1999年度日本オペレーションズ・リサーチ学会
秋季研究発表会

Lookahead Scheduling Requests for Multi-size Page Caching

01107924 神戸商科大学 木庭淳 KINIWA Jun
01503164 神戸商科大学 濱田年男 HAMADA Toshio
ネットワンシステムズ 溝口大輔 MIZOGUCHI Daisuke

1 Introduction

This paper studies the effects of reordering page requests for multi-size page caching. First, we consider semi-online model, where the input queue keeps requests which will be processed in the future. We develop an efficient page replacement algorithm which shifts some page requests in the queue under the model. Second, we analyze the miss ratios of two methods, non-shifting method and our shifting method. Third, we compare our method with other algorithms by simulation. The results indicate our method outperforms others.

2 Algorithm

Our system has a *cache* of size M, a *secondary storage* of large enough size, and an input *queue* for requests with infinite storage. Let $U = \{x, y, \ldots\}$ be the set of all *pages*, where each page size is uniformly distributed over $(0, m]$. If a requested page is not in the cache, called a *miss*, it incurs mean cost C. Otherwise, called a *hit*, it takes just 1. We consider *semi-online model*, where the queue length $L(t)$ at time t is determined by

$$\begin{aligned} L(t) &= |\text{arrived requests before } t| \\ &\quad - |\text{served requests before } t|. \end{aligned}$$

The *residence interval* of a page a is the maximal time interval during which a is in the cache.

For each future miss, our shifting method makes a plan which page should be replaced. Let $EC_j(r)$, called an *expected cache*, be a set of pages in the cache corresponding to the j-th future miss r. Let $a(EC_j(r))$ be the number of distinct pages in the $EC_j(r)$ for which requests have already arrived after the miss r. If $a(EC) \geq K$ for some constant K, the algorithm randomly selects a set of pages S to be evicted from $a(EC)$. Then the future requests for S in the queue are shifted to their residence intervals.

- Making a plan of evictions :
  1. If K pages $S \subset EC(q)$ are requested,
  2. Randomly select page(s) $a \in S$ to be evicted
  3. Shift every $a \in S$ before q
  4. Next EC is determined
- Carrying out the plan

3 Analysis of Miss Ratios

Non-shifting Method

Theorem 3.1 *If C is sufficiently large, the miss ratio f_n of the non-shifting method is approximately*

$$f_n \approx 1 - \frac{\rho^2}{C^2\{(1-\rho)L - \rho^2\}},$$

where L is the mean queue length and $\rho = \lambda/\mu$ is the ratio of arrival rate to service rate. ∎

Shifting Method

Lemma 3.1 *The expected number of evicted pages for each EC, denoted by α, is*

$$\begin{aligned} \alpha &= \sum_{j \geq 1} j \cdot \sqrt{\frac{2M/m - j + 1}{12}} \cdot \Phi(f(j)) \\ &\quad \times \{a(\Phi(a) - \Phi(b)) + (\phi(a) - \phi(b))\}, \end{aligned}$$

where $\Phi(x)$ is the probability that the standard normal is less than x, $\phi(x)$ is the density function of the standard normal distribution, $f(j) = (M - (2M/m - j + 1)m/2)/\sqrt{(2M/m - j + 1) \cdot m^2/12}$, $a = \sqrt{\frac{12}{2M/m-j+1}}(\frac{M}{m} - \frac{2M/m-j+1}{2} - 1)$ and $b = \sqrt{\frac{12}{2M/m-j+1}}(\frac{M}{m} - \frac{2M/m-j+1}{2})$. ∎

We can associate several states with each page, called *page state*. Each page state of page a and its transitions are illustrated in Figure 1. Let $\mathcal{S}$ be a set of *system states*. Any system state $\boldsymbol{x} \in \mathcal{S}$ is composed of the $(|U| + 1)$-dimensional vector.

