行政院國家科學委員會專題研究計畫　成果報告

P2P 定位系統與應用

計畫類別：個別型計畫
計畫編號：NSC91-2213-E-009-068-
執行期間：91 年 08 月 01 日至 92 年 07 月 31 日
執行單位：國立交通大學資訊科學學系

計畫主持人：楊武

報告類型：精簡報告
報告附件：出席國際會議研究心得報告及發表論文
處理方式：本計畫可公開查詢

中　華　民　國 92 年 11 月 10 日

行政院國家科學委員會補助專題研究計畫成果報告

※※※※※※※※※※※※※※※※※※※※※※※※※
※　　　　　　　　　　　　　　　　　　　　　　　　※
※　　　　　　　　　P2P 定位系統與應用　　　　　　　※
※　　　　　　　　　　　　　　　　　　　　　　　　※
※※※※※※※※※※※※※※※※※※※※※※※※※

計畫類別：個別型計畫
計畫編號： NSC 91－2213－E－009－068－
執行期間： 91 年 8 月 1 日至 92 年 7 月 31 日

計畫主持人：楊武
共同主持人：

本成果報告包括以下應繳交之附件：
□赴國外出差或研習心得報告一份
□赴大陸地區出差或研習心得報告一份
□出席國際學術會議心得報告及發表之論文各一份
□國際合作研究計畫國外研究報告書一份

執行單位：國立交通大學資訊科學學系

中　華　民　國　92　年　10　月　31　日

行政院國家科學委員會專題研究計畫成果報告
國科會專題研究計畫成果報告撰寫格式說明
Preparation of NSC Project Reports

計畫編號：NSC 91-2213-E-009-068-
執行期限：91 年 8 月 1 日至 92 年 7 月 31 日
主持人：楊武　國立交通大學資訊科學學系
計畫參與人員： 黃經緯　交通大學資訊科學學系

一、中文摘要

Server-Client 的架構已經成為現今各項網路應用軟體所採用的主要運作模式，然而在網路人口成長速度越來越快之際，此一模式遭遇到相當的瓶頸，在大量的要求(Request 集中的情況下，伺服器的處理速度或是頻寬越來越難以快速地處理數以千萬計的要求。

另外，點對點(Peer-to-Peer)的運作模式在 1998 年 Napster 檔案分享系統大受歡迎以來已經成為新一代網路服務的運作機制的解決方案之一。點對點的運作機制如同一個社會上，且在不考慮階層的情況下的人與人之間的溝通與交流。點對點的服務必須考慮到三件事情，一是點雨點之間如何進行有效率的訊息傳遞，二是整體網路系統如何自動進行維護與調整，三是如何確保整體系統的服務效能。

本計畫的目的是發展一套符合上述三項要求的演算法，同時利用模擬的方式加以驗證本演算法的正確性與效能。

關鍵詞：點對點、主從架構

Abstract

Server-Client model has become a major mechanism to be adopted in designing many network applications. However, with the fast growing users on the Internet, server-client model gradually become bottles of service systems. Under tremendous requests, servers is getting harder to handle them because the processor ability and the bandwidth are no enough.

Besides, peer-to-peer model became popular since a well-known file-sharing system – Napter got popular in 1998. The peer-to-peer model has become one of solutions to serve modern network services. The peer-to-peer works similar to the communication and activity between peoples in our society, without considering the hierarchy. There are three issues that we have to consider in peer-to-peer system. First is the efficient message delivery between nodes. Second is the automatic maintaining of the organization of the system. The third is the guarantee of the system performance.

The purpose of this project is to develop a set of algorithms which meet the above requirements. Also, we use simulation to verify the correctness and performance of our algorithms.

