Applied Mathematics & Information Sciences 5 (2) (2011),79S-88S
– An International Journal

A Cache Model of the Block Correlations Directed Cache Replacement Algorithm

Zhu Xudong
School of Computer Science and Information Engineering, Zhejiang Gongshang University
Hangzhou, China

Email Address: zhuxd@zjgsu.edu.cn

Received Oct 10, 2010; Revised Jan 19, 2011

This paper proposes a cache model of the Block Correlations Directed cache replacement algorithm (BCD). Since the predication success rate of accesses decides the BCD's hit ratio, the cache model analysis the utilization of the cache space improved by the correct predications and the penalty incurred by incorrect predications to estimate the hit rate of BCD. The model can also optimize the parameters of BCD to achieve theoretical optimum of such spatial locality based likely replacement algorithms. For most workloads of real systems the model of BCD can reduce deviation by 2.1%-21.8%.

Keywords: Cache model, Spatial locality.

1 Introduction

Block correlations are common semantic patterns in storage systems and can be used to direct performance optimization of storage systems. The correlated blocks usually are accessed by requests which close to each other in a stream. For example, there is a block correlation {*abc*}. When block *a* and *b* are accessed one after another, block c is likely to be accessed soon. References [1-2] try to explore the block correlations in a request stream.

Recently some studies [4-5], like BCD [3], propose cache replacement policies which use spatial locality in block correlations to increase the hit ratio. To achieve theoretical optimum of the replacement algorithms, BCD predicts the of blocks' spatial locality in the future by history and runtime information. BCD places blocks in the stack of cache by the prediction results to improve the utilization of the cache and reduce the penalty of prediction failure.

The predication success rate decides the hit rate of BCD. This paper proposes a cache model of BCD to achieve theoretical optimum of spatial locality based replacement algorithms. The cache model analyses the utilization of the cache space improved by the correct predications and the penalty incurred by incorrect predications to estimate the hit rate of BCD. For most of real system workloads the model of BCD can reduce deviation

by 2.1%-21.8%.

The paper is organized as follows. In the next section we introduce the BCD. Section 3 discusses the cache model. Section 4 presents our experimental results and Section 5 concludes the paper.

2. Basic policy and key problems

Figure 1 is a block correlations directed cache system structure which includes prefetching and cache replacement component. The cache system does not limit location in the system. It can be used in client end, storage server or storage area network. The prefetching component uses block correlations to direct prefetching policy [1]. This paper discusses the policy of cache replacement component.

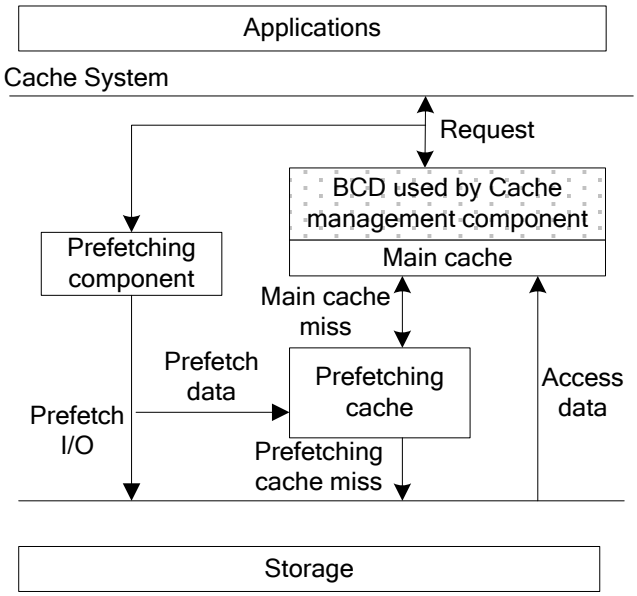


Figure 1. Block correlations directed cache system structure

To indentify the access characteristics we classify the requests as follow.

