

# 基于经济模型的网格 cache 的文件副本替换策略

沈 薇<sup>1</sup>,周 宇<sup>2</sup>,刘方爱<sup>1</sup>

SHEN Wei<sup>1</sup>,ZHOU Yu<sup>2</sup>,LIU Fang-ai<sup>1</sup>

1.山东师范大学 信息科学与工程学院,济南 250014

2.青岛高校软控股份有限公司 成型机研究所,山东 青岛 266045

1.School of Information Science and Engineering,Shandong Normal University, Ji'nan 250014, China

2.Tyre Building Machine Institute, Qingdao Mesnac Company, Qingdao, Shandong 266045, China

E-mail:swxq@163.com

SHEN Wei,ZHOU Yu,LIU Fang-ai.File replacement strategy for grid cache based on economic model.Computer Engineering and Applications,2007,43(21):112-114.

**Abstract:** In order to ensure efficient and fast access such distributed data,the grid system adopts data replication.However,when the space is limited,the files have the lowest value must be replaced.The paper proposes the file replacement strategy for grid cache based on economic model,fully used the way which in the economic auctioned to cause auctioneer's income to be biggest.The results simulated with Optorsim show that the proposed strategy results in a notable improvement over traditional replication strategies in a grid environment.

**Key words:** grid; replica; cache; auction; economic model

**摘 要:** 网格系统中,采用复制策略来保证有效快速地读取大量分布的数据。然而每个节点提供存储副本的空间是有限的,当空间不够时就要用替换策略把价值最低的替换出去。提出了一种基于经济模型的替换策略,充分利用了经济学中拍卖的方式来使拍卖人的收益最大。用 Optorsim 模拟实验结果表明在网格环境下,与其他传统的替换策略相比可以有效地提高系统的性能。

**关键词:** 网格;副本;cache;拍卖;经济模型

**文章编号:**1002-8331(2007)21-0112-03 **文献标识码:**A **中图分类号:**TP393

## 1 引言

网格计算是一个新出现的范例,它的目的是提供地理上分布的、有硬件和软件资源的用户组成的虚拟组织,并允许有效计算资源和存储资源。当今许多国际的项目正致力于实现数据网格<sup>[1,2]</sup>,通过计算密集型应用来分析大量分布的数据。数据网络连接地理上分布的各种计算资源和存储资源,以便有效地共享数据和资源。数据网格使不同大学和研究机构的科学家们相互合作来解决大规模计算问题。数据网格提供地理上分布的资源来解决一些大规模的数据密集型应用产生大量的数据集的问题。然而,高效地读取如此大的和广泛分布的数据由于 Internet 的高延迟而受到阻碍。

在网格中支持快速数据读取的主要阻碍是广域网和 Internet 的高延迟,并且影响整个网格系统的可扩展性。为了克制这些阻碍,大量数据需要复制在广泛分布的多个节点上。在分布式计算、Internet 和数据库系统中,复制是一个已经发展得很好的领域。

然而,节点中存储空间的数量和大小是有限的,因此一个好的替换算法对于复制技术的性能和有效性是非常重要的。在文中,提出了一种基于经济模型的副本替换算法,充分利用经

济学原理中的各种模型,运用到网格 cache 中资源的替换中,以最大化资源提供者的收益,尽量减少资源消费者的费用,以提高整个系统的性能。

## 2 相关工作

由于存储空间的限制,当节点上 cache 中剩余的空间越来越小,达到一定的阈值时,该节点的 I/O 性能将急速下降。因此当一个新的副本到来时,当节点上的 cache 剩余空间达到一定的阈值时,副本替换算法将被调用来删除 cache 中副本以腾出空间来存放新到达的副本。传统的副本替换算法有最少使用频率替换算法 LFU(Least Frequently Used algorithm),最近最少使用替换算法 LRU(Least Recently Used algorithm),先进先出替换算法 FIFO (First In First Out) 和随机替换算法(Random Replacement algorithm)。

LFU 算法:该算法在一定的时间窗口内,替换最少读取次数的副本。这个算法基于当前最近时间窗口的状态,因此不能严格地反映长时间的读取状态。

LRU 算法:该算法替换最近最长时间未使用的副本,这个算法对最近的副本有利。然而,同 LFU 一样,在一定的时间窗

**基金项目:**国家自然科学基金(the National Natural Science Foundation of China under Grant No.90612003, No.60373063)。

**作者简介:**沈薇(1981-),女,硕士研究生,主要研究方向:网格计算、数据网格、并行计算;周宇(1980-),男,主要研究方向:计算机网络,自动控制;刘方爱(1962-),男,博士生导师,主要研究方向:分布式处理、并行算法、光互联网络路由算法、网格计算。

