

Hot Mirroring ディスクアレイの提案

5G-10

茂木 和彦 喜連川 優

東京大学 生産技術研究所

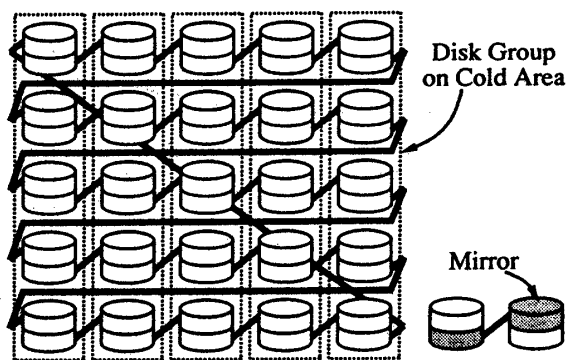
1 はじめに

2次記憶装置の高性能化・高信頼化を目的とした冗長情報を記録するディスクアレイ (RAID)[1] の開発が進められている。その中で、サイズは小さいが多数のアクセス要求があるような負荷では、ミラー (RAID Level 1) や RAID5 (RAID Level 5) が良いと考えられている。RAID5 ではパリティを用いた冗長化を行っており、データ書き込み時のパリティ更新のためのオーバーヘッドやディスク故障時のデータ復旧作業の影響による性能の低下が問題となっている。この点に関して優れているミラーでは、データのコピーを保持することによる冗長化を行っており、データ容量が少ないという問題点が存在する。これらの問題を解決するために、“Hot mirroring” と名付けた記憶管理法を提案する。本稿では、Hot mirroring を用いたディスクアレイの動作機構とその性能について述べる。

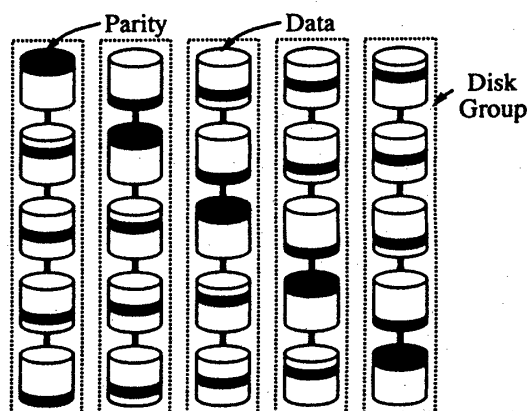
2 Hot mirroring

2.1 Hot mirroring の基本概念

ミラーと RAID5 を比較した時、RAID5 の利点は冗長情報の記録量が少ない点であり、ミラーの利点は動作性能が高い点である。これらの利点をうまく利用するためにミラーと RAID5 を組み合わせることを考える。通常、各ブロックのアクセス頻度には偏り (アクセスローカリティ) があり、頻度が高いもの (ホットブロック) と低いもの (コールドブロック) の2つに大別することができる。このとき、アクセスの大半がホットブロックに対してなされると考えられるため、ホットブロックへのアクセス性能がディスクアレイの性能の支配的要因となると考えられる。ホットブロックの量はさほど多くはないと考えられるので、これらを性能が高いミラーを用いて記録することが望ましい。一方のコールドブロックに関しては、性能よりも大容量性が重要であると考えられ、これらの記録に関してはパリティによる冗長化を行い、冗長情報量の削減を図ることが望ましい。そこで、Hot mirroring では各ディスクをそれぞれ2つの領域に分割し、アクセスローカリティ



(a) Hot mirrored area



(b) Cold parity protected area

図1: Hot mirroring におけるデータ配置

を利用してアクセス頻度が高いものはミラー、低いものは RAID5 と同様な方式で記録し、双方の特性をうまく引出すことにより高性能化・大容量化を図る。

2.2 ホットブロックの分離

Hot mirroring を用いる時、ホットブロックとコールドブロックの分離法が問題となる。通常、書き込みの大半はホットブロックであると考えられる。そこで、書き込みは全てホット領域に対して行うことにする。この方法を用いた時、コールドブロックもホット領域に対して書き込まれることになるので、適宜それをコールド領域に書き戻す必要がある。そこで、ホット領域のブロックに関しては、最後のアクセスが行われた時間を記録する。ホット領域の空きブロック数とその下

