網路實驗工具

吳明蔚　林盈達

國立交通大學資訊科學系
E-MAIL：gis89235@cis.nctu.edu.tw，ydlin@cis.nctu.edu.tw

摘 要

TCP/IP 網路協定透過網路實驗工具的觀察，能幫助了解其理論的運作方式，有不少工具都有其用途的獨特性及在網路多層架構中特別的觀察意義。茲引導使用者去了解並觀察 Data Link 層的 ARP 協定、Network 層的 ICMP 協定、及 Transport 層的 TCP 協定，並適時使用除錯工具進行網路運作的封包擷取。我們將實際操作六個工具，分別是有關名稱與地址的 arp 與 host，有關主機內部工作的 netstat 與 tcpdump，以及有關遠端主機的 ping 與 traceroute。

關鍵詞：TCP/IP、工具、arp、host、netstat、tcpdump、ping、traceroute

1. 工具分類

TCP/IP 網路協定的發展源自 80 年左右，當時觀察及分析網路的運作多半得透過昂貴的硬體設備，既受限於硬體操作的不便，也無法彈性調整相關參數。 到了今天，網際網路的興起讓 TCP/IP 通訊協定[1]普及至每個網路節點，同時公共軟體界有許多開放原碼程式，本文章將透過這些工具的觀察來了解TCP/IP 通訊協定的基本且重要的觀念[2]，這包括ARP (Address Resolution Protocol) 與 DNS (Domain Name System) 在 Data link 層的溝通原理；ICMP (Internet Control Message Protocol) 與 IP (Internet Protocol) 在 NetworkA 層的封包傳送；TCP (Transmission Control Protocol) 與 UDP (User Datagram Protocol) 在 Transport 層的運作流程。每個觀念或原理會介紹兩個開放原碼工具，茲整理分類於表一。第一類 Addressing 與 Naming 分別是指 ARP 及 DNS 兩個機制，ARP 將 TCP/IP 的 IP 位址 (IPv4 規範是 32 位元) 和硬體驅動程式的 Hardware 位址 (Ethernet 環境下是 48 位元)之間作 Mapping。DNS 是提供 TCP/IP 應用程式將 IP 位址與 Host 名稱之間作 Mapping。第二類 Internal Behaviors 著墨在內部運作的觀察，這包括網路連線狀況的分析，以及 IP 封包進出

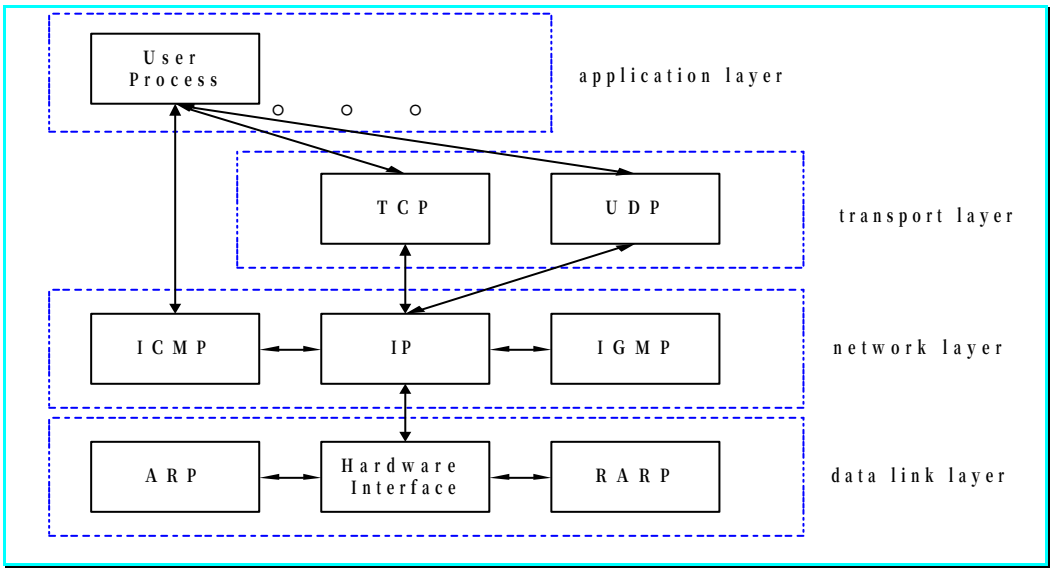
的觀察，藉以瞭解網路的運作情形。第三類 Probing 是利用使用者端的軟體，透過 ICMP 等協定來偵察探測網路節點的相關資訊，包括判斷目標主機是否 alive，以及追蹤所經過的 hop 節點。

