

文章编号: 1005-0523(2007)04-0071-04

# 移动数据库自适应缓存失效算法的研究

胡红<sup>1,2</sup>, 余为益<sup>2</sup>, 胡应平<sup>1,3</sup>

(1. 华东交通大学 信息工程学院, 江西 南昌 330013; 2. 上饶师范学院, 江西 上饶 334001;  
3. 南昌陆军学院, 江西 南昌 330103)

**摘要:** 在移动客户端建立缓存可以提高移动数据库系统的性能, 也会带来服务器上的数据和缓存中的数据不一致的问题, 即缓存一致性问题. 通常采用基于广播的缓存失效算法来维护缓存的一致性. 传统的采用同步广播方式或异步广播方式的缓存失效算法各有优缺点. 该文在分析传统算法优缺点和一些现有算法的基础上, 提出了一种改进的基于计数器的自适应缓存失效算法, 该算法结合了两种传统的算法, 可以提高缓存的利用率, 同时可以降低通信开销.

**关键词:** 失效报告; 缓存; 移动数据库; 缓存一致性

中图分类号: TP311.13

文献标识码: A

## 0 引言

缓存失效算法是解决缓存一致性问题的常用算法. 通常采用广播的方式发送缓存失效报告, 维护缓存数据的一致性. 传统的缓存失效算法有同步广播方式和异步广播方式两种. 两者相结合的缓存失效算法可以提高算法的效率. 本文讨论的是移动客户端的缓存管理问题, 本文后面部分的客户端或客户机如无特别说明, 均指移动客户端或移动客户机.

## 1 基于缓存失效报告的缓存管理

当客户端收到失效报告时, 会将缓存中的对应项失效, 通常缓存中的每一缓存项都有一个有效位, 使其失效就是将其有效位设为 invalid. 用户请求某一数据项时, 客户机要确定该数据项是否在缓存中, 如果在缓存中, 还要判断缓存中的数据项是否有效, 若有效, 则用缓存内容回答查询, 若无效, 说明缓存数据还未更新, 则把查询提交给服务器; 如果不在缓存中, 则将查询直接提交给服务器.

更新的数据项的内容通常要在失效报告之后采取广播的传送方式或经用户请求采取点对点的传送

方式. 当采取广播的方式传送数据项时, 客户端从广播中下载失效的数据项对应的数据项, 存入缓存中并将使有效位设为 valid, 这个过程叫做缓存的更新过程(Validate Cache). 当采取点对点的方式传送时, 数据项直接传送到客户端并执行缓存的更新过程.

## 2 传统的同步、异步广播方式的缓存失效算法

同步广播方式是指服务器跟踪最近一个广播周期内被更新的数据项, 定期广播失效报告 IR (Invalidation Report). 失效报告通常含有数据项 ID 和时间戳. 在连接的情况下, 客户机要知道缓存项是否有效, 必须等到下一个广播周期到来时才能确定, 因为它当前知道的是上一个广播周期内数据的更新, 而当前这个广播周期内数据的更新要等到下一个广播周期到来时才知道. 在断连的情况下, 当断连的时间  $t \leq W * L$  时, 仍然可以知道缓存的有效性, 其中  $L$  为广播周期,  $W$  为失效广播窗口大小; 而  $t > W * L$  时, 就要使所有缓存项失效.

异步广播方式是指, 当有数据更新时就立即广播失效报告. 在连接的情况下, 客户机可以在等待一个最大传输时延(从服务器知道数据更新到失效报

收稿日期: 2007-04-10

(作者简介: 胡红(1989-), 女, 安徽人, 华东交通大学信息工程学院硕士研究生, 研究方向: 移动数据库. <http://www.cnki.net>)

告传输到客户机)后根据缓存项中的有效位来判断该缓存项是否有效.在断连的情况下,就要根据时间戳和其他的一些策略(如估计数据项的更新时间)来判断缓存的有效性,有的情况也要将所有的缓存项失效.

这两种方式都各有优缺点.

同步方式下,广播周期是固定的(设为 $L$ ),而且是定期广播批量数据的失效报告,广播开销比较小,但它存在用户的查询等待时间较长的问题,因为用户发出查询时,客户机首先要看缓存中是否有需要的数据,然后是确定缓存中的数据是否是有效的,这必须要等到下一个广播周期到来时才能确定.在平均情况下,每个查询要经过 $L/2$ 的等待时间.失效报告到达客户端的平均时间也为 $L/2$ ,所以更新时间也较长.如果广播周期比较长,那么就不能满足某些用户的需求,缓存的更新也会严重滞后.

异步方式下,这样的方式虽然可以解决同步方式下的查询时间长的问题,但是一旦有数据更新就要广播失效报告,使得服务器频繁广播,广播开销较大.异步方式对于连接客户非常有效,可以及时更新缓存,但对于断连一段时间后重新连通的客户机,仅凭失效报告无法判断缓存的有效性.

