

仮想ストライピングを用いた RAID5 型ディスクアレイに於ける 更新処理方式の性能解析

3B-7

茂木 和彦 喜連川 優
東京大学 生産技術研究所

1 はじめに

2次記憶装置の高性能化と高信頼化を目的とした RAID5 型ディスクアレイの開発が進められている。RAID5 型では、高信頼化のためにパリティを記録しており、データの更新時にはデータブロックに加えてパリティブロックも更新する必要があるため、性能低下が生じ問題となっている。

この性能の低下を抑えるために「仮想ストライピング」[1]なる記憶管理法を提案した。仮想ストライピングではパリティストライプを単位とするパリティストライプの再編成により書き込み性能の向上を図っている。本稿では、シリンダを単位とするパリティストライプの再編成を行なう LFS[2]との性能比較を行なう。

2 RAID5 型ディスクアレイの高性能化手法

2.1 パリティストライプの動的割り付け

従来の RAID5 型ディスクアレイでパリティ更新のコストが高いのは、古いデータを用いてパリティを計算していることによる。そこで、書き込み時にストライプの動的割り付けを行ない、新たに書き込まれるデータのみで新しいパリティストライプを作成する。 n 台のデータディスクから構成される RAID5 型ディスクアレイに於いて、 n ブロックの更新が行なわれる場合、従来の方式では総計で $n \times (2D + 2P)$ アクセスが必要であった。一方、ストライプの動的割り付けを行なうと、書き込まれるデータのみでパリティは計算されるため、必要なアクセスは総計で $n \times D + P$ と大きく減少する。また、書き込むデータは場所を移動するため、同一シリンダへまとめて書き込むことが可能となり、書き込みの効率を向上させることが可能となる。

しかし、別の場所に移ったものの元のブロックのデータ(ダーティブロック)は、元のパリティストライプでパリティの計算に使われているため、内容を保存する必要があり、そこへデータを書き込むことはできない。このため、ダーティブロックを集めて新たに書き込み可能なパリティストライプを作成する(ガーベジコレクション)必要がある。本稿では、動的割り付けのためのパ

ティストライプの再編成を実行する方式として仮想ストライピングと LFS の2つの方式を取り上げる。これらは書き込み時に於けるストライプ生成ならびにガーベジコレクションの方式が異なっており対比させて説明する。

2.2 仮想ストライピング (VS)

仮想ストライピングでは、ストライプの仮想化を行う。これにより、物理的な場所の移動を伴うことなくストライプの組替えができ、効率的なガーベジコレクションが可能となる。動的割り付けはストライプ単位で行ない、1シリンダ分まとまったところで書き込みを行なう。ガーベジコレクションは次のように行なわれる。はじめに元となるストライプ(ヴィクティムストライプ)を決定し、その中のデータの存在するブロック(アクティブブロック)とパートナーストライプ(置換え相手を含むストライプ)のダーティブロックを置換える。この置換には、置換を行なうアクティブブロックとダーティブロックの読み出しとパリティ更新のためのパリティの読み書きの計4回のアクセスが必要である。しかし、ダーティブロックが多いものをヴィクティムストライプに選択することにより、置換に必要なアクセス数を減らすことができる。

2.3 LFS

LFS は高書き込みスループットを特徴としたファイルシステムとして開発されたが、今回は RAID5 型ディスクアレイへの適用を考える。LFS では、連続した空き領域を確保し、そこへ一括して書き込みを行うことにより、書き込みの高スループット化を図っている。この書き込みのための空き領域を確保するために、あるセグメント(データ管理の単位)内のデータを全て読み出し、別のセグメントに空きを詰めて書き直す作業を実行する。このガーベジコレクション操作はセグメントクリーニングと呼ばれ、この操作により動的割り付けのための領域を確保する。セグメントクリーニングは大きな負荷を伴うので、他への影響と効率を考慮してセグメントの大きさを決定する必要がある。

3 更新性能の評価

仮想ストライピング(シリンダ単位ガーベジコレクション)・LFS(1シリンダ/セグメント)・LFS(1/2シリンダ/セグメント)の3種類の方式について評価を行なった。

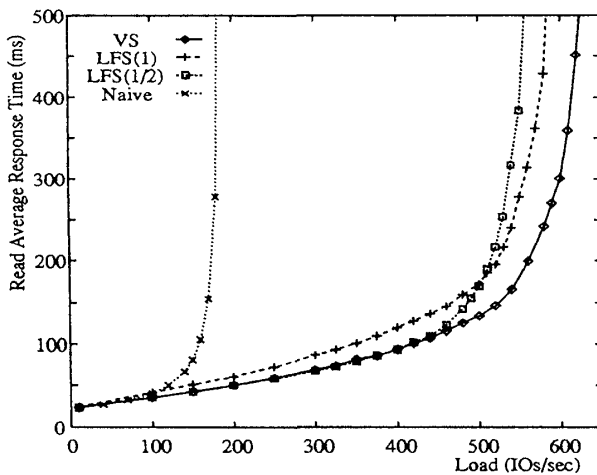


図 1: 静的負荷解析 (ディスク使用率 80%, R:W=1:1)

