第32卷 第20期 计 算 机 工 程 2006年10月
Vol.32 № 20 Computer Engineering October 2006
·软件技术与数据库· 文章编号：1000—3428(2006)20—0056—03 文献标识码：A 中图分类号：TP311

移动实时数据库系统多版本数据广播

雷向东，赵跃龙，陈松桥，袁晓莉
(中南大学信息科学与工程学院，长沙 410083)

摘 要：结合多版本乐观并发控制协议(MVOCC)提出了带失效报告多版本-有效性检查信息 (Multiversion- Validation Information, MV-VI)广播机制，并给出了多种多版本广播磁盘组织。MV-VI 广播机制支持移动主机(MHs)与网络断接。在广播周期中广播事务全局有效性检查信息 VI，IR 广播后插入多个快速失效报告(Fast Invalidation Report, FIR)，广播快速更新的数据。通过模拟仿真，对 MV-IV 广播机制进行了性能测试，实验结果表明，MV-IV 广播机制性能明显好于其它广播机制。

关键词：移动实时数据库系统；多版本数据广播；有效性检查信息；失效报告

Multiversion Data Broadcast in Mobile Real-time Database Systems

LEI Xiangdong , ZHAO Yuelong, CHEN Songqiao, YUAN Xiaoli
(College of Information Science and Engineering, Central South University, Changsha 410083)

【Abstract】 The multiversion-validation information(MV-VI)data broadcast mechanism with invalidation report is proposed that combines multiversion optimistic concurrency control protocol(MVOCC). Various multiversion broadcast disk organizations are introduced. The MI-VI data broadcast mechanism supports disconnection. During broadcast cycle, the global validation information(VI)of transactions is broadcasted. The fast invalidation reports(FIRs)are inserted after the IR to broadcast frequently updated data. Detailed simulation experiments are carried out to evaluate the proposed multiversion data broadcast mechanism. The results reveal that the performance of the MV-IV data broadcast mechanism is significantly better than other methods.

【Key words】 Mobile real-time database systems; Multiversion data broadcast; Validation information; Invalidation report(IR)

在移动实时数据库系统(RTMDBSs)中，数据广播是一个有效的数据传播方法，被频率请求的数据由服务器不断周期性地广播，广播数据立刻被移动主机(MH)收到，可获得的传输带宽得到有效利用。数据广播是提高RTMDBSs性能的一项关键技术。已提出许多数据广播机制[1,2]。但多数广播机制没有关注数据冲突问题，并且不适合实时环境。有些广播机制延时长，带宽利用率低，不适应数据更新快的应用。最近的许多研究工作表明失效报告(Invalidation Report，IR)[3]是一个有效的方法，MH为了检查 cache 中数据的有效性，不必直接查询服务器，只要监听IR。但IR方法也有许多缺陷。如果MH错过监听IR，MH不得不等到下一个IR广播，延时时间较长，不能满足MH的实时需要。本文结合多版本乐观并发控制(MVOCC)协议，提出带失效报告(Invalidation Report，IR)多版本-有效性检查信息 (Multiversion-Validation Information，MV-VI)广播机制。通过模拟仿真，对MV-IV数据广播机制进行了性能测试。实验结果表明，MV-IV数据广播机制性能比其它广播机制性能有明显的提高。

1 事务处理策略

MVOCC协议[4]对数据维护多个版本，每个数据项Q有一个版本序列$<Q_1, Q_2, \ldots, Q_m>$与之关联。每个版本Q_k包含3个数据字段：content表示Q_k版本值，$WTS(Q_k)$表示创建Q_k版本的事务的时间戳，$RTS(Q_k)$表示所有成功读取Q_k版本的事务的最大时间戳。每个事务T_i赋予一个时间戳$TS(T_i)$。更新事务执行分为3个阶段：读阶段，有效性检查阶段和写阶段，在读阶段，假设事务T_i发出read(Q)或 write(Q)操作。令Q_k表示Q的版本，其写时间戳小于或等于$TS(T_i)$的最大写时间戳。如果事务T_i发出 read(Q)，则返回Q_k的值。如果事务T_i发出write(Q)操作，且若$TS(T_i)<RTS(Q_k)$，则事务T_i回滚，否则，若$TS(T_i)= WTS(Q_k)$，则Q_k的内容被覆盖；否则创建Q的一个新版本。事务T_i的更新操作是在T_i的局部空间中。在有效性检查阶段，更新事务T_i在移动的主机上进行局部有效性检查。如果事务通过局部有效性检查，事务必须提交到服务器进行全局有效性检查。在写阶段，若事务T_i通过全局有效性检查，则实际的更新就可写入数据库中。只读事务读最新提交的数据版本，只有读阶段。

