

VODにおける耐故障並列ディスクの利用と パケット不配/遅配への対応

内川喜章

横田治夫

北陸先端科学技術大学院大学 情報科学研究科

〒923-12 石川県能美郡辰口町旭台15

あ ら ま し ビデオオンデマンド (VOD) サービスを実現するためのシステムの構成方法について提案を行う。システムは、ATM ネットワークとバッファを持つ複数クライアントと耐故障並列ディスクシステムからなる。並列ディスクシステムとしては、我々がこれまで提案してきた DR-net を用いる。DR-net は、複数ディスクを相互結合ネットワークで結合し、一緒に格納したパリティ情報を用いて、サブネットワーク内でデータ再構築を行う。VOD のために、DR-net の全てのディスクと全てのクライアントを ATM ネットワークでつなぐ。ディスクのアクセス時間や ATM ネットワークのデータ転送時間の遅延の揺らぎおよびパケットの廃棄に対応する手法を用いる。DR-net のパリティ情報もクライアントに転送してクライアント内でデータ再構築を行う。データが連続して描画装置に供給される必要条件について考察するとともに、シミュレーションによって提案する手法の効果を確認する。

和文キーワード VOD、DR-net、ATM、RAID、データ再構築

A VOD System Using Fault-Tolerant Parallel Disks and a Method for Handling Packet Delay/Loss

Yoshiaki Uchikawa

Haruo Yokota

School of Information Science

Japan Advanced Institute of Science and Technology, Hokuriku

15 Asahidai, Tatsunokuchi, Nomi, Ishikawa, 923-12 Japan

Abstract This paper propose a system for video on demand (VOD) services. The system is constructed from an ATM network, clients with buffers, and a fault-tolerant parallel disk system for storing digital video data. The parallel disk system, called DR-net, connects disks by an interconnection network, and reconstructs data in a subnetwork of the interconnection network by using parity data stored in disks. All disks and clients are also connected by the ATM network for VOD. A method used in the system can handle scatter in disk accesses and packet transfer delay and packet loss in the ATM network. The parity information in the DR-net is transferred to the clients to reconstruct data within the clients. We consider a necessary condition for providing data to display continuously, and confirm the effects by simulation.

英文 key words VOD、DR-net、ATM、RAID、Data Reconstruction

1 はじめに

米国防務省の「情報スーパーハイウェイ」の影響を受け、マルチメディアネットワークの研究が活性化されている [1]。その中でも、実現性の高いサービスの一つとしてビデオオンデマンド (Video On Demand: VOD) が注目を集めている。VOD サービスの実用システムの開発のためには、多数のクライアントの要求を高速に処理する事が可能な動画サーバと、高速大容量の通信回線の利用が不可欠である。

動画サーバの実現に関しては、従来のビデオテープ再生機等を利用していたのでは、同一ビデオプログラムに対する複数のクライアントからのリアルタイムの開始要求や、早送り、スローモーション、巻き戻しなどの多様な要求に答える事はできない。このため、動画データをデジタル化して磁気ディスクに格納して、複数のデータストリームとして取り出すことによって、これらの要求に答える必要がある。

しかし、動画データをそのままデジタル化して磁気ディスクに格納したのでは、データ容量が膨大になってしまう。このため、MPEG2 (Moving Picture Image Coding Expert Group Phase 2) [2] 等の高能率符号化 (データ圧縮) 技術の利用が重要となっている。MPEG2 の基本仕様 MP@ML (Main Profile at Main Level) では符号化データ速度は 15Mbit/sec 以下となっている。どの程度の画像の質を要求するかによってデータ転送速度が変わるが、ここでは 15Mbit/sec としてデータ容量を考えてみることにする。その場合、1 分間に供給しなければならないデータ量は、約 113MB になる。98% の映画は 135 分以内であると言われるため [3]、1 本の映画では約 15GB の容量が要求されることになる。

現在の磁気ハードディスクの容量は単体ではせいぜい数 GB である。更に、動画サーバとしては多数のプログラムをサポートする事が当然要求される。つまり、実用 VOD 局を実現するためには、数千台規模以上の超並列ディスクシステムの利用が前提となる。