### 3 算法的改进

在一个数据库中,不是所有的数据都会经常访问,通常它会遵循 20—80 规律,也就是经常访问的数据往往集中在整个数据库的 20% 的部分,可以根据数据的访问频率来决定哪些数据是热点数据,哪些数据是非热点数据,对不同的数据的采用不同的处理方式可以提高算法的效率,提高系统的性能.

可以用计数器来记录数据项的访问频率,进而确定数据是热点数据还是非热点数据<sup>[3]</sup>.对于热点数据,由于它经常被访问,访问的间隔也很短,缓存项要及时更新,否则数据项既使在缓存中也由于失效而不能使用,缓存的利用率会降低,所以对热点数据采用异步方式的失效算法,可以提高缓存利用率;而非热点数据不经常访问,更新时间可以相对较长,可以使用同步方式的失效算法.

### 4 基于计数器的自适应缓存失效算法

下面介绍基于计数器的自适应缓存失效算法.该算法的思想是利用计数器来指明数据的“热”度,

区分热点数据和非热点数据,对于热点数据采用异步广播失效报告的方式,对于非热点数据采用同步广播失效报告的方式.假设数据只能被服务器更新,且服务器不知道客户机的状态;客户机总是将刚刚访问过的数据缓存下来,而置换掉最久未访问的缓存项,即采用 LRU 置换算法.由于服务器不知道客户机的状态,在客户机连接到服务器上时,服务器会要求客户机将其上缓存的数据项和置换掉的数据项报告给服务器.

#### 4.1 计数器的管理

计数器的由服务器管理.

为服务器中每一个数据项增加一个计数器 Counter. Counter 的初值为 0,表示没有客户端缓存该数据项;系统中设置一个  $cnt_{\alpha}$  参数,  $cnt_{\alpha}$  值是区分热点数据和非热点数据的标志,主要是根据统计信息来设定.当 Counter 值大于  $cnt_{\alpha}$  时,认为该数据项是热点数据;否则,认为是非热点数据.

当客户端向服务器请求数据项时,将该数据项的计数器 Counter 加 1;当客户端将缓存中的数据项置换掉时,服务器上对应数据项的计数器 Counter 减 1.当有很多个客户端在一段时间内请求某一数据项时,该数据项的 Counter 值会迅速增大,并超过  $cnt_{\alpha}$ ,这个数据项就会成为热点数据.而当客户端将缓存中的某一数据项置换时,说明该数据项在一段时间内在该客户端未使用,所以它的“热”度会降低,相应的 Counter 值就会减小.

由于移动环境下经常会出现各种错误,可能会导致 Counter 值出现错误,服务器会定期检查 Counter 值,以保证其正确性.

#### 4.2 缓存更新过程

对于热点数据,更新的数据项会在失效报告之后立即广播,客户端接收到更新的数据项时,执行缓存的更新过程;对于非热点数据,更新的数据项在失效广播之后通过广播方式发送,客户端接收到更新的数据项时,再执行缓存的更新过程.

#### 4.3 服务器的广播模型

服务器中组织了一个异步广播队列、一个同步广播队列和一个更新数据广播队列.服务器将失效报告的优先级设置成最高优先级,将更新数据项的优先级设置为次高,其它信息的优先级都比它们低.服务器会收集数据更新信息,将更新的 Counter  $>$   $cnt_{\alpha}$  的数据项(热点数据)的失效报告加入到异步广播队列中,然后将更新的数据项也插入到异步广播队列中,只要该队列不空或同步广播的时刻未到,服

务器会一直广播异步广播队列中的失效报告或更新的数据项.将最近一个周期内更新的  $0 < \text{Counter} < \text{cnt}_a$  的数据项(非热点数据)收集在一个更新队列中,然后就这个更新队列形成失效报告加入到同步广播队列中,然后定期广播失效报告;将这个更新队列也插入到更新数据广播队列中.当异步广播队列为空,而同步广播时刻未到时,就广播更新数据广播队列中的数据项.

## 5 算法的分析

该算法利用了同步广播方式和异步广播方式的优点.对于占整个数据库较小部分的热点数据采用异步广播失效报告的方式,它的额外的广播开销可以从减少的上行通信量中得到抵销.对于热点数据采用异步广播方式,相对于采用同步广播方式,可以减少更新缓存项的时间.下面图1是一个缓存项采用异步广播失效报告方式和同步广播失效报告方式的更新时间的示意图.

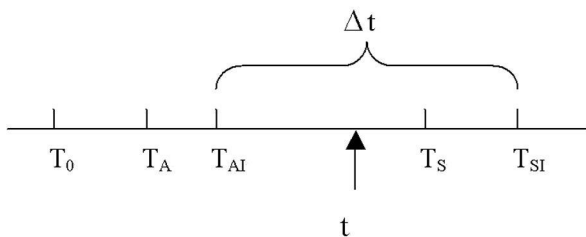


图1 缓存项更新时间示意图

其中,  $T_A$ 、 $T_{AI}$  是缓存该缓存项的客户端在异步广播失效报告方式收到失效报告的时刻和更新缓存项的时刻,  $T_S$ 、 $T_{SI}$  是缓存该缓存项的客户端在同步广播失效报告方式收到失效报告的时刻和更新缓存项的时刻.一般情况下,  $T_{AI} - T_A \leq T_{SI} - T_S$ .

