数据网格的存储资源预留方法

曲明成[1]，吴翔虎[1]，张 银[1]，廖明宏[2]，杨孝宗[1]，左德承[1]

（1. 哈尔滨工业大学 计算机科学与技术学院，哈尔滨 150001，qumingcheng@126.com；
2. 厦门大学 软件学院，福建 厦门 361005）

摘 要： 为了解决将二元法资源能力预留策略直接应用到数据网格的存储资源能力上所引起的接纳率较低、资源能力碎片过多、存储空间浪费严重等问题，通过分析数据网格应用中存储空间预留的特点，提出了四元法资源能力预留策略. 传统的时间槽算法无法适用于四元法资源能力预留，进而提出了使用基本阶梯法和阈值阶梯法能很好的解决此问题. 实验表明本文的资源能力预留方法具有更高的资源使用率和接纳率，而阈值阶梯法能够适应传输速度动态变化的网络特性.

关键词： 数据网格 QoS；存储资源预留；资源能力

中图分类号： TP311 **文献标志码：** A **文章编号：** 0367－6234(2010)03－0432－05

Storage resource reservation method for data grid

QU Ming-cheng[1], WU Xiang-hu[1], ZHANG Yin[1], LIAO Ming-hong[2], YANG Xiao-zong[1], ZUO De-cheng[1]

(1. School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China, qumingcheng@126.com; 2. Software School, Xiamen University, Xiamen 361005, China)

Abstract: To solve the problems of low acceptance rate, too much resource capacity debris, serious waste of storage space caused by directly applying binary resource capacity reservation strategy to storage resource capacity reservation (SRCR) in data grid, by analyzing the characteristics of SRCR for data grid application, a four-tuple reservation strategy (FTRS) is proposed. The traditional time slot algorithm for binary strategy can not be applied to FTRS, so the basic trapezoidal algorithm and threshold trapezoidal algorithm (TTA) are put forward. The experimental results show that, compared with binary strategy, FTRS achieves higher resource utilization and acceptance rate, meanwhile, TTA can adapt to network speed jitter.

Key words: QoS of data grid; storage resource reservation; resource capacity

现有的针对存储空间、计算能力和网络带宽的资源能力预留其主要解决方案为：通过在资源能力二维坐标空间中将一段时间内的资源能力预留抽象为矩形区域，最为有效的方法为基于时间槽的资源能力预留法[1-4]，本文中称其为二元法(REC). 数据网格资源能力预留可以表示成存储空间与时间的二元组，即 < StorageSpace, Time >，该资源能力同样为矩形区域，所以 StorageSpace 与 Time 为确定的数值. 如图 1 所示，网格的存储资源量总数为 M，有两个资源预留分别为 $R_1(h_1, t_1)$，$R_2(h_2, t_2)$，在图 1 中分别为两个矩形区域，而资源能力的大小为矩形区域的面积.

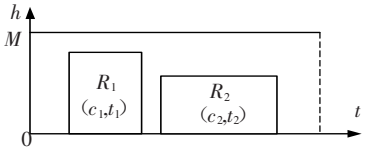


图 1 REC 法资源能力预留

REC 法在数据网格的资源能力预留中性能较差. 比如：欧洲原子能数据网格是为原子能实验特别设计的，其实验时以 10 M/s 的速度产生数据，并对其存储，而后以 1 M/s 的速度处理数据[5]. 而在其他很多时候也存在同样的问题，比

收稿日期： 2009－03－16.
基金项目： 黑龙江省重点科技资助项目(GC08A108).
作者简介： 曲明成(1980—)，男，博士研究生；
廖明宏(1966—)，男，教授，博士生导师；
杨孝宗(1939—)，男，教授，博士生导师；
左德承(1972—)，男，教授，博士生导师.

如目前将数据网格与计算网格相结合，在计算前从数据网格中预留一定的存储数据，将数据以一定的速度存储到预留的空间，待计算条件具备时以一定的速度进行处理，并随时删除处理过的数据，释放存储空间[6]. 为提升网络的下载速度，很多副本部署方法也呈现同样的特性[7-8].

