IEPM-BW 在 TWAREN 上的實現

陳振華　楊哲男　張宜正
國家高速網路與計算中心
chenhua@nchc.org.tw, yangcn@nchc.org.tw, andy@nchc.org.tw

摘要

IEPM-BW[1]是 SLAC 所主導的一個量測計畫，旨在建置一個國際性的網路效能觀測網；它用了多種網路效能量測工具來分別測量各段國際網路連線的頻寬、延遲等數據，並將結果作分析統計後呈現在網頁上以利觀看。 TWAREN 目前已加入此量測計劃，並成為了一個效能觀測網站(Monitoring Host)[2]，用來觀測分布在全球的 25 個測量點(Target Nodes)。此計畫可將 TWAREN 的國際網路量測範圍推的更廣，也更豐富了 TWAREN 在量測方面的數據資料與經驗。

關鍵詞：IPEM-BW、SLAC、TWAREN、效能量測。

Abstract

The goal of IEPM-BW, a famous measurement project led by SLAC, is to establish an international network performance observatory website. It uses several different network measurement tools to periodically measure all its networked institutions' performance, and then statically analyzes the results and shows them on its observatory website. TWAREN participated in the project as a measurement target node of the SLAC monitoring host in the beginning and then itself became a monitoring host. The TWAREN monitoring host now has 25 measurement target nodes. In so doing, TWAREN extends its international measurement range and also enriches its experience in network performance measurement.

Keywords: IEPM-BW, SLAC, TWAREN, performance measurement.

1. 前言

由於網路技術的快速發展，使得網路頻寬變得愈來愈大，網路的架構也越趨龐大及複雜，此外，隨著應用程式開發者以及使用者的頻寬需求上升，越來越多需要跑在高效能高頻寬網路上的應用也相繼出現，例如像是 Particle Physical Data Grig 及 GridPhysical Netowk 等計畫所開發出來的格網(Grid)應用就必須跑在像是 Esnet 及 Internet2 等高效能高頻寬的網路上；一般來說，這一些格網應用對於網路的效能好壞相當的敏感，因此有一個可以對高效能網路作效能監測的系統是相當需要的，故由美國史丹福線性加速中心(Stanford Linear Accelerator Center, SLAC)所主導的 IEPM-BW (Internet End-to-end Performance Monitoring – Bandwidth to the World)計畫相應而生。

IEPM-BW 旨在建置一個簡單但穩固的國際性網路效能觀測網，以提供高速網路的維運者與應用者一個平台去了解自己所使用的網路的狀況。IEPM-BW 計畫的量測範圍橫跨亞美歐三洲的許多國家，採用 6 種網路效能測量工具，以週期性測量的方式，分別測量出與各個測量點間的頻寬、延遲(latency)等網路效能數據，並將數據作統計分析，最後呈現在網頁上以供查閱。

TWAREN 是一個國際級的高品質網路，因深知網路效能測量的重要性，早就有建置自己的網路效能量測系統；為加強國際合作與增進經驗，TWAREN 加入了 IEPM-BW 計畫，一開始只是其中一個觀測站的一個測量點(註：IEPM-BW 目前共有 8 個觀測站，每一個觀測站旗下有 25~40 個測量點)，後來自己也成為了一個觀測站[2]，旗下有 25 個測量點；此舉不但將 TWAREN 的網路效能測量系統的範圍首次延伸到歐洲，且從與 IEPM-BW 合作的過程中也學習到了許多在網路效能測量方面的經驗。

第 2 章將一一介紹在這計畫中所使用的測量工具，以及其參數校調的意義，還有介紹我們在實驗中求取最佳參數的經驗；第 3 章將對 TWAREN 觀測站的測量的結果作介紹，以及如何觀察這些測量結果來判斷網路的狀態;第 4 章是結論與未來工作。

2. 測量工具的介紹與校調

一個網路效能觀測網之結果的精確度就取決於它所用來測量網路效能的工具以及維護管理人員之校調方式及經驗，因此相當重要。IEPM-BW 隨著時間的過去，先後更新替換了許多測量時所用的工具，目前所用的是 PathChrip、Ping、OWPing、Iperf、ThruLay、PathLoad 等六項測量工具，其中 PathChirp 跟 Pathload 是藉由送出 UDP 測試封包、而 Iperf 跟 ThruLay 是藉由送出 TCP 測試封包來測量網路頻寬；這些測量工具各自都有相當多的參數可供設定者依需求作設定；底下我們就來介紹這些測量工具並談一談 TWAREN 這個觀測站對於測量工具的參數設定(註：IEPM-BW 的各個觀測站所用的測量參數可依需求各自作調整)：