ところで、超並列ディスクシステムを構成する場合、信頼性に関して注意が必要が出てくる。ディスクアクセス時のシークなどの物理動作や回転機構は、半導体部品からなる並列システムと比較してその信頼性を低くする原因となっている。サーバのディスク台数を数千台規模にした場合に、ほとんど毎日のようにディスクが故障してシステム全体を止めていたのでは、使いものにならない。

複数のディスクからなるシステムにより性能を向上させるとともにその信頼性を保つために、RAID として知られる冗長ディスクレイが提案されている [4] [5] [6]。RAID は、データを何台かのディスク装置に冗長情報とともに格納し、ディスク装置故障時にその冗長情報を用いてデータを回復するもので、その冗長情報の構成方法により 5 つのレベルに分けられる [4]。レベル 5 の RAID で使われているパリティ計算手法は、ディスク台数が少ない場合には、性能とコストの面で効果的で、十分な性能と信頼性を提供できる。VOD サービスを実現するために RAID を用いる研

究も行なわれている [7]。

しかし、RAID はバスを用いて構成されるため、非常に多数のディスクを必要とする実用規模の VOD 局を実現するような場合は、そのバスがボトルネックとなりうる。また、RAID はパリティグループ内の単一故障を前提としているが、数千台規模になると、単一故障の前提は必ずしも十分とは言えない。つまり、実時間の応答性能が重要であるシステムでは、RAID は性能、信頼性の面で問題を含んでいると思われる。

我々は、RAID で用いられているパリティ計算の手法を相互接続ネットワークに適用して、性能、信頼性の問題を解決する方法として、データ再構築ネットワーク (Data Reconstruction Network: DR-net) を提案してきた [8] [9] [10] [11] [12] [13]。DR-net では、ディスクは相互結合ネットワークの各ノードに接続され、その相互接続ネットワークのサブネット内でパリティグループを構成する。ディスク故障時にデータを再構築する場合でも、パリティ計算のための通信がサブネット内でローカルに行なわれることから、ネットワークのサイズが大きくなった場合にもアクセス性能の劣化が少ないことを特徴とする。

ただ、RAID 同様、DR-net では、格納してあるデータを更新する場合には、対象のディスクとパリティディスクを変更する必要がある。このため、更新の少ない処理が、アプリケーションとしてより適しているといえる。VOD サービスでは、一度プログラムを格納してしまえば、動画データに対する更新の割合は非常に少ない。以上から、DR-net は、特にシステム規模が大きな実用 VOD サービスに適していると言える。

一方、クライアントとの間の通信回線としては非同期転送モード (Asynchronous Transfer Mode: ATM) のネットワーク利用が考えられている [14]。ATM は次世代の広域通信網の基幹技術として有望視されており、156Mbit/sec の伝送速度でデータを転送することが可能である。VOD サービスを実現するのに ATM ネットワークを用いた場合の問題点は、ATM の伝送単位であるセルの到達時間に大きな揺らぎがある点と、トラフィックが多いとセルが伝送中に失われる可能性があるという点である。

セルの到着時間の揺らぎの吸収のためには受信側にバッファを設け、再生開始時間に遅延を付加する方法が取られる [15]。しかし、その場合にも、遅延時間の設定と回線の状態によってはデータの必要となる時間内にセルが届かず、データが失われてしまう可能性はなくならない。

また、セルが伝送中に失われた場合 (不配の場合)、何らかのプロトコルを用いて再送を行なう必要があるが、VOD のように実時間の応答が必要なアプリケーションの場合は再送を行う事が困難である事が考えられる。

実用的には以上のような問題が起きた時、VOD サービスの場合には、前のフレームのデータをそのまま再生するか、あるいは補完アルゴリズムにより補完データを再生するような方法が取られる。しかし、そのような場合には、当然画像の品質の劣化を招くことになる。

本稿ではこの選配や不配によるフレームのデータが失われる率を減少させ、またより多くのクライアントの要求を満たすシステムのために DR-net が持つパリティデータ

クライアント側に送り、クライアント側でデータを再構築する方法を提案し、その方法による性能改善について考察し、シミュレーションによってその効果を確認する。

2 システムの構成/モデル

我々が提案するシステムは、耐故障並列ディスクシステムである DR-net と ATM ネットワーク、およびそのネットワークに接続された複数のクライアントからなる。動画データは予め MPEG2 でコード化され DR-net の各ディスクに分散配置される。各クライアントからの要求に従って、コード化された動画データがディスクから読み出され、ATM ネットワークを介してクライアントのバッファに送られる。デコードは各クライアントで行われることとする。