Keywords: peer-to-peer, server-client

二、緣由與目的

Pear-to-Pear (P2P)自 1997 年以來逐漸成為新一代的網路與分散式計算的運作模式，各種應用應運而生，例如 P2P 的檔案分享(Data Sharing)、P2P 搜尋引擎、P2P 分散式高效能計算等等。在浩瀚的千萬個網路節點中，運用智慧演算法，將其上的資源進行定位以及資源分享(Resource Share)是最為節省眾人成本的模式之一。

Peer to Peer (P2P)模式相當於人類之間的資源交換(Resource Sharing)，但發生的地點則是在網路上，由於網路上的節點數以萬計，而彼此資訊的交換有賴於彼此位置(Location)的雙方確認，在傳統的 Server-Client 架構下，僅需建構 Server 便可以完成此一動作，Server 不但身兼資訊的儲存，

也肩負節點的資訊交流任務。此一架構下，Server 成為流量負擔的瓶頸，同時，這也並不符合純粹的人類之間彼此的資源交換(並不需要藉助中央 Server 資料庫)。

另外，一個沒有中央 Server 所主控的 P2P 環境也具有資訊不易收集的優點，在 P2P 環境下，資訊隱私需要被保護，傳統 Server-Client 模式下，資料都經由中央 Server 控管，隱私權無法被保護，而純粹的 P2P 環境下，沒有任何 Server 可言，因此資訊隱私較受到保護。

本計畫的發展前提是在不存在中央伺服器的網路環境下，進行節點對節點之間的通訊，此處所謂的通訊包含了一對多(broadcast)與一對一(reply)兩種方式。由於沒有中央伺服器的幫助 ,因此節點的控制、連絡與管理成為相當棘手的難題，另外，訊息在節點之間的傳遞效率也成為另一個問題，而整體網路系統的堅固性亦是必須注重的。

本計畫的目的在於研究目前 P2P 環境下遭遇的三項問題 :

- 網路結構的自動最佳化
- 智慧型訊息傳送機制
- 保持堅固的網路連結性

問題 1: 網路結構的自動最佳化

所謂網路結構的自動最佳化是指讓一個具有上千上萬個節點的網路系統具有自動調整的能力，所有節點根據智慧型驗算法自動相互合作，將節點與節點之間的連結重新調整，使得整體網路上認兩點之間的通訊成本能夠降低。

問題 2: 智慧型訊息傳送機制

所謂智慧型訊息傳送機制是指在一個純粹的 P2P 環境下 , 訊息的廣播(broadcast)將免不了發生多重傳送的情形，如此將會增加整體系統不必要的負擔，智慧型訊息傳送機制設計一個智慧型的訊息傳送路徑，該路徑由訊息經過的節點共同決定，並且具有大幅降低重複傳送的優異表現。

問題 3: 保持堅固的網路連結性

一個純粹的 P2P 環境下，連結(connection)免不了會遇到中斷的情形，許多情形將可以能造成是個網路碎片(fragment)的發生，這種情形下，整體網路的服務性能大幅降低，我們發展初一套輕量級但卻有效的分散式演算法，讓系統自動維護整個的連結性(connectedness)，當系統出現碎片時，相關節點能夠主動負責將斷掉的部份重新連結起來，形成一個連通的系統。

本計畫已經發展出一個實作系統，稱為 DSE (Distributed Search Environment)，以下將以 DSE 作為本實作系統的簡稱。

三、結果與討論

1. 網路結構的自動最佳化:

在無中央伺服器的 P2P 環境下，整體系統是由數千甚至數萬個節點彼此連結而成，彼此之間的通訊亦須透過節點之間的連結(connection)來傳遞訊息，若節點的連結方式毫無結構時，容易形成訊息傳遞無效率的情形。DSE 發展出一個稱之為 NCC (neighbor clustering control) 的分散式演算法 ,所有的節點根據 NCC 演算法主動與自己距離相近的節點進行連結(connection)，此處所謂的距離之計算是以 ICMP 通訊協定來估計 ,若節點 A 欲得知其與節點 B 之間的距離時，則與 B 之間進行 ICMP 的 packet 傳輸，測得兩點間之頻寬的倒數即與距離成正比。

NCC 演算法讓每一個節點與 98%與自己相近的節點相連接 ,但保留 2%的節點與其他節點進行自由連結(arbitrary connection)，其用意是為了讓此系統產生 Small-World 效應，此效應已經被證實普遍存在於人類社會結構以及各種自然界的群體通訊之間。擁有 Small-World 效應的群體將具有傳遞訊息快速的優點，同時又能保持整體系統具有高度叢集化(clustering)的好處。