针对应用场景和存在的问题，将资源能力的预留表示成四元组 $R = < VP, VC, H, t >$，其中，VP 为数据生产速度，VC 为数据消耗速度，H 为存储空间总需求量，t 为起始时间. 这样资源能力的预留将可以表示成图 2(a)，称这种预留方法为四元法(TR). 从图 2(b)中可以看出，如果使用 REC 法，这两个资源能力预留是无法被同时接纳的. 显然 TR 使用的资源能力较 REC 小很多，如此必可以提升资源能力的使用率、资源能力预留请求的接纳率，同时可以减少资源能力碎片，从而提升数据数据网格资源能力预留的整体性能.

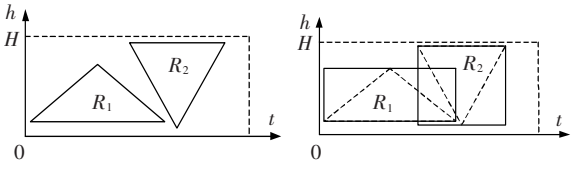


图 2　TR 资源能力预留与 REC 的比较

1　基本阶梯法(BTR)

1.1　基本定义

定义 1(资源能力空间)　在二维坐标系中，以时间 t 为横坐标，数据网格的可用存储空间 h 为纵坐标，称 $h = M$(存储空间总量) 与 $h = 0$(横坐标轴) 围成的无限空间为资源能力空间.

定义 2(资源能力三角形，资源三角形)　令 V_0，V_1 分别为一次资源能力预留中数据的生产速度和消耗速度，总存储空间需求为 M(总数据量)，已知 t_0, V_0, V_1, M. 则资源能力预留的几何图形可以表示为图 3(a)，其中 V_0, V_1 分别为三角形两条边的斜率，称这样的三角形为资源能力三角形. 其中，t_0 为预留起始时刻，t_1 为达到预留量时刻，t_2 为空间使用完毕时刻. 两个不与 t 坐标轴平行的边的函数分别为 $h_1 = V_0(t - t_0), h_2 = -V_1(t - t_2)$.

定义 3(资源能力阶梯三角形，阶梯三角形)　如图3(b)所示，将资源三角形以 t_1 为界分割成两个三角形，将区间 (t_0, t_1) 与 (t_1, t_2) 以时间单位 r 进行分割，其分割的份数分别为 n_1 和 n_2. 对于左侧三角形每个小区间的面积 $S_r^i = rV_0(t_1 + ir - t_0)$，而令 $S'_1 = \sum_{i=1}^{n_1} S_r^i$，右侧三角形每个小区间的面积 $S_r^j = -rV_0(t_1 + jr - t_2) = rV_0(t_2 - jr - t_1)$，令 $S''_1 = \sum_{j=1}^{n_2} S_r^j$，令 $S_T = S'_1 + S''_1$，称由 S_r^i 和 S_r^j 构成的区域为资源能力阶梯三角形.

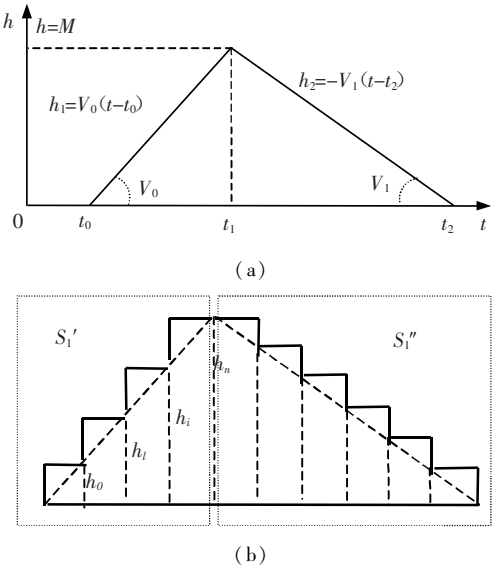


图 3　资源能力三角形与阶梯三角形

定义 4(资源能力使用率)　定义一段时间内所接纳的资源能力预留的总和与该段时间时间内资源能力空间大小的比值为资源能力使用率(U). 令某个阶梯三角形的大小为 S_T^i，REC 法资源能力大小为 S_R^i. 其对应的 $U_T = \sum S_T^i/\pi, U_R = \sum S_R^i/\pi$，$\pi$ 为某段时间内的资源能力空间大小.

