マルチメディア通信と分散処理 74－39
グループウェア 15－39
（1996. 1. 26）

分散ＲＡＩＤ方式ビデオサーバー（その２）

中村俊一郎　峯村治実　山口智久　（三菱電機・情報技術総合研究所）

清水洋　渡辺尚　水野忠則　（静岡大学・工学部）

ビデオサーバーの性能（同時ビデオ配信数）及び信頼性向上を目指した分散ＲＡＩＤ方式ビデオサーバーの実現法の提案を行う。前回はＲＡＩＤ０型の試作を行い、サーバー数に比例した性能向上が出来る結果が得られたが、今回は信頼性の向上をも実現するＲＡＩＤ４型及びＲＡＩＤ５型を考え、それぞれ試作ないしシミュレーションを行った。この結果ほぼ期待通りの性能が観測された。又数十ストリームのビデオストリームが転送されている最中に、突然１つのサーバーをダウンさせても、クライアントの映像の乱れ無しに処理が継続されることや、縮退時のコマ落ち率測定結果からも非常に良好な映像が供給されることを確認した。これらの結果から、本方式の実用化にに向けて大きく前進したと考える。

Distributed RAID Style Video Server （２）

Shunichiro Nakamura, Harumi Minemura, Tomohisa Yamaguchi
(Mitsubishi Electric Corp.)
Hiroshi Shimizu, Takashi Watanabe, Tadanori Mizuno
(Shizuoka Univ.)

In this paper, we present an implementation method of distributed RAID style video server that addresses the problem of increasing video stream supplying capability and reliability in VOD systems. In the former report we implemented RAID0 style videoserver and examined that performance improvement that is propotional to the number of servers was achieved. Here reliability improvement is also saught. We implemented a RAID4 style video server and made simulation for a RAID5 style video server. A precise performance evaluation was made in a near practical environment, with more than several tens of video streams. We tested the system, in both normal and degraded mode. We also made the failure insertion test, in which it was confirmed that a separation of the bad server was carried out instantaneously without exerting an influence in a video picture at the client. We think these results dramatically increase the feasibility of a distributed RAID style video server.

1．まえがき

分散RAID方式とは、通常ディスク群に対して適用されるRAID（ディスクアレイ）手法をサーバー群にまで拡張して適用を図り、全体として性能及び信頼性の向上を図ることを目的とするもので、いくつかの研究が報告されている。[3]，[4]我々はこの方式の有望な適用分野としてVOD用途のビデオサーバーを考えているが（図1）、この用途の実試作例は見当たらないため、これを試作する中で種々の問題を解決し、フィージビリティを確認する目的で研究を進めている。前回の本研究会において、性能向上のみを目的としたRAID0型の試作結果について報告した。[2]この結果、サーバー台数に比例してビデオサーバー全体の性能（同時供給ビデオストリーム数）が向上すること、コマ落ち率0.04%以下という非常に良好な映像が得られることが確認された。これにより、普及型の安価なサーバーを複数並べて高性能なビデオサーバーを実現出来る見通しが大いに強まった。

一方サーバー数が増えることにより、ビデオサーバー全体としての信頼性が低下するという問題があり、この解決策として、今回ディスクアレイと同様、RAID4型、RAID5型の分散RAID方式のビデオサーバーを考え、それぞれ試作ないしシミュレーションによる評価を実施した。

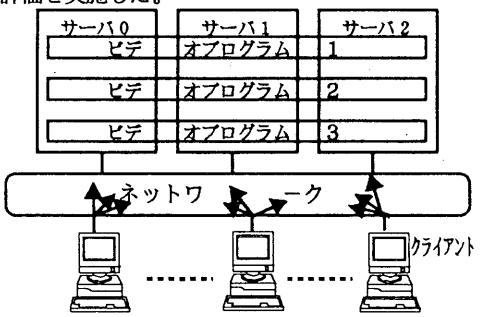

図1 分散RAID方式ビデオサーバーの概念図

2．ＲＡＩＤ４型の試作評価

2.1 概要

前回の本研究会で報告したRAID0型分散RAID方式ビデオサーバーはデータを単にストライピングして性能向上を図るものであった。今回はさらにデータに冗長性（パリティ）を持たせて、信頼性の向上を図ることを行った。即ちディスクアレイのRAID4と同様の方法によるRAID4型分散RAID方式ビデオサーバーの試作を行った。これにより複数のサーバの内の1台が故障してもデータの喪失が起こらず処理を継続出来る。この試作機の実現法はRAID0型の時と同様であり、クライアント側に図2に示されるソフトウェア構成で実現した。即ちMS-DOSのTSR(Terminate and Stay Resident)機能を使ってストライピングドライバを作成し、その中にRAID4型の機能を実現するという方法で実現した。

