

行政院國家科學委員會專題研究計畫 成果報告

行動計算環境中用戶端快取新架構之研發

計畫類別：個別型計畫

計畫編號：NSC91-2213-E-032-013-

執行期間：91年08月01日至92年07月31日

執行單位：淡江大學電機工程學系(所)

計畫主持人：莊博任

計畫參與人員：徐敬岳、邱育賢、趙惇豪、李柏逸

報告類型：精簡報告

處理方式：本計畫可公開查詢

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

行動計算環境中用戶端快取新架構之研發

The Design of Cache Invalidation Schemes in Mobile Computing Environments

計畫編號：NSC 91-2213-E-032-013

執行期限：91年8月1日至92年7月31日

主持人：莊博任 淡江大學電機系

計畫參與人員：徐敬岳、邱育賢、趙惇豪、李柏逸
淡江大學電機系

一、中文摘要

在行動計算環境之下，為了有效減少資料存取的時間及降低頻寬的使用量，因此在行動用戶端採取了快取暫存的機制。在伺服器與行動用戶端各保有資料的情況下，一旦伺服器將資料異動，則行動用戶端的資料便會產生不一致的情況。為確保行動用戶端快取資料項與伺服器資料維持一致性，可以從伺服器廣播無效驗證報告至行動用戶端，行動用戶端利用此報告對快取資料項做有效性驗證。

快取無效策略的效能由於伺服器與行動用戶端經常性斷線（為了節省頻寬與電源的使用）而深受影響。理想的快取無效策略必須有效率地使用有限的頻寬資源，使伺服器與行動用戶端重新連線後，猶能夠維持兩者之間資料項的一致性。

本研究計畫的重點即在參考多種現有快取無效策略後，去其缺失，擇其優點，加上自己的設計，產生更具效能的新快取無效策略。我們的新快取無效策略擁有相當精簡的無效驗證報告，利用非週期性廣播無效驗證報告與上傳 T_{ib} 值至伺服器的機制，加強系統的抗斷線能力，不會像其他快取無效策略般產生不必要的快取資料項無效設定，既節省頻寬的使用，又達到縮減資料取得時間的目的，其效能表現較之現有策略優越，成本且較低。

關鍵詞：行動計算環境、伺服器與用戶端、快取無效、正確性（一致性）、頻寬需求、通訊成本、模擬及效能評估。

Abstract

In a mobile computing environment, caching data items at the mobile clients is important as it reduces the data access time and bandwidth utilization. While caching is desirable, it may cause data inconsistency between the server and the mobile clients if their communication is disconnected for a period of time. To ensure information coherence between the cached items and their source items, the server can broadcast invalidation reports to the mobile clients, who then use the reports to update the cached data items.

Cache invalidation is indeed an effective approach to maintaining such data coherence. After careful investigation on the advantages and disadvantages of existing cache invalidation strategies, this research presents a new cache invalidation strategy with different designs and improved performance. Experimental evaluation shows that the new strategy is able to maintain data consistency between the server and mobile clients more effectively and efficiently than the other strategies.

Keywords: Mobile computing environments, servers and clients, cache invalidation, validity (consistency or coherency), bandwidth requirement, communication cost, simulation and performance evaluation..

二、計劃緣由與目的

在無線通訊計算環境之下，使用者 (user) 可以任意使用掌上型電腦來存取伺服器中的資料，並且不會受到時間與空間的限制。為了有效減少資料項存取的時間及降低頻寬使用量，在行動用戶端 (mobile client) 採取快取暫存 (caching) 的機制。由於伺服器 (server) 與行動用戶端各保有資料項，一旦伺服端的資料項有異動更新，卻未能及時知會行動用戶端時，便可能產生行動用戶端的資料項資訊與伺服端的資料項資訊不一致的情況 (指同一資料項)。

在有基礎架構的無線區域網路的情況下，為了協助行動用戶端確認本身快取記憶體內的資料項是否仍為有效，亦即是否仍與伺服端的資料項資訊一致，一般皆採用快取無效策略 (cache invalidation strategies, 如 [1-10] 等) 進行驗證：從伺服器廣播無效驗證報告 (Invalidation Report, 簡稱 IR) 至行動用戶端，此報告內容告知行動用戶端那些資料項已經更新，用戶端隨即據以對本身快取資料項的有效性進行驗證。驗證結果若發現快取資料項已經過時，便將其設為「無效」，當使用者需要引用此一資料項時，行動用戶端即經由上傳通道 (uplink channel) 向伺服器索取所需要的資料項，伺服器則透過下傳通道 (downlink channel) 將該資料傳送給行動用戶端，以供使用者取用。