2 断接

在移动计算环境一个重要的特性是MH频繁地自愿或不自愿与移动网络断开。在断接期间 MH 仍继续操作。MH 上事务对局部 cache 数据发出读操作和写操作，MH 只有在重新连接后才能发现它的 cache 中数据可能已过时。在 MH 上进行的更新只有在重新连接后才能传播给其它 MH。MH 与移动网络临时断开，如果断开时间小于检测窗口的大小，MH 用 IR 检测它的 cache 数据的有效性。如果 MH 与网络断开时间长，大于检测窗口的大小，MH 不能用 IR 检查它的 cache 中数据的有效性，必须将它的整个 cache 数据丢弃。由于多版本广播机制广播服务器中的每个数据项所有版本，

基金项目：国家“863”计划基金资助项目(511-910-092)；教育部博士点基金资助项目(20030533011)
作者简介：雷向东(1964－)，男，博士、副教授，主研方向：移动数据库技术，并行数据库技术；赵跃龙、陈松桥，博导；袁晓莉，工程师
收稿日期：2005-10-28 **E-mail：**leixiangdong@mail.csu.edu.cn

MH 可从广播通道中重新接收新的数据。

3 多版本广播磁盘组织

广播介质可以看作延迟时间很长的广播磁盘来建模，对于 MH 数据的请求可认为在所请求的数据广播时得到了服务。服务器周期性地对 MH 广播数据，不同访问频度的数据以不同的周期广播。热数据广播周期短，冷数据广播周期长，从而扩展了从服务器到 MH 之间的传输带宽。

3.1 广播磁盘

数据项按访问频度进行划分，访问频度相近的数据项放在同一个磁盘上。磁盘分成更小相等的数据单元，称为块。磁盘块的个数与磁盘的速度成反比。数据项所有版本都相继广播。热数据项位于快速磁盘上，冷数据项则位于慢速磁盘上。广播数据时，从每一个磁盘上取一块广播，磁盘上的块按顺序取。广播磁盘如图 1 所示。

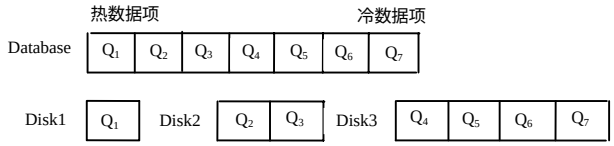

图 1 广播磁盘

广播周期分成多个微周期，在一个微周期中广播所有磁盘中的一块。假设每个数据项为一块，数据广播如图 2 所示。

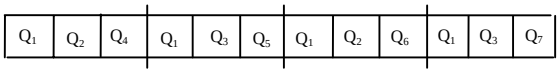

图 2 数据广播

在广播周期中，热数据项被频繁地广播，冷数据项广播次数则相对较少。

3.2 多版本广播磁盘组织

服务器维护每个数据项多版本，每个数据项所有版本都将广播。多版本广播磁盘组织必须决定数据项的老版本如何广播，以及数据项的老版本广播的最佳频率。多版本广播磁盘组织方式如下：

(1)聚类磁盘(Clustering Disk)

数据项所有版本放在同一个磁盘上，每个数据项所有版本将相继广播，热数据的老版本位于最新版本之后，都放在快速磁盘上。冷数据所有版本则位于慢速磁盘上。数据项所有版本以相同的频率广播，如果每个事务访问数据项的任何版本的概率是相同的，那么这种组织方式性能相当好。

(2)溢出磁盘(Overflow Disk)

数据项老版本放在溢出磁盘上，每一个广播周期之后附加多个微周期分配给数据项的老版本，数据项的老版本将在广播数据项最新版本之后广播。这种方法的缺陷是由于引入附加微周期，每个磁盘的相对速度将受到影响。解决的方法是将数据项老版本放在最慢的磁盘上。

(3)可调速磁盘(Adjustable-speed Disk)

每个数据项老版本放在可调速磁盘上，磁盘的相对速度是装载数据项最新版本磁盘的相对速度的 1/m。这种方法适应性强，当有许多长事务时，通过选择较小 m，使装载数据项老版本磁盘的相对速度加快，较快地广播数据项老版本，当多数事务需要访问数据项最近的版本时，通过选择较大的 m，使装载数据项老版本磁盘的相对速度变慢，较慢地广播数据项老版本。

