一類 , 而階層的分法則是將 domain name 由後往前看 , 如 tw → edu.tw → nctu.edu.tw → csie.nctu.edu.tw → **ccsun1.csie.nctu.edu.tw (real domain name)** , 簡單的說 , 就是由 tw(台灣) 衍伸出 edu.tw (台灣學術網路) 衍伸出 nctu.edu.tw (交通大學) 衍伸出 csie.nctu.edu.tw (交大資工系) 衍伸出 ccsun1.csie.nctu.edu.tw (交大資工系計算機中心的一部工作站) , 因此在這樣的模式下 , 除了在最底層 , 或說是 tree 中的 leaf 端點外 , 其他的點都不是一個完整的 domain name , 只能算是一個網域 , 下面是 Domain Base 時的分析圖



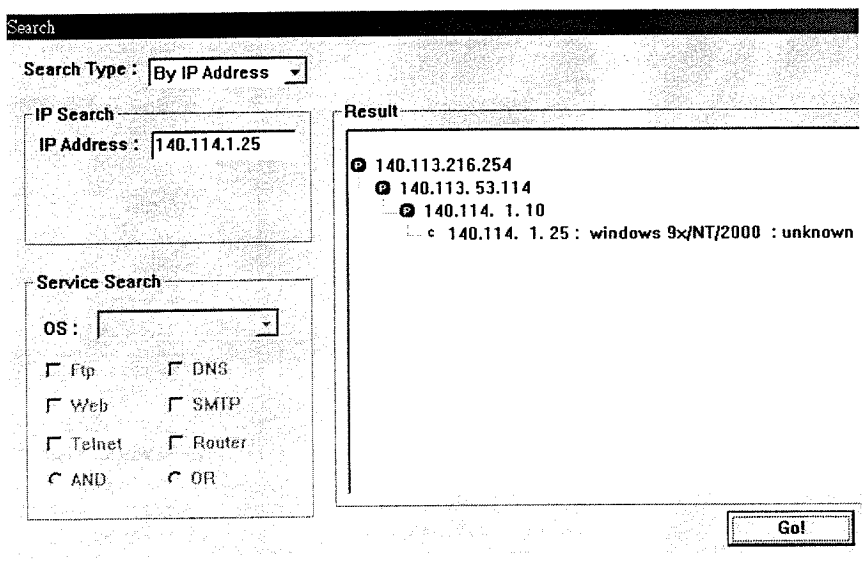
就圖上來看的話 , 我們可以得到 IP Base Mode 時所無法知道的訊息 , domain base 通常使用在我們要觀察一個特定的群組時所用 , 例如我們想要看交大資工所屬的所有電腦 , 但是交大資工的所有電腦不可能都在同一個 subnet 之下 , 可能分散在好幾個 subnet 之下 , 所以若使用 IP Base Mode 的話 , 要找到所有的交大資工電腦就有點費事了 , 可是由於我們在命名 domain name 時 , 有個共通點 , 就是會在主機名稱後面加上 **" csie.nctu.edu.tw "** , 因此我們使用 Domain Base Mode 時就可以輕鬆地找出所有的交大資工的電腦 , 另外 , Domain Base 的應用也蠻多的 , 例如我們可以從一個大單位中看出其中又有哪些小單

位，例如一個學校中，有很多的科系，而使用 Domain Base 的話，就可以一目了然，並且，我們可以就某個單位來做一些統計分析的工作，例如上圖中我們所點選的是整個清大的資料，則右上方的 property 視窗可以顯示出：

1. 有多少台電腦在其中 **(host number : 59)**
2. 全部電腦中 **OS** 的使用情形 **(OS Statistic)**
3. 提供了哪些服務及數量 **(Service Statistic)**

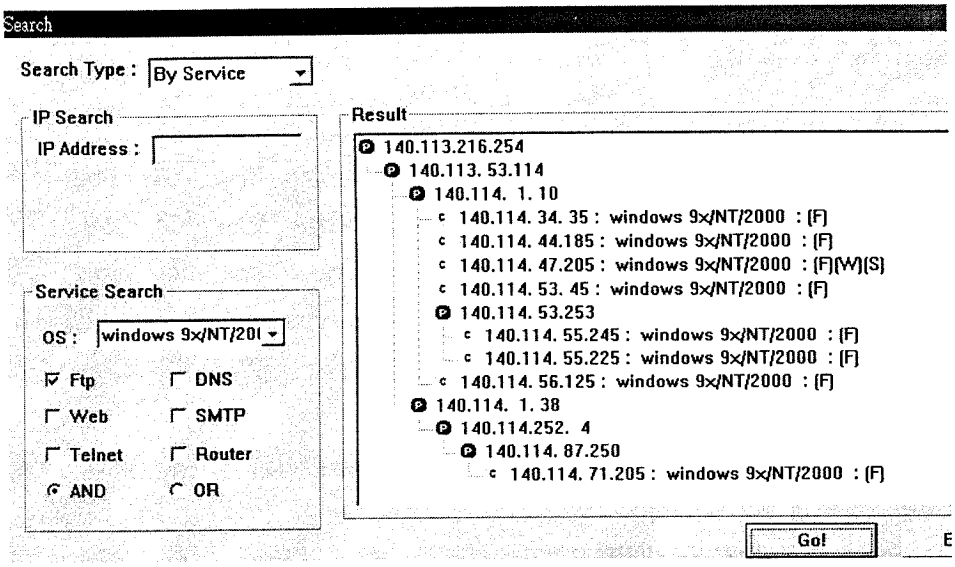
資料龐大的話，勢必會使用到搜尋的時候，因此我們提供了兩種搜尋的方法，一種是利用已知的 IP Address 來做搜尋，可以讓人知道這個 IP 到底有沒有人使用這個 IP 及列出封包要傳送到這個 IP 時的 Network Routing Flow，而另一種搜尋方式便是利用 OS 和提供的 Service 條件，來列出所有符合條件的機器

