

移动环境下多信道广播的有效数据访问

段红亮, 雷向东

(中南大学 信息科学与工程学院, 湖南 长沙 410083)

摘要:在移动计算环境中,数据广播是数据发布和数据获取的重要手段。为提高数据广播的性能,在服务器端采用了近似最优的多信道广播数据调度策略(TOSA),同时在移动客户端提出了高效的 Huffman 树索引策略。从服务器端和客户端同时提高数据广播性能。最后通过仿真试验以及推理,在访问时间、协调时间等方面证明了该模式的高效性。

关键词:移动计算;多信道;索引;访问时间;协调时间

中图分类号:TP311.13

文献标识码:A

文章编号:1673-629X(2009)04-0005-04

Efficient Data Access of Multi-Channel Broadcast in Mobile Environment

DUAN Hong-liang, LEI Xiang-dong

(College of Information Science & Engineering, Central South University, Changsha 410083, China)

Abstract: In mobile environment, data broadcasting is an important data dissemination and data retrieving approach. In order to improve performance of data broadcasting, adopted near-optimal multi-channel data broadcasting (TOSA) in server. At the same time, proposed Huffman tree data indexing scheme in mobile client. Improve the data broadcasting performance from both the server and the mobile client. Finally, the results of simulation and deduce reveal that this model improved the broadcast capability efficiently in access time, tune-in time and so on.

Key words: mobile computing; multi-channel; index; access time; tune-in time

0 引言

随着移动计算的硬件和软件的迅速发展,越来越多的人使用移动设备如手提电脑、手机、PDA等通过移动网络来访问以及存储各种数据。由于无线网络的带宽有限且具有非对称性(下行带宽远大于上行带宽)。所以数据广播技术是针对移动环境中无线网络特点的一种有效数据访问方式,以较小的代价来支持大量移动设备并行访问数据库,且具有很好的可扩展性。在移动数据广播中,有两个关键的执行标准,即移动客户端的访问时间以及电能消耗。近年来,为了改善数据访问的性能,提出了不少数据广播模式,参见文献[1~3],但很少同时从服务器端和移动客户端提高移动数据广播的性能。

为尽量减少移动客户端的等待时间和电能消耗,文中在增加广播信道的基础上,在服务器端采用了近

似最优的数据调度策略以减少访问时间。同时在移动客户端提出了高效的数据索引策略,在文献[4]中,有选择侦听概念的提出就是降低基于数据广播系统的电能消耗。通过有选择性的侦听,使其大部分时间处于休眠状态,仅仅当它想要的数据来到时才被激活,极大地降低了移动客户端电量的消耗。同时从服务器端和移动客户端提高移动环境下的数据广播效率。

1 数据项调度策略

在多信道数据广播系统中,评价多信道数据广播优劣的主要标准是多信道平均访问时间(MCAAD, multi-channel average access delay), $MCAAD = \sum_{i=1}^k \sum_{d_i \in C_i} w_i p_i$ 。为了尽量减少移动客户端的等待时间,近年来有很多关于多信道无线广播的数据调度算法,但很少能够确保达到理论最优或近似最优。由于此最优化问题是 NP 问题,在文中采用的两层调度策略(TOSA)近似达到最优。根据文献[5]中的研究工作中文不加证明地采用了 log-time 算法以及两层调度策略(TOSA)。

收稿日期:2008-07-31

基金项目:国家教育部博士点基金资助项目(20030533011)

作者简介:段红亮(1981-),男,湖南郴州人,硕士研究生,研究方向为移动数据库技术;雷向东,博士,副教授,研究方向为移动数据库技术、并行数据库技术。

为了便于分析,文中定义了表 1 所示的参数表。

表 1 参数表

参数	含义
K	可使用的信道数量
C_i	第 i 个信道, $(0 < i \leq k)$
B_i	第 i 个信道的带宽, $(0 < i \leq k)$
N	整个数据广播中数据项的数量
d_i	第 i 个数据项 $(0 < i \leq N)$
l_i	第 i 个数据项的长度 $(0 < i \leq N)$
N_i	分配到信道 C_i 的数据项的数量
p_i	数据项 d_i 的访问概率
w_i	获取数据项 d_i 的访问时间
s_i	数据项 d_i 的距离(两个连续广播周期中,数据项 d_i 的间隔)
A_i	信道 C_i 中所有数据项 d_j 的 $\sqrt{p_j \times l_j}$ 的总和

1.1 Log-time 算法

Log-time 算法由 Hameed 和 Vaidya 提出的对单信道多数据项的有效调度算法。Log-time 算法主要关注平均访问时间(AED),如果在单信道中每个数据