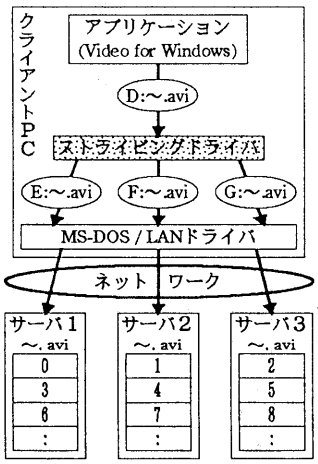

図2 RAID4型の実現方式

図3はこの試作システムの構成を示した図であり、これも前回報告のRAID0型の時と同様である。LANとして100Mbpsイーサネット、各サーバにはディスクを3台接続しソフトウェアストライピング構成とする。

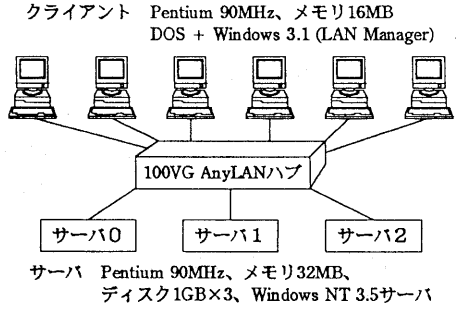

図3 試作機のシステム構成

2.2 ＲＡＩＤ４型ストライピング

図4は試作したRAID4型のストライピング方式を示す。サーバ当たり複数台（図4ではD0～D2の3台）のディスク接続が可能な方式である。図4ではストライピングされるブロックが順次サーバを移っていくが、図5のように1つのサーバ内のディスクを順次移っていく方法も考えられる。後述するストライプバッファ容量が少なくて済むこと、プログラムがサーバ

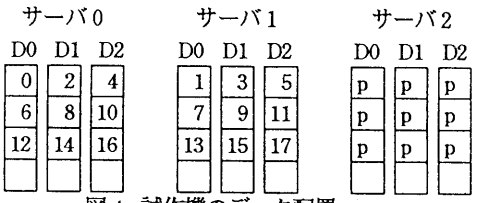

図4 試作機のデータ配置

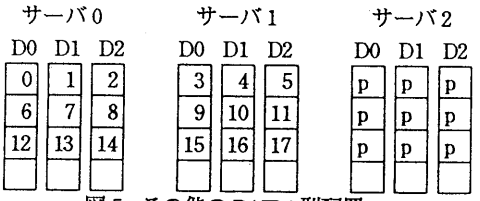

図5 その他のRAID4型配置

一当たりのディスク台数を意識しなくてよい等の長所により図4の方法を採用した。なお図4、5で横一列に1つだけパリティブロックを置く方法も考えられるが、この場合はディスク故障までしか対応不能のため除外した。以上のRAID4型構成の特徴は、正常時には働かないパリティ専用サーバが出来ることである。そのためこのパリティ専用サーバにビデオ放映の前後処理他を管理するコントロールサーバを兼用させる等の利用法も考えられる。

2.3 RAID4型の実現法と動作

前述のように図2のプログラム構成で、図2中のストライピングドライバにRAID4型の機能を作り込む形式で実現した。今回もサーバー側には何も手を加えていない。又今回もREAD機能だけを実装し、データの格納は別プログラムにより実施した。

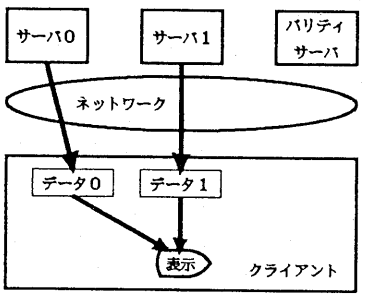

図6 正常時の動作

図6、7はRAID4型ビデオサーバの動作を示している。正常時の動作を示す図6では、パリティサーバにアクセスしないことを除いてRAID0型の時と同様である。図7はサーバ0が故障して切り離された状態（縮退運転）の動作を示しており、サーバ0の代わりにパリティサーバがアクセスされる。サーバ1のデータとパリティのExclisive OR演算を行い、サーバ0のデータを復元する。