2.1 PathChirp

PathChirp[3]以” self-introduced congestion”的觀念，藉由封包呼叫(packet chirps)的方式來動態的測量出網路的剩餘頻寬。packet chirps 是以上升指數的線型來調配每一次所送出的測試封包群中每一個測試封包的送出速率，然後將每一次 chirp 所測出的結果加以平均，藉以判斷出網路的剩餘頻寬。此方式較以前的 packet pairs 或 packet trains 方式都較為好。

Pathchirp_run －t 10 －J 6 －a 2 －l 100 －u 1000 為 TWAREN 觀測站對這個測量工具所下的參數，其中測量周期(-t)為 10 秒、一個 Jumbo 封包(-J)包含 6 個測試封包、發送 chirp 的平均速率(-a)為 2Mbps、在 chirp 中的測試封包的最低(-l)與最高速率(-u)為 100 與 1000Mbps。這些是適用於一般網路的參數設定，而對於頻寬過小(大)的測量點，則可用調整-l(-u)來因應即可。

2.2 Ping

Ping 是最簡單但是最常被用來測量網路效能的工具，可以測得測試封包往返某一點所需的往返時間(Round Trip Time, RTT)。經過一些交叉分析後，網路管理者或是有興趣的使用者就可透過網頁來分析這些資訊並判斷網路是否有異常。

Ping –c 10 –W 20 為 TWAREN 觀測站對這個測量工具所下的參數，其中一次測試只送 10 個封包(-c)、等待回覆封包時間(-W)為 20 秒。對於距離過長的測量點、可增加等待回覆封包的時間來因應。

2.3 OWPing

OWPing (One-Way Ping) [4]如其名為測試封包單向的延遲時間，可搭配 ping 工具來發現路由不對襯的問題。它能同時測量封包往跟返兩個單向所需花費的時間；當 ping 發現某個網段發生延遲時、就可利用 OWPing 來更近一步確認是這網段的哪一路由方向出現擁塞或是繞遠路。(註：因 OWPing 需要很精準的 time stamp 來標記時間，故需要網路時間協定(NTP)的協助才可順利運作，且同一套監測群下的所有機器最好設定同一台 NTP server。目前 TWAREN 共有兩台 NTP servers；其中一台為主，另一台為輔。)

對於 OWPing 的設定，我們皆採取預設值；不過我們有開啟認證模式、也就是 TWAREN 觀測站跟其測量點在作測量之前，須先認證彼此的身分。

2.4 Iperf

Iperf[5] 為 NLANR[6] 組織下之 DAST (Distributed Applications Support Team)小組所開發出來測試網路效能(Throughput)之工具。Iperf 最大的特色是可以藉由調整其兩個參數(-P&-w)來測出目前的最大頻寬，-P&-w 分別代表平行串流量(number of parallel threads)與視窗大小(window size)(註：一般在測試 TCP 之效能時，可利用-P 參數來模擬 multi-stream 的效果以測試出點對點間之最大 TCP 效能)。在不造成網路擁塞的情形下，調整此兩個參數的乘積所能到的最大值就是目前當下網路之最大可用頻寬(註：但在 UDP 測試參數下(-b)，容易造成網路的暫時擁塞)。

Iperf –i 5 –t 20 為 TWAREN 觀測站對這個測量工具所下的參數，其中結果呈現週期(-i)為每 5 秒，送測試封包時間(-t)為 20 秒、視窗大小為預設值。對於過遠的測量點，可增加視窗大小(-w)與送測試封包時間(-t)來因應。

2.5 ThruLay

ThruLay[7] 藉由採用計時較精確的 TSC register 來產生出大量的精確的 Possion 曲線測試串流，以測量出網路的 capacity；而且其產出結果即為繪圖工具 gnuplot 的標準輸入。

ThruLay –t 20 –i 5 為 TWAREN 觀測站對這個測量工具所下的參數，其中結果呈現週期(-i)為每 5 秒，送測試封包時間(-t)為 20 秒、視窗大小為預設值。對於過遠的測量點，可增加視窗大小(-w)與送測試封包時間(-t)來因應。

2.6 PathLoad

PathLoad[8]用來測量頻寬的基本概念是：逐漸增加送出的測試串流的速度，當單向的延遲逐時間漸增加時，表示測試串流的速度漸漸逼近剩餘頻寬數值。PathLoaad 已經過測試，絕不會影響目前網路的流量。