项 d_i 的实例是等距的,且距离 $s_i = (\sum_{j=1}^N \sqrt{p_j l_j}) \sqrt{\frac{l_i}{p_i}}$, 则此时的平均访问时间(AED)是最小的。

1.2 两层调度策略(TOSA)

TOSA 调度策略分为以下几步:

1)根据数据项的访问概率、数据项长度以及信道的带宽将 N 个数据项分配到 K 个信道中。

算法结构:

输入: N 个数据项的访问概率以及各个信道的带宽;

输出:将 N 个数据项分配到 K 个信道的初始分区结果;

Procedure:

```

1: sort items so that  $\forall i \leq j, \sqrt{p_i l_i} \geq \sqrt{p_j l_j}$ ;
2: sort Channels so that  $\forall i \leq j, B_i \geq B_j$ ;
3: let  $B = \sum_{i=1}^k \sqrt{B_i} / \sqrt{B_k}$ ;  $T_i = \text{sqrt} B_i / B_k$ ;
4: for  $(i = 1; i \leq N; i += 2B)$  do
5: for  $(j = 1, c_i = 0; j \leq K; j++, c_i += T_j)$  do
6: allocate items  $d_l (l \in [i + c_i, \text{MIN}(i + c_i + T_j, N)])$  to Channel  $C_j$ 
7: if  $(i + c_i + T_j) \geq N$  then
8: return;
9: end if
10: end for
11: for  $(j = K, c_i = B; j \geq 1; j--, c_i += T_j)$  do

```

```

12: allocate items  $d_l (l \in [i + c_i, \text{MIN}(i + c_i + T_j, N)])$  to Channel  $C_j$ 

```

```

13: if  $(i + c_i + T_j) \geq N$  then

```

```

14: return;

```

```

15: end if

```

```

16: end for

```

```

17: end for

```

2) 对上一步数据分配进一步调整。

算法结构:

输入: N 个数据项的初始分区;

输出:对初始分区的优化;

Procedure:

```

1: while true do

```

```

2: find two channels  $C_j$  and  $C_m$  such that  $\frac{A_j}{\sqrt{B_j}} \geq$ 
```

```

 $\frac{A_i}{\sqrt{B_i}} \geq \frac{A_m}{\sqrt{B_m}}, i, j, m \in [1, K]$ 

```

```

3: find  $d_{\min}$  from  $C_j$  such that  $\sqrt{p_{\min} l_{\min}} \leq \sqrt{p_0 l_0} \forall d_0 \in C_j$ ;

```

```

4: if  $(\frac{A_j^2}{B_j} + \frac{A_m^2}{B_m} > \frac{(A_j - \sqrt{p_{\min} l_{\min}})^2}{B_j} + \frac{(A_m - \sqrt{p_{\min} l_{\min}})^2}{B_m})$  then

```

```

5: move item  $d_{\min}$  from channel  $C_j$  to channel  $C_m$ ;

```

```

6: else

```

```

7: return;

```

```

8: end if

```

```

9: end while

```

3)根据 log-time 算法对每一个信道内的广播数

据进行优化调整。

在前两步中将各个数据项分配到不同的信道中,属于高层数据调度,而最后一步是在各个信道内部对数据项进行调整,属于底层数据调度。

在两层调度策略(TOSA)中提出了两个假设:

(1)假设在整个广播过程中,数据项 d_i 的访问概率 p_i 是可知的,同时也是固定不变的。

(2)假设每一个客户端请求只需一个数据项。

TOSA 算法是近似最优的数据调度策略。

1.3 TOSA 运行实例

一个基于表 2 的 8 个数据项在两个信道中广播的 TOSA 实例如表 3 所示。

表 2 访问概率

d_i	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8
p_i	0.5	0.2	0.1	0.1	0.07	0.01	0.01	0.01

表3 TOSA 运行实例

initialization: $C_1: d_1, d_4, d_5, d_8$ $C_2: d_2, d_3, d_6, d_7$
Permutation:
first iteration: $C_1: d_1, d_4, d_5, d_8$ $C_2: d_2, d_3, d_6, d_7$ Search: $A_1^2/B_1 > A_2^2/B_2$ $d_{\min} = d_8$ evaluation: $\frac{A_1^2}{B_1} + \frac{A_2^2}{B_2} > \frac{(A_1 - \sqrt{d_8})^2}{B_1} + \frac{(A_2 + \sqrt{d_8})^2}{B_2}$ action: move item d_8 from C_1 to C_2 continue Permutation step
second iteration: $C_1: d_1, d_4, d_5$ $C_2: d_2, d_3, d_6, d_7$ Search: $A_1^2/B_1 > A_2^2/B_2$ $d_{\min} = d_5$ evaluation: $\frac{A_1^2}{B_1} + \frac{A_2^2}{B_2} > \frac{(A_1 - \sqrt{d_5})^2}{B_1} + \frac{(A_2 + \sqrt{d_5})^2}{B_2}$ action: permutation step is stopped Log - time algorithm to schedule item in each channel $s_i = \frac{d_i \in c_i \sqrt{p_i}}{\sqrt{p_i}}, i \in [1, k] \text{MCAAD} = 1.475$