Categories	Name of the tool
Addressing and Naming	ARP
	Host
Internal Behaviors	Netstat
	Tcpdump
Probing	Ping
	Traceroute

表一. 將工具依功能作三個主要分類

2. 協定運作觀察點

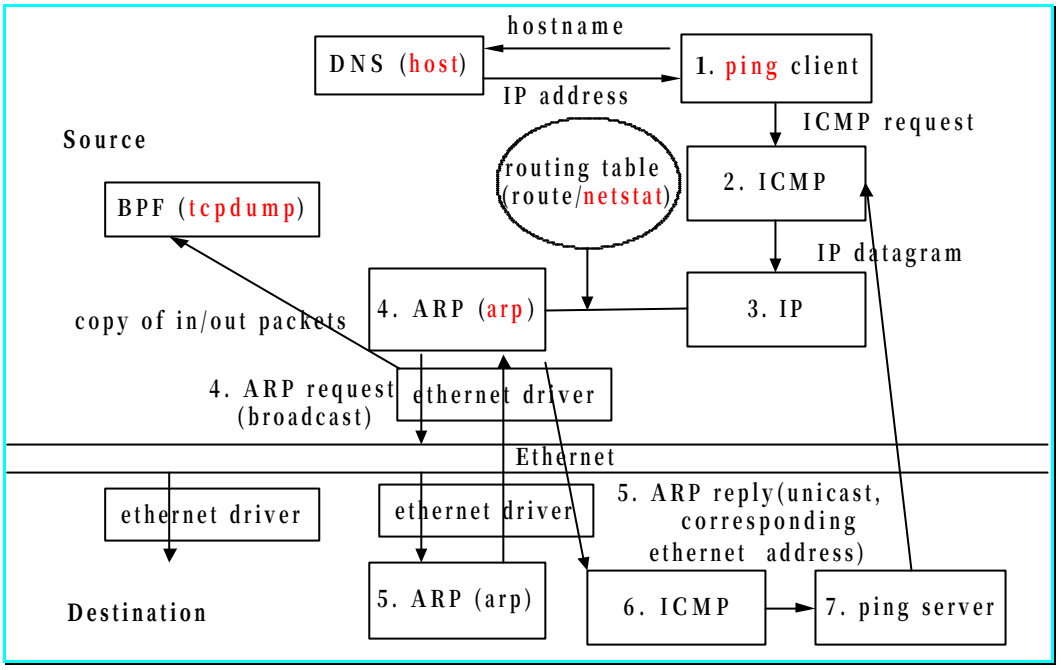
TCP/IP 是五層式的協定架構，包括 Physical 層、Data Link 層、Network 層、Transport 層，以及 Application 層。Data Link 層通常包括網路卡及相關驅動程式。Network 層處理網路上封包的運作，譬如封包的 Routing。Transport 層負責端點至端點(end-to-end)的資料流傳送，其中 TCP 和 UDP 各有各的特性。Application 層負責應用軟體 Client/Server 的處理，如本文中的各工具。示意圖請參考圖一[1]。在 Linux 系統中，Network 及 Transport 層位於 kernel，而 Application 層則是 user space。



圖一. TCP/IP 的各協定的溝通示意圖

從學習的角度而言，知道網路實驗工具的操作方式並不是首要工作，使用者應該先瞭解該工具在整個 TCP/IP 協定運作所扮演的角色，也就是清楚的知道這些工具的觀察點在哪。圖二舉

