

行政院國家科學委員會專題研究計畫 成果報告

子計畫二：行動電子商務系統中之資訊勘測與資訊存取演算法之研發與實作(3/3)

計畫類別：整合型計畫

計畫編號：NSC92-2213-E-002-010-

執行期間：92年08月01日至93年07月31日

執行單位：國立臺灣大學電機工程學系暨研究所

計畫主持人：陳銘憲

報告類型：完整報告

處理方式：本計畫可公開查詢

中 華 民 國 93 年 12 月 13 日

行政院國家科學委員會專題研究計畫成果報告

行動電子商務系統關鍵技術之研發與實作 - 子計畫二：

行動電子商務系統中之資訊勘測與資訊存取演算法之研發與實作

Design and Implementation of Data Mining and Data Retrieval Algorithms in a
Mobile Commerce System

計畫編號：NSC 92-2213-E-002-010

執行期限：92 年 8 月 1 日至 93 年 7 月 31 日

主持人：陳銘憲教授 臺灣大學電機工程學系暨研究所

一、中文摘要

本計畫為整合型計畫『行動電子商務系統關鍵技術之研發與實作』之一子計畫『行動電子商務系統中之資訊勘測與資訊存取演算法之研發與實作』之第三年計畫。目前新一代行動商務技術正在研發，未來的行動商務環境將能夠提供許多個人化的商品消費和查詢服務。許多研究中顯示，使用者的移動是有其規律性的。善用使用者移動的規律性能夠花費較少的代價，獲取提昇行動通訊網路服務品質的大幅提昇。在第一年的研究中，我們於行動電子商務系統中研發了使用者行為搜集機制及設計資訊勘測的演算法以分析使用者在行動電子商務之行為模式。在第二年的研究中，我們探討如何利用使用者移動模式來決定個人化資料與共享資料配置機制以提昇行動電子商務系統的服務品質，以及使用者對行動電子商務的接受度。根據找出的使用者之移動行為模式，我們針對個人化資料與共享資料提出不同的資料配置演算法，並以系統模擬來評估這些方法的效能與特性。

而在本年度的研究中，我們著眼於使用行動電子商務之客戶，在存取其資料庫時之行為模式。由於行動電子商務之客戶，所使用之無線網路，具有頻寬之限制，根

據使用者存取資料的特性，決定相關資料該如何散播，可以有效的降低使用者等待之時間。我們針對使用者存取的特性，將資料劃分為熱門與冷門，分別以不同的機制傳送給使用者，實驗結果顯示，我們所提之演算法，可有效的降低使用者所等候之時間。

關鍵詞：行動電子商務、資訊勘測、資料配置、使用者移動行為模式、使用者之存取模式、資料散播

英文摘要

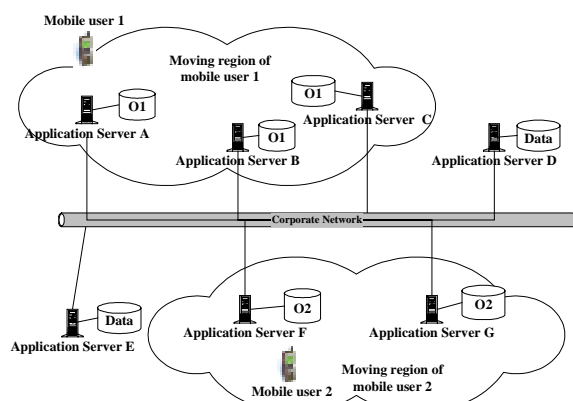
As a sub-project of the joint project entitled "A Mobile Electronic Commerce System," this project, named "Design and Implementation of Data Mining and Data Retrieval Algorithms in a Mobile Commerce System," attempts to design data allocation schemes for personal and shared data to enable mobile E-commerce. We will carry out this project in a three-year span.

