4P-4

二重化ボリューム方式における I／O性能

古川 知幸, 稗田 隆, 永井 啓喜, 大南 正人
(NTT情報通信処理研究所)

1. はじめに

フォールト・トレラント技術において、二重化ボリューム方式は欠かすことができない機能である。データ保全のために、磁気ディスクのハードウェア・コストが2倍かかる事を厭わないユーザが少なくない。

二重化ボリュームは、情報処理システムの信頼性向上を狙った技術であるが、オンライン・システムに適用する場合は、その性能についても考慮する必要がある。

二重化ボリュームのI/Oレスポンス特性については、一般に以下のようになる。

①WRITEイベントは2台のボリュームに対し発行し、完了同期を取らなければならないため、2台のボリュームの遅い方に合せられる。このため、レスポンス時間は、シングル構成に比べ多少悪化する。

②READイベントはどちらか有利な側のボリュームに対してのみアクセスすればよいため、工夫次第でレスポンス時間の向上が可能である。

そこで本稿では、後者のREADイベントの最適スケジューリング方式に関して新しい方式を提案し、従来方式との比較考察を行う。

なお、以下では二重化ボリュームを構成するペアとなる2台のボリュームを物理ボリュームと呼び、物理ボリューム2台1組で利用者に見せる概念的なボリュームを論理ボリュームと呼ぶ。

2. 従来方式

従来シングル・ボリュームで検討されていた、SSTF、SCAN等の最適スケジューリング方式は、二重化ボリュームに対しても有効である。

SSTF[1](Shortest Seek Time First)というのは、各ボリュームのヘッドの位置を常時把握し、ヘッド移動時間(シーク時間)が最も小さいイベントを優先してスケジュールする方式である。

SCAN方式は、SSTFの欠点であるイベントの沈み込みを回避する方式で、さらにいくつかのバリエーションがある[2].

これらの方式は、シングル構成では高負荷時にのみ有効であるが、二重化構成では、シーク時間の評価／イベントの選択時に物理ボリュームの選択の自由度があるため、低負荷時でも有効となる。

この他に各物理ボリュームの、ヘッドの受持ちトラックの範囲をそれぞれ半分づつ固定的に決めておき、各READイベントを受持ちの物理ボリュームに割当てることにより、ヘッドの動作時間を削減する方式が、実現されている(スプリット・シーク方式[3])。

3. 提案方式

図1に磁気ディスクに対する典型的なI/Oのレスポンス時間の内訳を示す。

シーク時間はレスポンス時間中半分以上を占めるが、回転待ち時間も3割近くあり、無視できない。また、チャネルの負荷が高くなれば、チャネルでの待合わせ時間(特に再結合時間)も増大する。

現状の二重化ボリュームのREADイベント最適スケジューリング方式としては、シーク時間に着目したものが主流であり、他の要素(回転待ち時間、チャネルの負荷状況等)を考慮したものは見当らない。この理由は、他の要素は制御プログラムから把握しにくい点が考えられる。

ここで提案する方式は、ヘッドの位置やドライヴの回転位置、チャネルの負荷状況を予め知らなくても、最適な物理ボリュームを選定できる。具体的には、二重化ボリュームに対するREAD要求が到着した時に、2台の物理ボリュームに対し同時にI/O要求を発行する。どちらか先にSet Sectorが完了したのを受けて、該物理ボリュームの動作は継続し、もう一方の物理ボリュームのI/O動作は中止する。

Set Sector完了の認知方法、I/O動作の中止方法については、ハードウェアで実現する方法を含めていくつか考えられるが、ここではPCI割込みと、チャネル・プログラムの

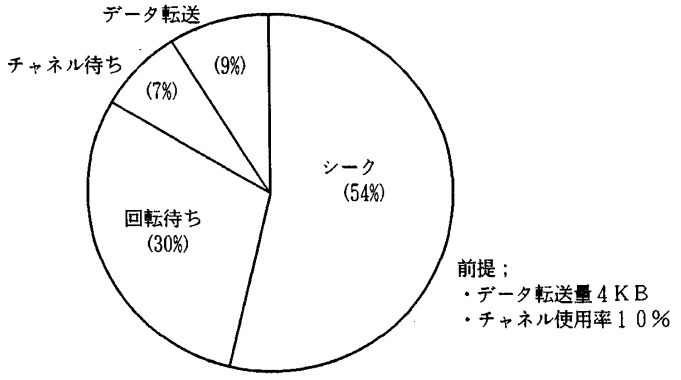

図1 典型的磁気ディスクの応答時間内訳

Performance consideration of dual DASD volumes
Tomoyuki FURUKAWA, Takasi HIEDA, Hiroki NAGAI, Masato Ohminami
NTT Communications and Information Processing Laboratories

動的変更による中断方式を例示する(図２)。

４．考察

提案方式の効果について、レスポンス時間の各項目毎に考察する。

（１）シーク時間

二重化ボリュームでシーク時間を予測して物理ボリュームのスケジューリングを行う場合、次の２点の問題がある。

①WRITE時には２台の物理ボリュームに同一データを書込むため、ヘッドは両面とも同じ場所へ位置付けられる。このためREAD/WRITEを交互に繰返すようなアプリケーションでは、２台の物理ボリューム間で有意差が生じない。