Ping 這個工具做為例子，整合本文的其他工具，協力完成一個完整的 Source 與 Destination 的溝通，期使使用者瞭解該工具的觀察點。



圖二. 使用 Ping 工具之各協定的觀察點

1. 假設在 Source 端執行 Ping，對象是 Destination 端。當輸入的位址是 Host 名稱，而非 IP 位址時，需要藉由 DNS (或稱 Resolver)來作 Mapping。Ping 這程式已包含 resolve hostname 的功能。
2. ICMP 會接受 ICMP echo request 並將此 request 送至 Destination 的 IP 位址。
3. 去查 Routing table 的 entry，如果 Destination 端是 locally connected，則封包會送至該 local host；如果 Destination 端是 remotely connected，則封包會送至指定的 default gateway 或 local next-hop router。
4. 不管此封包是要送至 host 或 router，都必須知道 IP 位址和 Hardware 位址的 Mapping，而這正是 ARP 的工作。ARP 會送出 Broadcast 的 ARP request 給 local 網路的每個 host。
5. 符合條件的 host 會回應 Unicast 的 ARP reply。
6. 則 Source 端根據此 Hardware 位址將 IP 封包送出。
7. 此例中，Source 與 Destination 在同一 local 網路，所以 Destination 直接收到封包，

Destination 端的 ARP cache 也會有 Source 端的 entry，所以 Ping 可直接回送封包給 Source 端。

整個 Ping 程式的運作，仰賴許多不同協定的相互合作，每個觀察點的工具實際操作方式會在稍後各節介紹。

3. 名稱與地址之工具

A. Host

此"host"工具允許去查詢 DNS 伺服器，輸入 host 名稱會回應 IP 位址；或輸入 IP 位址作反查回應 host 名稱，另一個與"host"類似的工具是"nslookup"。圖三是 host 工具在 debug 模式下運作的細節。其中 Type 為 PTR 代表 Pointer query，IP 位址會以 domain 名稱呈現在 in-addr.arpa 的 domain，而 Class=IN 代表是 Internet 位址。

```
從 host -d 140.113.88.190 的輸出...

;; res_nmkquery(QUERY, 190.88.113.140.IN-ADDR.ARPA., IN, PTR)
;; res_send()
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 51850
;; flags: rd; QUERY: 1, ANSWER: 0, AUTHORITY: 0, ADDITIONAL: 0
;;        190.88.113.140.IN-ADDR.ARPA, type = PTR, class = IN
;; Querying server (# 1) address = 140.113.179.250
;; got answer:
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 51850
;; flags: qr rd ra; QUERY: 1, ANSWER: 1, AUTHORITY: 2, ADDITIONAL: 3
;;        190.88.113.140.IN-ADDR.ARPA, type = PTR, class = IN
190.88.113.140.IN-ADDR.ARPA.  22h10m46s IN PTR  speedserver.cis.nctu.edu.tw.
88.113.140.in-addr.arpa.  22h10m46s IN NS  cisserv.cis.nctu.edu.tw.
88.113.140.in-addr.arpa.  22h10m46s IN NS  cissol1.cis.nctu.edu.tw.
cisserv.cis.nctu.edu.tw.  17h6m51s IN A  140.113.207.5
cisserv.cis.nctu.edu.tw.  17h6m51s IN A  140.113.23.1
cissol1.cis.nctu.edu.tw.  17h6m51s IN A  140.113.23.101
rcode = 0 (Success), ancount=1
190.88.113.140.IN-ADDR.ARPA domain name pointer speedserver.cis.nctu.edu.tw
```

圖三. Host 的除錯模式

另外一種使用方法是作 DNS 伺服器的 Zone transfer，所謂 Zone transfer 即是 Secondary DNS 從 Primary DNS 那獲得該 Zone 的資訊。現在大部分的 DNS 伺服器會去過濾這樣的 DNS Query，確定是否真的是來自 Secondary DNS 的要求。不過也是有例外，從圖四可以看出，透過 Zone transfer 獲得約 370 筆的可能 host 名稱與 IP 位址。其中，grep A 是去找出有 A 這個字，並 wc –l 算出行數。

```
從 host -l -v -t any eic.nctu.edu.tw | grep A | wc -l 的輸出...

    373
```

圖四. Host 的 Zone transfer

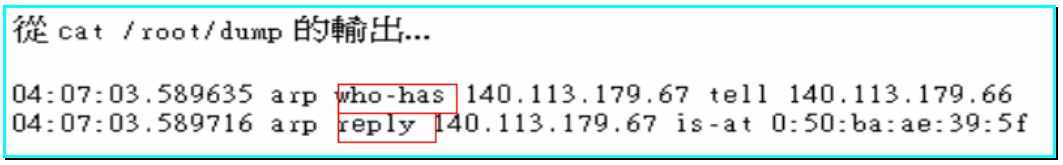
所有的參數及意義請見表二。

host [-l] [-v] [-w] [-r] [-d] [-t querytype] [-a] host [server]	
-l 列出完整的 domain	-d 開啟除錯模式
-v 輸出冗長的細節	-t 指定查詢型態
-w 無限等待對方伺服器的回應	-a 相當於-v –t any
-r 關閉遞回式查詢，無其他 DNS 協助	

