映像情報サーバMAMIにおける連続データ転送方式の検討

4B－1

高倉 健　　櫻井 紀彦

NTT 情報通信網研究所

1 はじめに

同一デバイスに格納した動画像情報に対し，複数端末からの多重アクセスを可能とする映像情報サーバでの，定常的な実効転送速度を向上させる方式を検討する．

動画像情報のサーバでは，映像を途切れることなく再生しつつ，複数クライアントの要求をリアルタイムに処理している．これは，動画像情報はシーケンシャルにアクセスされるのが普通だが，ランダムアクセス性能も保持する必要があることを意味する．我々は，動画の長大データ格納と転送性能向上のために，ディスクアレイを用いて記憶システムを構成した．当面の課題として，映像情報サーバのディスクアレイ記憶システムを，より適切な構成にすることが挙げられる．

本報告では，複数ディスクで連続データ転送を行う方式を提案した．ディスクアレイの中でも映像情報の格納に適当だと言われているRAID3を念頭に置き，その中の個々のディスクに着目した場合に，十分な利用が行われていないI/Oバスの使用率向上を目指す．シーク・回転待ちで生じるデータ転送の空き時間の削減が狙いだ．前もってシーク(Pre-Seek)し，連続して(Successive)データ転送するということから，PreSS転送方式(Pre-Seek & Successive Data Transfer)と呼ぶことにする．複数ディスクによる実効転送速度向上法の中で，PreSS転送が効果をあげる単位データ長について考察する．

2 従来のディスクの問題点

映像情報サーバの同一映像への多重アクセスには，記憶媒体からのデータ転送速度と映像情報の表示速度の比が利用されている．このため実効転送速度の向上は重要課題の1つである．

ディスクの実効転送速度低下の原因は，シーク・回転待ちのアクセスタイム T_{acc} の存在である．バースト時の転送速度を V_{dk} とすると，実効転送速度 V_{eff} と単位データ長 L との関係は次式のようになる．

$$V_{eff} = \frac{L}{T_{acc} + \frac{L}{V_{dk}}}$$

L は次の転送までの間(サイクル幅 T_{cyc})の映像を表示するのに必要なデータ量を意味する．映像表示速度を V_{play} とすれば $L = V_{play}\, T_{cyc}$ となるが，同じ映像を連続して見続ける場合には，データアクセスはシーケンシャルな読み出しになるから，$L \to \infty$ で $V_{eff} \to V_{dk}$ となる．一方CD-ROM等に見られるインタラクティブな動画のアプリケーションを，サーバ上で動作してアクセスする場合，応答性能は重要なファクタであり L は短い方が望ましい．

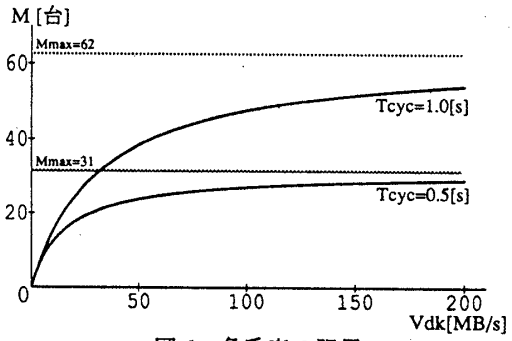

図 1: 多重度の限界

同時に多重アクセス可能なクライアント数を，我々は多重度と呼んでいる．多重度 M と V_{dk} の関係は次式で表せる．

$$M = \frac{V_{eff}}{V_{play}} = \frac{T_{cyc}}{T_{acc} + \frac{V_{play} T_{cyc}}{V_{dk}}}$$

転送速度 V_{dk} のディスクを用い，$T_{cyc} = 1s, 0.5s$ とした場合を図1に表す．$V_{play} = 0.5MB/s, T_{acc} = 16ms$ として評価した．図から解るように，単純に V_{dk} を上げるだけでは多重度を大きく向上させることはできない．V_{eff} の式から，V_{dk} の増加に伴い T_{acc} の影響が顕著に表れることがわかる．また，応答性能を高めるために単位データ長 L を小さくした場合は，その影響がことさら大きいものとなる．

映像情報を扱うのに適当だと言われるRAID3では，これを構成する n_{RAID3} 台のディスクからデータを並列に読み出し，パラレルに転送することで転送速度を高めている．従って転送速度の向上には極めて有効なRAID3も，小さな単位データ長で大きな実効転送速度を得るための本質的な改善にはならない．

近年のI/O装置の性能向上の動向を見ると，T_{acc} の減少傾向よりも V_{dk} の増加傾向の方が勝る．すなわち問題となるのは T_{acc} の影響である．次章では，RAID3を構成する個々のディスク装置の1システムに着目し，T_{acc} の影響を抑えるデータ転送方式について述べる．

Successive Data Transfer for MAMI
(Multiple Access server for Moving picture Information).