4 处理频繁更新的数据

IR 广播机制对于数据变化频率不大的情况下是有效的。可是，如果数据变化频率大，MH不得不等很长的时间才得到所需的数据[5]，不适合实时环境。为了减少查询延迟时间，在广播周期中，IR广播后插入多个快速失效报告(Fast Invalidation Report，FIR)，广播快速更新的数据。每个FIR只包含在上一个FIR广播后的快速更新数据项，MH立即能收到这个更新的数据，减少了查询延迟时间。由于大多数快速更新的数据项很小，FIR的大小与IR相比要小得多，因此广播这些数据不需增加太多的开销。

5 广播通道

IR 包含前一个广播周期更新的数据项。服务器周期地对所有 MH 广播 IR，使每个 MH 能更新它的 cache 中的数据项。

在MVOCC并发控制协议中，只读事务在MH上有效处理。更新事务在MH局部有效性检查。当事务T_{val}在MH上通过局部有效性检查，事务T_i必须提交到服务器进行全局有效性检查。服务器对所有MH广播有效性检查信息(VI)。有效性检查事务T_{val}的VI由(T_{val}，TS(T_{val})、ReadSet(T_{val})、 WriteSet(T_{val}))组成。广播通道如图 3 所示。

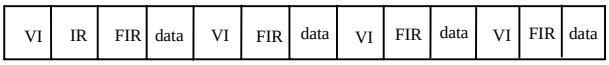

图 3 广播通道

6 性能评价

通过仿真模拟测评本文提出的 MV-VI 数据广播机制性能。模拟的模型由一个服务器，广播磁盘和多个 MHs 组成。广播磁盘传输数据项和控制信息，数据装载在服务器各磁盘上，数据库系统在虚拟页式存储器系统上实现，事务处理的数据在数据库中呈均匀分布。模拟实验参数如表 1 所示。

表 1 模拟系统参数设置

参数	含义	值
MH:	移动	
Think Time	事务到达间隔时间	2 s
Translength	事务长度	5
ReadOperationProbility	读操作概率	0.8
Slack Factor	松驰因子	2.0-8.0
NumberofMobileClients	MH 数	10
Local Cache Size	局部 cache 大小	100 data items
服务器:		
DBSize	数据库的页数	2 400 data items
Think Time	事务到达间隔时间	2 s
Translength	事务长度	10
ReadOperationProbility	读操作概率	0.5
PageHitRatet	页命中率	0.75
MessageDelay	消息延迟时间	10ms
DiskAccessTime	磁盘访问时间	10ms
CPUComputerTime	CPU 计算时间	20ms
Number of Broadcast Disks	广播磁盘数	3
Relative Frequent Broadcast Disks	广播磁盘相对频率	5, 3, 1
Broadcast bandwidth	广播带宽	40 data items/s
Hot data access probability	热数据访问率	0.75
Hot data update probability	热数据更新率	0.35

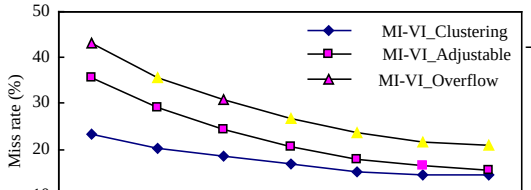

图 4 延误截止时间率－更新时间间隔

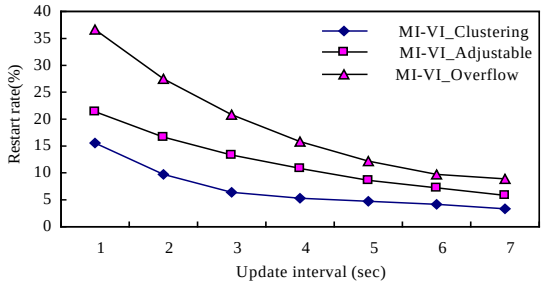

图 5 重启动率－更新时间间隔

图 4、图 5 显示更新时间间隔对延误截止时间率和重启动率的影响。设定可调速磁盘速度是装载数据项最新版本磁盘的相对速度的 1/3。从图中可看出，使用聚类磁盘 MI-VI 广播机制要优于其它组织方式。主要原因是热数据的老版本位于最新版本之后，放在快速磁盘上，冷数据所有版本则位于慢速磁盘上。热数据项的老版本访问频率较冷数据项老版本访问频率高，由于热数据项的老版本广播频率较冷数据项老版本广播频率高，自然这种组织方式性能相当好。使用溢出磁盘 MI-VI 广播机制，由于引入附加微周期，每个磁盘的相对速度将受到影响。使用可调速磁盘 MI-VI 广播机制，每个数据项的老版本位于可调速磁盘上。由于可调速磁盘速度较慢，广播老版本的频率相对较慢。