2.1 DR-net

DR-net は RAID レベル 5 に対応する並列ディスクアレイで、RAID のバスの代わりに相互結合ネットワークを用いて、そのサブネットワーク内でデータの再構築を行う。相互結合ネットワークのトポロジーは特に限定していないが、現在主に検討を行っているモデルは $5n \times 5m$ のトラス結合で、その上で十字型および卍型の 2 種類のパリティグループを構成している。その場合、1 つのパリティグループは 4 台のデータディスクと 1 台のパリティディスクからなり、2 つのパリティグループを重ね合わせることでより複数の故障にも対応できる。5 × 5 のトラスネットワークの場合、2 台までの故障ディスクのいかなる配置パターンでもデータを回復でき、4 台までなら 95% 回復でき、最大 9 台の故障まで対応できることが分かっている [10]。

今回のモデルでは、トラス結合と同時に、各ノードは直接 ATM ネットワークに接続されるものとする。また、各ノードはディスクアクセスとネットワーク転送を同時に行うためのダブルバッファを各々持つこととする。動画データは DR-net の全ディスクに分散して配置される。

2.2 クライアントにおけるデータ再構築

前述したように、ATM ネットワークは、セルの到着時間に揺らぎがある。また、場合によってはセルがネットワーク内で廃棄されることもありうる。さらに、ディスクアクセス時間にも、シーク時間や回転待ち時間によって、データが用意できるまでの間隔に大きな揺らぎが生じる。我々は、この ATM ネットワークデータ転送やディスクアクセスの遅延の揺らぎに対応するため、DR-net のパリティディスクの冗長情報をそのまま送信し、クライアント側でデータの再構築を行う方法を提案する。

受信側の伝送チャネルは 1 クライアントに対して DR-net におけるパリティグループの構成ディスク数に相当するだけの回線数を持つ。本稿においては、現 DR-net の構成 [11] を前提に、1 クライアント当たり 5 回線とする。1 つのクライアントの受信バッファは、描画装置と伝送回線双方から同時にアクセス可能なダブルバッファ構成とし、それぞれのバッファが各回線からは並列に受信が可能のように、5

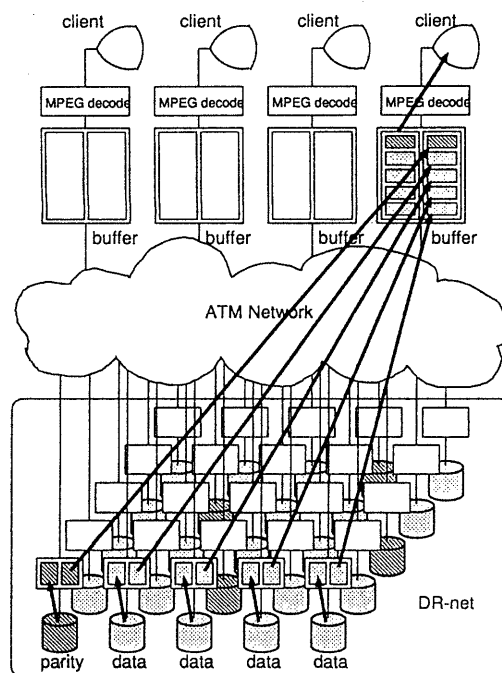


図 1: システムの構成

つに分割されているものとする。提案するシステムの構成を図 1 に示す。

各クライアントは、自分の受信バッファの中の 4 つの部分にデータが揃ったところでデータを MPEG デコーダを含む描画装置へ送る。パリティ部分を除いた部分が 4 つ揃った場合には、そのままその 4 つの部分を送ればよい。パリティ部分以外の 1 つがまだ到着していない場合には、その部分をパリティと残りの 3 つのデータから再構築 (排他的論理和を計算) して、デコーダへ送る。

以上のようにシステムを構成することにより、5 つのセルの内、1 つのセルが大きく遅れたり、あるいはネットワーク内で捨てられた場合でも、元の画像品質を保つことが可能となる。次に、連続再生の条件について考察する。