Let $e(\boldsymbol{x}, t)$ be the number of ECs at time t, $p(\boldsymbol{y}, \boldsymbol{x})$ the transition probability, and $\Delta g(\boldsymbol{y}, \boldsymbol{x})$

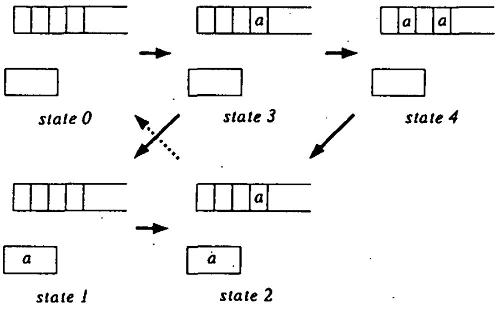

Figure 1: Page state transition

the increment of ECs when the system state moves from y to x. Then we have

$$e(x, t + \Delta t) = \sum_{y \in S} p(y, x) \cdot \{e(y, t) + \Delta g(y, x)\}.$$

If $\Delta t \to 0$, we obtain

$$e(x) = \sum_{\substack{y \neq x \\ y \in S}} \frac{q(y, x)}{\lambda + f_s \mu} \{e(y) + \Delta g(y, x)\}, \qquad (1)$$

where $q(y, x)$ is equal to $p(y, x)$ except for containing Δt as a factor.

Let $s(x)$ denote the probability of x. Then we have

$$s(x) = \sum_{\substack{y \neq x \\ y \in S}} \frac{q(y, x) s(y)}{\lambda + f_s \mu} \quad \text{with} \quad \sum_{x \in S} s(x) = 1. \qquad (2)$$

From (1) and (2), the expected number of ECs, denoted by β, can be obtained ;

$$\beta = \sum_{x \in S} s(x) e(x).$$

From the lemmas above, the shifting rate is $\alpha \cdot \beta / L$. Now we can get a miss ratio f_s for our method.

Theorem 3.2 *If C is sufficiently large, the miss ratio of our shifting method is approximately*

$$f_s \approx f_n - \frac{\alpha \cdot \beta}{L},$$

where L is the mean queue length. ∎

4 Simulation

To evaluate our shifting method, we execute simulation experiments and compare with other two methods. One is Randomized Marking algorithm, proposed by [1], and the other is its slight variation, called Revised Marking.

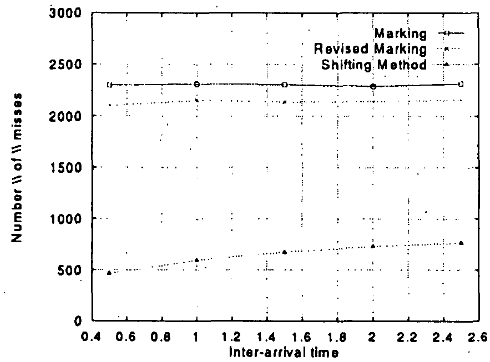

Figure 2: Varying mean inter-arrival time

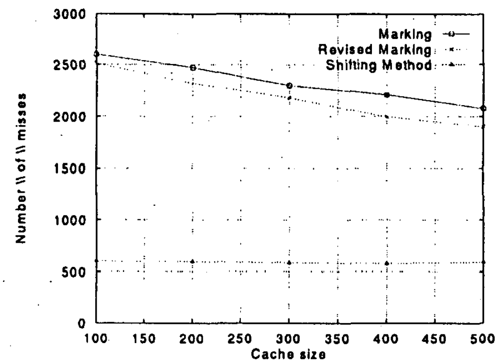

Figure 3: Varying cache size

5 Conclusion

We investigated effects of lookahead scheduling requests on the caching problem. From the analytical or experimental points of view, it turned out our method ourperforms non-shifting methods.

References

[1] S.Irani, "Page Replacement with Multi-Size Pages and Applications to Web Caching," *In Proceedings 29th ACM Symposium on Theory of Computing*, pp.701–710, 1997.