在我們的模擬系統中，NCC 演算法能夠大幅降低整

體系統中單一連結(connection)的平均距離，如圖 Figure 1. 所示，從將近 1000 個單位降低到 10 個單位左右。Figure 2. 顯示了 NCC 演算法的主要功能部份。

2: 智慧型訊息傳送機制

在 P2P 環境中，由於沒有中央伺服器的幫助，因此每個節點必須產生一個獨特的編號加以區別，DSE 系統採用 DHT (Dynamic Hash Table)的方式，利用 SHA 演算法，輸入每個節點所獨有的資訊(例如 IP address、時間以及使用者個人資訊等)，產生一組獨一無二的編號，稱之為該節點的 ID，不同節點的 ID 可以互相比較大小，並且用於訊息的傳遞。

訊息的傳遞必須依靠節點之間的連結作為橋樑，同一個節點可能具有多個連結向外延伸，也可能有多個連結連入本機，甚至數個節點可能形成循環(cycle)，因此，訊息有非常高的可能會重複傳遞。DSE 採用兩種方式解決這個問題。

首先每一個訊息中包含一個 HOPS 計數器，訊息每經過一個節點，其 HOPS 計數器就增加一；另外，每個節點都有一個統一個 HOPS Limit，當流入的訊息的 HOPS 計數器超過此 Limit 時，此訊息即停止傳遞。HOPS 計數器的設計可以避免訊息落入無限循環中。

另外，DSE 發展出一套新的訊息傳遞機制，稱之為 SMRB(smart message routing and broadcasting)演算法。SMRB 演算法包含兩個 Stage，第一個 Stage 考慮訊息在某一個節點上的送出(send)，第二個 Stage 考慮某數個節點送出同一個訊息到某一個單一節點(receive)。

第一個 Stage 的運作方式如下：每個訊息擁有一個記錄稱之為 BCTL (broadcast travel list) 用以記錄該訊息最近曾經造訪的節點以及接下來欲造訪的節點，當該訊息傳遞到某一個節點時，該節點比對該訊息的 BCTL 與該節點的鄰近節點(neighbor)，以過濾出將會重複傳遞的節點，並加以略過。第二 Stage 的運作方式如下：若兩個節點皆欲傳送同一個訊息給共同的一個節點時，兩個節點將會自行進行彼此 ID 的比較，由較大的節點負責進行傳送即可。

我們設計了一個可供觀察的項目稱之為 AMOR (average messages originates form one requetst)，即一個 request 在進行廣播(broadcast)傳遞中平均會產生多少個訊息(message)。我們希望觀察 BCTL 的容量與 AMOR 之間在 SMRB Stage 1 的關係，其結果如 Figure 3 與 4。我們可以看到 Stage1(參考 Figure 3)下，AMOR 大幅降低(由藍色曲線變成其他顏色的曲線)。至於 Stage 2(參考 Figure 4)方面則可以將 Stage1 中的 AMOR 再降低將近一半左右。

3. 保持堅固的網路連結性

一個沒有中央伺服器的 P2P 網路環境中，節點(Node)之間的斷線(Disconnection)是不可避免的情形，斷線可能發生在節點的下線(Logout)、區域網路的局部故障、甚至是大規模的網路骨幹崩潰等等皆有可能。在網路斷線的情形發生下，互相連結而成的 P2P 網路有可能發生碎片(Fragment)的情形，也就是整個系統形成不相連(Disconnection)的許多碎片。當系統產生碎片時，整體的搜尋效能將大幅下降，若整體網路斷裂成 n 個不相連且大小均等的碎片時，搜尋效能將驟降為原本的 1/n。我們的 DSE 系統具有因應此問題的容錯機制，當網路發生斷線時，會自動嘗試將碎片重新連結起來。

我們的演算法稱之為 RAL(Ring Around Leader) 演算法，每個節點內部持續地執行 RAL 演算法。先前說過，每個訊息(Message)中包含了該訊息的來源節點的 ID，RAL 演算法將節點本身的 ID 與所有經過該節點的訊息中的來源 ID 加以比較，取出最大值，並將該值散播給鄰近節點。鄰近節點獲得該值後，亦與本身所計算的 ID 最大值作比較，取其大者，並且亦將該值散播給鄰近節點。在各節點的相互合