2.3 連続必要条件

動画サービスを行うシステムにおいては連続再生が可能でなければならない。そのためには、特定の時間間隔で決まった量の画像フレームのデータが描画装置で利用可能になっている必要がある。つまり、本システムの場合には、描画装置が要求した時にダブルバッファの一方に必ず 1 画面分のデータが有効になっていることが要求される。これを連続必要条件と呼ぶ [15]。このためにたとえ平均的な画像サーバの応答時間が再生に十分な性能であっても、ディスクの応答の揺らぎや ATM のセルの到着の揺らぎによって、描画装置がデータを要求するタイミングに必要なデータが到

着せず、連続必要条件を満たさない場合がある。

システムのパラメータを、

符号化データ速度	V_{coding}	bit/sec
ATM 伝送速度	V_{atm}	bit/sec
ディスク転送速度	V_{disk}	bit/sec
フレームサイズ	F	bit
ATM 遅延	D_{atm}	sec
ディスクのアクセス遅延	D_{disk}	sec
パリティグループ内ディスク数	n	
パリティグループ数	N	
クライアント数	C	

とし、まず、データ再構築をクライアント側で行なわない場合の1つのクライアントについて、連続必要条件を考慮してみる。

連続必要条件は1フレームを符号化データ速度でデコードに渡している間に次のデータが受信バッファに揃うことであるから、

$$t_{get} \leq \frac{F}{V_{coding}}$$

となる。ここで、コード化された動画データをディスクから ATM ネットワーク経由で取ってくる時間 t_{get} は、

$$t_{get} = \max\{t_1, \dots, t_i, \dots, t_{n-1}\}$$

$$t_i = D_i + \frac{F}{\min(V_{atm}, V_{disk}) \times (n-1)}$$

$$D_i = D_{atm_i} + D_{disk_i}$$

となる。つまり、 t_{get} は、1フレームの容量をデータディスクの数で割った分を、ディスクの読みだし時の転送速度と ATM ネットワークの伝送速度の遅い方によって割って転送に要する時間を算出したものに、ディスクアクセスのための遅れの時間と ATM における遅れの時間を加えたもののうち、パリティグループ内で一番大きなものとなる。ただし、パリティを構成するディスクの数が n とした時、パリティは n 番目のディスクに格納されており、データディスクの数は $n-1$ であるとする。

以上の内容を、 N 個のパリティグループからなるサーバーから、 C 個のクライアントにデータを送るシステム全体に適用すると、連続必要条件は、

$$C \cdot \max\{D_i\} + \frac{C \cdot F}{\min(V_{atm}, V_{disk}) \times (n-1)} \leq \frac{N \cdot F}{V_{coding}}$$

となる。ここで、 V_{atm} 、 V_{disk} が十分大きければ、条件成立のためには $C \cdot \max\{D_i\}$ の項が支配的になる。

次に、クライアント側で再構築を行う場合を考える。データの再構築に必要な時間 $t_{reconst}$ は、実際には排他的論理和を計算する時間であり、 t_{get} に比較して十分小さいと仮定することができるため、 $t_{reconst} \ll t_{get}$ とすると、

$$t_{get} = \max\{t_1, t_2, \dots, t_n\} \cap \sim \max\{t_1, t_2, \dots, t_n\}$$

ここで、 t_i は、再構築を行なわない場合と同じであるから、その最大のものを無視することが出来ることを意味している。 D_i の揺らぎが大きいつきには、その効果は大きいと思われる。更に、ATM ネットワークでは、負荷の状態によってはセルが廃棄され、 D_{atm_i} が無限大になる可能性がある。

クライアント側で再構築しない場合は描画内容に影響が出るが、本手法を使うことにより、その確率が低い場合くパリティグループ内の2つ以上のセルが廃棄されることがない限り、描画内容に影響がない。

つまり、クライアント側でデータの再構築を行う場合はグループ内で最も遅く届くデータを除いた残りのデータから必要な情報を取り出す事が出来る。従ってディスクのシーク時間や回転待ち時間の揺らぎ、ATM のセルの遅延によって起こるデータの到着時間の揺らぎを低減する事が可能である。また、ATM のセルが伝送中に失われた場合であっても再構築アルゴリズムにより正しいデータを回復できる可能性があることを意味している。

3 シミュレーション

クライアントにおけるデータ再構築の効果を調べるため、コンピュータシミュレーションを行った。

前節で述べたシステムにおいて、複数のクライアントから同時に動画データをストリームとして要求する。送られた要求に対してデータをディスクから読み出し、ATM ネットワークを介して、クライアントの受信バッファに送り、再構築されて1フレームの有効なデータとなるまでの時間をシミュレーションにより求める。