1. **Prefetched Request, PR**: Besides the first request requests in a block correlations are prefetched requests.

2. **Non-Prefetched Request, NPR**: (a) request that does not access in a pattern. (b) The first request of a pattern.

1. How to identify request type online. Traditional spatial-locality based algorithms can identify only sequential access and random access. One method is to get information of the current request by communication between replacement component and

prefetching component. It is not general for a cache system. A simple and efficient request identification mechanism is an issue to be solved in the study.

Table 1. Influence of request-type prediction on cache-miss ratio
{ABCD} is frequent access pattern and uses LRU and MRU policies respectively for the blocks with NPR and PR predicted

Req	Pref	LRU		BCD (inaccurate prediction)			BCD (accurate prediction)		
		Stack state	Miss times	Stack state	Miss times	Type	Stack state	Miss times	Type
E			1		1	NPR		1	NPR
F		E	2	E	2	NPR	E	2	NPR
A		FE	3	FE	3	NPR	FE	3	NPR
B	B	AFE	3	AFE	3	PR	AFE	3	PR
C	C	BAFE	3	AFEB	3	PR	AFEB	3	PR
D	D	CBAF	3	AFEC	3	PR	AFEC	3	PR
E		DCBA	4	AFED	3	NPR	AFED	3	NPR
F		EDCB	5	EAFD	3	NPR	EAFD	3	NPR
B		FEDC	6	FEAD	4	PR	FEAD	4	NPR
C		BFED	7	FEAB	5	PR	BFEA	5	NPR
B		CBFE	8	FEAC	6	PR	CBFE	5	NPR
C		BDFE	9	FEAB	7	PR	BCFE	5	NPR

2. How to predict accurately the future request type of the block. A block can be accessed by different types of requests. Accurate prediction of the future-request type for each block can effectively reduce the cache-miss ratio. On the contrary wrong prediction increases the cache-miss ratio. As shown in Table 1, if blocks to be accessed by NPR in future are mistakenly predicted to be accessed by PR, the block is soon replaced from the cache. It directly results in cache miss. Similarly, if a PR is predicted as NPR, the block accessed by it stays in the cache for long time and reduces the utilization of cache space. Table 1 shows the influence of accurate prediction of request type on cache-miss ratio. In addition prediction of the algorithm must have very low overhead. How to predict the future access type of blocks accurately and rapidly is a key problem to be solved.

3. How to reduce penalty incurred by failure of prediction. If BCD chooses LRU or MRU policy simply based on prediction results, failure in prediction results in high cache-miss ratio. BCD should consider how to reduce the penalty in case of prediction failure. The overhead of the algorithm should be also taken into account. For example, DULO's sorting operation produces a larger overhead. How to place blocks based on the prediction results and tolerate the failure of prediction is another important goal of this paper

3. The BCD Scheme

BCD adjusts the residence time of the block in the cache to maximize the NPR hit ratio and to reduce the penalty of failure of prediction. The blocks that are predicted being accessed by different request types should have different residence times in the cache.

The block is replaced out of the cache from the stack bottom (LRU end). If the block is not accessed again, the minimum residence time of a block in the cache is equal to the distance from the current position to the stack bottom. There are two sections in BCD's stack and the depths of them are respectively T_h and T_l. The blocks depart through the bottom of high section and enter the top of the low one. The blocks are replaced at the bottom of the low section. Therefore blocks at the high section have longer residence times than blocks at the low section.