	VS	LFS(1)	LFS(1/2)
Read (ms)	19.97	19.51	19.93
Write (ms)	6.92	3.51	3.92
GC (ms)	8.06	13.13	14.39
W+GC (ms)	14.98	16.64	18.31

表 1: 1 アクセスあたりの占有時間 (400 IOs/sec)

8D+P 構成に於いて、ディスク使用率が 80%、R:W = 1:1 の静的負荷を与えた場合の読み込みの平均レスポンスタイムを図 1 に示す。仮想ストライピングは、低負荷では LFS(1/2) と同程度の性能を示し、高負荷への耐久性では全ての方式の中で一番良い。以下、このような特性が得られる理由について考察する。

3.1 書き込みコスト

それぞれの方式について、上記の場合の 400 IOs/sec の負荷に於いて、1 つの空ブロックの作成と、1 ブロックのデータの読み込み・書き込み (パリティ書き込みも含む) に必要なディスクの占有時間を表 1 に示す。全ての方式で、書き込みのコストは空きの作成を含めても読み込みより小さくなる。

仮想ストライピングはダークブロックが多いストライプを優先的に用いて空きストライプを作成するため、空きストライプ作成のためのアクセス数は少なく、1 ブロック当たりの占有時間は小さい。しかし、作成された空きブロックは連続して配置されていないために書き込みのコストは大きくなる。

一方の LFS は空きブロックを作成するためにセグメント内のアクティブデータをすべて読み出し、更に書き込む必要がある。そのため、空きブロックの作成に大きな時間を取られる。一方、連続領域に書き込みを実行するため、書き込みに必要な時間は小さい。なお、セグメントが大きいほど書き込みと空き領域の作成は効率的

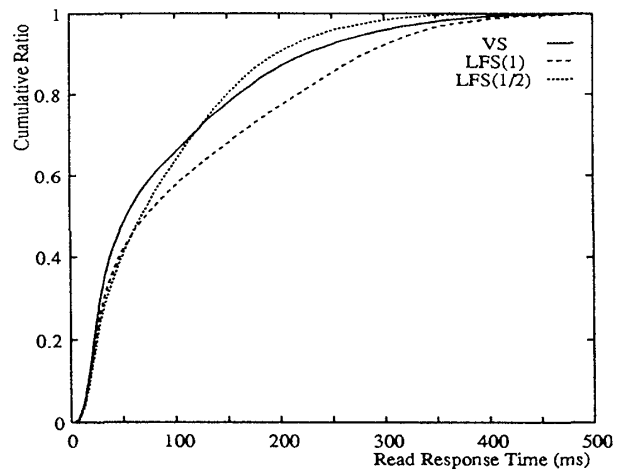


図 2: リードレスポンスタイムの分布 (400 IOs/sec)

に実行され、このことは表から確認される。

書き込みを実行するにはそのための空き領域を作成しておく必要がある。従って、空き作成と書き込みの占有時間の総和が小さいほどより高い負荷に耐えることができることを意味する。

3.2 レスポンスタイムの分散

400 IOs/sec の負荷に於いて、それぞれの方式のリードレスポンスタイムの分布を図 2 に示す。それぞれの方式は、書き込みまたは空き領域の作成のために読み込みのアクセスを待たせている。その時間が長いと遅いレスポンスのもの割合が増え、レスポンスタイムが悪化する。LFS(1) の場合、セグメントが 1 シリンダと大きいため待ち時間が大きい。このため、低負荷での性能の悪化を招いている。

4 まとめ

ストライプの動的割り付けにより RAID5 型ディスクアレイの高性能化を図る方式である仮想ストライピングと LFS について、データの更新性能を比較した。LFS はガーベジコレクションのためにセグメント内にある生きているデータを読み出す必要があるため、その負荷が大きい。その影響を小さくするためにセグメントを小さくすると、効率が落ちて高負荷に耐えられなくなる。一方、仮想ストライピングはガーベジコレクションの効率が高く、また、通常負荷への影響も小さい。

参考文献

- [1] 茂木和彦, 喜連川優: 仮想ストライピングによる RAID5 型ディスクアレイの性能評価. DE93-45, 電子情報通信学会 データ工学研究会, Sep. 1993.
- [2] M. Rosenblum and J. Ousterhout: The design and implementation of a log-structured file system. *Proc. of ACM SOSP*, pp. 1-15, Oct. 1991.