表二. Host 指令的主要參數

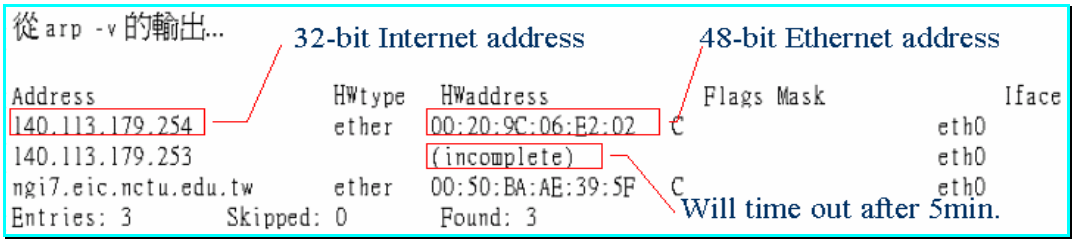
B. ARP

此"arp"工具提供使用者檢視系統的 ARP cache，並可新增與刪除 ARP entry。先前提過，ARP 會負責 IP 位址與 Hardware 位址之間的 Mapping，從圖五 (透過 tcpdump 所觀察到的結果) 可以看出實際的 ARP request 與 ARP reply 狀況。ARP 先作 who-has 的廣播，而有一個 reply 回應。



圖五. ARP 的要求與回應

觀察 ARP cache 的方法十分簡單，如圖六的輸出，IP 位址在左邊，HWType 表示是 Ethernet 網路，HWaddress 即是該 IP 位址的 Hardware 位址 (即網路卡卡號)，而 Flag 為 C 表示 Complete 完整的 arp entry。當去 Ping 一個不存在於網路上的機器十，會造成 incomplete 的 arp entry，因為資料並不完整，incomplete 的 entry 會在短時間內被刪除 (5 分鐘)。



圖六. ARP 在 Verbose 參數下的輸出

主要的參數及意義請見表三。

arp [-v] [-n] [-i] [-H type] [-a hostname] [-d hostname] [-s hostname hw_addr]	
-v 輸出冗長的細節	-a 顯示指定 host 名稱
-n 顯示數字位址	-d 刪除該 host 的 arp entry
-i 顯示指定 Interface	-s 新增該 host 的 arp entry
-H 顯示指定網路型態(type)	