作下,所有節點將可獲得一最大的 ID 值,擁有該值的節點即為整體網路的 Leader 節點。另外,在散播 ID 值的過程中,屬於該 ID 值的節點的 IP Address 以及當時的時間標籤(Timestamp)亦附加於其 ID 之後,我們稱此一特別的字串(ID + IP + Timestamp)為 Component Identifier (CID)。在一個連通的系統中,所有節點最終都將獲得 Leader 節點的 IP Address。

另外,當某一節點得知自己為 Leader 節點時,將立刻通知其鄰居節點(Neighbor Node)進行互相連結,藉以形成一個環狀結構(Ring)。而 Leader 節點會定時產生新的 CID 並且加以散播給其他節點。

當某一節點發現本身的 CID 的 Timestamp 太久沒有變動時,即表示整體系統可能已經發生碎片,以致於新的 CID 無法傳遞過來,因此該節點則立即嘗試連結 Leader 節點。

四、成果自評

我們已經發展出一個純粹 P2P 環境的基礎建設,也就是 Self-Organization 機制,本機制提供了三項功能: 網路結構的自動最佳化、智慧型訊息傳送機制 以及 保持堅固的網路連結性。利用這相機制,我們已經可以在完全沒有伺服器的狀況下,自動維護整個網路系統的連結性、並且讓節點根據彼此之間的距離而自動群聚化,讓訊息傳遞使用低成本的路徑,而智慧型訊息傳送機制則可以大幅降低訊息傳遞中所發生的多餘傳遞次數,除去不必要的頻寬耗費。

在未來 Client-Server 架構逐漸無法應付日漸增多的網路服務的情況下,Peer-to-peer 架構已經成為新一代的解決方案,但是如何將採用 Client-Server 架構的眾多服務逐漸改用 Peer-to-peer 的運作方式則將會是下一個必須解決的問題。

五、參考文獻

[1] Karl Aberer, Magdalena Punceva, Manfred Hauswirth and Roman Schmidt. Improving Data Access in P2P. IEEE Internet Computing, 6(1), 2002.

[2] Andy Oram. Peer-to-Peer Harnessing the Power of Distributed Technologies. O'Reilly 2001.

[3] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of `small-world' networks. Nature, vol. 363, pp. 202-204.

[4] Sylvia Ratnasamy, Scott Shenker and Ion Stoica. Routing Algorithms for DHTs: Some Open Questions. First International Workshop on Peer-to-Peer Systems (IPTPS), 2002.

[5] Matei Ripeanu, Ian Foster and Adriana Iamnitchi. Mapping the Gnutella Network: Properties of Large-Scale Peer-to-Peer Systems and Implications for System Design. IEEE Internet Computing Journal, 6(1), 2002.

[6] Ian Clarke, Oskar Sandberg, Brandon Wiley and Theodore W. Hong. Freenet: A Distributed Anonymous Information Storage and Retrieval System. Lecture Notes in Computer Science, vol. 2009, pp. 46+ ,2001.

[7] Ion Stoica, Robert Morris, David Karger, M. Frans Kaashoek, Hari Balakrishnan. Chord: A Scalable Peertopeer Lookup Service for Internet Applications. Technical Report TR-819, MIT, March 2001.

[8] Kunwadee Sripanidkulchai. The popularity of Gnutella queries and its implications on scalability. The O'Reilly Peer-to-Peer and Web Services Conference, September 2001.

[9] Andy Oram. Peer-to-Peer Harnessing the Power of Distributed Technologies. O'Reilly 2001. pp. 94-122.

[10] Proceedings of Designing Privacy Enhancing Technologies: Workshop on Design Issues in Anonymity and Unobservability, July 2000.

[11] The Free Network Project. http://freenetproject.org/

[12] The free haven project. http://freehaven.net/.

[13] References H. Zhang, A. Goel, R. Govindan, Using the Small-World Model to Improve Freenet Performance, Proceedings of IEEE Infocom, 2002. 14.