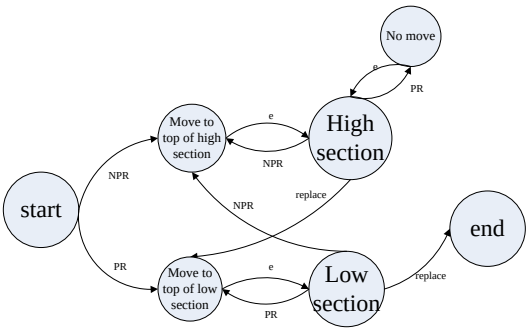


Figure 3. State conversion of a block in BCD

If a block is predicted to be accessed by NPR, BCD places it at the top of the high section. BCD uses two policies to reduce the penalty of the prediction failure. Firstly BCD keeps temporal locality of NPR requests. If the block is hit at the high section and is predicted to be accessed by PR next, BCD does not change the position of the block in the stack. If the type of next request for this block is NPR, the prediction fails, but the block still can stay in the cache for a longer time. Secondly, If the block is not in the high section and it is predicted to be accessed by PR next time, BCD places it at the top of the low section to keep the block for a period to reduce the penalty of prediction failure.

In general Figure 3 shows the state conversion of a block in the BCD cache.

4. Cache model of BCD

In this section we build the BCD model to analyze the influence of various factors on cache effect and direct configuration of BCD parameter. Let cache capacity T be a constant greater than 0. Let d be a reference distance of the block which represents the number of other blocks accessed between two adjacent accessors to the same block (the block repeatedly accessed during the interval is counted once only).

Theorem 1: Let T be the size of the cache, $f(d)$ be the proportion of requests that reference them at a distance d. Then miss ratio of LRU algorithm is

$$M_{\text{lru}} \approx 1 - \int_0^T f(d)\mathrm{d}(d) .$$

Proof: According to LRU, the block can be hit in the cache when the reference distance is smaller than T. Therefore the hit ratio of LRU is

$$H_{\text{lru}} = \sum_{0}^{T} f(d)$$

When the number of requests is large enough, $f(d)$ tends to be 0, so

$$H_{\text{lru}} = \sum_{0}^{T} f(d) \approx \int_{0}^{T} f(d)\mathrm{d}(d)$$

and accordingly

$$M_{\text{lru}} = 1 - H_{\text{lru}} \approx 1 - \int_{0}^{T} f(d)\mathrm{d}(d) .$$

Let *S* be the total count of requests in the storage system. Let h_{main} be the hit times in the main cache, $h_{\text{main-pr}}$ be the hit times of PRs in main cache and let $h_{\text{main-npr}}$ be the hit times of NPRs in the main cache. Then $h_{\text{main}} = h_{\text{main-pr}} + h_{\text{main-npr}}$. Let S_{pr} be the number of PRs in S. The cache system miss ratio is

$$\begin{aligned} M &= (S - h_{\text{main}} - (S_{\text{pr}} - h_{\text{main-pr}}))/S \\ &= 1 - (h_{\text{main-npr}} + S_{\text{pr}})/S \end{aligned} \tag{1}$$

Since S_{pr} is irrelevant to the replacement policy, *M* is decided by $h_{\text{main-npr}}$. The block of NPRs resides in the main cache as long as possible to increase $h_{\text{main-npr}}$. It is more complicated to use BCD in the storage system. The following problems and challenges are encountered.

All the PRs are hit in the main cache. Let $f_{\text{npr}}(d)$ be the probability density function of the reference distance. Let PRs of the workloads be $P_{\text{pr}} = S_{\text{pr}}/S$, as $P_{\text{npr}} = 1 - S_{\text{pr}}/S$. According to Formula (1), if T_1=0 and the predictions are all correct, BCD has the theoretical optimum cache miss ratio (OPT):

$$\begin{aligned} M_{\text{opt}} &= 1 - (S_{\text{pr}} + h_{\text{main-npr}})/S \\ &= 1 - P_{\text{pr}} - P_{\text{npr}} \int_{0}^{T} f_{\text{npr}}(d)\mathrm{d}(d) \end{aligned} \tag{2}$$