まず、各クライアントにおける応答時間の分布を測定する。次にこの応答時間の平均がクライアントの数を変えた時にどのような変化を示すかを測定する。この時、ATM ネットワークの遅延時間の揺らぎの割合を変えながら変化を見る。更に、クライアントにおいてダブルバッファの切替時に次のデータの準備が出来ずに、連続必要条件を満たされない場合の比率が、クライアント数を変化させた場合にどのように変わるかを測定する。

比較のために DR-net 内でのみパリティ情報を扱い、クライアント側での再構築を行わない場合の応答時間も求める。その場合のデータ転送は、パリティ情報を含まないデータディスクのデータのみ、つまり $n-1$ 個のブロックを対象とする。フレームを構成している全てのデータが受信側のバッファに揃った時がそのフレームの受信の完了時となる。この点以外は、冗長情報を使用する場合と条件は等しいとする。

3.1 シミュレーションの条件

今回のシミュレーションで想定するディスクの諸元を、表1に示す。

ディスクアクセスの遅延 D_{disk} は、シーク時間と回転待ち時間の和、

$$D_{disk} = D_{seek} + D_{sync}$$

とする。ここで、シークにおけるポジショニング時間 D_{seek} の計算は直線モデルを用いている。つまり、ある時点でのヘッドの位置から、次の要求のヘッドの位置までのヘッド移動距離を h_{move} とすると、

$$D_{seek} = (s_{max} - s_{min}) \cdot \frac{h_{move}}{c} + s_{min}$$

表 1: シミュレーションにおけるディスクの諸元

シリンダ数 (c)	1893	
ヘッド数 (H)	21	
トラック数 (T)	52836	
回転数 (r)	5400	rpm
シーク時間 最大 (s_{max})	22.0	ms
シーク時間 平均 (s_{avg})	12.0	ms
シーク時間 最小 (s_{min})	2.0	ms
ヘッド変更時間 (h_{change})	3.0	ms
データ転送速度 (V_{disk})	10	MB/s

とする。また、平均回転待ち時間 D_{sync} は、

$$D_{sync} = \frac{60000}{2 \times r}$$

とする。

動画データは並列ディスク全体に分散して配置してあるものとする。分散の方法は、パリティを含むグループを1つの単位とする。従って、(ディスクのブロックサイズ) × (パリティグループのデータディスク数) 単位で並列ディスク全体に分散されるものとする。パリティグループの構成数 (n) は5台、つまり1パリティグループ内のデータディスクは4台とし、パリティグループの数 (N) を20とする。システム全体としては、100台のディスクでシミュレーションを行なう。

また、アクセス最小単位はページサイズ (P) で、4KB、8KB、16KBとする。ページの配置はシリンダ、セクタ位置、ヘッドいずれもランダムとする。

符合化データ速度は、MPEG2を想定し、15Mbit/secとする。1秒当たり25枚(40ms/Frame)のフレームを使用するとして1枚当たりのフレームのデータ量を均一にして1フレーム当たり75kBとする。1クライアントあたり1フレームのデータが40msごとに並列ディスクアレイに対して要求される。

なお、クライアント側では受信ダブルバッファが両方とも満たされた状態になると送信側に対して受信不能である事を知らせる。送信側は1フレーム消費される時間だけ次のフレームの送待を待つこととしておく。

ATMは、伝送速度 (V_{atm}) を156Mbit/secとし、セルのデータ長53byte、不配確率を 10^{-6} とする。ATMにおける遅延 (D_{atm}) は、固定分の遅延時間 ($D_{atm,fix}$) と、遅延時間の揺らぎ分 ($D_{atm,scatter}$) の和とし、

$$D_{atm} = D_{atm,fix} + D_{atm,scatter}$$

とする。とりあえず、ここでは広域網を想定して、 $D_{atm,fix}$ を10msとする。セルのデータ遅延の揺らぎはポアソン分布であると仮定し、分布の平均値 λ をパラメータにして、 λ を5ms,10ms,20msで、 $D_{atm,scatter}$ を求める。

3.2 シミュレーション結果

ページサイズ (P) が4KBで、クライアント数 (C) が10の場合に、各々のクライアントが100フレームづつの要求を出した場合の特定の1つのクライアントの応答時間を図