快取無效策略的效能深受伺服器與行動用戶端經常性的斷線的影響。這是因為行動用戶端為了節省頻寬與電源的使用量，會暫時中斷與伺服端的連線，斷線期間，行動用戶端會遺失一些伺服器所廣播的無效驗證報告，造成重新連線之後難以判斷快取資料項有效性的困擾。因此，如何在行動用戶端與伺服器重新連線後，有效維持雙方資料項的一致性，值得深入探討，本計畫之研究目的即在詳細探討現有之快取無效策略，充分了解其優缺點，加以截長補短並佐以新的思考與設計，以產生一個成本得以降低、效能卻能超越既有策略的快取無效新策略。

三、結果與討論

我們的新策略包含許多特點，其演算內容如下：

1. 伺服器非週期性廣播無效驗證報告至行動用戶端：當行動用戶端與伺服器保持連線狀態時，伺服器會廣播即時更新的 IR 至行動用戶端，使行動用戶端獲得最新資訊，與伺服端的資料項維持一致性。無效驗證報告的內容包含更新資料項的 ID 及更新時間戳記 T (僅含一個)。收到無效驗證報告的行動用戶端需將本身的 T_{ib} 值改為 T 值。
2. 行動用戶端記錄其 T_{ib} 值：當 T_{ib} 後行動用戶端發生斷線，且重新連線後使用者提出資料項詢問，行動用戶端將 T_{ib} 值與詢問的資料項 ID 上傳至伺服器。
3. 使用者對行動用戶端提出的資料項詢問會有以下兩種情況會發生：
 - (1) 該資料項並沒有儲存於行動用戶端快取記憶體中，必須到伺服器索取資料項供使用者使用。
 - (2) 該資料項儲存於行動用戶端快取記憶體中，由於無法確定其有效性，必須利用無效驗證報告對快取資料項做有效性判斷。此時行動用戶端透過上傳通道發 T_{ib} 至伺服器，由伺服器依據 T_{ib} 判斷行動用戶端遺漏多少無效驗證報告內容。如果判定沒有遺漏 (表示斷線期間內伺服器並沒有做資料項更新)，伺服器不需廣播無效驗證報告至行動用戶端，只需下傳詢問的資料項內容即可；若判定有遺漏，伺服器即必須馬上廣播無效驗證報告與詢問的資料項內容至行動用戶端，供行動用戶端據以逐一檢驗快取資料項的有效性。
4. 當行動用戶端曾經發生斷線情況時，伺服器的無效驗證報告內容為 T_{ib} 後所有的更新資料項的 ID 與到目前為止最後一個更新資料項的時間戳記 T ，此時間戳記即成為行動用戶端新的 T_{ib} 值。

我們利用上傳 T_{ib} 值至伺服器彌補非週期性廣播無效驗證報告的漏洞，也就是使用少量的上傳通道 (上傳 T_{ib} 值)，換取所

有未接收到的 IR 內容。

此外，我們考量到一重要的議題：當伺服器端廣播無效驗證報告給此地區的某一個上傳 T_{ib} 的行動用戶端時，這個地區的所有行動用戶端都會接收到這份無效驗證報告，有可能因此造成混淆，造成其他行動用戶端無法確定這無效驗證報告的可信度。為解決此一問題，我們將所有的資料項加上一個序號（sequence number），這個序號取自資料項更新的次數，也就是資料項出現在無效驗證報告中的次數，採取累加的計算方法，且序號到了某一個程度就會重置（reset）。

這個做法的優點有二：

1. 當其他行動用戶端收到此無效驗證報告時，可以利用它判斷部分快取資料項的正確性。如果收到的無效驗證報告中有資料項也儲存於自己的快取記憶體中且無效驗證報告中的序號大於快取資料項的序號時，表示該快取資料項已經過期，即可將其設為無效，若序號相等，則表示該快取資料項並未過期。
2. 使用簡單的序號為資料項作註記，可以進一步縮小無效驗證報告的內容，因為序號本身所佔的空間很小，與資料項的大小比起來甚至可以忽略不計，有助於節省頻寬。以上說明顯示，我們的新策略不但相當程度地縮減了無效驗證報告內容、減少頻寬的浪費，且同一份無效驗證報告可供多個行動用戶端進行快取資料項有效性的驗證，極具效能優勢。

