プライマリ・バックアップ構成を有効利用したストレージシステムの省電力効果見積

引田 諭之† 横田治夫†,††

† 東京工業大学大学院情報理工学研究科計算工学専攻
†† 東京工業大学学術国際情報センター

1 はじめに

近年，インターネット等のIT技術の普及により個人や企業で扱うデータ量が急増している．それに伴い，データセンター等では大容量のストレージが要求されており，その消費電力量の増加が大きな問題となっている．

このような状況から，従来よりストレージの省電力化については様々な研究がなされてきた．我々も，キャッシュメモリとディスクドライブの双方においてプライマリ・バックアップ構成を有効利用することで，信頼性を高めつつ，さらに省電力効果が期待できる手法を提案している．これまでに，提案手法の有効性を検証するための消費電力量の概算式を構築し，その効果の見積もりを行った[3]．

本稿では，その概算式で用いた各パラメータについて省電力化に与える影響を調べ，その影響度合に関する比較を行う．

2 提案手法

2.1 構成

本提案手法は，キャッシュメモリ，データディスク，キャッシュディスクから構成される（図1）．

キャッシュメモリは，プライマリのデータを保持するプライマリ層と，バックアップデータを保持するバックアップ層を構成する[2]．あるノードのキャッシュメモリのプライマリ層に書き込まれたデータは，それとは別ノードのキャッシュメモリのバックアップ層にデータのコピーを書き込むことにより，キャッシュメモリの信頼性を確保している．ただし，各ノードは個別の電源を有し，UPS等の断電対策が施されているものとする．

ディスクドライブは実際のデータを保持し，データの配置はキャッシュメモリと同様にプライマリ・バックアップ構成をとる．ディスクアクセスがある一定の閾値時間を超えて発生しなかった場合，そのディスクドライブの回転を停止しスタンバイ状態に移行する．

キャッシュディスクはデータをキャッシュするためのディスクドライブ群であり，少数のディスクドライブ

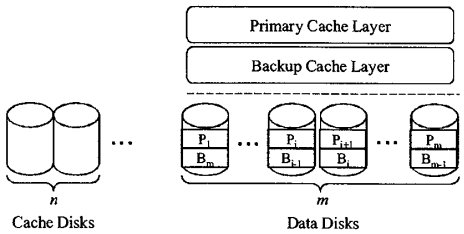

図1: 提案手法の全体構成

を読み出し処理専用のキャッシュとして用いることにより，データディスクへのアクセス頻度を低く抑える．省電力化のためにキャッシュディスクを用いる手法はMAID[1]と同様である．

2.2 動作

2.2.1 書き込み処理

書き込みデータはまずキャッシュメモリに書き込まれる．キャッシュメモリに書き込む際は，該当キャッシュメモリのプライマリ層と，それに対応する別のキャッシュメモリのバックアップ層にデータを書き込む．もしキャッシュメモリに設定されているバッファ容量の閾値を超えてしまう場合は，データディスクに書き込みを行う．該当データディスクが回転中の場合，そのままキャッシュデータを書き込むが，停止中であれば，データディスクをスピンアップさせた後でデータを書き込む．データディスクに書き込む際は，プライマリ層のデータとバックアップ層のデータを同期させて書き込む．

2.2.2 読み出し処理

読み出し要求に対する処理としては，まずキャッシュメモリに該当データがあるかを確認し，あればキャッシュメモリから読み出す．もしキャッシュメモリに存在しなければ次にキャッシュディスクを確認し，キャッシュディスク中に該当データが存在すればそこから読み出す．キャッシュディスク中にもデータが存在しない場合は，データディスクにアクセスする．該当データはプライマリディスクとバックアップディスクの両方に存在しているので，其々のディスクが回転中かどうかを確認し，回転中のディスクからデータを読み出す．両方とも回転中の場合，メモリバッファのキューが長い方のディスクから読み出す．両方とも停止中だった場合は，回転停止期間の長い方のディスクからデータ読み出す．

データを読み出した後は，対応するメモリバッファのデータをディスクに書き込み（プライマリ層，バッ

An Estimation of Disk Storage Systems Utilizing the Primary-Backup Configuration
Satoshi HIKIDA† and HaruoYOKOTA†,††
†Dept. of Computer Science, Graduate School of Information Science and Engineering, Tokyo Institute of Technology
††Global Scientific Information and Computing Center, Tokyo Institute of Technology
†hikida@de.cs.titech.ac.jp
††yokota@cs.titech.ac.jp

表 1: 概算式で用いる記号とその説明

記号	説明
n	キャッシュディスク数
m	データディスク数
$P_{standby}$	スタンバイ状態時のディスク消費電力
P_{idle}	アイドル状態時のディスク消費電力
P_{tran}	スタンバイ状態からアイドル状態へ遷移する際の消費電力
P_{read}	アクティブ状態（read）時のディスク消費電力
P_{write}	アクティブ状態（write）時のディスク消費電力
$P_{dataDisk}$	データディスク全体の消費電力
$P_{cacheDisk}$	キャッシュディスク全体の消費電力
P_{normal}	従来方式のストレージ全体の消費電力
h_c	読み出しアクセスに対するキャッシュメモリのヒット率
h_d	読み出しアクセスに対するキャッシュディスクのヒット率
b_w	メモリバッファに対する書き込み可能率
r_d	ディスクアクセス時にディスクが回転している確率