1.2　公式推导

为了能够有效地计算某个资源能力预留请求是否可以被接纳，理想的情况下是资源能力空间在资源能力预留时刻处有足够的空间可以容纳下资源能力三角形(REC 法为矩形). 给出 S_T, S_R, U_T, U_R 的求解过程，求解时参考图 3 进行几何推导.

令时间单位长度为 r，将 $t_1 - t_0$ 分成 $n_1 = \frac{t_1 - t_0}{r}, n_2 = \frac{t_2 - t_1}{r}$ 份，则 $h_i = ir \cdot \frac{h}{t_1 - t_0}$，并假设区间 $(t_0, t_1), (t_1, t_2)$ 之间有整数个单位时间间隔. t_1, t_2 的求解为

$$\begin{cases} \frac{h}{t_1 - t_0} = V_0 \\ \frac{h}{t_2 - t_1} = V_1 \end{cases} \Rightarrow \begin{cases} t_1 = t_0 + \frac{h}{V_0} \\ t_2 = t_0 + \frac{h}{V_0} + \frac{h}{V_1}. \end{cases} \tag{1}$$

由式(1) 可以得出 S_R 为

$$S_R = h(t_2 - t_0) = h\left(t_0 + \frac{h}{V_0} + \frac{h}{V_1} - t_0\right) = h^2\left(\frac{1}{V_0} + \frac{1}{V_1}\right). \tag{2}$$

$$h_i = irV_0,\ h_j = (n_2 - j)rV_1. \tag{3}$$

为了求解 S_T,需先求出 S_1',S_1'',由式(1)推出:$n_1 = \frac{t_1 - t_0}{r} = \frac{h}{V_0 r}$,$n_2 = \frac{t_2 - t_1}{r} = \frac{h}{V_1 r}$,由此可以推出:

$$S_1' = r\sum_{i=1}^{n_1}\left(ir \cdot \frac{h}{t_1 - t_0}\right) = \frac{r^2 h}{t_1 - t_0}\sum_{i=1}^{n_1} i = \frac{r^2 h}{t_1 - t_0} \cdot \frac{n_1(n_1+1)}{2} = \frac{r^2 h}{t_1 - t_0} \cdot \frac{1}{2}\frac{t_1 - t_0}{r}\left(\frac{t_1 - t_0}{r} + 1\right) = \frac{1}{2}rh\left(\frac{t_1 - t_0}{r} + 1\right) = \frac{1}{2}h(t_1 - t_0) + \frac{1}{2}rh.$$

$S_1'' = \frac{1}{2}h(t_2 - t_1) + \frac{1}{2}rh$,由此得出

$$S_T = S_1' + S_1'' = \frac{1}{2}h(t_2 - t_0) + rh = \frac{h^2}{2}\left(\frac{1}{V_0} + \frac{1}{V_1}\right) + rh. \quad (4)$$

由定义 4 和式(2),式(4)得出

$$U_R = \sum\left(h_i^2\left(\frac{1}{V_0^i} + \frac{1}{V_1^i}\right)\right)/\pi,$$

$$U_T = \sum\left(\frac{h_i^2}{2}\left(\frac{1}{V_0^i} + \frac{1}{V_1^i}\right) + rh_i\right)/\pi. \quad (5)$$

2 阈值阶梯法(VTR)

由于网络速度的动态变化,基本阶梯法的资源能力预留存在一定的局限性(要求网络速度平稳),因此应为生产与消耗速度引入一定的速度变化区间.

定义 5(可靠阈值 ε) 生产与消耗速度实际中会存在偏差,在基本阶梯法中引入速度偏差,即生产与消耗速度分别位于区间$(V_0 - \varepsilon, V_0 + \varepsilon)$与$(V_1 - \varepsilon, V_1 + \varepsilon)$中,称 ε 为可靠阈值.

定义 6(阈值梯形) 在基本阶梯法中引入可靠阈值后,即相当于在图3(a)的几何图形上将非平行 t 轴的两条三角形边的斜率分别 $\pm\varepsilon$,而后形成的图形为一个梯形,如图 4 中的由 t_0, t_2, A, B 4 个点构成的图形,称其为阈值梯形.