The reference distance distribution of workloads and NPR sequence can be approximately obtained by linear conversion of the exponential distribution. Figure 4 and Figure 5 are respectively reference distance distribution of Cello92 and OLTP. Analysis of Cello96, Cello99 and TCP-C can also obtain the similar results. Let the probability density function of reference distance of I/O stream be $f(d) = b_1 f'(d) = b_1 \lambda_1 e^{-\lambda_1 d}$, the distribution function of reference distance of I/O stream can be obtained as:

$$F(d)=\begin{cases} a_1-b_1e^{-\lambda_1 d}, & d>0 \\ 0, & d=0 \end{cases} \tag{3}$$

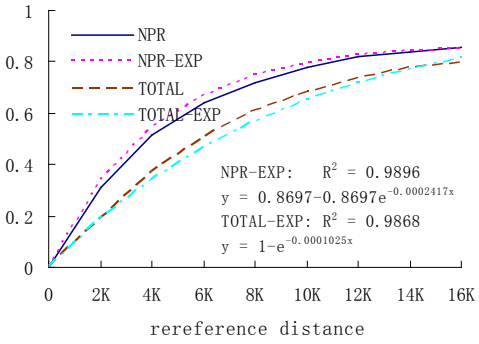


Figure 4. Reference distance distribution of Cello92

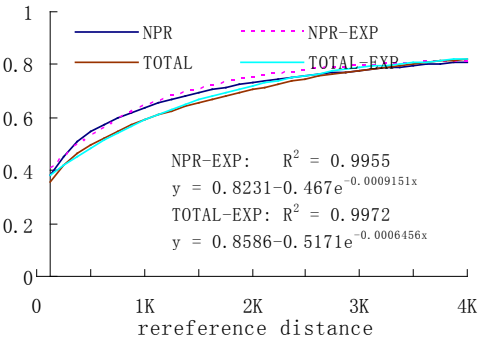


Figure 5. Reference distance distribution of OLTP

Similarly let the distribution function of Reference distance of NPR sequence be

$$F_{\text{npr}}(d)=\begin{cases} a_2-b_2e^{-\lambda_2 d}, & d>0 \\ 0, & d=0 \end{cases} \tag{4}$$

The LRU's miss ratio is $M_{\text{lru}}=1-(a_1-b_1e^{-\lambda_1 T})$ according to Theorem 1 and the OPT's miss ratio is $M_{\text{opt}}=1-P_{\text{pr}}-P_{\text{npr}}(a_2-b_2e^{-\lambda_2 T})$ according to Formula 2.

Considering failure of prediction in the predicted NRP request sequence the probability density function of NPR reference distance is $f(\gamma,d)$, of which $\gamma=\text{P}_{\text{npr}}*(1-\alpha)/(\text{P}_{\text{npr}}*(1-\alpha)+\text{P}_{\text{pr}}*\beta)$, representing the proportion of correctly predicted NPR in all the predicted NRP requests. In the sequence formed with predicted NPR requests, simply suppose reference distance distribution of the real NPR has an approximately linear relationship with the proportion of real NPRs in the sequence:

$$F(\gamma,T_h)=\int_0^{T_h} f(\gamma,d)=F_{npr}(T_h)-\frac{1-\gamma}{1-P_{npr}}*(F_{npr}(T_h)-F(T_h))$$
$$=\frac{(\gamma-P_{npr})*F_{npr}(T_h)+(1-\gamma)*F(T_h)}{1-P_{npr}}$$

In case of prediction failure let Type 1 error be a mistake of predicting NPR by PR and let Type 2 error be a mistake of predicting PR by NPR. Let the probability of Type 1 error be α and the probability of Type 2 error be β. The requests hit in BCD cache are from four parts: PRs, correctly predicted NPR requests, NPR request that is predicted RP, but the block was not moved and NPR request that is predicted RP, but the block was placed in low section. The BCD's miss ratio is

$$\begin{aligned}M_{BCD}(T_l)\approx 1-&P_{pr}\\&-P_{npr}(1-\alpha)(F(\gamma,T_h)+F(T)-F(T_h))\\&-P_{npr}\alpha P_{npr}F(T)(1-P_{pr}F(T_h))\\&-P_{npr}\alpha P_{pr}F(T_l)\end{aligned}\qquad(5)$$