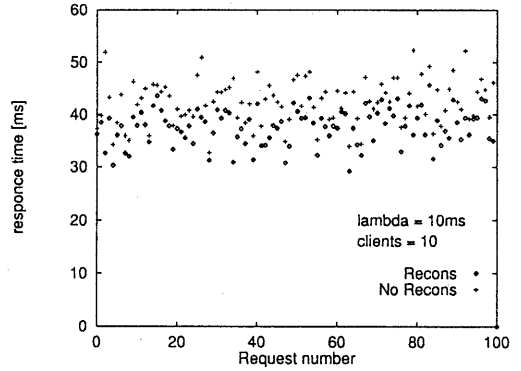


図 2: $C = 10$ の時の応答時間の分布 ($P = 4KB$)

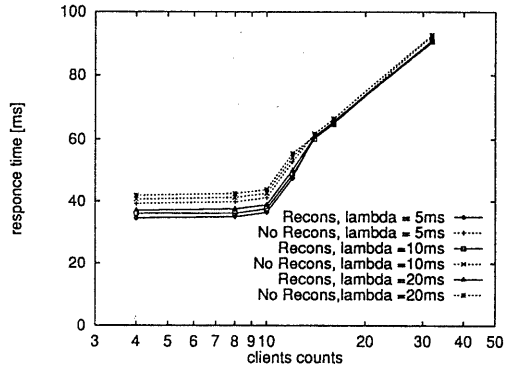


図 3: C の変化による平均応答時間の推移 ($P = 4KB$)

2に示す。受信側でデータ再構成を行った場合と行わない場合をそれぞれ別の印でプロットした。ここでは、ATMネットワークのセル配送のポアソン分布の平均 λ を10msとした。この条件において応答時間は40msを中心として30msから50msの範囲に分布している事が分かる。

次に、同様にページサイズを4KBとして、クライアントの数を変化させた場合のシステム応答時間の変化を図3に示す。各々のクライアントが100フレームのデータを要求した場合の全クライアントの応答時間の平均を求めてプロットを行った。クライアント数10以下においてはいずれの条件でも応答時間はほぼ一定である。クライアント側での再構成を行った場合、再構成を行わない場合に対して、このクライアント数10以下(4,8,10)の平均で13%応答時間が短くなっている。クライアント数が12以上で応答時間が悪くなっているのは、ディスクにおけるアクセス競合が生じるようになったためで、それにより連続必要条件を満たすことも困難となる。

図4は上と同様の条件で、本来フレームを描画する時間になってもそのフレームが届かない割合、つまり連続必要条件を満たさない要求の割合をプロットしたものである。

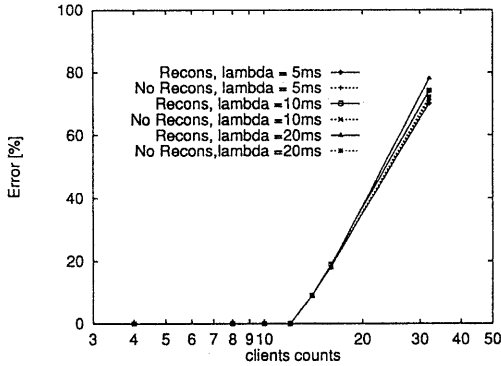


図 4: C によるフレーム遅れの割合の推移 ($P = 4\text{KB}$)