表三. Arp 指令的主要參數

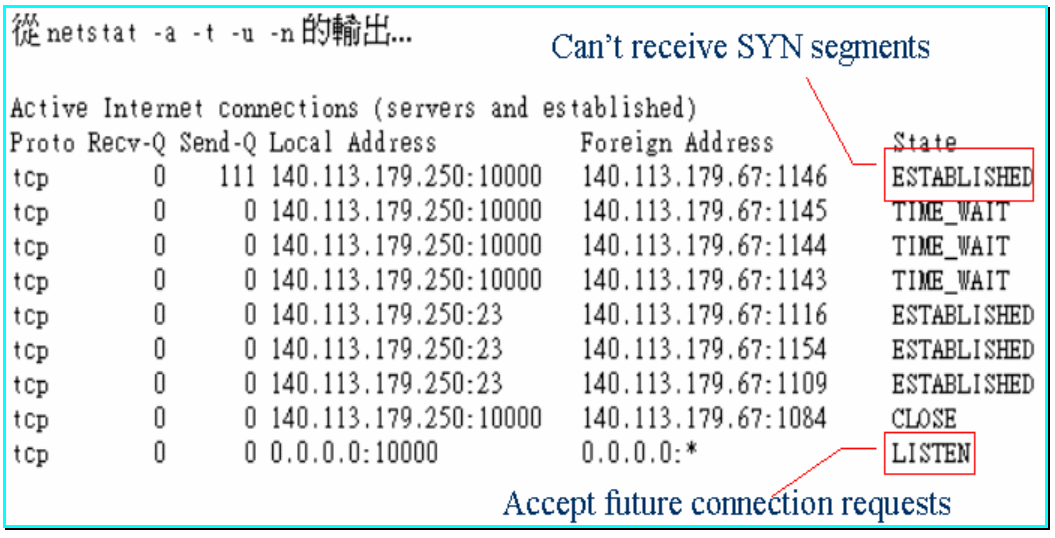
II. 主機內部運作之工具

A. Netstat

此"netstat"工具會顯示各種不同網路相關的統計，因此 netstat 可以說是網路監控最基本的工具之一。從圖七可以看出此 Interface 上的一些資料傳輸統計，包括其 MTU (表示最大傳輸單位，Ethernet 的封包格式依 RFC 894 規定其資料是從 46~1500 bytes)、已收到封包數量、已收到錯誤封包數量、已傳送封包數量等等資訊。

圖七. 此機器上 Interface 的統計資料

在圖八則顯示目前 All connections (-a)中，屬於 TCP (-t)及 UDP (-u)的連線狀況，並且以 Numeric (-n) IP 位址的方式呈現 Local 位址。其中 Local Address 是以 host(or IP):port 來呈現，每一個 active 的 server connection，其 port state 一定會有一個是處於 LISTEN，以作為接收連線要求(SYN_REQUEST)。詳細的 TCP state 請參考表四。



圖八. 目前機器的所有 TCP 及 UDP 的 Active 連線

Connection State	Explanation
ESTABLISHED	Socket 已建立連線

SYN_SENT	Socket 正嘗試主動建立連線
SYN_RECV	從網路上收到連線要求
FIN_WAIT1	Socket 已關閉，連線正在結束中
FIN_WAIT2	連線已關閉，Socket 正等待遠端的結束
TIME_WAIT	Socket 正等待中，結束處理仍在網路中的封包
CLOSED	Socket 已沒有在使用

表四. TCP 連線狀態及說明

主要的參數及意義請見表五。

netstat [-v] [-n] [-i] [-g] [-s] [-N] [-A family]	
-v 輸出冗長的細節	-s 顯示每個協定的摘要統計
-n 顯示數字位址	-N 支援 netlink 並觀察其建立與刪除
-i 顯示指定 Interface	-A 顯示指定協定種類
-g 顯示 IGMP multicast group 成員關係	

表五. Netstat 指令的主要參數

B. Tcpdump

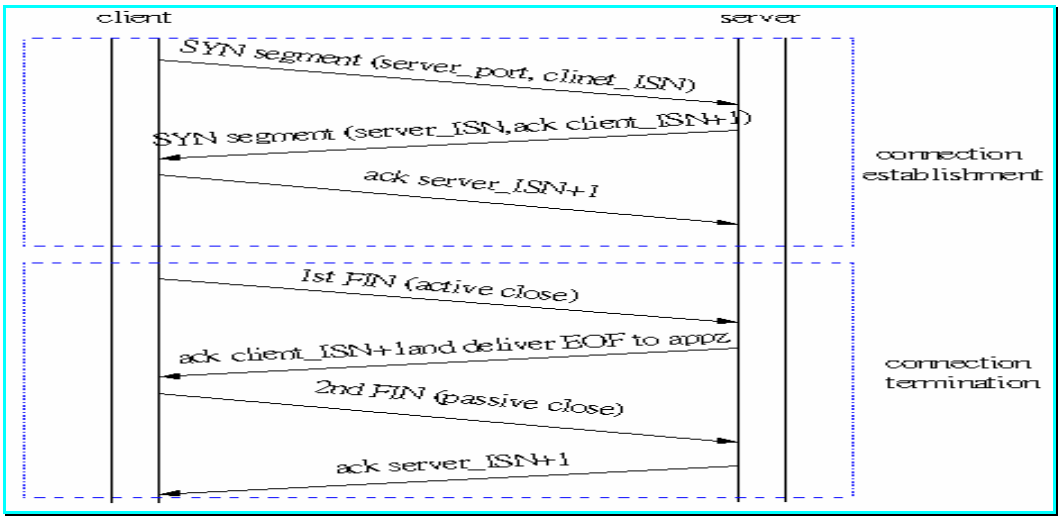
此"Tcpdump"工具可以擷取封包並顯示其內容，以進行網路流量監控。在了解網路協定的過程中，tcpdump 是很適合的協定分析工具。原因是 tcpdump 可以擷取任何送出和傳入的封包，並可以指定 host 及指定 port 號碼等條件式敘述。例如只對有關 140.113.88.190 這個 IP 位址的 port 21 的封包有興趣，而且只要擷取 10 個封包，則可以寫成如圖九，而 tcpdump 即會開始收集符合這些條件的封包。

```
從 tcpdump -c 10 host 140.113.88.190 and port 21 > dump 的輸出...
tcpdump: listening on eth0
```