Takeshi TAKAKURA, Norihiko SAKURAI.
NTT Network Information Systems Labs.
1-2356 Take, Yokosuka, Kanagawa 238-03, Japan

3 PreSS 転送方式とは

先に述べたように，実効転送速度向上の妨げと成るのは T_{acc} の存在である．そこで我々は，RAID3 を構成する 1 ディスクについて，T_{acc} 時間のタイムロスを隠蔽するため，同一 I/O バスに複数台のディスクを使用することを考えた．

通常 1 つのディスクがシーク・回転待ちをしてデータの読み出しを行なうと，次のディスクは，先のデータ転送が終了して初めてシーク・回転待ちを始める．すると T_{acc} の影響が，データ転送の行われない「バス上の空き時間」として表れてしまう．この空き時間を有効に利用するため，データ読み出しの処理命令を複数のディスクに対し予め発行しておき，バス上で間断なく転送を行おうというのが PreSS 転送の考え方である．

映像情報はシーケンシャルであるから，次に読み出すべきデータを容易に指定することができる．1 つのディスクの転送終了までに次のディスクの転送準備を済ませておけば，1 台目が終われば直ちに 2 台目，そして 3 台目と，連続したデータ転送が行われる．PreSS 転送での I/O バス上の転送イメージを図 2に示した．

図 2: PreSS 転送のイメージ

PreSS 転送で I/O バスをフルに活用するには，一連のデータ転送処理時間を隠し切るだけのディスク数が必要である．このディスク数を n_{ideal} とすると，

$$n_{ideal} = \frac{T_{acc} + \frac{L}{V_{dk}}}{\frac{L}{V_{bus}}}$$

と表せる．ここで V_{bus} はバスの転送速度である．

n_{ideal} 台のディスクで PreSS 転送を行うと，実効転送速度は，

$$V_{eff}^{ideal}_{(PreSS)} = \frac{n_{ideal}L}{T_{acc} + \frac{L}{V_{dk}}} = n_{ideal}V_{eff}$$

まで上げることができる．また，通常の RAID3 において n_{RAID3} 台のディスクによるパラレル転送と，PreSS 転送とを併用して，ディスクシステムの V_{eff} を増加させることも考えられる．

実際は n_x 台のディスクで PreSS 転送を行うとして，制御上のオーバーヘッドについても加味すると，これに伴う所要時間の増加率 α を用いて，

$$V_{eff}_{(PreSS)} = \frac{n_x L}{(T_{acc} + \frac{L}{V_{dk}})(1+\alpha)} = \frac{n_x}{1+\alpha}V_{eff}$$

の転送速度が見込まれる．係数 $n_x/(1+\alpha)$ は PreSS 転送の効果を示している．

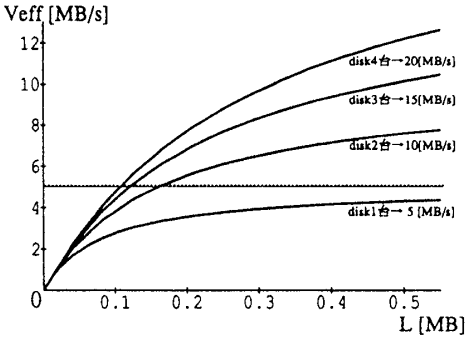

図 3: 単位データ長と実効転送速度

4 有効な利用方法

PreSS 転送が有効なのは，単位データ長 L が非常に小さい時である．どの程度の L に対して効果があるのか，具体的に調べてみる．

例として V_{dk} = 5MB/s のディスク 1〜4 台を用いて構成される RAID3 を考えてみる．パラレル転送を行うと、理論上 20MB/s までの転送速度が可能になる．しかし実際は図 3のように，1 アクセスで読み出す単位データ長 L に依存して，システムとしての実効転送速度は異なる．

PreSS 転送では，L と関係なく V_{dk} に近い実効転送速度が得られる．従って ~100KB 程度のデータブロックの転送を行う場合には、4 台のディスクを使ったパラレル転送よりも有効であると言える．パラレル転送を行うディスク数に応じて，PreSS 転送が有効になる L の範囲は変化するが，単位データ長が小さければ小さい程，PreSS 転送が威力を発揮することが解る．

映像情報サーバにはある程度の応答性能が求められるため，L の値は小さい方が望ましい．RAID3 の個々のディスクで PreSS 転送を行い，RAID3 全体としてパラレル転送で実効転送速度を保証すれば，映像情報サーバの記憶記憶装置として適切な構成となるだろう．

5 まとめ

SCSI ディスクのデータ・バッファに，次にアクセスすると予想される映像情報を先読みしておくことで、連続データ転送を行なう方法を提案した．転送単位のデータ長に適応したシステム構成を行うことで，PreSS 転送方式が，バスの使用率 (データ転送に使われる時間の割合) を高めるのに有効なことが解った．PreSS 転送とパラレル転送とを組み合わせることにより，バス性能・データのアクセス形態に応じた，柔軟なデータ転送システムの構築が期待される．

参考文献

- D.A.Patterson, G.Gibson, and R.H.Karz, "A Case for Redundant Arrays of Inexpensive Disks.(RAID)", Report no.UCB/CSD 87/391, Computer Science Div. University of California, Berkeley, 1987.
- A.Ishikawa, J.Kishigami, N.Sakurai and N.Kotani, "Multiple-access moving picture information system (MAMI)", IEEE GLOBECOM'92, 1992