クアップ層の両方から），その後に読み出したデータおよびデータディスクに書き込んだデータをキャッシュディスクに書き込む．

3 消費電力量の見積もり

3.1 消費電力量の概算式

本提案手法および従来方式について消費電力量を見積もる為の概算式を構築した [3]．概算式で用いる記号とその説明を表 1 に示す．本提案手法における総消費電力量 P_{total} は，キャッシュディスクの消費電力 $P_{cacheDisk}$ とデータディスクの消費電力 $P_{dataDisk}$ の合計であり，それぞれの概算式は以下のようになる．

$$P_{total} = P_{dataDisk} + P_{cacheDisk} \quad (1)$$

$$\begin{aligned} P_{dataDisk} = {} & m(h_c f_r + (1-h_c)h_d f_r + b_w f_w + (1-(f_r+f_w))(r_d P_{idle} + (1-r_d)P_{standby}) \\ & + f_r m(1-h_c)(1-h_d)(r_d P_{read} + (1-r_d)(P_{read} + P_{tran})) \\ & + f_w m(1-b_w)(r_d P_{write} + (1-r_d)(P_{write} + P_{tran})) \end{aligned} \quad (2)$$

$$\begin{aligned} P_{cacheDisk} = {} & n(h_c f_r + (1-f_r))P_{idle} \\ & + f_r n(1-h_c)h_d P_{read} \\ & + f_r n(1-h_c)(1-h_d)P_{write} \\ & + f_w n(1-b_w)P_{write} \end{aligned} \quad (3)$$

また，本提案手法を用いない従来方式における消費電力量の概算式は以下のようになる．

$$P_{normal} = f_r m P_{read} + f_w m P_{write} + m(1-(f_r+f_w))P_{idle} \quad (4)$$

3.2 消費電力量の見積もり結果

h_c, h_d ,b_w, r_d の各パラメータについて，アクセス頻度 F （$F = f_r + f_w$）が高頻度（$F = 1$），中頻度（$F = 0.5$），低頻度（$F = 0.1$）の場合における，従来方式に対する消費電力量の削減率を比較する．

図 2 は，当該パラメータを変化させた時の，アクセス頻度に対する最大消費電力削減率の値をグラフに表したものである．グラフより，省電力化に一番影響を与えるパラメータは r_d であることが分かる．これは，ディスクアクセスが発生した際に該当ディスクが回転中である確率を表している．その次に影響を与えるのがメモリキャッシュのヒット率 h_c であり，キャッシュディスクのヒット率 h_d もほぼ同程度の影響を与える．

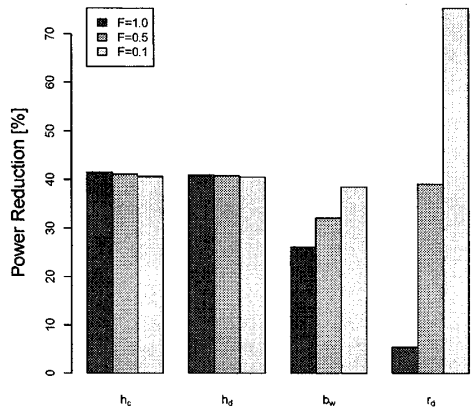

図 2: 各パラメータにおける消費電力削減率（$n = 5, m = 95$ 固定）．対象のパラメータの値を 0 から 1.0 まで変化させて消費電力削減率が最大になるようにした．それ以外のパラメータは 0.5 に固定して算出している．

4 まとめと今後の課題

プライマリ・バックアップ構成を有効利用してディスクストレージの信頼性を確保しつつ，ディスクの回転状態を考慮して省電力化する手法の提案を行った．提案手法の効果を検証するために消費電力量の概算式を用いてパラメータの影響の比較を行った．今後はシミュレーション環境を構築し，パフォーマンスに関しても検証を行う予定である．また，信頼性に関する定量的な評価も今後の課題である．

謝辞

本研究の一部は文部科学省科学研究費補助金特定領域研究（#21013017）の助成により行われた．

参考文献

[1] Dennis Colarelli and Dirk Grunwald. Massive arrays of idle disks for storage archives. In *Supercomputing '02: Proceedings of the 2002 ACM/IEEE conference on Supercomputing*, pp. 1–11, Los Alamitos, CA, USA, 2002. IEEE Computer Society Press.

[2] Hui-I Hsiao and David J. DeWitt. Chained declustering: A new availability strategy for multiprocessor database machines. In *Proceedings of the Sixth International Conference on Data Engineering*, pp. 456–465, Washington, DC, USA, 1990. IEEE Computer Society.

[3] 引田諭之, 横田治夫. プライマリ・バックアップ構成を有効利用したストレージシステムの省電力化手法の提案, *DEIM*, 2010.