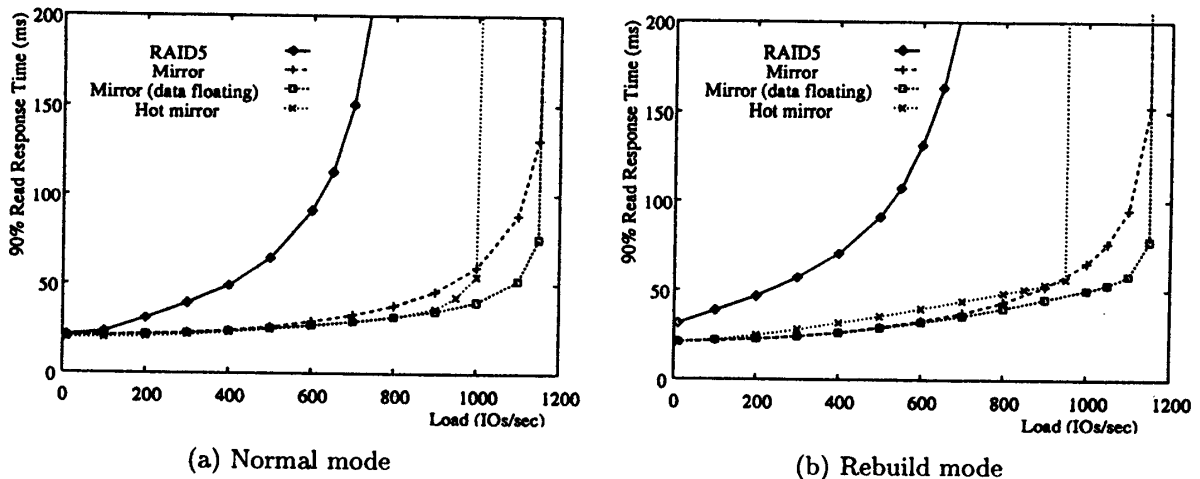


図 2: Hot mirroring の性能 (90-10 access locality, R:W = 7:3)

限値を超えた時、最終アクセスからの経過時間が長いものをコールドブロックと見做し、コールド領域へ書き戻すことにする。

2.3 Hot mirroring のデータ管理

Hot mirroring のデータ配置法を図 1 に示す。ミラー領域でのデータのコピーは、負荷分散と信頼性を考慮して Chained Declustering[2] を基にしたデータ配置を行う。これにより読み出し時の負荷分散も容易に行える。また、ホットとコールドの分離を考えると、ホット領域内でのデータの存在場所は一定ではなくなる。そこで、ミラー領域への書き込み時には各ディスクの負荷を調べてその場所を移動させる。パリティプロテクト領域においては、ディスクを幾つかのグループに分割し、その中でパリティストライプを形成させる。

さて、ディスクが故障した時のことを考える。このとき、信頼性と負荷分散の観点から、故障ディスクを含むディスクグループをできるだけ復旧動作に専念させ、その他のディスクでアクセス要求を処理することが望ましい。そこで、図 1 のようにミラー領域とパリティプロテクト領域では直交したデータ配置を行い、アクセスの大半を占めるホットブロックへのアクセスを故障したディスクを含むグループでできるだけ行わないようにする。

3 性能評価

Hot mirroring の有効性を調べるためにシミュレーションによる性能評価を行った。5 台のデータディスクに対して 1 台のパリティディスクをもつグループが 4 つある (4*(5D+P)) 構成を仮定する。各ディスクの 20% を

ホット領域に割り当て、コールド領域のキャッシュ的な動作をさせる。(ホット領域はコールド領域のデータブロックの 15% を保持できる。) 90% のアクセスが 10% のブロックに集中し、読み出しの割合が 70%、書き込みの割合が 30% の負荷を用いた時の通常動作時と復旧動作時の性能を図 2 に示す。図の横軸は単位時間当たりの I/O 要求数を、縦軸は、(a) 10 万アクセス中 (b) 復旧開始から終了までの復旧動作中の 90% の読み込みアクセスが満たすことが可能なレスポンスタイムを示す。また比較のため、図中に RAID5・ミラー・ミラー(書き込み位置を変更) の性能も示した。このように Hot mirroring はミラーにかなり近い性能を示す。

4 まとめ

新しいディスクアレイの記憶管理方式である Hot mirroring について説明した。本方式は、アクセスローカリティを利用してミラーと RAID5 の融合を図るものである。また、その性能は、ミラーに近いものであることを示した。今後、アクセスローカリティが予測不可能な時により高い性能を得るための手法について検討を加える必要がある。

参考文献

- [1] David A. Patterson, Garth Gibson, and Randy H. Katz. A Case for Redundant Arrays of Inexpensive Disks (RAID). In *Proc. of ACM SIGMOD*, pp. 109-116, Jun. 1988.
- [2] H. Hsiao and D. DeWitt. Chained Decrustering: A New Availability Strategy for Multiprocessor Database Machines. In *Proc. of IEEE Data Engineering*, pp. 456-465, Feb. 1990.