As addressed in other studies, the moving behavior of users is regular, and the quality of service of the mobile commerce system can be greatly improved by employing the user moving patterns with slight trade-off. In the first year, we had

designed a user moving pattern analysis algorithm to capture the user moving patterns from the historical user moving logs. In the second year, we develop data allocation schemes for personal and shared data to increase the quality of service according to the obtained user moving patterns. In this year, we focus on the access pattern of the client in a mobile Commerce System. First, we devise an efficient algorithm to partition the mobile database into popular and unpopular. Moreover, we propose an effective algorithm to disseminate the popular and unpopular data items in different approaches. The experimental results show that the average waiting time of the mobile users can be significantly improved. **Keywords:** Mobile commerce, data allocation, data mining, data allocation, user moving pattern, user accessing pattern, data dissemination

二、計劃緣由與目的

隨著無線網路與網際網路技術的成熟，許多國內外之廠商紛紛投入行動商務服務之領域 [7,11]。目前新一代行動商務技術正在研發，未來的行動商務環境（如圖一所示）將能夠提供許多個人化的商品消費和查詢服務 [2,3,5,9]。

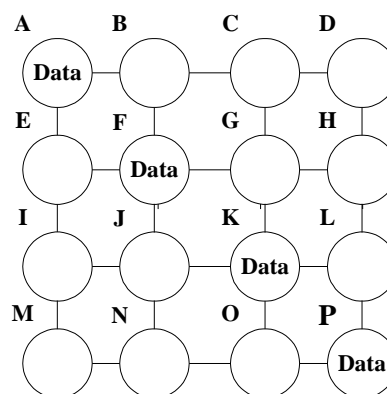


圖一. 行動資訊系統架構圖

大體而言，使用者的移動具有相當強烈的規律性，在學術界的研究中，常用 Markov Chain 或 Round-trip Model 來描述使用者的移動行為模式。在 [6] 的研究中指出，利用使用者移動的規律性將可以花費較小的代價以獲取系統服務品質的大幅度提昇。這些資訊之勘測將有助於系統資源的規畫與提供更多的加值性服務。因此，在上年度的研究中，我們利用所搜集的使用者移動行為資料進行資訊勘測 [1, 3]，並設計出一資訊勘測演算法來分析出使用者的移動行為模式 (User Moving Pattern) [8]。

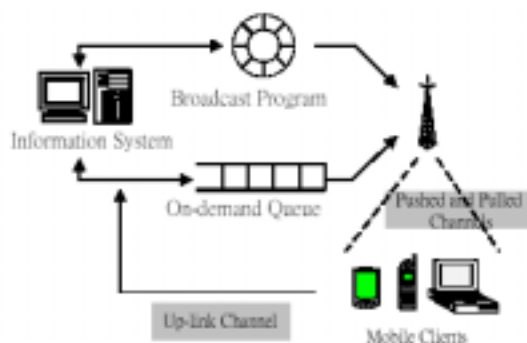
圖二是一個資料配置的示意圖。如果我們得知大部分的使用者傾向在 Cell A, F, K 和 P 中移動，我們應該將資料複製成四份並分別放在 Cell A, F, K 和 P 中。如此一來，大部份的使用者將能夠快速地存取

資料。從這個例子中我們發現，資料的配置方式將大大地影響使用者存取資料的時間。



圖二、資料配置示意圖

除了使用者的移動具有強烈的規律性外，在同一個區域內，從使用該無線網路之全部使用者，所表現出來存取資料之行為模式，亦可找出相關資料被使用的特性。許多研究顯示，觀察所有使用者的存取模式，根據不同筆的資料作圖，我們可以得到一個近似 Zipf Distribution [17] 的機率分佈。我們本年度所研究的內容，便是根據此一資料使用之機率分佈，設計出一套有效之資料散播演算法。