圖九. 針對指定 Host 及 port 的 Tcpdump 方法

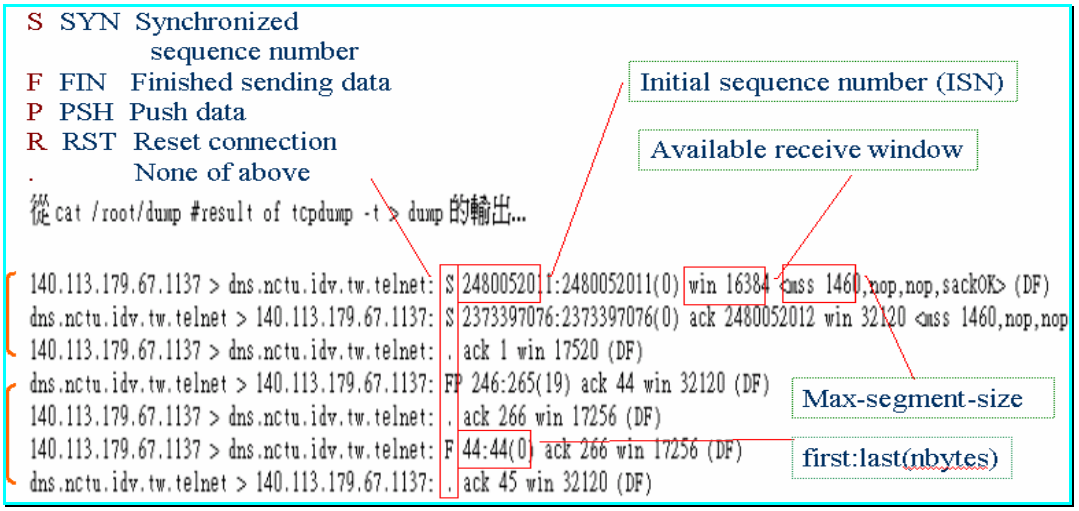
一個完整的 TCP connection 會有建立與結束(請參考圖十)。當 TCP connection 在建立時，會進行 Three-way 的 Handshaking 及 connection 的 Synchronization。由使用者端送出 SYN (Synchronized Sequence Number) 封包，裡面有它自己的 ISN(Initial Sequence Number)及伺服器端的 port 號碼。而伺服器端該 port 收到此 SYN 要求後，也會回送一個 SYN 封包，裡面有它自己 ISN 及剛剛收到的使用者端 ISN+1(作為確認)。等使用者端也收到伺服器端回送的這個 SYN 封包後，使用端會將伺服端的 ISN+1 作為 acknowledge 訊息回送回去，Handshaking 就此成功結束。

不像 TCP connection 的建立是 Three-way Handshaking，在 TCP connection 的結束需要四個步驟，其原因是 TCP 為 Full duplex，封包的傳送是雙向的，所以兩端都需要 ack 過後才能確認結束。



圖十. TCP **連線的建立與結束**

從圖十一可以觀察的到剛剛所述的 TCP connection **的建立與結束，前面三行是建立的過程，後面四行是結束的過程。其中** S **表示** SYN **要求，而** F **表示** FIN **結束訊息。每個封包的資料長度由** first:last(n bytes)**來表示，**first**是起始位元，**last**是結束位元，而** n bytes **即是** last-first **的長度。所以如果是** control message **的話，其資料長度多為** 0**，而** data **封包則不為** 0**。**



