

行政院國家科學委員會專題研究計畫 成果報告

行動計算環境下之省電與服務品質感知 Transcoding Proxy 之設計

計畫類別：個別型計畫

計畫編號：NSC94-2218-E-009-036-

執行期間：94年10月01日至95年07月31日

執行單位：國立交通大學資訊工程學系(所)

計畫主持人：黃俊龍

報告類型：精簡報告

處理方式：本計畫可公開查詢

中 華 民 國 95 年 10 月 17 日

行政院國家科學委員會專題研究計畫成果報告

行動計算環境下之省電與服務品質感知 Transcoding Proxy 之設計

A QoS-Aware and Energy-Conserving Transcoding Proxy Using On-demand Data
Broadcasting

計畫編號：NSC94-2218-E-009-036

執行期限：94 年 10 月 1 日至 95 年 7 月 31 日

主持人：黃俊龍教授 交通大學資訊工程學系

一、中文摘要

研究結果指出，在行動資訊系統與行動設備間架設 transcoding proxy 能在少量增加行動資訊系統管理者工作量的情形下，有效地調合行動資訊系統與行動設備之間的不協調(mismatch)。在本計畫中，我們研發了適用於行動計算環境的 transcoding proxy，其主要的成果如下：一是研發了整合 on-demand broadcast 技術之 transcoding proxy 架構：我們已提出一整合資料廣播技術之 transcoding proxy 架構，以提昇 transcoding proxy 的延展性 (scalability) 與系統效能；二為研發了適用於 transcoding proxy 的服務品質提供機制：我們已研發系統效能評估機制，針對評估出之系統效能動態調整版本決定機制 (version decision policy) 與服務允諾控制 (service admission control) 機制之參數，藉以將使用者感受到之服務品質控制在一合理的範圍內；三是研發了省電的資料索引機制：研究實驗結果顯示，在資料廣播環境中加入適當的資料索引能夠有效地減少行動設備的能源消耗。因此我們亦研發適合 transcoding proxy 之資料索引機制，並將其與 transcoding proxy 整合，以減少行動設備的能源消耗。

關鍵詞：Transcoding proxy、資料廣播、QoS、行動計算

英文摘要

Recent advances in wireless communication have greatly changed our daily life. Users are able to access Internet at anytime, from anywhere via various mobile devices. However, the high diversity in the capabilities of various mobile devices such as display capabilities and computation power makes the design of mobile information systems more challenging. A transcoding proxy is placed between a client and an information server to coordinate the mismatch between what the server provides and what the client prefers. Therefore, in this project, we designed a QoS-aware and energy-conserved transcoding proxy. Our primary efforts include (1) designing a transcoding proxy architecture to improve system scalability and performance by integrating the technique of on-demand broadcast, (2) developing a QoS provision mechanism to adjust version decision policy and service admission control scheme according to the estimated system workload to keep the QoS between an acceptable range, and (3) proposing a data indexing method to reduce power consumption of clients by inserting data indices into the broadcast program.

Keywords: Transcoding proxy, data broadcast, QoS, mobile computing

| 版本 | 版本一 | 版本二 | 版本三 |
|------|---|---|---|
| 範例 |  |  |  |
| 圖片規格 | 154x154, 24 bits | 154x154, 4 bits | 77x77, 1 bit |

圖一. 同一物件的不同版本

二、計畫緣由與目的

近年來，無線網路技術的進步已經大大地改變我們的生活。然而相較傳統有線的網路環境，在無線網路環境下開發行動資訊系統將面臨以下幾個挑戰：

1. 有限的網路頻寬
2. 有限的電力容量
3. 行動設備間的異質性
4. 使用者的移動性(mobility)

研究中指出，在無線網路環境中，利用 on-demand broadcast 技術可有效地提高網路頻寬的使用率 (utilization)。透過 on-demand broadcast 技術傳送熱門資料一次，即可滿足大多數使用者的要求，因此 on-demand broadcast 已被認為是建置大型行動資訊系統的重要技術。

此外，為了解決行動設備能力之異質性所產生的問題，行動資訊系統必須能根據使用者所使用的行動設備的特質，對資料進行調整，而這個調整的動作稱為 content adaptation。圖二是一張圖片的不同版本的規格。假設行動資訊系統接到一個要求該圖片的資料要求時，伺服器必須根據發出此資料要求的行動設備的特質，決定回傳圖片的版本。