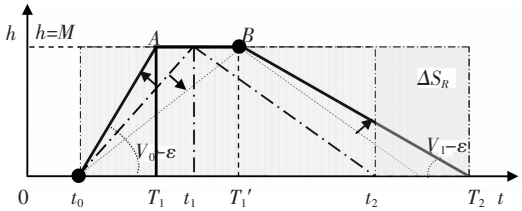


图 4 阈值梯形图

阈值梯形的面积为资源能力预留的数量,令其为 S_T',其面积可以分解为3部分,如图5(a),即三角形 $\Delta t_0 T_1 A(S_a)$,矩形 $T_1 T_1' AB(S_b)$,三角形 $\Delta T_1' T_2 B(S_c)$. 将3个区域重新组合成图5(b),并对新的大三角形按照基本阶梯法进行处理,同时根据图 4 进行推导:

$$\Delta t_2 = T_2 - t_2 = \left(t_0 + \frac{h}{V_0 - \varepsilon} + \frac{h}{V_1 - \varepsilon}\right) - \left(t_0 + \frac{h}{V_0} + \frac{h}{V_1}\right) = \frac{h\varepsilon}{V_0(V_0 - \varepsilon)} + \frac{h\varepsilon}{V_1(V_1 - \varepsilon)} \cdot$$

$$T_2 = t_0 + \frac{h}{V_0 - \varepsilon} + \frac{h}{V_1 - \varepsilon}. \quad (6)$$

$$\Delta t_1 = T_1' - T_1 = \left(t_0 + \frac{h}{V_0 - \varepsilon}\right) - \left(t_0 + \frac{h}{V_0 + \varepsilon}\right) = \frac{2h\varepsilon}{V_0^2 - \varepsilon^2}.$$

根据式(4)和图5(b)可以得出 $S_T' = (S_a + S_c) + S_b$,由此推出:

$$S_T' = \left(\frac{h^2}{2}\left(\frac{1}{V_0 + \varepsilon} + \frac{1}{V_1 - \varepsilon}\right) + rh\right) + \Delta t_1 h = \frac{2\varepsilon h^2}{V_0^2 - \varepsilon^2} + \frac{h^2}{2}\left(\frac{1}{V_0 + \varepsilon} + \frac{1}{V_1 - \varepsilon}\right) + rh. \quad (7)$$

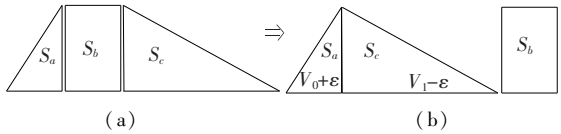


图 5 阈值梯形分解图

定理 1 由于为在基本解题法中引入可靠阈值后致使资源能力预留时间增加了 $\Delta t_2 = T_2 - t_2$,当 $V_0 - V_1 + \varepsilon > 0$ 时,如果对 REC 法进行 t_2 可靠阈值扩展,其资源能力 S_R' 增加量 $S_R > S_T = S_T' - S_T$.

证明:

$$\Delta S_R = \Delta t_2 h = h^2\varepsilon\left(\frac{1}{V_0(V_0 - \varepsilon)} + \frac{1}{V_1(V_1 - \varepsilon)}\right).$$

$$\Delta S_T = S'_T - S_T = \frac{h^2}{V_0 - \varepsilon} - \frac{h^2}{V_0 + \varepsilon} + \frac{h^2}{2}\left(\frac{1}{V_0 + \varepsilon} + \frac{1}{V_1 - \varepsilon}\right) + rh - \frac{h^2}{2}\left(\frac{1}{V_0} + \frac{1}{V_1}\right).$$

$$\Delta S_T - \Delta S_R = \frac{\varepsilon h^2}{2}\left(\frac{1}{V_1(V_1 - \varepsilon)} - \frac{1}{V_0(V_0 + \varepsilon)}\right) + rh = \frac{\varepsilon h^2}{2} \cdot \frac{(V_0 + V_1)(V_0 + \varepsilon - V_1)}{V_1 V_0 (V_1 - \varepsilon)(V_0 + \varepsilon)} + rh.$$

因为 $V_0 - V_1 + \varepsilon > 0$,所以 $\Delta S_T - \Delta S_R > 0$,证毕.