Search By IP Address :



如上圖所示，我們輸入 140.114.1.25， Result 視窗會顯示出我們找到這台電腦，而且會列出由本機一直到 140.114.1.25 過程中所有的 Router，其實就是以 TraceRoute 的結果作爲依據

Search By Service and OS :



如圖所示，我們可以用來尋找符合 **OS** 是 **Windows** 系列的並且有提供 **FTP** 服務的機器，當然我們也可以尋找符合眾多條件中的其中一種就可以的需求，只要把 AND 改才 OR 就可以了

(Note): 在 **IP address** 後面有附加資料的才是我們所找到的符合條件的機器，沒有附加資料的只是顯示出 **Routing Flow** 而已

輸出資料

無

八.結果與統計

最後,我們將資料匯入excel中,先錄製一份巨集,並利用該巨集對資料作整理.我們將excel資料分成幾部分:TR, summary, statistic, statistic2,及圖表部分.TR和summary分別是之前所整理的tr.txt與summary.txt. statistic則統計所用os分布與所提供service分布. statistic2則統計os與service間的關係.最後利用上述資料作成圖表.

我們分別對台大,中央,交大,與成大作測試,並將搜集的資料以excel作處理,得到一些結果如下:

(1) 台大:共偵測826台機器.其中:

os部分: windows系列共677台

solaris 2.x共149台

service部分: 提供ftp有156台

提供telnet有108台

提供smtp有116台

提供dns有35台

提供web有132台

不提供服務或未知共581台

os與service間的對應關係如下:

	ftp	telnet	Smtp	Dns	web
FreeBSD	0	0	0	0	0
Windows	71	9	47	18	70
Solaris2.x	85	99	69	17	62

(2) 中央:共偵測236台機器.其中:

os部分: windows系列共194台

solaris 2.x共42台

service部分: 提供ftp有67台

提供telnet有34台

提供smtp有35台

提供dns有6台

提供web有68台

不提供服務或未知共118台

os與service間的對應關係如下:

	ftp	telnet	Smtp	Dns	web
FreeBSD	0	0	0	0	0
Windows	35	1	14	4	52
Solaris2.x	32	33	21	2	16

(3) 交大:共偵測312台機器.其中:

os部分: freebsd 3.x共19台

windows系列共230台

solaris 2.x共63台

service部分: 提供ftp有97台

提供telnet有53台

提供smtp有62台

提供dns有15台

提供web有58台

不提供服務或未知共168台

os與service間的對應關係如下:

	ftp	telnet	Smtp	Dns	web
FreeBSD	7	5	10	8	9
Windows	48	2	18	5	30
Solaris2.x	42	46	34	2	19

(4) 成大:共偵測189台機器.其中:

os部分: windows系列共137台

solaris 2.x共52台

service部分: 提供ftp有55台

提供telnet有41台

提供smtp有42台

提供dns有11台

提供web有74台

不提供服務或未知共82台

os與service間的對應關係如下:

	ftp	telnet	Smtp	Dns	web
FreeBSD	0	0	0	0	0
Windows	20	3	16	7	46
Solaris2.x	35	38	26	4	28

綜合來看,在這1563台電腦中,

(1) os的普及率依次爲windows(79.2%), solaris 2.x(19.6%), freebsd 3.x(1.2%).

(2) 較常提供的服務依次爲ftp(375台), web(332台), smtp(255台), telnet(236台), dns(67台).

(3) 較常用來提供服務的os依次爲solaris 2.x(56.1%), windows(40.8%), freebsd 3.x(3.1%)

九.參考資料

書籍

[1] Windows 程式開發設計指南. Charles Petzold 著.
[2] ICMP 能幫你做什麼？談 OS 偵測技術（一） 鮑友仲著

網站

[1] Hackland 駭客資訊網 http://www.hackland.idv.tw/

User: jhliou
Host: ccsun6
Class: ccsun6
Job: q.prn