根據執行版本轉換的場合，transcoding 可大致區分為 client based,

server based 與 proxy based 三類，而 proxy-based transcoding(又稱 transcoding proxy)已被認為是實現 content adaptation 的重要技術。和傳統的 proxy 比較，transcoding proxy 在決定是否快取某筆資料時，除了考慮目前可用空間以及該筆資料的熱門程度外，尚須考慮相同資料不同版本間的轉換關係。

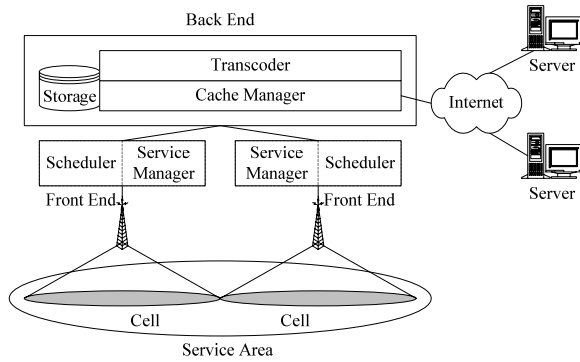
Transcoding proxy 的相關研究議題，包括系統架構、cache replacement 機制、pre-adaptation、協同合作等，已在近幾年被廣泛地討論。然而，目前的研究仍採用傳統 unicast 的資料傳輸模式，因此未能充份利用無線網路的頻寬。此外，先前的研究也未討論服務品質、節省能源消耗與 user mobility 等在行動計算環境中重要的議題。有鑑於此，在本計畫中，我們將著重於省電及服務品質感知的 transcoding proxy 之設計，透過使用 on-demand broadcast 技術以提高網路頻寬的使用效率。

三、問題與結果討論

3.1 系統架構

圖二為本計畫提出之 transcoding proxy 架構 QETP(為 QoS-aware and Energy conserved Transcoding Proxy 的簡稱)。QETP 由 front-end 和 back-end 兩個要素構

成。



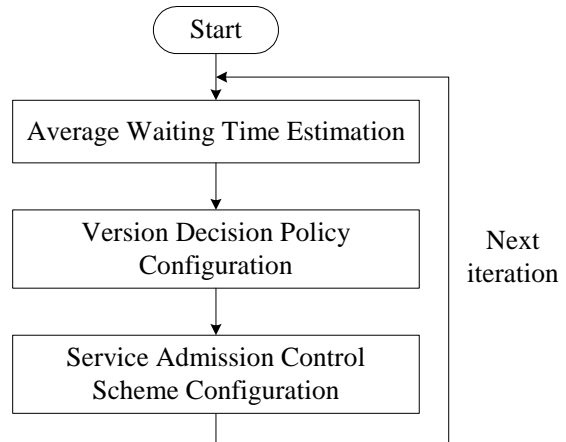
圖二. transcoding proxy 架構

Front-end 架構於每一個 cell 之上，並由 service manager 與 scheduler 所組成，其說明如下：service manager 主管所有關於服務操作的問題，例如：處理服務換手與新使用者註冊等等，以控制服務品質與減少 user mobility 所造成的問題。此外，每一個 service manager 中擁有一個資料庫，紀錄在此 cell 中的使用者簡介以及其所使用的通訊裝置。在 scheduler 方面，當收到一個資料要求時，scheduler 會檢查在資料要求佇列中是否有相同的資料要求得以合併，並且決定所收到資料要求的優先順序，再根據使用者的通訊裝置以及網路狀態，決定適合的版本給使用者。

Back-end 中包含的元素，即 on-demand broadcast 技術中傳統的 cache manager 與 transcoder，其說明如下：當 cache manager 收到一自 scheduler 傳來的資料要求後，cache manager 負責回傳要求的版本給 scheduler，如果此版本已在 cache 中，則立即將答案傳給 scheduler；否則，cache manager 則會檢查(required)版本是否可以被 cache 中的其他版本所產生，若可以，則產生之；否則，cache manager 則會將資料要求提交給 data server，並要求回傳正確的原始版本，進而回應答案給 scheduler。而 transcoder 便根據上述 cache manager 中