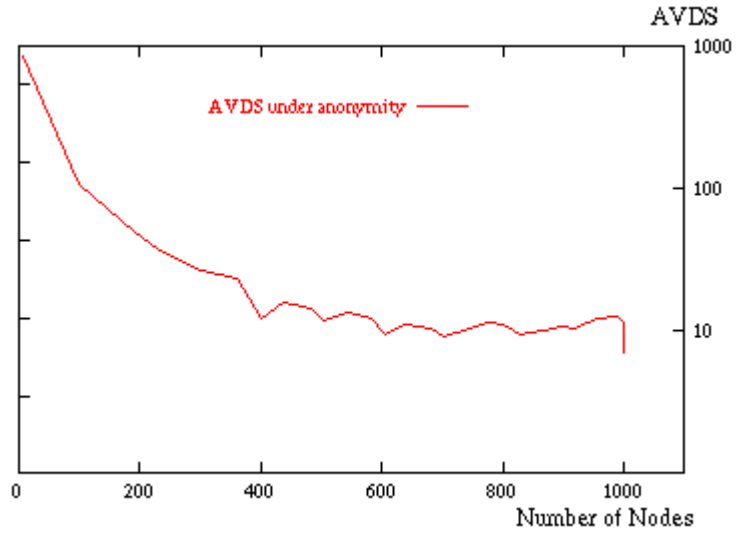

AVDS
1000
100
10
AVDS under anonymity
0
200
400
600
800
1000
Number of Nodes

Figure 1.

```
function insertNBR(node N) {
    insert(NBRQ, N);                    // insert the new node N into the set NBRQ
    };

function NCC() {
    if (sizeof(NBRQ)<NBUB) return;
    Q=sort_by_distance(NBRQ);           // sort the NBRQ by their distances (from local node)
    R=0.98;                             // 98% are closer nodes
    S=get_closer_nodes(Q, NBLB*R);      // get closer nodes from Q (e.g. 98%)
    Q=Q-S;
    V=get_random_nodes(Q, NBLB*(1-R)); // get nodes from Q randomly (e.g. 2%)
    NBRQ=S \bigcup    V;
    };
```

Figure 2.

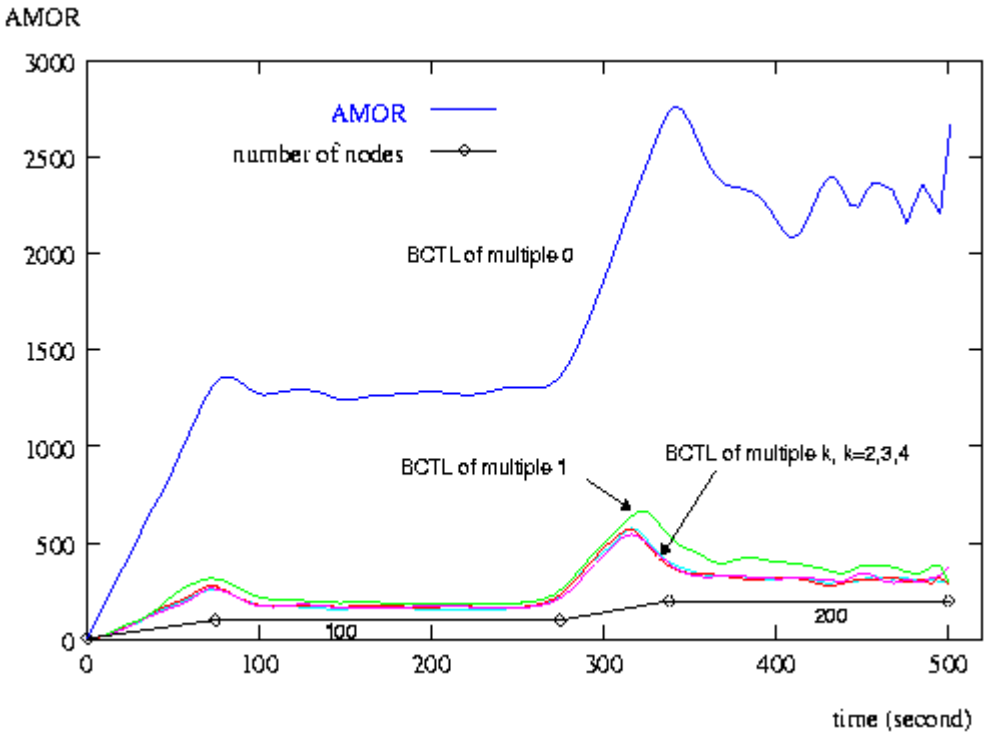


Figure 3.