定理 1 说明,在时间增加相同的情况下针对 REC 法进行可靠扩展的代价要大于阈值阶梯法.

由式(2),式(5)推出:

$$S_R' = (T_2 - t_0)h = h^2\left(\frac{1}{V_0 - \varepsilon} + \frac{1}{V_1 - \varepsilon}\right). \quad (8)$$

由式(7),式(8)推出:

$$U'_R = \sum \left(\frac{h_i^2}{V_0^i - \varepsilon} - \frac{h_i^2}{V_0^i + \varepsilon} + \frac{h_i^2}{2} \left(\frac{1}{V_0^i + \varepsilon} + \frac{1}{V_1^i - \varepsilon} \right) + rh_i \right) / \pi,$$

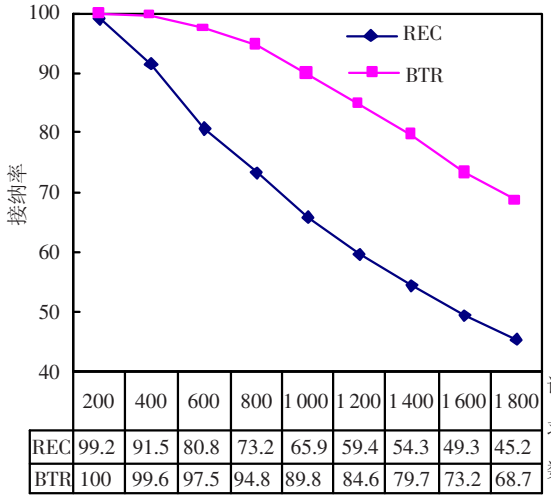
$$U'_T = \sum \left(h_i^2 \left(\frac{1}{V_0^i - \varepsilon} + \frac{1}{V_1^i - \varepsilon} \right) \right) / \pi. \quad (9)$$

3　仿真实验

令资源能力空间具有 2 500 个时间单位，资源最大数量为 4 000，在一次测试中，针对该段资源能力空间随机的产生不同数量的资源能力预留请求. 判断某请求是否可以被接纳的条件为：该请求的任意一个 h_i 都能被其所对应的时间单位内的资源能力空间所接纳. 实验中将测试 REC 法与 BRT\VTR 法的资源能力预留请求的接纳率、资源的使用率，并对其进行对比分析. 同时模拟速度动态变化，以检测不同的 ε 对已预留的资源能力使用失效的影响.

3.1　基本阶梯法性能测试

在一定的资源能力空间中，对 REC 法与 BTR 法接纳率与资源能力使用率进行比较. 从图 6 可以看出随着请求数的增加两种方法的接纳率都成下降趋势，但是 BTR 法接纳率明显较 REC 法的高. 而从图 7 中可以看出，在接纳率相同时 BTR 法的资源能力使用率明显高于 BTR 法.

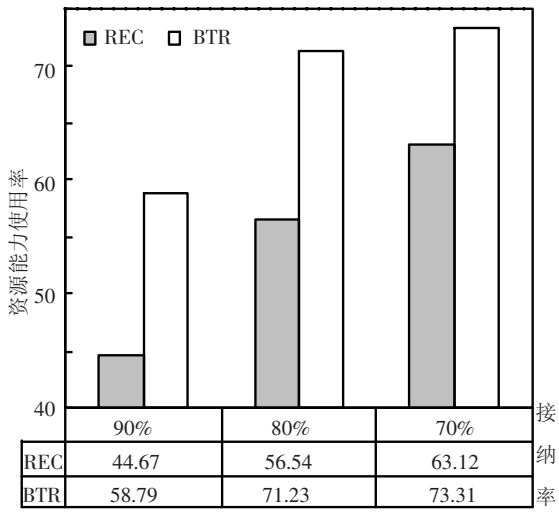


	200	400	600	800	1 000	1 200	1 400	1 600	1 800
REC	99.2	91.5	80.8	73.2	65.9	59.4	54.3	49.3	45.2
BTR	100	99.6	97.5	94.8	89.8	84.6	79.7	73.2	68.7