的行為做不同版本之間的轉換工作。

3.2 服務品質提供機制

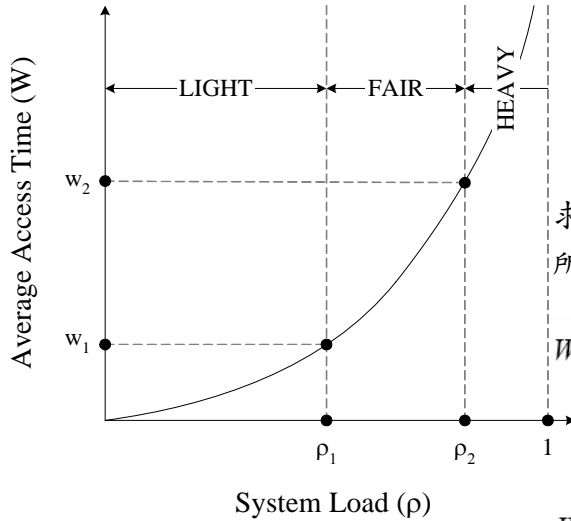


圖三. 服務品質提供機制之流程圖

我們提出的服務品質提供機制為一個週期性執行的演算法，其流程如圖三所示。在第一個步驟中，我們將系統細分為三個元素，並利用數學分析來估算每個元素的平均等待時間，藉此評估系統的負載狀態。而在第二個步驟中，版本決定機制將根據每個元素的負載狀態，來決定版本的提供策略。在最後一個步驟中，服務允諾機制負責管控服務範圍內的使用者數量，以降低因“使用者註冊”及“服務換手”造成服務品質的下降。

在本計畫中，我們採用“每一個資料要求的平均等待時間”做為服務品質的評量標準。本系統會先要求系統管理員先輸入兩個平均等待時間的門檻值： w_1 與 w_2 ，而 QETP 會將使用者的平均等待時間控制在 w_1 與 w_2 之間。圖四為系統負載與平均等待時間的關係圖。從圖四中我們可以觀察到，平均等待時間會隨著系統負載的上升而增加。我們首先將 w_1 與 w_2 轉換成相對應的系統負載 ρ_1 與 ρ_2 ，並藉由 ρ_1 與 ρ_2 將系統負載分成 LIGHT、FAIR 與 HEAVY 三個區域。在每次執行時，QETP 會先利用先前所提出之系統負載評估機制評估系

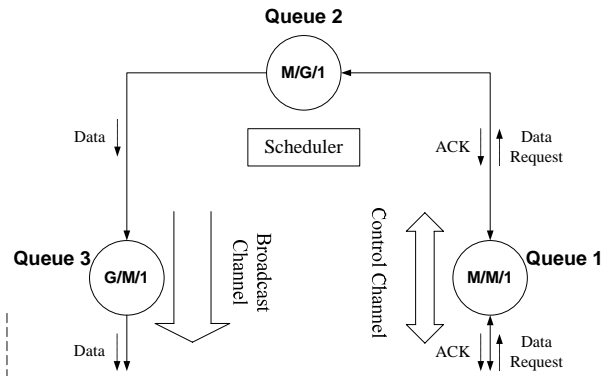
統目前的負載，再藉由其與 ρ_1 、 ρ_2 的關係找出系統負載的狀況(即 LIGHT、FAIR 或 HEAVY)。最後 QETP 再根據系統的負載狀況調整版本決定機制與服務允諾機制的運作，將系統的負載情形控制在 LIGHT 或 FAIR 的狀態下。



圖四.系統負載與平均等待時間的關係圖

3.3 系統負載評估機制

為了評估系統的負載，我們採用如圖五所示的系統負載評估模型。根據先前的研究，我們將上傳頻道 model 成一 M/M/1 queue，因此，我們可以輕易地評估 scheduler 所收到的資料要求的特性，以及傳送一個資料要求到 scheduler 平均所花費的時間為何。