圖三 個人化及共享資料之散播

在本年度的研究中，沿續之前的研究，我們將資料區分為個人化資料與共享資料兩種。接著分別針對這兩種資料，觀

察其被使用者所存取之頻率。我們發現到，個人化之資料，由於較具獨特性，因此，使用者之存取頻率較少，而共享資料，較具共通性，被存取的頻率相對而言，則高於個人化資料。因此，給定有限的無線網路頻寬，如何將個人化資料及共享資料作有效的散播，則成為一個重要的課題。在最後一年的研究過程中，我們提出更有效的方法，來區分個人化資料及共享資料，如圖三所示，我們將共享資料利用資料廣播 (Broadcast) 的方式傳送 [12]，而個人化資料，則是根據使用者所送之要求訊息，來作即使處理 (On-demand)；接著，根據不同種類的資料，我們預測出使用者使用在存取資料時，所可能造成的等候的時間。將此一等候時間回授給資料配置系統作參考，當系統達到平衡時，我們便會得到一個最佳資料配置之方式。同時，有鑑於在未來的行動網路中，多媒體資料將會普遍地被使用，因此我們所提出的方法，都假設其資料大小是可變的，因此，相較於傳統之資料散播演算法而言，我們所提出的方法，會更具延展性以及擴充性。

三、研究方法

在傳統的資料散播的演算法中，每一筆資料都假設為固定大小。因此，在傳統的資料散播環境中，當共享資料採用 Broadcast 的方式作散播時，由下列公式，我們可以預測出使用者存取共享資料的平均時間：

$$W_b = \frac{nz}{2bc_{push}} + \frac{z}{b}$$

其中，n 和 z 所代表的，分別是共享資料的數量，和每一筆共享資料的資料大小；而 c 和 b 所代表的，則是傳送共享資料的頻道數目和頻寬大小。

至於個人化資料，在傳統的資料散播

的環境中，由於它們是採用 On-Demand 的方式作傳播，我們可以將其看成是一個 M/M/C Queue，根據 M/M/C Queue 的特性，我們可以預測出使用者使用個人化資料的平均時間如下：

$$W_{oM/M/c} = \frac{1}{\mu} + \left(\frac{r^{c_{pull}}}{c_{pull}!(c_{pull}\mu)(1-\rho)^2} \right) p_c, \text{ where}$$

$$\mu \approx \frac{b}{z}, \rho = \frac{\lambda_{pull}}{c_{pull}\mu}, r = \frac{\lambda_{pull}}{\mu}, \text{ and}$$

$$p_c = \left(\sum_{n=0}^{c_{pull}-1} \frac{r^n}{n!} + \frac{r^{c_{pull}}}{((1-\rho)c_{pull}!) } \right)^{-1}$$

根據共享資料以及個人化資料的存取時間，我們可以決定出傳播共享資料的資料量以及頻道數，來使得使用者的平均存取時間降至最低。

然而，傳統的資料散播環境，所假設的是每筆資料的資料大小相同。但對日新月異的行動通訊技術而言，在未來的資料散播環境中，每一筆的資料，由於可能包括圖片，聲音甚至視訊等多媒體影像，每一筆的資料，其大小將會不盡相同。若使用傳統演算法作資料散播，則勢必會降低系統的效益以及增加使用者存取資料的時間。有鑑於此，我們所提出的 GDS 演算法，便將多媒體資料之特性列入考慮，將傳統演算法作大幅度之修改，以滿足未來環境中，系統以及使用者之需求。

和傳統的演算法 [13],[16],[18] 相比，GDS 仍然是將共享資料以 Broadcast 的方式作傳送，而個人化資料，同樣的也是以 On-demand 的方式作傳送，但不同於傳統演算法而言，我們所採用的預測使用者存取共享資料之方式，如下列公式所示：

$$W_b = E[W_b^{(i)}]$$

$$= \sum_{i=k+1}^M p_i W_b^{(i)}$$

$$= \frac{\left(\sum_{j=k+1}^M z_j \right)}{2bc_{push}} + \sum_{i=k+1}^M p_i \frac{z_i}{b}$$

其中 Z_i 和 P_i 所代表的，分別是每一筆共享資料的資料大小和使用機率。至於個人化資料，在考慮每筆資料大小皆不相同的環境下，原本的 M/M/C Queue 的模式將不再適用，因此，我們在預測個人化資料時，所採用的，是 M/G/C Queue 行為模式，M/G/C Queue [14], [15] 的預測方式，如下列公式所示：

$$W_{oM/G/c} = E[S] + W_{qM/G/c}$$

$$\approx E[S] + \frac{Q_{M/M/c} E[S^2]}{2E^2[S] \lambda_{pull}}$$

其中 $E[S]$ 和 $E[S^2]$ 所代表的，分別是 M/G/C queue 的服務時間和變異數，而 $Q_{M/M/c}$ 所代表的，則是相對應的 M/M/C Queue 中所等待的資料佇列的大小。