②対象ボリュームを他系と系間共用している（すなわちＬＣＭＰ構成）場合、他系からのアクセスによるヘッド位置の乱れは、自系の制御プログラムからは、認識できない。

①の問題については、提案方式においても事情は同じであり、更新を伴うオンライン・トランザクション処理等では、シーク時間について所詮期待通りの効果は得られないと考えられる。

②については、SSTFおよびSCAN方式においての問題であり、スプリット・シーク方式および提案方式では問題ない。

（２）回転待ち時間

二重化ボリュームでは、ドライヴの回転位置により、各物理ボリューム間で回転待ち時間について有意差がある。提案方式では、短い方の待ち時間が採用され、平均待ち時間は次式のようになる。

$$Tw(\theta) = \{ (\frac{\theta}{2\pi})^2 - \frac{\theta}{2\pi} + \frac{1}{2} \} R$$

ここでθは物理ボリュームのＸ面とＹ面の回転位相差(rad)、Ｒはドライヴが１回転するのに要する時間(sec)である。

従来方式の平均待ち時間はＲ／２であるから、提案方式の改善効果はθの値によって、最大でＲ／４（$\theta = \pi$の時）、最小で０（$\theta = 0$の時）となる。現在のハードウェアでは、ボリューム間の回転位相差θは偶然が支配するが、仮に回転位相差を1/2回転に保つような制御が実現できれば、本方式の効果は大きい。

（３）チャネル待ち時間

チャネルでのイベント競合による待ち時間は、実行中のイベント数から確率的に予測する事はできるが、実際に実行した結果と一致するとは限らない。特にデバイスのオフライン動作後の再結合待ちは、チャネル高負荷時には、しばしば極端に長くなるが、これを事前に予測／回避するのは困難である。提案方式ならば、片方の物理ボリュームに再結合待ちが発生しても、もう一方の物理ボリュームに発生しなければ、レスポンス時間を低下させずに済む。

ところで提案方式は、READイベントの動作を２台の物理ボリュームに対して同時に行うため、チャネルの負荷を高めるのではないかという懸念があるが、２台同時に動作するのはSet Sectorまでの動作であり、最もチャネル負荷に影響を及ぼすデータ転送は含まれないため、問題無いと考える。

（４）ディスク・キャッシュ

提案方式またはスプリット・シーク方式では、キャッシュのヒット率を２倍にすることができる。但しそのためには、キャッシュの対象領域を２台の物理ボリューム毎に分ける（例えばＸ面は前半のトラック，Ｙ面は後半とする）制御をIOCで行う事が前提である。

この場合、スプリット・シーク方式では受持ちトラックの分割方法とキャッシュの対象領域を一致させておく必要があるのに対し、提案方式では、そのような注意は不要である。

５．まとめ

二重化ボリュームに対するREADイベントの性能改善方式を提案し、その有効性について述べた。本方式は、シーク時間など特定の項目に着目したものではなく、トータルなレスポンス時間を改善できる点が最大のメリットであると考える。

文献

[1]Hofri,M. Disk scheduling:FCFS vs SSTF revisited. Commun. ACM 23,11(Nov.1980),p645-653

[2]Teorey,T.J他 A comparative analysis of disk scheduling policies. Commum. ACM 15,3(Mar.1972),p177-184

[3]グレイ他，ﾌｫｰﾙﾄ･ﾄﾚﾗﾝﾄ･ｼｽﾃﾑ，ﾏｸﾞﾛｳﾋﾙﾌﾞｯｸ社刊(1986),p107-109

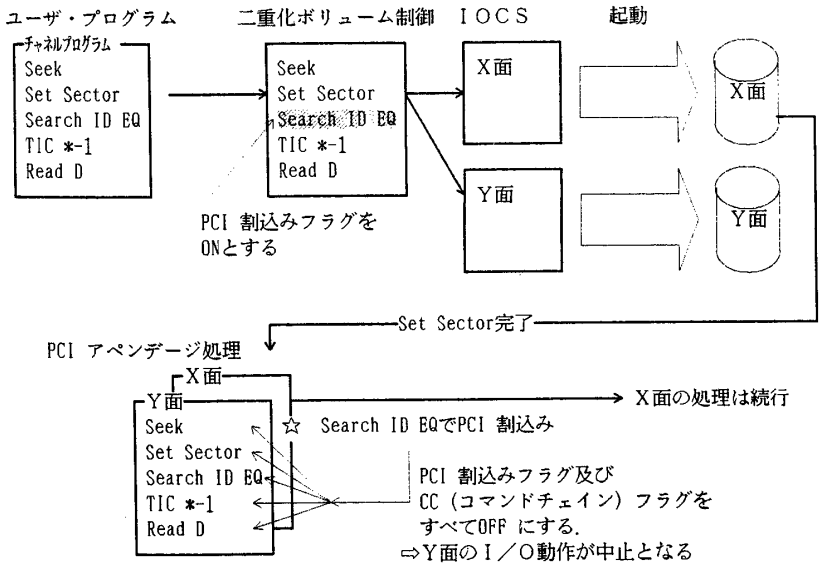

図２　Ｓｅｔ　Ｓｅｃｔｏｒ完了認知およびＩ／Ｏ動作中止方法