口内不能预测以后的副本访问情况。该算法仅仅辨别较早到来的副本。

**FIFO 算法:**该算法把副本按照到来的先后顺序组织成队列,先到来的副本放在队列头部,后到的副本放在队列的尾部,每次有新的副本到来时,替换队列头部最先到达的副本。同 LRU 一样辨别较早到来的副本。同上两种方法一样,不能预测将来副本的访问情况。

**随机替换算法:**该算法从 cache 中已有的副本中随机选择一个副本。该算法的优点是硬件实现比较容易,缺点是以后访问次数较多的副本可能被替换出去,从而引起整个网格系统的吞吐量下降和大量的带宽消耗,有时还会引起颤动。

**理想型淘汰算法 (OPT, Optimal Replacement algorithm):**该算法淘汰在访问串中将来再也不出现的或者是在离当前最远的位置上出现的副本。这样,淘汰掉该副本将不会造成因需要访问该副本又立即把它调入的现象。遗憾的是,这种策略无法实现,但它可以用来衡量其他策略的优劣。

基于上述几种替换算法的优点和缺点,提出了一种基于经济模型的副本替换算法。在基于经济模型的网格 cache 替换策略中,网格中的资源相当于市场中的货物,它们被需要访问它们的代理所购买,目标是本地代理的收益尽量大。

### 3 基于经济模型的替换策略

#### 3.1 网格 cache 的引入

网格系统是一种无缝、集成的资源共享和协作环境,其目标是协调网格中资源的使用,以便及时地响应网格用户的资源请求。但是,网络传输会造成相当大的延迟,网络传输成了网格性能提升的瓶颈,为了尽量减少由于传输所造成的延迟,提高网格用户资源请求的响应时间,在网格系统中引入了 cache 技术<sup>[3]</sup>。对常用的资源进行缓存,从而减少传输所造成的时延,使网格的性能得以提高。同时,网格中不同的资源,其访问代价不同。对于访问代价高的资源在本地进行缓存,则减少了再次访问该资源时的二次花费<sup>[4]</sup>。

网格 cache 指存储能力强于本地磁盘弱于主存结点的暂时性存储<sup>[5]</sup>。使用 cache 可以减少多个任务使用相同资源时的重复传输,既减少了带宽占用、缓解中央服务器的负载,也降低了异构系统间的传输造成的处理开销。

#### 3.2 基于经济模型的替换算法

本文提出的经济模型的目的是利用拍卖协议来选择 cache 中最没有利用价值的副本来进行替换。因此我们使用 Vickrey 拍卖协议<sup>[6]</sup>。Vickrey 拍卖是第二价格秘密投标的拍卖形式,它包含一个独自の协商机制,每个投标者都提交一个投标价格给拍卖人。其他的投标者并不知道每个投标者提交的竞标价格。拍卖的胜利者是提交最高投标价格的投标者,但只支付第二高的竞标价格。与其他拍卖方式相比,这种拍卖方式的一个优点是每个投标者都投出他们真实的价格。

在网格中的每个节点上有一个 Replica Manager,用来负责副本的管理。Replica Manager 负责副本在本节点上的创建、删除、替换等工作。在文中,它起到拍卖人的作用,而本节点中的 cache 中存放的每个文件副本都有一个存储代理 (Storage Broker),存储代理充当投标人的角色。然而 Replica Manager 启动拍卖线程在本文中并不是拍卖一种商品,而是要买一种商品。商品在文中指 cache 中存放的文件副本。网格中各个节点采用

的结构如图 1 所示。Cache 中的每个文件副本提出它们想要卖的竞标价格,而获得胜利的文件副本则被 Replica Manager 支付第二低的价格,也就是说,我们采用的是反 Vickrey 拍卖。