Scheme GDS:

```

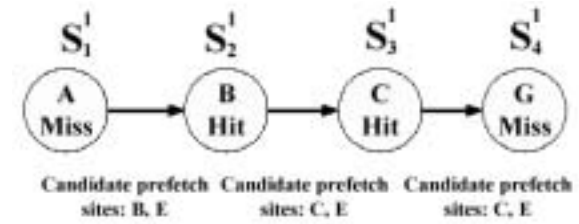
Input:  $\{D\}, c$ 
Output:  $U_{push}^*, U_{pull}^*, c_{push}^*, c_{pull}^*, \{U_i | 1 \leq i \leq c_{push}^*\}$ 
1. begin
2.  $GlobalMinTime \leftarrow \infty$ 
3.  $U_{push}^* \leftarrow \{\emptyset\}, U_{pull}^* \leftarrow \{\emptyset\}$ 
4.  $c_{push}^* \leftarrow 0, c_{pull}^* \leftarrow 0$ 
5.  $c_{push}^* \leftarrow c, c_{pull}^* \leftarrow 0$ 
6. while  $c_{push}^* \geq 0$ 
7.   if  $c_{push}^* == c$ 
8.      $U_{push}^* \leftarrow \{D\}, U_{pull}^* \leftarrow \{\emptyset\}$ 
9.      $LocalMinTime \leftarrow EvaluatePushTime(U_{push}^*, c_{push}^*)$ 
10.  else if  $c_{pull}^* == c$ 
11.     $U_{push}^* \leftarrow \{\emptyset\}, U_{pull}^* \leftarrow \{D\}$ 
12.     $LocalMinTime \leftarrow EvaluatePullTime(U_{pull}^*, c_{pull}^*)$ 
13.  else
14.     $(U_{push}^*, U_{pull}^*) \leftarrow DataClassification(\{D\}, c_{push}^*, c_{pull}^*)$ 
15.     $t_{push} \leftarrow EvaluatePushTime(U_{push}^*, c_{push}^*)$ 
16.     $t_{pull} \leftarrow EvaluatePullTime(U_{pull}^*, c_{pull}^*)$ 
17.     $t \leftarrow EvaluateTime(t_{push}, t_{pull}, U_{push}^*, U_{pull}^*)$ 
18.     $LocalMinTime \leftarrow t$ 
19.  end if
20.  if  $LocalMinTime < GlobalMinTime$ 
21.     $U_{push}^* \leftarrow U_{push}^*, U_{pull}^* \leftarrow U_{pull}^*$ 
22.     $c_{push}^* \leftarrow c_{push}^*, c_{pull}^* \leftarrow c_{pull}^*$ 
23.  end if
24.   $c_{push}^* \leftarrow c_{push}^* - 1, c_{pull}^* \leftarrow c_{pull}^* + 1$ 
25.  continue
26. end while
27.  $\{U_i\} \leftarrow MinErrorScheduling(U_{push}^*, c_{push}^*)$ 
28. return  $U_{push}^*, U_{pull}^*, c_{push}^*, c_{pull}^*, \{U_i\}$ 
29. end

```

圖四、GDS 演算法之內容

圖四所條列出來的，是 GDS 演算法的內容。根據前兩筆公式，GDS 可以反覆的預測使用者的平均使用時間，同時，預測的結果也會加以回授，根據回授的結果，GDS 亦可將效能作進一步之修改，當系統取得平衡時，GDS 所達到的，則是一個近乎最佳值的使用者等候時間。