图 6 显示 MH 断接对延误截止时间率的影响。IR 广播机制性能较差，其原因是如果事务错过当前 IR，事务必须等到下一个 IR，延时时间较长，事务可能错过截止时间。而 MV-VI 方法由于广播数据多个版机本，只要事务所需数据项的版本还在广播，事务就能访问这些数据项。因而，MV-VI 广播机制更能容忍 MH 断接。

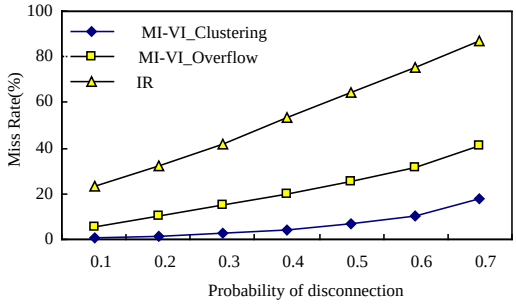

图 6 延误截止时间率－断接率

7 结论

本文结合多版本乐观并发控制 MVOCC 协议提出 MV-VI 广播机制，并给出了多种多版本广播磁盘组织。MV-VI 广播机制支持 MHs 与网络断接。在广播周期中广播事务全局有效性检查信息 VI，在 IR 广播后插入多个 FIRs，广播快速更新的数据。通过模拟仿真，对 MV-IV 广播机制进行了性能测试。实验结果表明，MV-IV 广播机制性能明显好于其它广播机制。

参考文献

1 Cao Guohong. Proactive Power-aware Cache Management for Mobile Computing Systems[J]. IEEE Trans. on Computers, 2002, 51(6): 608.

2 Pitoura E, Chrysanthis P K. Multiversion Data Broadcast[J]. IEEE Trans. on Computers, 2002, 51(10): 1334-1230.

3 Tan K, Cai J, Ooi B. Evaluation of Cache Invalidation Strategies in Wireless Environments[J]. IEEE Trans. on Parallel and Distributed Systems, 2001, 12(8): 789-807.

4 Lei Xiangdong, Zhao Yuelong, Chen Songqiao, et al. Transaction Processing in Mobile Real-time Database Systems[J]. Chinese Journal of Electronics, 2005, 14(3): 491-494.

5 Elmagarmid A, Jin Jing, Helal A, et al. Scalable Cache Invalidation Algorithms for Mobile Data Access[J]. IEEE Trans. on Knowledge and Data Engineering, 2003, 15(6): 1498-1511.

(上接第 50 页)

体占据一定优势，MBPM-BM 其次，MCounting 最差。MBPM 查找时间较均衡，基本不随 k 增加而变化，而在α≤1/4 情况下，MBPM-BM 效率还是最高的。

MBPM-BM 总体占据优势，且在汉字等大字符集情况下效率最高，运行时间要比紧跟其后的 MCounting 减少 40%以上，即使在中小字符集情况下，只要错误率不高，表现依然很好。

4 总结

MBPM-BM是一种适合汉字等大字符集的高效多模式近似字符串匹配算法，在过滤阶段，扩充了BPM-BM[2]单模式过滤思路，进行多模式并行过滤；在匹配阶段，采用MBPM进行并行匹配。

虽然在多模式匹配中 k 对于所有模式均相同，但只需简单修改算法中用到的状态变量 S 的初值，就可适应不同模式。算法其余部分无须作任何改变。

参考文献

1 Navarro G. Multiple Approximate String Matching by Counting[C]. Proc. of the WSP'97. Carleton University Press, 1997: 125-139.

2 陈开渠, 赵　洁, 彭志威. 快速中文字符串模糊匹配算法[J]. 中文信息学报, 2004, 18(2): 58-65.

3 Hyyrö H, Fredriksson K, Navarro G. Increased Bit-parallelism for Approximate String Matching[C]. Proc. of WEA'04. Berlin: Springer Verlag, 2004: 285-298.

4 Myers G. A fast Bit-vector Algorithm for Approximate String Matching Based on Dynamic Programming[J]. Journal of the ACM Archive, 1999, 46(3): 395-415.

5 Hyyrö H. Explaining and Extending the Bit-parallel Algorithm of Myers[R]. University of Tampere, Finland, Technical Report: A-2001-10, 2001.