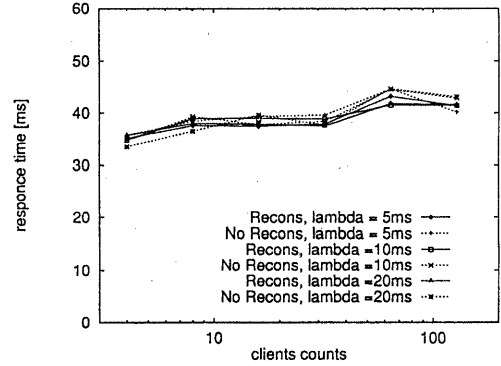


図 6: $P = 16\text{KB}$ の時の平均応答時間の推移

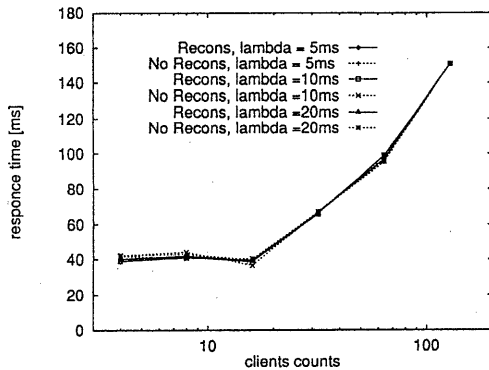


図 5: $P = 8\text{KB}$ の時の平均応答時間の推移

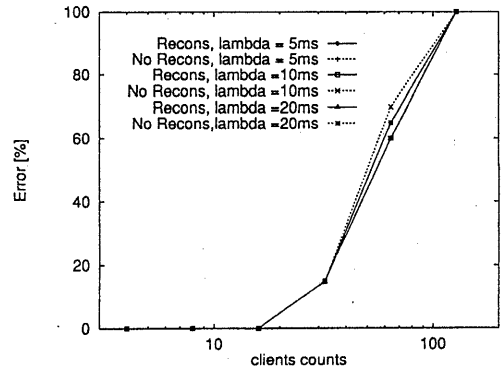


図 7: $P = 8\text{KB}$ の時のフレーム遅れ割合の推移

クライアント数 12 を越えるとクライアント数が増えるに従ってフレーム遅れが多発するようになる。これは、ページサイズが 4 KB とフレームサイズに比べてかなり小さいために、シークの発生が頻繁で、 D_{disk} が大きくなっているのが原因と思われる。

そこで、ページサイズを 8KB と 16 KB に変化させた場合の応答時間の推移をそれぞれ図 5 と図 6 に示す。ページサイズを大きくすることにより、応答時間をほぼ一定に保てるクライアント数が増えることが分かる。その場合、フレームの遅れの割合を見てみると、ページサイズ 8KB の推移は図 7 に示すようになっており、少なくともクライアント数 16 まで遅れが発生していないことが分かる。ページサイズを 16KB にすると、フレームの遅れはクライアント数 128 まで 0 である。

3.3 シミュレーションに関する考察

シミュレーションの結果より、本稿で提案したクライアント側でデータの再構築を行う方法はディスクアクセス時間や伝送回線の遅延の大きな揺らぎのあるシステムにおい

て性能の改善効果がある事が示された。

ページサイズが 4KB の場合はクライアント数 10 以下の領域で、ページサイズが 8KB の場合はクライアント数 16 以下の領域で、ページサイズが 16KB の場合にはクライアント数 128 までの領域で、クライアント数の変化に対し、応答時間の変化は微小である。これは、ディスクのアクセス時間と ATM ネットワークの遅れが連続必要条件を満たしているのと同時に、描画装置の速度が一定であるために受信バッファにデータがたまり、送信側が待たされるためであると考えられる。

ページサイズ 4KB の場合を見てみると、図 2 においてクライアント数 14 以上の領域で応答時間の増加の割合が減少している。これは図 3 におけるフレーム遅れが起き始める領域と一致する。今回のシミュレーションにおいては、描画に間に合わないフレームが発生した場合、その遅れたフレームの到着時点を基準に取り直し再生を再開するような方法をとった。この方法では描画に間に合わなくなるフレームが発生すると 1 フレーム停止して再生が再開されることになる。クライアント数 14 以上における応答時間の増加の割合の減少は、このシミュレーションの方法が原因

であると考えられる。

4 おわりに

VOD サービスを実現するためのシステムの構成方法について提案を行い、ATM ネットワークのバケット不配/遅配に対応するための手法について考察し、DR-net 利用を前提とした並列ディスク利用のシミュレーションによってその効果について確認を行った。

クライアント側でデータ再構築すること方法は、シミュレーションから安定状態で 13% の応答時間の向上であり、セルの損失などに対する信頼性の向上を考慮すれば有効性の高い方法である事を示す事ができた。

今回の報告では、時間の関係で、ディスク数（パリティグループの数）を変化させたデータを用意することが出来なかった。100 台という数字は、本システムが目的とする実用 VOD サービスに対しては必ずしも十分な大きさとはいえない。今後、より大きなシステムに関して、データを取っていきたいと考えている。それに伴い、クライアント数もより大きな数字を対象にすべきであると考えている。また、符合理化データ速度やディスク諸元に関しても、一つの値を想定してシミュレーションを行なったが、要求される画像品質や利用するディスクを変化させてみる必要もあると思われる。