TWAREN 觀測站對這個測量工具所下的參數皆使用預設值並以 90 秒為週期去執行它。

為了找出最適合用於 TWAREN 跟 SLAC 之間的測量工具的參數，我們曾經跟 SLAC 觀測站作過實驗；首先是針對 Iperf，結果如圖 1 所示，由實驗結果我們可以知道：圖中的淡藍色的曲線有最好的測量值，且又以 P 值為 126 時的點有最大的測量值，即”iperf –w 149 –P 126”可以得到最大的throughput；但由於過大的平行串流量(-P)會容易導致 TCP throughput 的不穩，因此”iperf –w 578 –P 32”(註：同樣也是在淡藍色線上的點)會是較適合的選擇。我們後續也對 ThruLay 作過實驗，發現”Thrulay –w 149000 –m 126”可以得到最大的測量值。

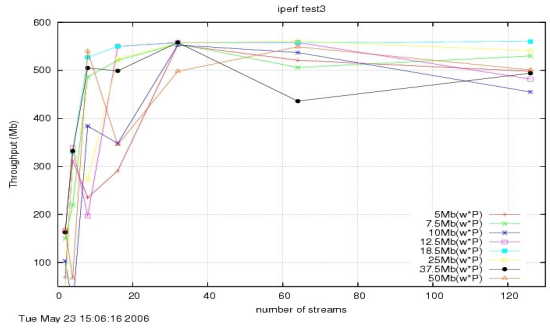

圖 1 Iperf 實驗結果

為避免可能會造成的短暫網路擁塞，我們尚未調整 TWAREN 觀測站的 Iperf 的-P&-w 兩參數或是 ThruLay 的-w&-m 兩參數到我們實驗出來的值，故所測量出來的頻寬不會是最大值；但現階段尚可接受，因為只要測量出來的頻寬值不要忽然過大或過小，就表示網路環境穩定，那也就夠了。此外，由經驗我們知道可以透過調整 TCP windows size 或是應用層下之 buffer size 來增加 TCP 網路之傳輸效能。

3. 成果的介紹與應用

在本計畫裡的每一種測試皆呈現在 IEPM-BW 網頁上，它是一個公開、透明的網頁資料呈現。因此除了網管人員可以查詢之外，任何有興趣之使用者皆可透過網頁來查詢他（她）有興趣之資料。它除了呈現立即的數據之外，並且也儲存了歷史資料以供查詢。以工程師的角度來看，當使用者反應網路品質不良時，工程師可以透過網頁上記錄的 RTT 值、one-way latency、throughput 值、route paths 資訊及所有歷史紀錄來判斷，是否是光纖品質不良、或是與我們對等連結的網路(peer network)發生了問題、或者封包遺失率太大、或者是發生了網路來回不同路的現象等等問題。

以經營者的角度來看，因為 IEPM-BW 可記錄長期的資料，透過以往的歷史資料，系統會自動轉換成趨勢圖，經營者可依此記錄來要求線路提供者改善網路的實體層品質，經營者也可以依長期的觀察紀錄，來做為與其他網路間比較的根本，並可做為下一次線路重新佈建時的依據。以使用者的角度來說，就不必像以前一樣必須清楚的下一些指令，或是只能查詢單方向的資訊，而無法查詢來回的資訊。透過網頁畫面，使用者也可以輕易的查詢到所有加入 IEPM-BW 之網路狀態歷史資料，以方便比較過去與現在的數據。

底下我們將一一介紹各量測數據值所代表的意義為何，以及各趨勢分析所要注意的地方為何。

3.1 即時測量值

如圖 2，透過網頁顯示出所有測量工具的即時測量值跟測量時間；由測量時間點之連結點下去，則可以得到最近一個月的測量值統計圖；以圖 3 為例，就是 PathChrip 最近一個月的統計圖，圖中的箭頭表示當時的封包路徑有更動過，以方便我們確認測量值的變化是否是因為封包路徑的變動而起。圖 2 中的空白格子表示那個測量點尚沒有參加此測量工具的測量活動，圖中的粗紅顏色(例如 snv.nl 測量點之 OWPING 資料)表示此測量工具目前因某事停擺，故所呈現的測量值不是最新的。