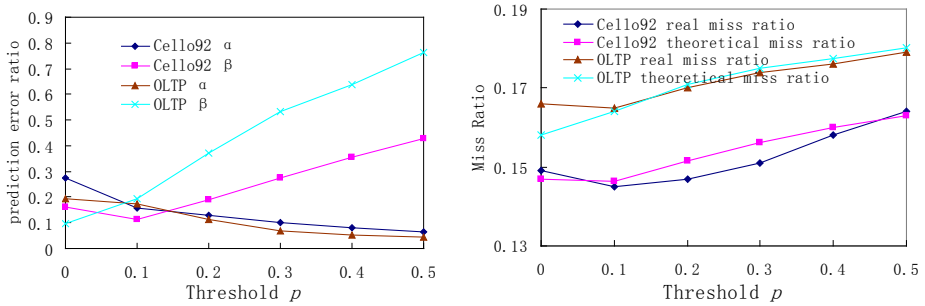
With Formula 5 the low section cache capacity can be calculated to minimize the miss ratio of BCD.

5. Test and evaluation

Our simulator uses block correlations explored by C-Miner [1] and uses LRU to manage the main cache. The simulated disk specification is similar to that of the 10000 RPM IBM Ultrastar 36Z15.

In our experiments we use only the first half part of the trace to mine block correlations. We uses block relations to direct prefetching and evaluate the performance of BCD.

This section discusses influence of predicting on BCD performance and validate BCD model with the example of Cello92 and OLTP. According to discussion in Section 4, threshold p is used to distinguish block type. Let $p\in[0,0.5]$. If $p_i \le p$, block i is defined as LPB. If $p_i > (1-p)$, block i is defined as MPB.



(a) Influence of threshold p on prediction error (b) Influence of threshold p on miss ratio

Figure 6. Influence of prediction on BCD miss ratio.

BCD can accurately predict request type. The ratio of Type 1 prediction error α and the ratio of Type 2 prediction error β can both be smaller than 0.2. Figure 6(a) shows prediction error ratio for Cello92 and OLTP under different thresholds p. It can be seen in the figure that α and β of Cello92 decrease with increase of p when p <0.1. When p is greater than 0.1, α continues decreasing to 0.1 and below. Most of NPR requests are accurately predicted. When p is greater than 0.1, β rises to 44% with increase of p. Lots of PRs are predicted as NPR so that blocks accessed by these requests are placed to the high section of the cache and the residence time of these blocks is reduced. α of OLTP slowly decreases with the increase of p and β rapidly increases with the increase of p.

Miss ratio of BCD rises with increase of prediction error ratio. Figure 6(b) shows miss ratio of BCD under different values of p. Miss ratio of Cello92 is the lowest when $p = 0.1$ because α is decreased with the increase of p when $p \le 0.1$. When $p \ge 0.1$, α decreases slowly while β increases rapidly with increase of p. Since p becomes the dominant factor affecting miss ratio, the miss ratio of BCD rises. Similarly miss ratio of OLTP is the lowest when $p = 0.1$.

Figure 7 shows the influence of low section size on BCD miss ratio. With the increase of low section capacity the miss ratio of Cello92 rises because BCD can accurately predict request type on Cello92 and the failure possibility is very small. Therefore increase of low section capacity cannot reduces many miss times in case of prediction failure while decrease of high section capacity reduces residence time of blocks accessed by NPR, resulting in an increase of miss ratio. When low section capacity is equal to main cache capacity, BCD retrogresses into an approximate LRU algorithm. Differently from Cello92, when low section capacity for OLTP is 0, miss ratio of BCD is greater than that of LRU and rapidly decreases with increase of low section capacity. This is because many reference distances of OLTP are all small. Type 1 prediction error directly results in cache miss while a smaller capacity of low section can significantly decrease l misses caused by Type 1 prediction error. When low section capacity continues to increase, the miss ratio of BCD decreases more slowly because the number of requests the reference distances of which are smaller than low section capacity increases slowly. When capacity of low section is greater than 1/8 of the main cache capacity, miss resulting from decrease of high section cache capacity becomes the dominant factor and miss ratio begins increasing again.