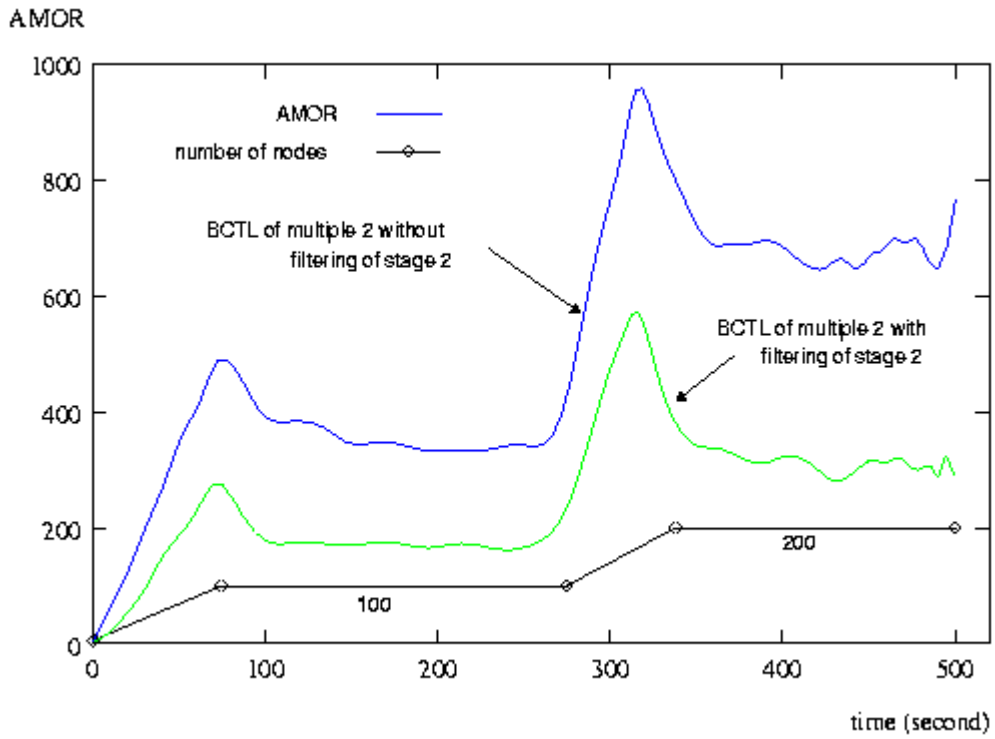
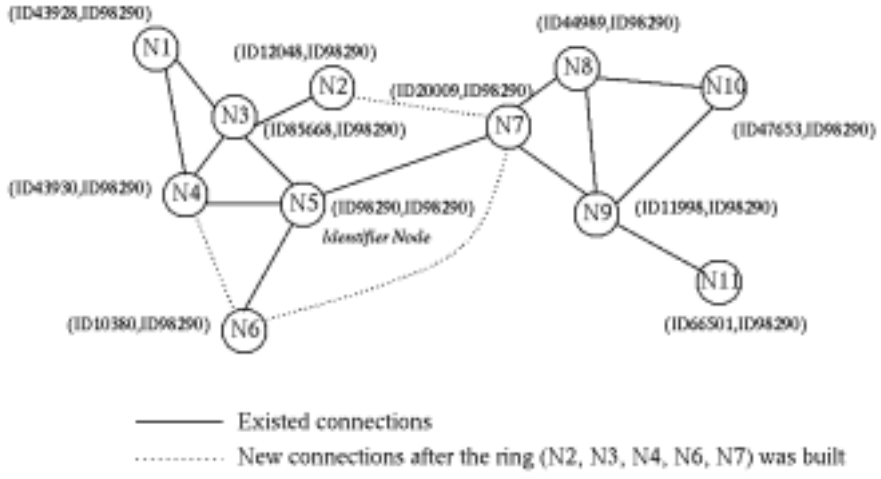


Figure 4.



(圖中的 Identifier Node N5 即為 Leader 節點，(N2,N3,N4,N6,N7) 形成一個環狀)

Figure 5.