圖五. 系統負載評估模型

經過推導，我們可以得出一個資料要求在上傳頻道、scheduler 與廣播頻道平均所花的等待時間分別為：

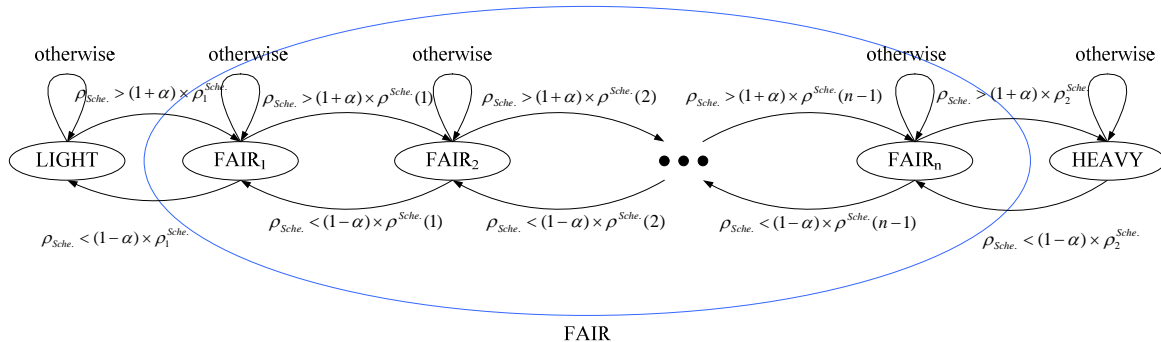
$$W_{Ctrl.} = \frac{1}{\mu_{Ctrl.} - \lambda_{Ctrl.}} = \frac{1}{\frac{B_{Ctrl.}}{s_{Ctrl.} + s_{Ack.}} - \lambda_{Ctrl.}}$$

$$W_{Sche.} = \frac{L_{Sche.}}{\lambda_{Ctrl.}} = \frac{1}{\mu_{Sche.}} + \frac{\frac{\rho_{Sche.}}{\mu_{Sche.}} + \lambda_{Ctrl.} \sigma_{Sche.}^2}{2(1 - \rho_{Sche.})}$$

$$W_{BCast} = \frac{1}{\mu_{BCast}(1 - r_0)}$$

而在 QETP 中，每一個資料要求的平均等待時間則為：

$$W_{Sys.} = W_{Ctrl.} + W_{Sche.} + W_{BCast}$$



圖六.狀態轉換圖

3.4 版本決定機制

版本決定機制主要根據系統的負載狀態，分派適當的版本給需求者。其細分為下列三個階段：

1. 第一個階段為狀態決定階段，我們的研究結果顯示，scheduler 的狀態轉換圖如圖六所示，根據使用者所給與的參數 (ρ_1 與 ρ_2)，我們可以由圖六之計算轉換，決定此 scheduler 的狀態處於 LIGHT、FAIR 或是 HEAVY。

2. 接下來是候選版本選擇階段，根據 scheduler 的不同狀態，選出多個依品質優劣排序的候選版本，以供系統在合乎負載的條件下提供版本給需求者。

3. 最後一個階段為版本階段決定，系統根據在佇列中的資料要求，在眾多候選版本中選出一個最適合的作為最後的決定版本，將其傳送給需求者並且將此版本儲存在快取中，以供下次利用。

3.5 服務允諾機制

| | | curState _{agg.} | | | | |
|---------------------------|-------|--------------------------|--------|--------|-----------|-------|
| | | LIGHT | FAIR | | | HEAVY |
| curState _{ctrl.} | LIGHT | 0/0 | 0/0 | 0.33/0 | 0.66/0.15 | 1/0.3 |
| | FAIR | 0/0 | 0.25/0 | 0.5/0 | 0.75/0.3 | 1/0.6 |

Prob_{Block}/Prob_{Drop}

圖七. 決定 Prob_{Block} 及 Prob_{Drop} 值的範例

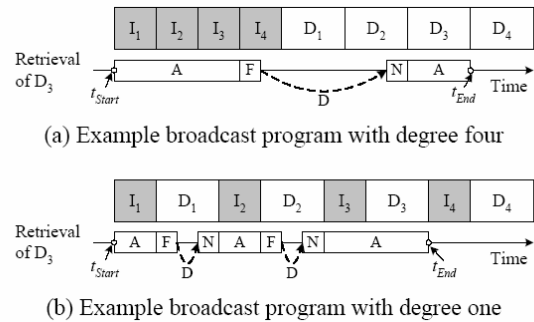
當服務範圍內的使用者逐漸增多時，系統必須考量服務品質的問題，不允許使用者註冊進入系統，甚至，當系統負荷過重時，中斷服務換手中的使用者。因此，在服務允諾機制中，分為兩個主要階段，即狀態決定階段以及控制允諾階段，其說明如下：

1. 在狀態決定階段中，根據系統中正在服務的使用者數量、以及網路狀態等資訊，來決定控制頻道的負載狀態，其決定狀態的方法亦運用圖六來判別。

2. 在控制允諾階段中，當控制頻道的

負載為 HEAVY 時，系統會封鎖使用者註冊並中斷所有服務換手中的使用者；當控制頻道在負載為 FAIR 或 LIGHT 時，系統會決定兩個機率值 Prob_{Block} 和 Prob_{Drop}，分別代表封鎖使用者註冊及中斷服務換手中使用者的機率，圖七為一個範例。我們設計的方法傾向於讓 Prob_{Block} 大於 Prob_{Drop} 是因為我們寧可不允許使用者進入系統，也不願讓使用者在正在服務的狀態下被中斷通訊，如此較符合使用者的期待。