圖十一. **以** Tcpdump **觀察** TCP **連線的建立與結束**

主要的參數及意義請見表六。

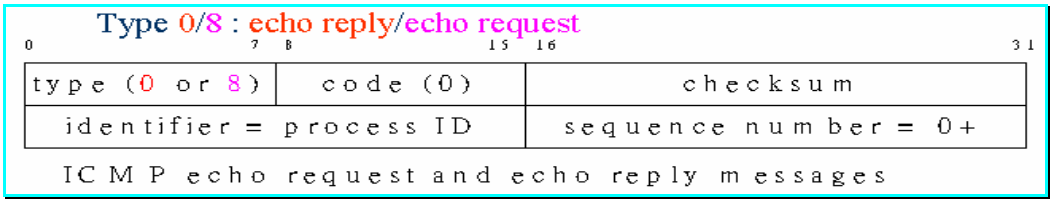
tcpdump [-v] [-n] [-i] [-c count] [-t] [-s snaplen] [expression]	
-v **輸出冗長的細節；**-vv **為更冗長**	-t **不顯示時間單位**
-n **顯示數字位址**	-s **改變擷取封包大小，預設為** 68bytes
-I **監聽指定的** Interface	Expression **為可以使用** AND, OR **等邏輯表示法**
-cdump **特定的封包個數**	

表六. Tcpdump **指令的主要參數**

III. 偵察探測之工具

A. Ping

此"Ping"是很基本的主機診斷工具，提供資訊包括主機是否能正常連線，以及到目的主機的往返時間(Round-trip time)。Ping 工具是利用 ICMP 的查詢要求(echo request)與回覆(echo reply)，詳細封包格式請見圖十二。其中 identifier 會設為處理該 ICMP 訊息的程序 ID，Sequence 值從 0 開始，每送出一個封包就會遞增。



圖十二. ICMP echo reply 與 echo request 的格式

在圖十三中是顯示 ping 對目的位址安靜地(quiet mode)以氾濫方式(flooding)傳送 5000 個封包。安靜模式下不會顯示每一個封包的細節，只會輸出最後的診斷結果，而氾濫方式則以約每秒 100 個 ICMP 封包(或更高)來送至目的地。因此氾濫模式也適合用在短時間內製造可觀的網路流量。

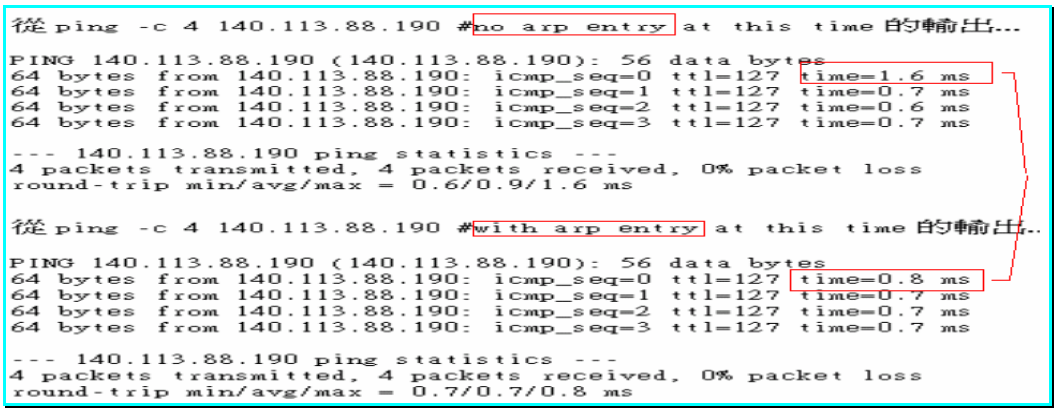
```
從 ping -q -f -c 5000 140.113.88.190 #quietly flood 5000 packets to Speedsrv 的輸出...

PING 140.113.88.190 (140.113.88.190): 56 data bytes

--- 140.113.88.190 ping statistics ---
5071 packets transmitted, 5000 packets received, 1% packet loss
round-trip min/avg/max = 0.6/10.3/30.5 ms
```

圖十三. 以安靜模式來 Flood 對方主機

稍早章節提到的 ARP Cache，可以使用 Ping 來驗證此 Cache 的效果。在圖中是第一次傳送 ICMP 要求至 140.113.88.190，此時 ARP Cache 並沒有此 ARP entry，因此可以見到第一個 ICMP 封包的 Round-Trip Time 比其他剩下的封包足足多了 2~3 倍的時間；而在第二次執行 Ping 時，因為先前此時 ARP Cache 已經有 140.113.88.190 的 ARP entry，故無須再去做 ARP request/reply，可以發現其第一次的 ICMP 封包的 Round-Trip Time 和其他 ICMP 封包花差不多的時間。足可見 ARP Cache 的存在是有差別的。