四、結果與討論

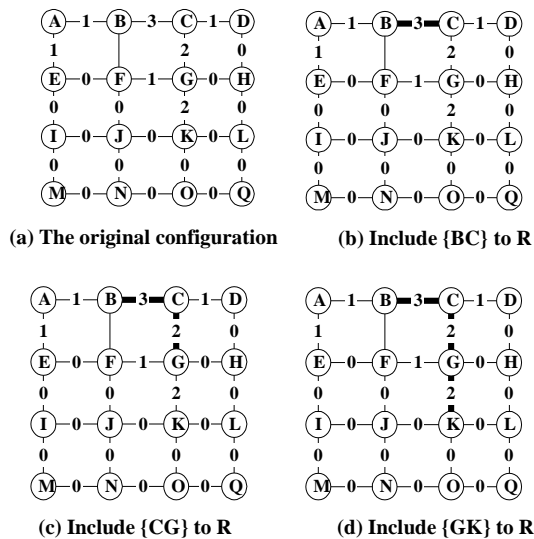


圖五、DP 演算法之範例

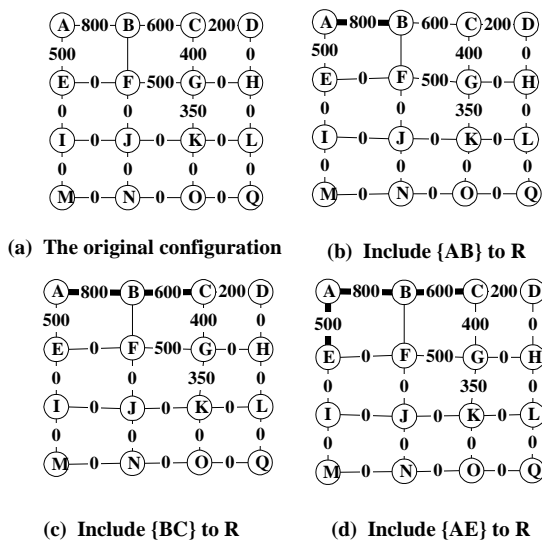
在前一年的執行成果中，針對個人化資料，我們提出了一個套用使用者移動行為模式的個人化資料 Prefetch 演算法-DP [8]。我們用以下的一個範例來描述 DP 的執行過程。假設我們分析出使用者 U_1 有兩個移動行為模式： $\{AE\}$ 和 $\{ABC\}$ 。因此，當使用者在 Cell A 時，我們可以從 U_1 的移動行為模式中得知 U_1 的下一個進入的 Cell 可能為 Cell B 和 Cell E。因此我們預先將 U_1 的個人化資料 Prefetch 至 Cell B 與 E 中。當使用者移動到 Cell B 後，我們可以根據 U_1 的下一個可能進入的 Cell 來 Prefetch U_1 的個人化資料 Prefetch。DP 演算法的執行範例如圖五所示。

針對共享資料，我們提出了兩個資料配置演算法：SD-Local 和 SD-Global [10]。在 SD-Local 演算法中，首先先將使用者的移動行為模式切成數個 Cell 間的移動。接著計算每個相鄰的 Cell 間有移動的使用者人數，並以使用者人數為此移動的權重值。以圖六為例，其中 AB 的權重值為 1 代表有一個使用者的移動行為模式中有

從 Cell A 到 Cell B 或從 Cell B 到 Cell A 的移動。SD-Global 演算法的流程和 SD-Local 演算法的流程大致相同。如圖七所示，不同之處在於在 SD-Global 演算法中，相鄰 Cell 移動的權重值是此移動在使用者移動行為模式之出現的次數。