3.6 資料索引機制



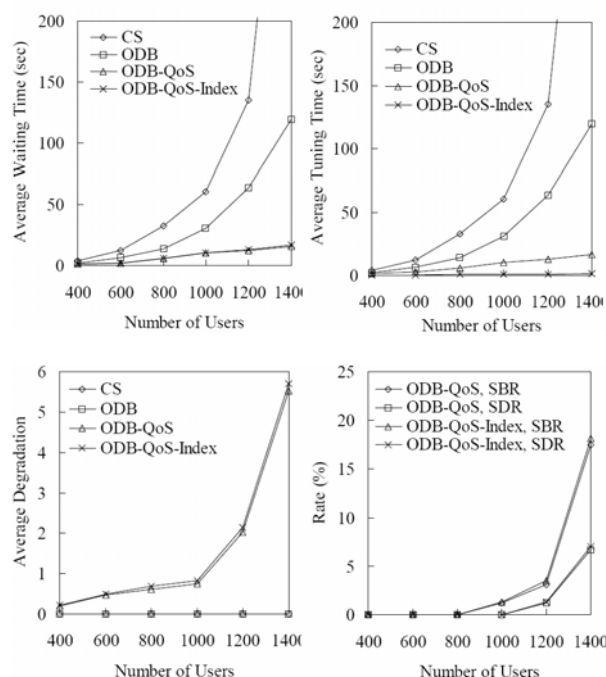
圖八. 索引及資料之廣播排序方法

在參考文獻[18]中指出，將 degree of broadcast programs 值設得愈小會使得行動裝置愈快收到索引的傳播資訊，進而節省能源消耗。然而，此機制並不考慮無線網路介面(Wireless Network Interfaces)開啟及關閉時所消耗的電源與時間。有鑑於此，本計畫將無線網路介面在開啟及關閉時，對行動裝置所造成的影響考慮在內。圖八為兩種不同索引及資料項目的廣播排列程序，圖中 A 和 D 分別代表行動裝置在聽取廣播時的活動模式與休眠模式時間，而 F 和 N 則表示行動裝置在關閉和開啟無線網路介面時所花費的時間。當 degree of broadcast programs 值愈小時(如圖 8(b)所示)，行動裝置會愈頻繁地切換活動與休眠模式，導致 F 與 N 的大量增加，造成行動裝置消耗更多的電源。

有鑑於此，我們提出了一個選擇適當 degree of broadcast programs 值的資料索引機制，將行動裝置的能源消耗降到最低。根據系統負載分為兩個階段來動態調整 degree of broadcast programs。第一個階段為數值統計階段，系統收集每個資料要求的抵達時間、完成時間及其他統計資訊。第二個階段為調整階段，伺服器根據第一個階段所收集到的資訊，選取適當的 degree of broadcast programs 值，藉此達到節省能源消耗的效果。

3.7 實驗結果評估

我們設計了數個實驗以評估我們所提出的方法的效能，部份實驗結果如圖九所示：



圖九. 實驗結果

由實驗結果得知，和傳統的方法相較，我們所提出的方法(即 ODB-QoS-Index) 能夠大大降低使用者的平均等待時間。此外，透過 indexing 技術，ODB-QoS-Index 能在少量增加使用者的平均等待時間的時形下有效地減少行動設備的能源消耗。由

於行動設備的能源有限，因此 ODB-QoS-Index 很適合用於行動資訊系統中。

由於無線網路的進步，使用行動資訊服務的人口快速增加。若行動資訊系統不對服務品質進行控管，將很容易因為大量使用者的使用而造成系統負荷過重。為了解決這個問題，當系統使用人數超過系統負荷時，ODB-QoS-Index 能夠適時地降低物件品質與拒絕新使用者的加入，藉以保障現有使用者的服務品質。實驗結果顯示，在使用人數增加至一定程度時，ODB-QoS-Index 仍能有效地保障正在使用服務的使用者所享有的服務品質。

四、結果自評

透過無線網路提供行動資訊服務之研究，近年來無論在電信服務業及資訊網路業皆為日漸重要之課題。有鑑於此，發展相關之技術和培育這一方面的研究人才已是刻不容緩之課題。參加本計劃之研究人員，在學理上和實務上都能受到紮實而有效的訓練，相信對於國家行動服務的推動將大有助益。