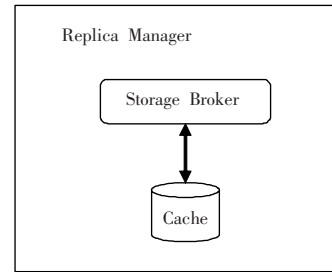


图 1 网格中节点的结构

当有新的副本到来,而 cache 的剩余空间不够时,Replica Manager 就启动拍卖线程。该线程发送一个 Callbid 消息给 Storage Broker,一旦发送了请求拍卖的消息,拍卖线程就等待各个文件副本竞争投标。经过一段固定的时间后,Replica Manager 就选择拍卖的胜利者(提交最小价格的文件副本)。当选中竞标的胜利者后,Replica Manager 就会用新到来的副本替换竞标的胜利者。

Storage Broker 收集本地节点 cache 中存放的文件副本信息,然后使用预测函数来估算文件副本的以后的收益。这个预测函数近似估算一个文件在以后一定的时间窗口内的将会被访问的次数,并假定文件过去读取的历史记录遵循 random walk。在 random walk 中,通过给定的二项式分布得到一个值,该值是通过附加的一步从当前的识别器中计算出下一个被请求文件。从过去的一个时间段  $T_1$  内,该函数使用历史的文件请求(读取方式)来预测将来一个时间段  $T$  内文件的请求次数。在我们经济模型中使用的预测函数,更多的详细资料可参见[6]。

上述过程的流程如下:

- (1) 有新的副本到达,而 cache 中剩余的空间不够时,Replica Manager 启动拍卖线程。
- (2) Storage Broker 通过预测函数计算出 cache 中每个文件副本的将来的收益,并参加竞标价格给 Replica Manager。
- (3) Replica Manager 选中胜利者,并用新副本将其替换。

### 4 算法分析与测试

网格中数据副本的放置与删除是高度动态的,由于网格 cache 的空间是一定的,当文件的副本数目达到一定值时,必须选择一个文件副本替换出去。为了使网格的整体性能达到最优,选择被替换的副本不仅是当前代价最小的,也应该是在以后的时间内代价最小的。经济模型是利用现实生活中各种拍卖机制来提出的,采用经济模型主要有两个动机:一个原因是使复制优化在分布式方式下;另一个原因是网格是高度动态的环境,资源的可用性在没有任何预警的情况下是可以改变的。通过利用经济模型,在作业执行时通过开发市场机制的动态性来制定决策。

#### 4.1 模拟配置

OptorSim 用来研究经济模型的网格模拟器。在网格环境中,替换副本的选择直接影响到用户作业的执行时间和代价,在文中,假设文件大小为 50 MB,网络拓扑如图 2 所示,为了简

化模拟,假设没有竞争的网络流量。实验的目的是证明基于经济模型的替换策略可以有效地减小用户作业的执行时间和代价,并能提高整个网络系统的性能。

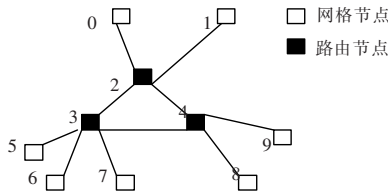


图2 网络的拓扑结构

各节点的配置如下：

1	2	200	0	0	50	0	0	0	0	0	0	0	0	0	0
1	3	100	0	0	50	0	0	0	0	0	0	0	0	0	0
1	1	200	50	50	0	50	50	0	0	0	0	0	0	0	0
1	1	200	0	0	50	0	50	50	50	50	0	0	0	0	0
1	3	200	0	0	50	50	0	0	0	0	0	50	50	0	0
1	2	200	0	0	0	50	0	0	0	0	0	0	0	0	0
1	2	300	0	0	0	50	0	0	0	0	0	0	0	0	0
1	1	100	0	0	0	50	0	0	0	0	0	0	0	0	0
1	2	300	0	0	0	0	50	0	0	0	0	0	0	0	0
1	1	100	0	0	0	0	50	0	0	0	0	0	0	0	0

第1列表示每个节点上CE的个数,第2列表示每个节点上SE的个数,第3列表示每个节点上SE的大小,单位为MB,以后表示各个节点之间的带宽矩阵,假定各节点间的带宽均为50 MB/s。令 initial file distribution=1,说明开始时文件的副本在节点1上。用户作业请求文件顺序由作业的读取方式决定,我们考虑以下方式:顺序方式(所有文件按预先规定的顺序访问)。模拟运行作业从1个到100个,每隔5秒提交一个作业。