Nodenames (25)	PATHCHIRP [ref] pathchirp mbps	PING min | % loss	OWPING [ref] tmin | fmin	iperf [ref] mbps (Window)	thrulay TCP [ref] mbps (Window)	pathload [ref] min mbps
iepm-bw.bnl.org	07/05 09:15 1083.8	07/05 09:32 211.33 | 0	07/05 09:32 93.9 | -1	07/05 08:58 10.8 (40 MB)	07/05 08:17 4.0 (41.9 MB)	
iepm-bw.caltech.ul-org	07/05 09:11 576.4	07/05 09:25 219.473 | 0	07/05 09:32 128 | 91.4	07/05 08:16 71.8 (40 MB)	07/05 08:31 26.2 (41.9 MB)	07/04 21:19 285.7
iepm-bw.cern.ch	07/05 09:13 121.9	07/05 09:26 294.427 | 0	07/05 09:32 141 | 153	07/05 08:37 31.3 (40 MB)	07/05 08:51 12.2 (41.9 MB)	07/03 09:57 89.6
iepm-bw.fzk.de	07/05 09:29 647.0	07/05 09:28 300.393 | 0	07/05 09:28 143 | -1		07/05 09:22 1.6 (40.0 MB)	
iepm-bw.slac.stanford.edu	07/05 09:14 583.2	07/05 09:34 167.939 | 0	07/05 09:32 67.7 | 99.9	07/05 08:34 134.0 (40 MB)	07/05 08:41 57.1 (40.0 MB)	07/05 06:05 571.6
iepm-bw.snv.ul-org	07/05 09:08 545.6	07/05 09:34 226.733 | 0	06/25 23:17 136 | -1	07/05 09:00 38.6 (40 MB)	07/05 08:19 17.3 (40.0 MB)	07/05 09:33 600.0
node1.binp.nsk.su		07/05 09:32				

圖 2 即時測量值

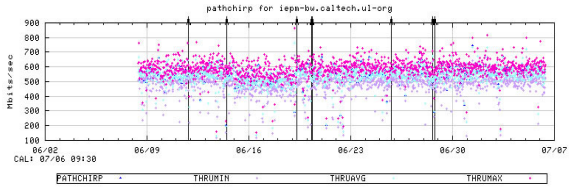

圖 3 PathChirp 統計圖

3.2 測量統計圖

由圖 2 的測量點名字點下去，即會出現此測量點的測量統計圖；以圖 4 為例，是 TWAREN 這個觀測站對 SLAC 這個測量點的一個測量結果統計圖，它包括了上述的所有的測量工具的測量值，分別以不同的顏色被顯示在圖上，讓我們可以宏觀地藉由觀察各種顏色線條的連續狀況，就可以判別網路目前的效能狀況與趨勢情形；以此圖為例，中間有一塊地方的測量數據忽然起了很大的變化，起因是因為當時 TWAREN 正在調整它的 BGP 協定跟國際線路，導致 SLAC 往返 TWAREN 測量點的網路路線變的很長，故黃色部份所代表的 RTT 忽然加大，而且 PathLoad、PathChirp 等測量出來的值變的小很多。

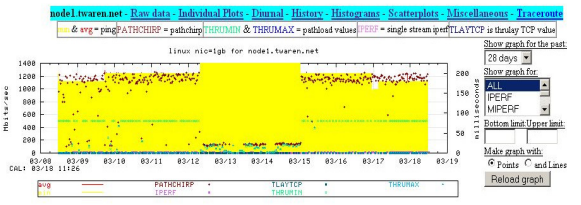

圖 4 測量統計圖

3.3 測量值頻率統計圖

以圖 5 為例，顯示出 PathChirp(圖 5 左)跟 Ping(圖 5 右)各自所測量出來的值的出現頻率；由圖我們可以知道在當時的網路狀況下，Ping 所測出來的值幾乎集中在兩個數字而已，表示此線路的 RTT 相當的穩定；反觀 PathChirp 的值就比較分散，因為 PathChirp 的值本來就會隨網路的流量不同而有所變化，但我們可以期待的是：大部分的時間裡，網路的剩餘頻寬會是頻率最高的那一個值。