參考文獻

- [1] S. Acharya and S. Muthukrishnan. Scheduling On-demand Broadcasts: New Metrics and Algorithms. In *Proceedings of the 4th ACM/IEEE International Conference on Mobile Computing and Networking*, pages 43–94, October 1998.
- [2] M. Agrawal, A. Manjhi, N. Bansal, and S. Seshan. Improving Web Performance in Broadcast-Unicast Networks. In *Proceedings of the IEEE INFOCOM Conference*, March-April 2003.
- [3] D. Aksoy and M. J. Franklin. Scheduling for Large-Scale On-Demand Data Broadcasting. In *Proceedings of IEEE INFOCOM Conference*, pages 651–659, March 1998.
- [4] D. Aksoy, M. J. Franklin, and S. Zdonik. Data Staging for On-Demand Broadcast. In *Proceedings of the 27th International*

- Conference on Very Large Data Bases*, pages 571–580, September 2001.
- [5] V. Cardellini, P. S. Yu, and Y.-W. Huang. Collaborative Proxy System for Distributed Web Content Transcoding. In *Proceedings of the 9th ACM International Conference on Information and Knowledge Management*, November 2000.
- [6] C.-Y. Chang and M.-S. Chen. Exploring Aggregate Effect with Weighted Transcoding Graphs for Efficient Cache Replacement in Transcoding Proxies. In *Proceedings of the 18th IEEE International Conference on Data Engineering*, February 2002.
- [7] R. Han, P. Bhagwat, R. Lammaire, T. Mummert, V. Perret, and J. Rubas. Dynamic Adaptation in an Image Transcoding Proxy for Mobile Web Browsing. *IEEE Personal Communications*, 5(6), December 1998.
- [8] J.-L. Hsiao, H.-P. Hung, and M.-S. Chen. Versatile Transcoding Proxy for Internet Content Adaptation. *to appear in IEEE Transaction on Multimedia*.
- [9] J.-L. Huang and W.-C. Peng. An Energy-Conserved On-Demand Data Broadcasting System. In *Proceedings of the 6th International Conference on Mobile Data Management*, May 2005.
- [10] S. Lee, D. P. Carney, and S. Zdonik. Index Hint for On-demand Broadcasting. In *Proceedings of the 19th IEEE International Conference on Data Engineering*, March 2003.
- [11] W. Y. Lum and F. C. M. Lau. A Context-Aware Decision Engine for Content Adaptation. *IEEE Pervasive Computing*, 1(3), July-September 2002.
- [12] W. Y. Lum and F. C. M. Lau. On Balancing Between Transcoding Overhead and Spatial Consumption in Content Adaptation. In *Proceedings of the 8th ACM International Conference Mobile Computing and Networking*, September 2002.
- [13] C. Poellabauer and K. Schwan. Energy-Aware Media Transcoding in Wireless Systems. In *Proceedings of the 2nd IEEE International Conference on Pervasive Computing and Communications*, March 2004.
- [14] E. Shih, P. Bahl, and M. J. Sinclair. Wake onWireless: An Event Driven Energy Saving Strategy for Better Operated Devices. In *Proceedings of the 8th ACM/IEEE International Conference on Mobile Computing and Networking*, September 2002.
- [15] A. Singh, A. Trivedi, K. Ramamritham, and P. Shenoy. PTC: Proxies that Transcode and Cache in Heterogeneous Web Client Environments. *Kluwer World Wide Web: Internet and Web Information Systems*, 7, March 2004.
- [16] J. R. Smith, R. Mohan, and C.-S. Li. Content-based Transcoding on Images in the Internet. In *Proceedings of IEEE International Conference on Image Processing*, October 1998.
- [17] M. A. Viredaz, L. S. Brakmo, and W. R. Hambrgen. Energy Management on Handheld Devices. *ACM Queue*, 1(7):44–52, October 2003.
- [18] J. Xu, W.-C. Lee, and X. Tang. Exponential Index: A Parameterized Distributed Indexing Scheme for Data on Air. In *Proceedings of the 2nd ACM/USENIX International Conference on Mobile Systems*, June 2004.
- [19] J. Xu, X. Tang, and W.-C. Lee. Time-Critical On-Demand Data Broadcast: Algorithms, Analysis, and Performance Evaluation. *IEEE Transactions on Parallel and Distributed Systems*, 2006.