从图1中可以得到,如果忽略  $T_{AI} - T_A$  和  $T_{SI} - T_S$  的差异,减少的时间(设为  $t$ )等于同步方式下的查询等待时间,在平均情况下,更新时间可以减少  $L/2$  ( $L$  为同步广播周期).在  $t$  这段时间里,可能有很多个客户端要访问这个数据项,如果是在同步方式下,当它发现缓存项失效后,就会向服务器提出请求,而异步方式下可以直接用缓存项回答,提高了缓

存的利用率,同时减少了上行通信量,而且对多个热点数据都会出现上述的这种情况,可以减少大量的上行通信量.所以对热点数据采用异步广播方式可以提高缓存的利用率,同时也减少了通信开销.

对于非热点数据采用同步广播方式可以在节省广播开销的同时也能满足大多数客户端的需求,因为对于非热点数据,访问的客户端很少,访问的频率也不高,在缓存项更新的这段时间,客户端访问的机会很小,影响的客户端也很少,所以能满足大多数客户端的需求.

另外,计数器不仅可以量化数据项的“热”度,还可以节省有限的带宽,因为对于  $\text{Counter}$  为 0 的数据项,表示没有客户端缓存该数据项,这样的数据项的失效报告可以不包括在失效广播中.

## 6 结束语

本文提出的这种基于计数器的自适应算法,利用了同步广播方式和异步广播方式的优点,提高了缓存的利用率,同时也降低了通信开销.这种从分析数据特性入手改进算法的思想,正符合以数据为导向的信息时代的特性和要求,对不同的数据采取不同的处理策略往往能取得事半功倍的效果.

### 参考文献:

- [1] 陈婵颖,薛贺.缓存技术在移动数据库中的应用研究[J].计算机工程与设计,2006,27(9):1615-1617.
- [2] 罗军,陈洪波.基于异步数据广播技术的移动数据库缓存策略[J].计算机工程,2005,31(24):102-104.
- [3] GuoHong Cao. On Improving the Performance of Cache Invalidation in Mobile Environments [J]. Mobile Networks and Applications. 2002, Vol. 7, 291-303.
- [4] LI YONG CHUNG, CHONG SUN HWANG, JINGON SHON. Adaptive Concurrency Control Schemes Supporting Mobile Client Caching [J]. International Journal of Computational Engineering Science. 2001, Vol. 2(3):537-555.
- [5] 吴劲,卢显良,任立勇,周旭.移动计算环境中缓存失效策略的归类研究法[J].计算机科学,2004,31(1):39-41.
- [6] 邵雄凯,何瑜.移动计算环境中基于广播的数据缓存策略[J].计算机应用,2006,26(2):364-367.

# The Research of Adaptive Cache Invalidation Algorithms for Mobile Database

HU Hong<sup>1,2</sup>, YU Wei-yi<sup>2</sup>, HU Ying-ping<sup>1,3</sup>

(1. School of Information Engineering, East China Jiao tong University, Nanchang 330013; 2. Shang Rao Normal College, Shangrao 334001;  
3. Nanchang Military Academy, Nanchang 330103, China)

**Abstract:** Mobile Client Cache can improve the performance of the mobile database system. However, it leads to the problem that data in server and in cache are not consistent, that is, cache inconsistency problem. People often maintain the cache coherency through the cache invalidation algorithms based on broadcasts. Conventional cache invalidation algorithms based on synchronous broadcasts or asynchronous broadcasts have both advantages and disadvantages. The paper discusses the conventional algorithms and some existing algorithms and puts forward an improved adaptive cache invalidation algorithm based on counters, it integrates the two conventional algorithms, and it can improve the cache utilization rate and reduce the communication overheads.

**Key words:** invalidation report; cache; cache coherency; mobile database

(上接第 53 页)在开放式布局的广场而言,除了在设计规划阶段尽量避免外,对于已建成的广场只能依靠交通引导设计、交通管理手段对进出广场的交通进行引导,从而缓解交通拥堵状况。

参考文献:

- [1] CJJ37—90, 城市道路设计规范[S].
- [2] GB50220—95, 城市道路交通规划设计规范[S].
- [3] DGJ08—7—2005, 建筑工程交通设计及停车库(场)设置标准[S].
- [4] 周商吾. 交通工程[M]. 上海: 同济大学出版社, 1987.
- [5] HCM 1985.

## The Comparative Analysis of Traffic Impact of Different Plazas in Different Road Net Model

PAN Xiao-dong, LIN Tao

(Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai 200092, China)

**Abstract:** This paper studies two plaza projects in NingBo and SuZhou to do some research of their different aspects in the process of traffic impact analysis, such as traffic arrangement, exits and entrances of parking, determination of traffic volume influence area, lots of inner intersections, use of the underground space. Based on the analysis, we discuss the extent to which the two type plazas influence the around traffic environment, and the reason that leads to those problems.

**Key words:** close plaza; on-limits plaza; road net model; traffic impact analysis