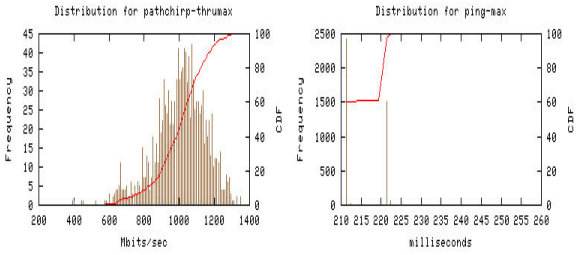

圖 5 測量值頻率統計圖

3.4 測量值相依統計圖

此種統計資料是將兩種以上之測量工具的測量值統計在一張圖上，以利觀察 TCP throughout、UDP throughput、RTT 等之間的相依關係。以圖 6 為例，是 ThruLay(橫軸)跟 PathChirp(縱軸)之間的相依關係圖，由圖我們可得知此網路的 TCP throughput(由 ThruLay 測量出)跟 UDP throughput(由 PathChirp 測量出)間沒有相依性存在(因為數值相當分散)，也就是說當 TCP throughput 的值變大(小)時，UDP throughput 的值不見得會跟著等比例變大(小)。又如圖 7 所示，我們可以發現 RTT 跟 tcp throughout 或 udp throughput 之間也是沒有相依性存在的，這也許是因為當 RTT 為 210ms 跟 220ms 時不會有什麼太大差別，但理論上當 ping 的 RTT 值變大時，利用 TCP 封包來測試的 iperf、thrulay 的測量值應該會相對的變小，利用 UDP 封包來測試的 pathload 應會不受影響。

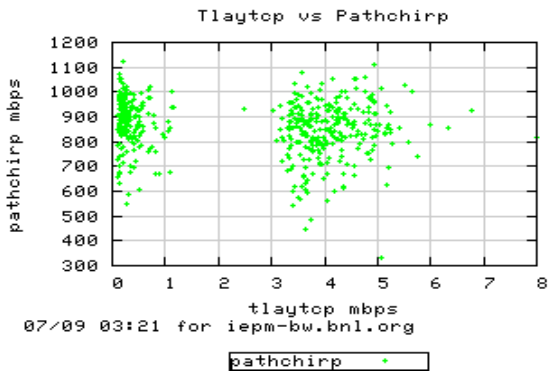

圖 6 ThruLay vs PathChirp

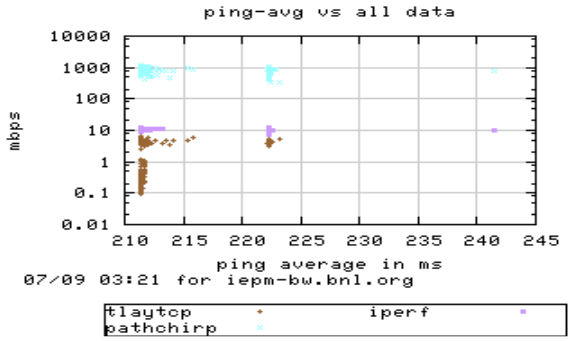

圖 7 Ping-avg vs all data

4. 結論與未來工作

Internet 是一個所謂的 Global Community 型態之互動行為[9]。因此透過國際間之網路組織來做大規模的量測互動是一種難得的機會，也是一個重要的工作。TWAREN 加入 IEPM-BW 這個計畫的最大收穫就是獲得了一個與國際接軌之量測平台，透過本次機會可與來自世界各地的網路效能測量團隊探討網路效能方面的議題；且由於此平台所涵蓋的地域範圍很廣，讓我們有機會可以接觸到各種不同的網路效能的設定，大大刺激且提升了我們的經驗，也讓各國的測量團隊有機會初次認識 TWAREN，以開啟後續可能的合作。其次，藉由此計畫也熟悉了許多的測量工具，知道如何使用與校調它們，也學到呈現測量結果的各種方式，這些東西都定會對 TWAREN 往後的網路效能測量的能力有所提升。

未來的工作方面，除了繼續維運與觀測現有的測量點外，並將積極爭取擁有更多測量點跟測量項目的機會。期望透過這些量測工具的觀測以及 TWAREN 維運中心的 7x24 全年無休之高品質網路運作來提供 IT 及應用開發者一個高效能之網路環境。

參考文獻

[1] IEPM-BW 計畫網站
(https://confluence.slac.stanford.edu/display/IEPM/IEPM-BW)

[2] IEPM-BW@TWAREN.NET 觀測站
(http://iepmbw.twaren.net/~iepm/iepm-bw.twaren.net/slac_wan_bw_tests.html)

[3] PathChirp 官方網站
(http://www.spin.rice.edu/Software/pathChirp/)

[4] OWPing 官方網站
(http://e2epi.internet2.edu/owamp)

[5] Iperf 官方網站
(http://dast.nlanr.net/Projects/Iperf)

[6] National Laboratory for Applied Network Research (NLANR), http://www.nlanr.net

[7] ThryLay 官方網站
(http://www.internet2.edu/~shalunov/thrulay)

[8] PathLoad 官方網站
(http://www.cc.gatech.edu/fac/Constantinos.Dovrolis/pathload.html)

[9] ICFA SCIC Network Monitoring Report, January 2007 Report of the ICFA-SCIC Monitoring Working Group.