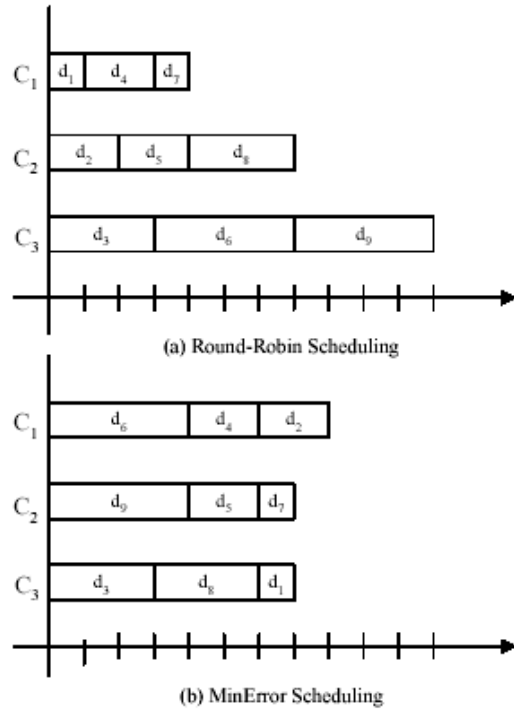
圖六、SD-Local 演算法的執行範例



圖七、SD-Global 演算法的執行範例

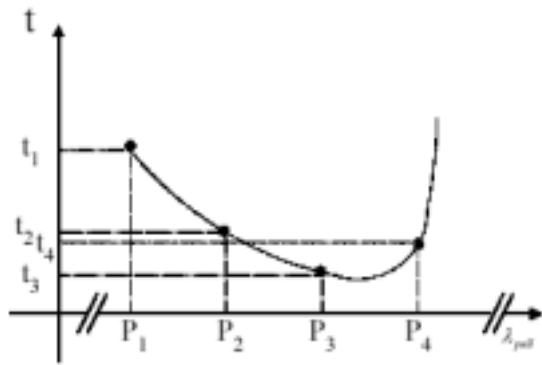
至於我們在最後一年所提出的 GDS 演算法，則是以第二年所提出的 DP 演算法 SD-Local 演算法，和 SD-Global 演算法

為基礎，所作的資料配置。上述演算法，是將不同的資料指派給不同區域的基地台，而 GDS 之演算法，則是將指派給某一特定區域之基地台的資料，作適當之頻寬分配。相關執行，仍有一些議題需加以考慮，其執行實例如下：

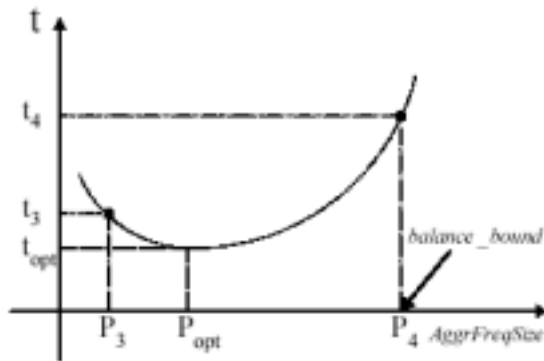


圖七、共享資料之頻道分配

對共享資料而言，雖然我們能夠用公式準確預測出共享資料的平均使用時間，然而，不同的資料排程方式，對預測的效果，也會造成不同的影響，圖七所振示的，便是我們所提出來的排程效果，和其他的排程效果，所造成的差異，對一般的排程效果而言，其演算法不會考慮到資料大小的差異，只會考慮到資料輸入之順序；而我們所提出來的 MinError Scheduling 的排程方式，則將資料輸入順序列入考慮，因此，每個頻道所造成的差異將會減到最低，而能準確地預測出共享資料的存取時間。



(a)



(b)

圖八、GDS 之最佳化問題

接下來所要探討的，是 GDS 執行時，所解決最佳化問題的方法。GDS 的執行過程，可約略分成兩個層次作調整：即粗調和微調，以達到最佳化的目的。在粗調部份，如圖八之(a)所示，GDS 採用的是以使用者的使用頻率，將資料和頻寬作分配，而在微調部份，GDS 會根據粗調所得到的結果，進一步將料重新分配，如圖八之(b)所示。在粗調的過程中，我們可以找出，使用者等候時間的最小值，會落在 P3 和 P4 兩點之間，而在微調的過程中，我們則根據粗調所找出的範圍，作進一步的搜尋，而得到具有最小等候時間的 P_{opt} 點。