圖十四. ARP cache **的存在之影響**

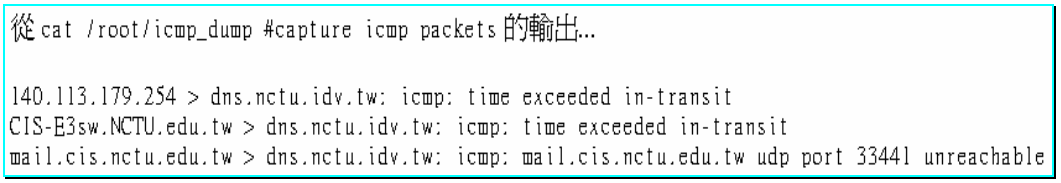
主要的參數及意義請見表七。

ping [-dfnqrvR] [-c count] [-i wait] [-s packetsize]	
-v **輸出冗長的細節**	-f **泛濫模式**
-n **顯示數字位址**	-q **安靜模式**
-i **每個** request **之間的時間** interval	-s **指定封包大小**
-c **指定封包個數**	

表七. Ping **指令的主要參數**

B. Traceroute

此"Traceroute"**工具提供使用者觀察** IP **封包在兩端機器間傳送所經過的路徑，協助除錯參考以及更深入了解** TCP/IP **通訊協定。**Traceroute **使用** ICMP **裡頭的** TTL(Time-to-Live)**欄位，此欄位最初是為了防止資料封包在網路傳輸過程進入無窮迴圈，當封包每經過一台機器其** TTL **的數值就會減** 1**，一但數值變成** 0**，則該機器會** discard(**丟棄**)**此封包並回傳** ICMP **時間逾時**(time exceeded)**，則** Traceroute **得以獲知該機器的位址，並藉由控制** TTL **的數值，測第一台主機時的設** TTL **為** 1**，然後第二台主機** TTL **是** 2**，直到該封包抵達目的地。特別注意的是如何辨別是否抵達目的地？基本上，**Traceroute **會將** UDP **的** port **號碼設為非常大的值(大於** 30,000**)，而目的地的主機根本不可能使用這樣大的** port **號碼，以至於會產生** port **無法抵達**(unreachable)**的錯誤訊息。從這兩種不同的錯誤訊息，發送端可以判斷** Traceroute **的** IP **封包是否已抵達目的地。請參考圖十五。**



圖十五. 以 Tcpdump **觀察** Traceroute **的** ICMP **錯誤回應**

主要的參數及意義請見表八。

traceroute [-dFInrvx] [-f first_ttl] [-g gateway] [-i iface] [-m max_ttl] [-p port] [-q nqueries] [-s src_addr] [-t tos] [-w waittime] host [packetlen]	
-v 輸出冗長的細節	-a 顯示指定 host 名稱
-n 顯示數字位址	-d 刪除該 host 的 arp entry
-i 顯示指定 Interface	-s 新增該 host 的 arp entry
-H 顯示指定網路型態(type)	

表八. Traceroute 指令的主要參數

IV. 結論

網路上的工具繁多，本文章以基本的工具（一些進階的套件，也值得讀者自行試試，請參考表九）去觀察了解許多基本的 TCP/IP 協定觀念。工具用的多，不如用的巧，透過實證去操作，越能幫助使用者清楚工具和協定之間的關係與用途。這也對日後網路相關問題的診斷與分析是有相當幫助的。

Naming and Addressing	
dnswalk	A DNS database debugger
Lanlord	A dhcpd lease reporting program
KcmBind	A front-end utility to configure bind
Internal Behaviors	
Funlog	A colorful IP logger of the network traffic
Sniffit	A complete packet sniffer
Perro	Logs incoming IP:TCP, UDP and ICMP packets
Probing	
penemo	PErl NEtwork Monitor performs several check functions
OpenNMS	Java-based network & systems management
BINDMON	Run on a DNS server to do system health monitoring

表九. GPL 版權為主的部分相關套件

V. 參考資料

[1] W. Richard Stevens, "TCP/IP Illustrated Volume 1: The Protocols", Addison Wesley, 1994

[2] Steve Maxwell, "Red Hat Linux Network Management Tools", McGraw-Hill, 2000