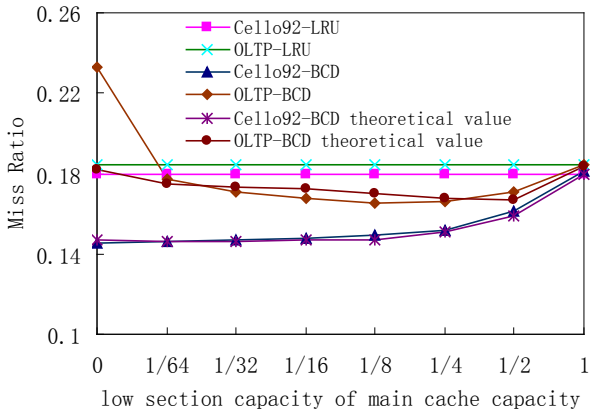


Figure 7. Influence of low section capacity on BCD

To validate BCD model Figure 6 and Figure 7 show the theoretical miss ratio calculated by the model. It can be seen that maximum relative errors between theoretical and actual test values are no more than 21.8% and mean relative errors are 2.1%. BCD model can effectively describe relations between factors of BCD cache replacement algorithm and miss ratio. BCD model can be used to direct the optimization of the low section capacity. In Figure 10, when capacity of low section is 0, there is a large error between the OLTP theoretical miss ratio and actual miss ratio because lots of reference distance is very short in OLTP. When size of low section approaches zero, reference distance distribution is no longer fitted with the exponential distribution function.

6. Conclusions and future work

This paper proposes a cache model of a Block Correlations Directed cache replacement algorithm. To achieve theoretical optimum of spatial locality based replacement algorithms BCD tries to increase NPR's hit ratio by predicting the request type and reducing the penalty caused by the failure of prediction. Experiments show that BCD can effectively reduce miss ratio of cache and average I/O response time.

There is still much work in the future: Firstly we will carry out the BCD on a practical system and verify its performance by more applications. Secondly self-adaption of the low section capacity to accommodate the characteristics of the workload should be learned. Finally research to reduce the average response time of BCD with optimized data layout.

Acknowledgements

This work was supported by Zhejiang Sci & Tech Project (2010C33045), Zhejiang Provincial NSF China (Y1101316).

References

[1] Zhenmin L., Zhifeng C., Sudarshan M. S., Yuanyuan Z. and *C-Miner: Mining block correlations in storage systems.* FAST'04, San Francisco, 2004, 173--186.

[2] Hsu Windsor W., Smith Alan Jay and Young Honesty C., *The automatic improvement of locality in storage systems.* ACM Trans. on Computer Systems. 2005. 23(4): 424–473.

[3] Zhu Xudong, Ke Jian, Xu Lu, *BCD: To achieve the theoretical optimum of spatial locality based cache replacement algorithm*. NAS09, Zhang Jia Jie, 2009, 269-272.

[4] Kim J. M., Choi J., Kim J. et al. *A low-overhead high-performance unified buffer management scheme that exploits sequential and looping references.* OSDI'00 San Diego, CA, Oct. 2000, 119--134.

[5] S. Jiang, X. N. Ding, F. Chen et al. *DULO: An effective buffer cache management scheme to exploit both temporal and spatial locality*. FAST'05, 2005.

Zhu Xudong received the PhD degree in Computer Architecture from the Institute of Computing Technology, Chinese Academy of Sciences (ICT/CAS) in 2009. He is currently an Assistant Professor in the Zhejiang Gongshang University. His research interests are in the areas of Computer Architecture, Distributed Systems and Storage Systems.