基于经济模型的副本替换算法与LRU(Least Recently Used)算法相比较,LRU中文件总是被复制到作业执行的节点,如果节点上的存储空间满,则删除最近最少读取的文件。图3显示作业数变化时,两种算法的平均作业的执行时间,图4显示随着时间的推移,网络cache的使用情况。

实验结果表明,虽然在开始的一段时间内基于经济模型的

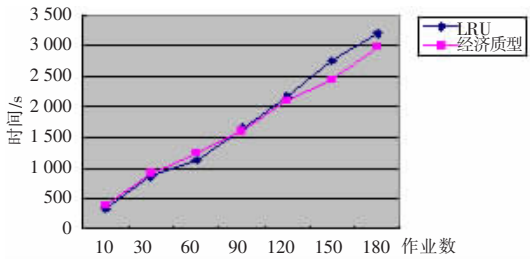


图3 两种算法的平均作业运行时间

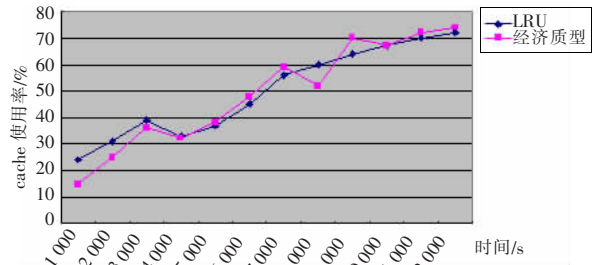


图4 网络cache的使用情况

替换策略在作业的执行时间没有 LRU 好,基于经济模型的替换算法的网络 cache 的利用率由于数据集和存储能力的特点的原因其稳定性没有 LRU 算法好,但随着运行时间的增加,基于经济模型的副本替换算法选择被替换的副本是传输代价和存储代价最小的,大大减少了后继作业的运行时间。在没有抖动的情况下,基于经济模型的副本替换算法有很好的鲁棒性和可靠性。

### 5 结论

文中,讨论了基于经济模型的网络 cache 替换管理策略。利用经济学中的拍卖机制,使拍卖人的收益最大化,竞标人的花费最小化。使用反 Vickrey 拍卖机制来避免投标人投出假的价格来获得拍卖人的信任,并通过使用预测函数来估算文件的将来收益。模拟结果显示通过系统不断地学习,使整个系统的总体性能不断提高。(收稿日期:2007年3月)

### 参考文献:

- [1] Cervenak A,Foster I,Kesselman C,et al.The data grid:towards an architecture for the distributed management and analysis of large science datasets[J].Journal of Network and Computer Applications, 2001,23.
- [2] Hoschek W,Jean-Martinez J,Samar A,et al.Data management in an international data grid project [C]//IEEE/ACM Int Workshop on Grid Computing(Grid'2000),Bangalore,India,2000-12.
- [3] 陈梅,都志辉.网络 cache 若干问题分析[J].计算机科学,2004,31(5):15-17.
- [4] Carman M,Zini F,Serafini L.Towards an economy-based optimisation of file access and replication on a Data Grid(CCGRID'02)[C].Berlin,Germany,May 21-24,2002.
- [5] Vickrey W.Counterspeculation,Auctions,and Competitive Sealed Tenders[J].The Journal of Finance,1961,16(1):8-37.
- [6] Capozza L,Stockinger K,Zini F.Preliminary evaluation of revenue prediction functions for economically-effective file replication DataGrid-02-TED-020724[R].CERN,Geneva,Switzerland,2002-07.

(上接 102 页)

for PCC:dynamic storage allocation [C]//Proc 2003 European Symposium on Programming(ESOP'03).Warsaw,Poland,2003.

[5] Feng X,Shao Z,Vaynberg A,et al.Modular verification of assembly code with stack-based control abstractions [C]//Proceedings of the 2006 ACM SIGPLAN Conference on Programming Language Design and Implementation,2006:401-414.

[6] Hoare C A R.An axiomatic basis for computer programming[J].

Communications of the ACM,1969,12(10):576-580.

[7] Bertot Y,Casteran P.Coq'Art:the calculus of inductive constructions [M].Berlin:Springer-Verlag,2004.

[8] The Coq development team.The Coq Proof Assistant Reference Manual[EB/OL].http://coq.inria.fr/.

[9] Design of a Certifying Compiler Supporting the Proof of Program Safety[EB/OL].http://sbg.ustcsz.edu.cn/lss/.