3J-8

複数サーバ環境における上位キャッシュを考慮した下位キャッシュ置換手法の性能評価

長廻　雄介†　　山口　実靖†

†工学院大学大学院 工学研究科 電気・電子工学専攻

1. はじめに

近年，ストレージ容量は増大を続けており，扱うデータ量も増加の一途を辿っている．それに伴う管理コスト増加が情報システムの問題の一つとなっている．そこで NAS (Network Attached Storage)や SAN (Storage Area Network)といったネットワークストレージを用いたデータの集約・一元管理により，管理コストを削減する手法が提案され広く使われている．

ネットワークストレージへのアクセスは，サーバ計算機上のキャッシュとストレージ機器のキャッシュを介して行われる．この場合，ネットワークストレージに到着する I/O 要求には従来の局所性とは性質の異なる負の参照の時間的局所性が存在し，参照の局所性を期待している LRU キャッシュ置換手法は効果的に機能しないことが既存研究[1]から分かっている．

本稿では，LRU を用いた複数のサーバ計算機がネットワークストレージにアクセスする環境に適したキャッシュ置換手法を提案する．

2. 負の参照の時間的局所性

ネットワークストレージへのアクセスは，図 1 のようにサーバ計算機キャッシュとストレージ機器キャッシュを介して行われる．サーバ計算機のキャッシュ置換手法は多くの場合 LRU 置換手法が用いられており，この場合，最近参照されたデータはサーバ計算機のキャッシュに格納される．このため，最近参照されたデータへのアクセスはサーバ計算機上で処理され，ストレージ機器に参照要求が届くことはない．従って，ネットワークストレージにおいては同じデータが近い将来再度アクセスされることはなく，通常の参照の時間的局所性とは逆の，負の参照の時間的局所性が存在することになり，従来の参照の時間的局所性を期待している LRU キャッシュ置換手法は効果的に機能しない．

図 2はシミュレーションにより計測した再参照間隔と参照確率の関係性を示したものである．

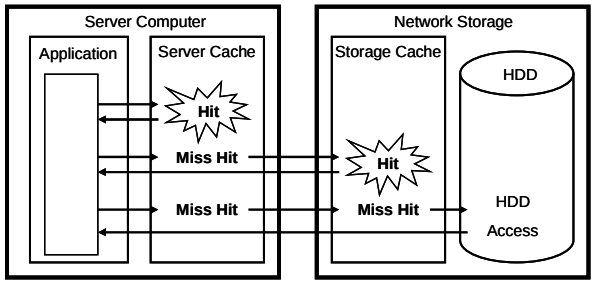

図 1 二重キャッシュ構造

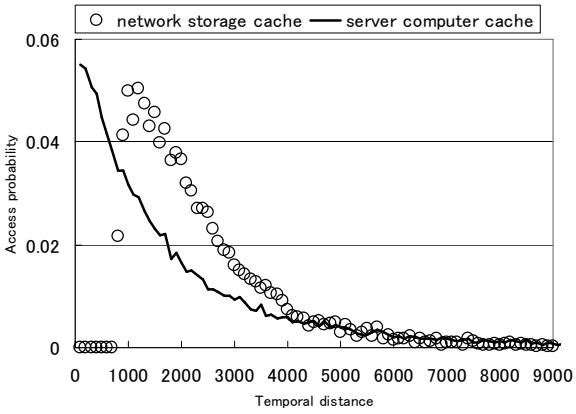

図 2 負の参照の時間的局所性

横軸は再度参照されるまでの時間を表している（単位はアクセス数）．サーバ計算機上には従来の時間的局所性が見られるが，ストレージ機器への参照要求は近い将来（約 1000 アクセス）到達しないことが分かる．これはシミュレーション上で設定したサーバ計算機のキャッシュサイズが，約 1000 アクセスで満たされるためであり，サーバ計算機のキャッシュと同等のアクセスがこない限り，一度格納されたデータはサーバ計算機のキャッシュに残り，同一データへの再アクセスはサーバ計算機で処理される．

3. 提案手法

負の参照の時間的局所性を考慮したキャッシュ置換手法として INTE を提案する．INTE ではアクセス履歴を保持し，履歴から再アクセス時間ごとのアクセス発生確率を算出し，図 2のような確率密度関数を作成する．そして，確率密度関数を積分して近い将来の再アクセスの発生確率を計算し，2 つの破棄対象候補の中で再アクセ

Performance Evaluation of Server Cache Size Aware Cache Replacement Method
Yusuke NAGASAKO†, Saneyasu YAMAGUCHI†
†Graduate school of Electrical and Electronics Engineering, Kogakuin University

ス確率が最も低いものを破棄する．破棄対象候補は最後にアクセスされてからの時間が最長のものと，最短のものの 2 個とする．最後にアクセスされてからの経過時間が a である場合，再アクセス確率は図 2の横軸 a 地点から定められた区間だけ積分することで得られる．保持しているアクセス履歴中に当該データのアクセスが 2 個以上記録されている場合は，最新の 2 アクセスを選択し両再アクセス確率の合計を当該データの再アクセス確率とする．

