7H－6

仮想ストライピングを用いたRAID5型ディスクアレイのストライプ管理方式

茂木 和彦　喜連川 優
東京大学 生産技術研究所

1　はじめに

外部記憶装置の高性能化・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このパリティも更新する必要があるため、性能低下が生じ性能上の問題点となっている。

更新性能向上を目的とし、「仮想ストライピング」[1]という記憶管理法を考案した。仮想ストライピングではパリティストライプを動的に組み変えることにより、更新前のデータを用いずにパリティを計算する。このため、性能低下の原因となるデータ更新時のデータとパリティを読み出しを行なう必要がなくなる。一方、パリティストライプの組み変えを実行するために、空きストライプの確保するためのガーベジコレクションが必要である。本稿では仮想ストライピングの機構について述べる。

2　仮想ストライピング

仮想ストライプの概念図を図1に示す。

2.1　パリティストライプの仮想化

仮想ストライピングはパリティストライプを仮想化し、データとパリティストライプの関係を動的に変更する。また、データの物理アドレスと論理アドレスも分離する。

パリティストライプの組合せを管理するために仮想ストライプテーブル(図2)を用いる。このテーブルには、それぞれのストライプを区別するストライプ番号と、そのストライプを構成するデータのシリンダ番号とブロック番号が記録される。各ストライプのパリティディスクは、ストライプ番号から計算される。ガーベジコレクションのために、各ストライプのダーティブロックの数も記録する。

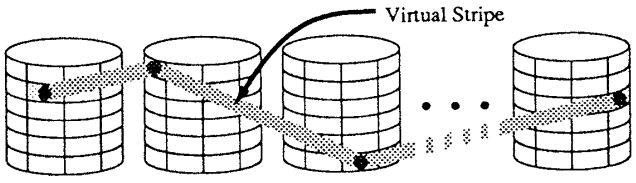

図1: 仮想ストライプの概念図

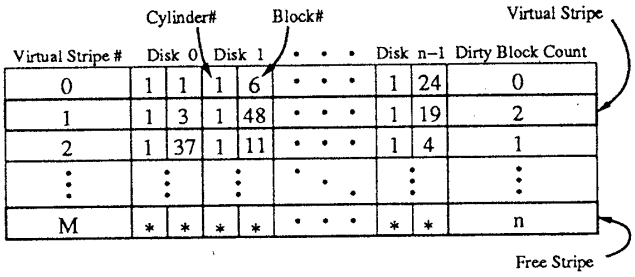

図2: 仮想ストライプテーブル

2.2　パリティストライプの動的割り付け

新たに書き込まれるデータはその属するパリティストライプを変更し、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。パリティは新しく書き込まれるデータのみで計算され、新しいストライプは空きパリティストライプに記録される。n台のデータディスクでパリティが計算され、nブロックの更新が行なわれる場合を考える。従来の方式では、1ブロックの書き換え毎に古いデータやパリティを読み出す必要があり、総計で$n \times (2D + 2P)$アクセスが必要であった。一方、パリティストライプの動的割り付けを行なうとnブロックに対して1つのパリティを計算・書き込むので総計で$n \times D + P$アクセスで済み、データ更新時の処理が軽減される。

2.3　ガーベジコレクション

パリティストライプの動的割り付けを実行して

Stripe management scheme of RAID5 disk arrays with Virtual Striping
Kazuhiko Mogi and Masaru Kitsuregawa
Institute of Industrial Science, University of Tokyo
7-22-1, Roppongi, Minato, Tokyo 106, Japan

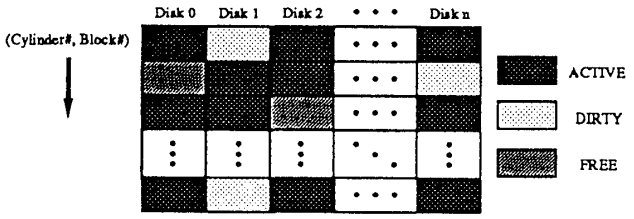

図 3: 物理ブロック状態テーブル

いくと、最終的に新しく作成されたストライプを記録する場所はなくなる。データが別の場所に移ったブロック (ダーティブロック) は、あるパリティストライプのパリティの計算で使われている。その内容は保存する必要があり、そのままではデータを書き込むことはできない。そこで、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必要がある。(ガーベジコレクション)

新しい空きストライプを作るには、その元となるストライプ (ヴィクティムストライプ) を決め、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必要がある。パリティストライプの仮想化によりデータの移動を行なう必要はないが、パートナーストライプのパリティを計算し直す必要があり、1ブロックの置換えには、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アクセスが必要である。置換を行なうブロックをできるだけ減らすために、仮想ストライプテーブル用いて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。

ガーベジコレクションを実行する時には各ブロックがアクティブ、ダーティ、フリーのどの状態にあるかを知る必要がある。これは物理ブロック状態テーブル (図 3) により管理する。

2.4 フローティング / アクセススケジューリング

フローティング [2]、アクセススケジューリング [3] 共にこれまでに高性能化手法として知られているものである。仮想ストライピングでは、仮想化を行なう際にフローティングの考えを利用すると共にアクセススケジューリングを組み込んでいる。

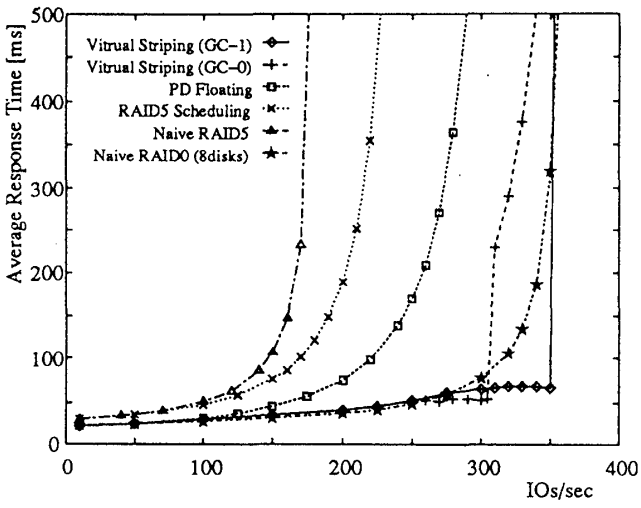

図 4: 静的負荷解析

3 シミュレーションによる評価

8D+P 構成において、ディスク使用率が 80% とした時の平均レスポンスタイムを従来の方式と比較する (図 4)。仮想ストライピングを用いた場合には、通常の RAID5 型やフローティングを用いた RAID5 型に比べて性能が大きく改善されることがわかる。

4 まとめ

仮想ストライピングを用いた RAID5 型ディスクアレイの機構について説明した。パリティストライプを組み替えることにより従来の RAID5 型ディスクアレイに比べて大幅な性能の向上が期待できることを示した。現在、ガーベジコレクタの高効率化について検討を進めている。

参考文献

[1] 茂木和彦、喜連川 優, 「動的パリティストライプの再編成による RAID5 型ディスクアレイの高性能化手法 (仮想ストライピング) に関する基本検討」, コンピュータシステム研究会, 電子情報通信学会, 1993 年 8 月.

[2] J. Menon et.al, "Methods for Improved Update Performance of Disk Arrays", Proc. of 25th Hawaii Int. Conf. on System Science Vol. I, pp. 74–83, January 1992.

[3] M. Seltzer et.al., "Disk Scheduling Revisited", Proc. of Winter 1990 USENIX Technical Conf., January 1990.