2 数据项索引策略

在移动环境中,服务器端采用高效的数据调度策略的同时,应采用高效的数据索引策略,使移动客户端能有效地访问到所需数据项。这样才能大幅地减少移动客户端的访问时间以及电能消耗。在移动客户端的索引策略中,衡量其索引性能有以下指标:

(1) 谐调时间 (Tuning Time, TT): 客户为获得数据侦听广播的总时间。

(2) 访问时间 (Access Time, AT): 从客户开始侦听到完成下载数据的时间间隔, 平均访问时间 = 探测时间 + 等待时间 + 下载时间。

(3) 额外开销 (Additional Space, AP): 索引 (不包括数据项) 所占信道空间。

文中采用了 Huffman 树索引技术, 并将整个索引作为单独的数据项在同一信道中进行广播。

2.1 Huffman 树索引技术

传统的树索引结构采用的是固定扇出的平衡树, 没有考虑数据的访问概率, 基于表 2 的传统树索引结构如图 1 所示。在该索引结构下, 移动设备访问任何一个数据记录需要遍历 3 个索引节点。事实上, 各个数据记录的访问概率存在很大的倾斜。因此, 如果建立基于访问概率的索引树, 则可以降低平均访问代价。

由于 Huffman 树是带权路径长度 (WPL) 最小的二叉树, 也称最优二叉树, 所以文中提出了基于数据访问概率的 Huffman 树索引, 基于表 2 的 Huffman 树索引结构如图 2 所示。Huffman 索引树构造包括以下步骤:

- (1) 获得具有 n 棵子树的森林, 每个节点标明相应的访问概率;
- (2) 将 2 棵具有最小访问概率的子树合并, 形成一个新的节点, 该节点的访问概率等于所有子树之和;
- (3) $n = n - 2 + 1$, 如果 $n = 1$, 则停止, 否则转到第 2 步。

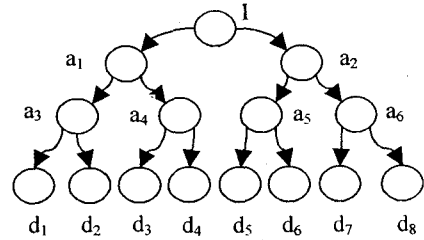


图1 传统树索引

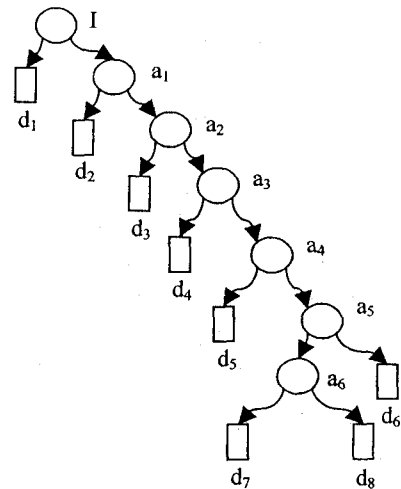


图2 Huffman 树索引

根据图 1 与图 2, 使用 Huffman 索引树, 平均遍历索引节点的数目从 3 降到了 2, 从而节省了移动客户端电量的消耗。试验证明, 若数据访问概率的倾斜性越大, 则 Huffman 树索引技术越优。

2.2 获取数据项

对每一个数据项, 使用信道号以及信道内的顺序号两个参数进行区分。移动客户端根据数据项索引指针, 每次访问数据最多只在信道间跳转一次。

移动客户端从广播中访问数据分为一下几步:

- (1) 移动客户在同一的信道中监听广播, 得到索引树的起始地址, 即索引桶。
- (2) 遍历该索引桶, 找到需要访问的数据项的相关指针。
- (3) 跳转到正确的信道, 找到所需的数据项。

3 性能分析

在服务器端的数据调度策略中,假设数据项访问概率服从 Zipf 分布,描述为:

$$p_i = \frac{(1/i)^\theta}{\sum_{i=1}^N (1/i)^\theta}, \text{其中 } \theta \text{ 为斜率(Skewness)。经过}$$