图 6　不同资源能力请求 REC 与 BTR 接纳率比较

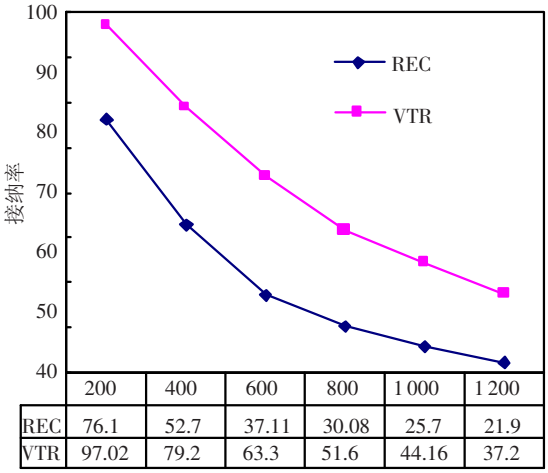
3.2　阈值阶梯法性能测试

在一定的资源能力空间中，当 $\varepsilon = 0$ 时，对 REC 法与 VTR 法接纳率与资源能力使用率进行比较. 从图 8 可以看出随着资源能力预留请求数的增加两种方法的接纳率都成下降趋势，但是 VTR 法接纳率明显较 REC 法的高. 而从图 9 中可以看出，在接纳率相同时 VTR 法的资源能力使用率明显高于 BTR 法.



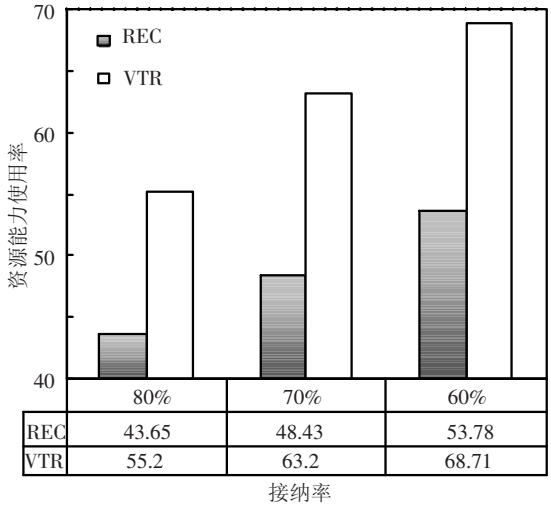
	90%	80%	70%
REC	44.67	56.54	63.12
BTR	58.79	71.23	73.31

图 7　相同接纳率 REC 与 BTR 资源能力使用率比较



	200	400	600	800	1 000	1 200
REC	76.1	52.7	37.11	30.08	25.7	21.9
VTR	97.02	79.2	63.3	51.6	44.16	37.2

图 8　不同资源能力请求 REC 与 VTR 接纳率比较



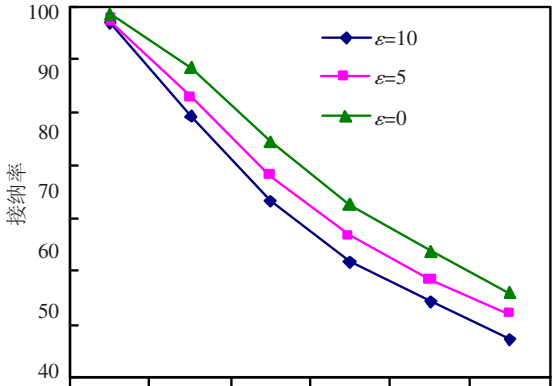
	80%	70%	60%
REC	43.65	48.43	53.78
VTR	55.2	63.2	68.71

图 9　相同接纳率 REC 与 VTR 资源能力使用率比较

3.3　综合比较与预留失效检测

令 $\varepsilon = 0$、5、10，观测请求量与接纳率的变化

曲线,从图 10 中可以看出较小的 ε 产生较高的接纳率. 令生产与消耗速度分别在(0.8 V,1.2 V)之间随机的变化. 在整个资源能力空间中,如果某个时间单位的实际资源能力需求量超出网格的资源总量那么算作一次失效,搜索请求量为 600 时的失效时间单位数,并将总数与总的资源能力空间时间单位数进行比值,得出失效单位数 f. 针对不同 ε 取值分别进行 10 次实验,取平均值,绘出柱形图如图 11 所示,从图 11 中可以看到,当 ε = 4 时,f 已经为 0,效果较好.