4. 評価

4.1. 評価手法

複数サーバ環境における提案手法の有効性を確認するため，シミュレーションによる評価を行った．本シミュレーションでは無作為に選択したサーバからランダムにデータアクセス要求を発生させ，それぞれのキャッシュ置換アルゴリズム使用時のストレージキャッシュヒット率を調査した．評価対象となる置換手法は，LRU, FIX，MQ，INTE である．LRU は最後のアクセスからの時間が最長のものを置換対象とする手法であり，FIX はキャッシュに格納したデータを置換せず保管し続ける手法であり，MQ(Multi Queue)[2]は 2 次キャッシュを考慮した置換手法で，複数の LRU リストを管理し頻度ごとに各リストに割り当て最下位リストの最も古いデータを破棄する手法であり，INTE は 3 章で提案した置換手法である．

シミュレーション条件は以下の通りである．サーバ台数は 1 台，4 台，16 台であり，各サーバのデータ領域は独立しておりサーバ間でのストレージ領域の共有は行われていない．アクセスサーバはランダム回(平均 100 の指数分布)ごとに変更し，次のアクセスサーバは平均 2 の指数分布に従う乱数により決定する．ストレージサイズは 1,000,000 ブロックであり，ストレージキャッシュサイズは 512～4096 ブロックである．サーバキャッシュサイズは，奇数番目のサーバにおいて 1024 ブロック，偶数番目のサーバにおいて 2048 である．

サーバ計算機上のアプリケーションは指数分布に従う乱数によりアクセス対象ブロックを選択しアクセス要求を発行する．アクセス確率分布のピーク領域は一様分布乱数により決定され，50,000 アクセスごとに変更される．アプリケーションによる総発行アクセス数は 500,000 アクセスである．

4.2. 性能評価

図 3，図 4，図 5にキャッシュヒット率を示す．

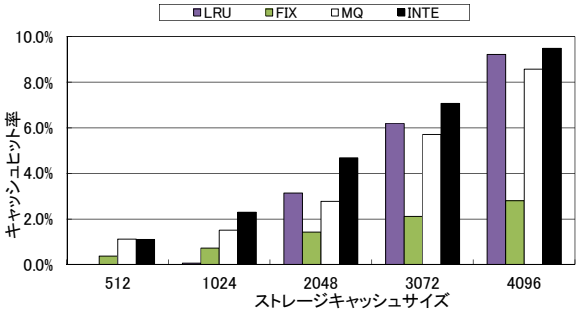

図 3キャッシュヒット率(サーバ 1 台)

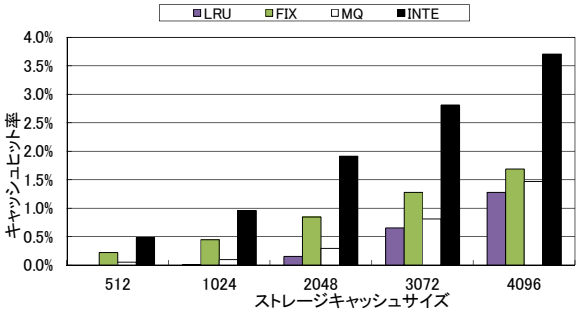

図 4キャッシュヒット率(サーバ 4 台)

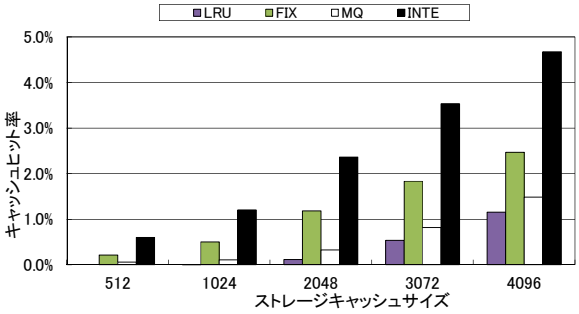

図 5　キャッシュヒット率(サーバ 16 台)

提案手法である INTE が最もキャッシュヒット率が高い事が確認できる．また LRU はストレージ機器キャッシュサイズが小さい場合キャッシュヒット率はほぼ 0 であり，サーバキャッシュサイズと比較してストレージ機器キャッシュサイズが大きくならない限り，LRU は効果的に機能していないことが分かる．

5. おわりに

本稿では，負の参照の時間的局所性を考慮した手法として INTE 置換手法を提案し，シミュレーションによりその有効性を示した．今後は，書き込み処理に関する考察，キャッシュ置換手法の OS への実装などを行う予定である．

謝辞

本研究は科研費（22700039）の助成を受けたものである．

参考文献

[1]宮野 晋平, 山口 実靖,浅谷 耕一, “多段キャッシュ型ネットワークストレージへのアクセスの時間的局所性を考慮したメモリキャッシュ制御”, 情報処理学会研究報告 マルチメディア通信と分散処理研究会報告， 2009 年 3 月

[2] Yuanyuan Zhou and James F. Philbin, Kai Li “Second-Level Buffer Cache Management,” IEEE Transactions on parallel and distributed systems, vol. 15, no. 7, JULY 2004