大量仿真试验表明 TOSA 是近似最优调度策略。在相同长度的数据项长度以及信道带宽情况下,TOSA 算法常用的 GREEDY 算法的比较,如图 3 所示,仿真参数表如表 4 所示。

表 4 仿真参数

参数	含义
广播数据项的个数	10 000
信道数	2~5
数据项的长度	512Bytes

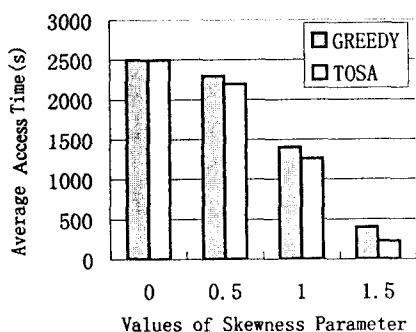


图 3 平均访问时间比较

从图 3 中可知,在同等条件下,随着数据项访问倾斜率的不断增加,TOSA 越能显示其优越性。同时 TOSA 还考虑到了数据项长度的变化,更进一步地适应了将来不断变化的需要。

在移动客户端的索引策略中,假设索引树中,叶子节点数为 n ,则

● 在传统索引树中:

平均谐波时间: $TTT = [\log_2 n] + 1$, 其中 $[\log_2 n]$ 为索引树的高度。

额外开销: $TSP = n + \sum_{k=1}^L [\frac{n}{2^k}]$, $L = [\log_2 n]$ 。

平均访问时间: $TAT = n + \sum_{k=1}^L [\frac{n}{2^k}] + \text{Data} + 1$, 其中 Data 为等待时间。

● 在 Huffman 索引树中:

平均谐波时间: $HTT = (WPL + n)/n$, WPL 为带权路径长度。

额外开销: $HSP = 2n - 1$

平均访问时间: $HAT = 2n - 1 + \text{Data} + 1$, 其中 Data 为等待时间。

虽然 Huffman 树的 WPL 不能使用公式计算,由 Huffman 树的 WPL 的最优性可以得出: $HTT \leq TTT$, 又因为 $HSP \leq TSP$, $HAT \leq TAT$, 故 Huffman 索引树优于传统索引树。从而大大减少了移动客户端的平均电能消耗以及平均访问时间。

4 结束语

文中综合考虑了在移动环境中从服务器端和移动客户端同时提高数据广播效率,尽量减少移动客户端的等待时间和节约电能。在服务器端,文中使用了 TOSA 数据调度策略;在移动客户端,文中使用了 Huffman 树索引技术。并证明了其优越性。

在移动环境中,移动客户端如何同时请求多个数据项,在索引策略中如何减少索引树的高度,如何提高广播数据的安全性等等,有待进一步研究。

参考文献:

- [1] Shivakumar N, Venkatasubramanian S. Efficient indexing for broadcast based wireless systems[J]. Mobile Network and Application, 1996(12):433-446.
- [2] Lo S-C, Chen L P. Optimal index and data allocation in multiple broadcast channels[C]//In Proceedings of the 16th International Conference on Data Engineering. San Diego, CA, USA: [s. n.], 2000.
- [3] Lee G, Yeh M S, Lo S C, et al. A Strategy for efficient access of multiple data items in mobile environments[C]//In Proceedings of 3rd International Conference on Mobile Data Management. Singapore: [s. n.], 2002:71-78.
- [4] Imielinski T, Viswanathan S, Badrinath B R. Energy efficient indexing on air[C]//the 4th International Conference on Extending Database Technology. [s. l.]: [s. n.], 1994:254-258.
- [5] Zheng Baihua, Wu Xia, Jin Xing, et al. TOSA: a near-optimal scheduling algorithm for multichannel data broadcast [M]. New York, NY, USA: [s. n.], 2005:29-37.

(上接第 4 页)

- [5] 程红莉,周宁,肖爽.文本驱动的商务智能研究[J].情报科学,2007(10):1525-1529.
- [6] 朱德利. SQL Server 2005 数据挖掘与商务智能完全解决方案[M].北京:电子工业出版社,2007:10-11.
- [7] 陈思睿,张永,杨志勇.基于粗糙集的特征选择方法的研究[J].计算机工程与应用,2006(21):159-161.

- [8] 梁开健.基于 DCSSM 的文本特征提取及文本挖掘研究[J].自动化技术与应用,2005,24(5):54-56.
- [9] 王珍珍.关于文本挖掘中文本分类与文本聚类研究[J].计算机与信息技术,2007(6):55-56.
- [10] 黄迎春,李晓晔,邓文新.文本挖掘技术的研究[J].齐齐哈尔大学学报,2006,22(3):53-55.