	200	400	600	800	1 000	1 200
ε=10	97.02	79.2	63.3	51.6	44.16	37.2
ε=5	97.2	82.98	67.92	56.82	48.18	41.78
ε=0	98.6	88.32	74.46	62.54	53.92	45.96

图 10 ε 对 VTR 接纳率的影响

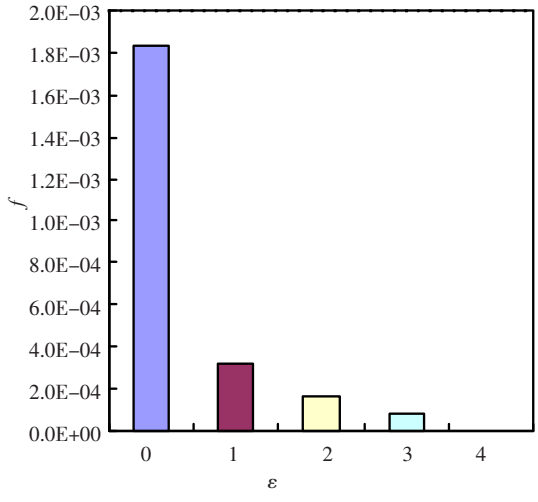


图 11 ε 对 VTR 预留失效的影响

3.4 实验小结

① 采用 BTR 与 VTR 方法进行资源能力预留较传统的 REC 法有更好的性能;② 随着 ε 的增加 VTR 的性能逐渐降低;③ 在 ε 达到合理的数值时预留失效可以降低为 0;④ 采用合理的 ε 值,可以大幅的提升资源能力预留请求的接纳率和资源能力的使用率,并可以降低或消除预留失效.

4 结 论

1)提出了基本阶梯法和阈值阶梯法很好的实现了四元法预留策略,而阈值阶梯法更能适应传输速度动态变化的网络特性.

2)从实验可以看出:与二元法相比四元法能够有效降低资源能力碎片,提升存储空间的使用率,进而提高资源能力预留的接纳率和网格系统的整体吞吐量.

参考文献:

[1] BURCHARD L O. On the performance of computer networks with advance reservation mechanisms[C]//Proceedings of the 11th IEEE Intl. Conference on Networks (ICON'03). Washington: IEEE Computer Society, 2003:449 - 454.

[2] HU Chunming, HUAI Jinpeng, WO Tianyu. Flexible resource reservation using slack time for service grid [C]//Proceedings of the 12th International Conference on Parallel and Distributed Systems (ICPADS'06). Washington: IEEE Computer Society, 2006:327 - 334.

[3] 胡春明, 怀进鹏, 沃天宇. 一种基于松弛时间的服务网格资源能力预留机制[J]. 计算机研究与发展, 2007, 44(1):20 - 28.

[4] XIAO Peng, HU Zhigang, LI Xi, *et al.* A novel statistic-based relaxed grid resource reservation strategy [C]//Proceedings of the 9th International Conference for Young Computer Scientists. Washington: IEEE Computer Society, 2008:703 - 707.

[5] Large Scale System Configuration Workshop. CERN and the DataGrid Project[EB/OL]. http://homepages.inf.ed.ac.uk/group/lssconf/files2001/datagrid-edinburgh.pdf.

[6] MCCLATCHEY R, ANJUM A, STOCKINGER H, *et al.* Data intensive and network aware (diana) grid scheduling[J]. Journal of Grid Computing, 2007, 5(1):43 - 64.

[7] SHI Ke. A replication and cache based distributed metadata management system for data grid[C]//Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing. Washington: IEEE Computer Society, 2007: 20 - 25.

[8] YUAN Yulai, WU Yongwei, YANG Guangwen, *et al.* Dynamic data replication based on local optimization principle in data grid[C]//The 6th International Conference on Grid and Cooperative Computing (GCC 2007). Washington: IEEE Computer Society, 2007: 815 - 822.

(编辑 张 红)