更に、今回シミュレーションのモデルは、各所を単純化してある。ディスクのシークにおけるアクセス時間の計算は、今回はヘッドの移動距離に対して直線的に変化するとしている。これに関しては、より精密なモデルが示されている [16]。また、ATM の誤り率、伝送の遅延に関しては伝送されるデータ量や回線利用頻度には依存しないとしてある。これらのモデルをより精密にする事でシミュレーションの精度の向上が期待できる。

また、今回はディスクの故障時の動作についての検討は行われていない。ディスク故障時にはサーバである DR-net 内で故障の検出、データの再構成が行われ、アクセス性能が低下する [12]。特に、サーバ内でパリティデータが使われるためにクライアント側での再構築の効果が失われる事が考えられる。これらの原因でディスク故障時にはサービス可能なクライアント数の減少が起きると考えられる。この点について定量的に検討する必要がある。

今後の以上のような課題に対して検討を進めると同時に、本研究室で現在試作中の 5 × 5 の 25 台のノードを用いた DR-net の実験システム [11] を、ATM ネットワークに接続し、今回シミュレートを行った VOD の実験システムの実機への実装が行なっていく予定である。この実験システムへの実装を行い、実証を行いたいと考えている。

参考文献

- [1] A. L. Narasimha Reddy and James C. Wyllie. I/O Issues in a Multimedia System. *IEEE Computer*, 27(3):69-74, Mar 1994.
- [2] 加藤雅浩、田中正晴、飯塚智弘. 画像圧縮方式 MPEG2

次世代インフラの要に. 日経エレクトロニクス, 3.14 1994.

- [3] ハードディスクに映像が入る、光磁気と相変化が追う. 日経エレクトロニクス, 7.18 1994.
- [4] D. A. Patterson, G. Gibson, and R. H. Katz. A Case for Redundant Arrays of Inexpensive Disks (RAID). In *Proc. of ACM SIGMOD Conference*, pages 109-116, Jun 1988.
- [5] G. A. Gibson. Performance and Reliability in Redundant Array of Inexpensive Disks. In *Proc. of 1989 CMG Annual Conference*, 1989.
- [6] G. A. Gibson. *Redundant Disk Arrays*. The MIT Press, 1992.
- [7] 梶谷 浩一. 動画サーバのためのディスクアレー管理法についての考察. 電子情報通信学会論文誌, J77-D-I(1):66-76, Jan 1994.
- [8] 横田治夫. RAID のネットワーク上への展開と信頼性向上. 信学技法 CPSY 93-11, 電子情報通信学会, Apr 1993.
- [9] 横田治夫. データ再構築ネット (DR-net) における不均衡対策. 信学技法 FTS 93-20, 電子情報通信学会, Aug 1993.
- [10] H. Yokota. DR-nets: Data-Reconstruction Networks for Highly Reliable Parallel-Disk Systems. In *Proc. of 2nd Workshop on I/O in Parallel Computer Systems*, pages 105-116, Apr 1994.
- [11] S. Tomonaga and H. Yokot. An Implementation of a Highly Reliable Parallel-Disk System using Transputers. In *Proc. of the 6th Transputer/Occam Intn'l Conf.*, pages 241-254. IOS Press, Jun 1994.
- [12] Haruo Yokota and Seishi Tomonaga. The Performance of a Highly Reliable Parallel Disk System. In A. De Gloria, M. R. Jane, and D. Maini, editors, *Proc. of the World Transputer Congress '94*, pages 147-160. IOS Press, 1994.
- [13] 横田治夫、友永誠史. 高信頼並列ディスクプロトタイプのアクセス性能. 信学技法 FTS 94-37, 電子情報通信学会, Jul 1994.
- [14] Reza Rooholamini, Vlandimir Cherkassky, and Mark Garver. Finding the Right ATM Switch for the Market. *IEEE Computer*, 27(4):16-28, April 1994.
- [15] P. Venkat Rangan and Harrick M. Vin. Designing File Systems for Digital Video and Audio. In *Proc. of the 13th ACM Sympo. on OS Principles*, 1991.
- [16] Chris Ruemmler and John Wilkes. An Introduction to Disk Drive Modeling. *IEEE Computer*, 27(3):17-28, March 1994.