五、結果自評

本子計畫的研究人員透過這個計畫，學習新的行動計算網路的應用技術，在行動計算網路的環境上利用行動者移

動模式針對個人化資料與共享資料設計資料配置演算法。我們由研究發展中所得到的經驗和成果可以提供行動網路上相關資訊系統參考使用。我們可以預期行動電子商務資訊系統之重要性必將與日俱增，而藉由電信國家型計畫之執行，我們深信這可為國家實驗網路提供同時深具前瞻技術與實用價值之成果。

五、參考文獻

- [1] R. Agrawal and R. Srikant, "Mining sequential Patterns," *Proceedings of the 11th International Conference on Data Engineering*, March 1995.
- [2] C.-Y. Chang and M.-S. Chen, "On Exploring Aggregate Effect for Efficient Cache Replacement in Transcoding Proxies," *IEEE Trans. on Parallel and Distributed Systems*, Vol. 14, No. 7, July 2003.
- [3] M.-S. Chen, J. Han, and P. S. Yu. "Data Mining: An Overview from a Database Perspective." *IEEE Transactions on Knowledge and Data Engineering*, Vol. 8, No. 6, pp. 866-883, December 1996.
- [4] M.-S. Chen, P. S. Yu and K.-L. Wu, "Optimizing Index Allocation for Sequential Data Broadcasting in Wireless Mobile Computing," *IEEE Trans. on Knowledge and Data Engineering*, Vol. 15, No. 1, pp. 161-173, February 2003.
- [5] C.-H. Lee and M.-S. Chen, "Processing Distributed Mobile Queries with Interleaved Remote Mobile Joins," *IEEE Trans. on Computers*, Vol. 51, No. 10, pp. 1182-1195, October 2002.
- [6] D. A. Levine, I. F. Akyildiz, and M. Naghshineh, "A Resource Estimation

- and Call Admission Algorithms for Wireless Multimedia Networks Using the Shadow Cluster Concept,” *IEEE/ACM Transactions on Networking*, Feb. 1997, vol. 5, No. 1.
- [7] F. Muller-Veerse, “Mobile Commerce Report,” Durlacher Corp., London, <http://www.durlacher.com/downloads/mcomreport.pdf>.
- [8] W.-C. Peng and M.-S. Chen, “Developing Data Allocation Schemes by Incremental Mining of User Moving Patterns in a Mobile Computing System,” *IEEE Trans. on Knowledge and Data Engineering*, Vol. 15, No. 1, pp. 70-85, February 2003.
- [9] W.-C. Peng and M.-S. Chen, “Efficient Channel Allocation Tree Generation for Data Broadcasting in a Mobile Computing Environment,” accepted by *ACM Wireless Networks*, 2002.
- [10] W.-C. Peng and M.-S. Chen, “Exploring User Moving Patterns to Improve the Allocation of Shared Data in a Mobile Computing Environment,” *Proc. of the Thrid International Conference on Mobile Data Management (MDM-02)*, January 8 - 10, 2002.
- [11] U. Varshney, R. J. Vetter and R. Kalakota, “Mobile Commerce: A New Frontier,” *IEEE Computer*, October 2000.
- [12] S. Acharya, R. Alonso, M. J. Franklin, and S. B. Zdonik, “Broadcast Disks: Data Management for Asymmetric Communications Environments” *In Proc. of ACM SIGMOD*, 1995.
- [13] C.-L. Hu, and M.-S. Chen, “Adaptive Information Dissemination” *In Proc. of IEEE ICC*, 2002.
- [14] C.-L. Wang, and R. W. Wolff, “The M/G/C queue in light traffic.” *Queueing Systems: Theory and Applications*, 1998.
- [15] D.-D. Yao, “Refining the Diffusion Approximation for the M/G/C Queue,” *IEEE Computer*, October 2000.
- [16] Xu, J., Lee, D.L., and Li, B “On Bandwidth Allocation for Data Dissemination in Cellular Mobile Networks” *ACM WINET*, 2003
- [17] G. K. Zipf “Human Behaviour and the Principle of Least Effort” *Addison-Wesley*, 1949
- [18] S. Acharya, M. J. Franklin, and S. B. Zdonik. “Balancing push and pull for data broadcast” *In Proc. of ACM SIGMOD*, 1997