四、計畫成果自評

綜合言之，我們的快取無效策略具有下列特點：

1. 在未發生斷線的情況下，採用伺服器端非週期性廣播無效驗證報告至各行動用戶端，也就是當伺服器端有資料項更新時，即馬上通知行動用戶端。
2. 若曾有斷線情況且行動用戶端重新連線後有使用者詢問資料項，行動用戶端即上傳 T_{ib} 值至伺服器端，伺服器端隨即廣播 T_{ib} 後所有的更新資料項資訊，供行動用戶端驗證之用。

3. 我們的策略一方面可以讓行動用戶端達到即時驗證快取資料項，另一方面不會將不必要的快取資料項設為無效。
4. 我們採用行動用戶端保存 T_{ib} 值的機制，同一份無效驗證報告可供所有的行動用戶端比對使用。

為了評比我們所設計的快取無效策略與既有之相關策略的效能表現，我們在周延的環境設定下，進行深入的模擬分析。基於以上特殊設計，我們的策略在所模擬的多項效能評比中，表現皆優於其他數種策略，例如我們擁有最簡短的無效驗證報告內容、不會產生不必要的設定無效資料項數、長時間斷線下需至伺服器擷取之資料項數最少、耗費最少頻寬（下傳通道）、以及快取失誤率低等。

本研究計畫所預計達成的目標皆如期圓滿達成，其成果已初步整理成篇並投付國際研討會審理中。我們並已擬定在相關領域更進一步研究的方向與大綱，期望在既有基礎上，繼續努力耕耘，尋求突破。

五、參考文獻

- [1] D. Barbara and T. Imielinski, "Sleepers and Workaholics: Caching Strategies for Mobile Environments," *Proc. 1994 ACM SIGMOD Conf. on Management of Data*, pp. 1-12, 1994.
- [2] D. Barbara and T. Imielinski, "Sleepers and Workaholics: Caching Strategies for Mobile Environments (extend version)," *MOBIDATA: An Interactive Journal of Mobile Computing*, Vol. 1, No. 1, Nov. 1994.
- [3] Q. Hu, and D. L. Lee, "Adaptive cache invalidation methods in mobile environments," *Proc. Int'l Conf. on High Performance Distributed Computing*, 1997, pp. 264-273.
- [4] J. Jing, A. Elmagarmid, A. Helal, and R. Alonso, "Bit-sequences: an adaptive cache invalidation method in mobile client/server environments," *Mobile Networks and Applications*, Vol. 31, No. 2, pp.115-127, 1997.

- [5] K.-L. Tan, J. Cai, and B. C. Ooi, "An evaluation of cache invalidation strategies in wireless environments," *IEEE Tran. on Parallel and Distributed Systems*, Vol. 12, No. 8, pp. 789-807, Aug. 2001.
- [6] E. Chan, C.-H. Yuen, K.-Y. Lam, and H.W Leung, "An adaptive AVI-based cache invalidation scheme for mobile computing systems," *Proc. Int'l Conf. on Database and Expert Systems Applications*, 2000.
- [7] K.-L. Wu, P. S. Yu, and M.-S. Chen, "Energy-efficient caching for wireless mobile computing," *Proc. Int'l Conf. on Data Engineering*, Feb. 1996, pp. 336-343.
- [8] S. H. Nam, I. Y. Chung, and C.-S. Hwang, "An efficient cache invalidation scheme for mobile wireless environments," *Proc. Int'l Conf. on Parallel and Distributed Systems*, 2001, pp.289-296.
- [9] Q. Hu and D. L. Lee, "Cache Algorithms Based on Adaptive Invalidation Reports for Mobile Environments," *Cluster Computing*, Vol. 1, pp. 35-90, 1998.
- [10] A. Kahol, S. Khurana, S. K. S. Gupta, and P. K. Srimani, "A Strategy to Manage Cache Consistency in a Disconnected Distributed Environment," *IEEE Trans. on Parallel and Distributed Systems*, Vol. 12, No. 7, pp. 686-700, July 2001.