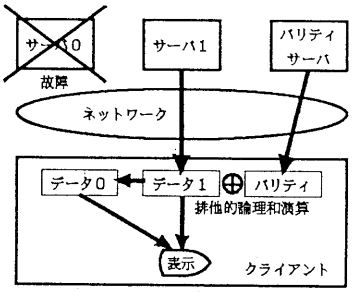

図7 縮退運転時の動作

2.4 ストライピングバッファ

この縮退運転時にビデオデータの特性を生かした高速化の手段が存在する。トランザクション処理等ではディスク上のデータブロックはランダムにアクセスされることが多いため、データを復元するために読みとられた同一パリティグループの他のデータブロックが再利用されることはほとんどない。これに対しビデオデータは連続的アクセスのため、上記パリティグループのデータブロックはパリティを除いて再利用されるという特徴がある。図7においてデータ1は最初データ0の復元を行う目的で読み出されるが、次にはこれに続く放映データとして読み出される。クライアント上に1パリティグループ分のストライピングバッファを設けることにより、上記2回の読み出しを1回に削減出来る。この機能により、図6、7の両者においてサーバからの読み出し、ネットワーク上の転送がそれぞれ2回ずつと同一であり、正常時と縮退時の性能（ビデオストリーム供給能力）に差が出ないことが判る。

ストライピングバッファは図8に示すようにクライ

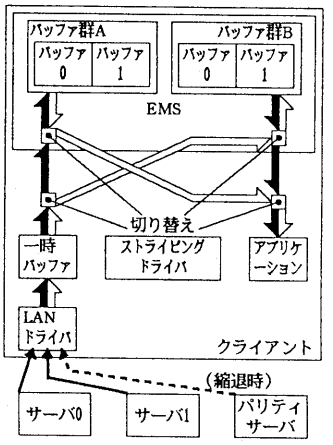

図8 ストライピングバッファの動作

アントのMS-DOS・EMS(Expanded Memory Specification)メモリ上に2つのバッファ群として構成される。サーバ数をSとすると、バッファ群は(S-1)個のバッファからなり、サーバから読み出された1ブロック(16KB)のデータが1バッファに読み込まれていく。MS-DOSの制約からこれは1バッファづつシリアルに行われ、バッファ群にデータが揃ったところでアプリケーションへの転送が可能となる。この間もう一つのバッファ群からはアプリケーションにデータが転送されダブルバッファリング処理が行われる。パリティサーバ以外のサーバが故障した縮退運転中は、故障サーバからのデータが入るべきバッファにパリティサーバからのパリティが読み込まれ、(S-1)個のバッファ間でExcusive OR演算を行って故障サーバのデータが復元され、結果が上記パリティが読み込まれたバッファに上書きされ、アプリケーションプログラムに渡される。

以上のようなストライピングバッファの働きにより、サーバ側の縮退時の性能を正常時と同一に保つことが出来る。

2.5 性能評価

以上述べたRAID4型分散RAID方式ビデオサーバを試作し性能評価を行った。性能評価環境はRAID0型の時と同様である。即ち図3のシステム構成で、ビデオデータはVideo for Windowsに付いてきたAVI形式の、1.2Mbpsの約5分のビデオデータを使用した。又同様に1台のクライアントではビデオデータをサーバー群から読み出して放映し、コマ落ち率の測定を行うが、他のクライアントでは疑似プログラムを走らせて、1つのクライアントからあたかも多数のクライアントがアクセスしているようにサーバー群に対して読み出しを行うようにし、クライアント数の不足を補った。コマ落ち率の測定は同じくVideo for Windowsに付いて来たテストプログラムを使用した。

図9がこの結果であり、正常時と縮退時の2つのケースにつき、ビデオストリーム数を増やしていった時のサーバCPUの使用率が示されている。このグラフから正常時と縮退時の性能がほとんど変わらないことが達成されていることが判る。最大供給ビデオストリーム数（サーバCPU=100%の点）はそれぞれ40.4、40.0と近く、これはRAID0型1サーバの20.2に対して約2倍の性能となっている。以上、当初の目標である、縮退時性能が落ちないこと、サーバ数に比例した性能向上の2点を達成していることを確認した。

表1は図9中の各測定点に対応した、放映クライアントでのコマ落ち率を示したものである。この表から上記測定時のコマ落ち率は0%ないし、多くとも0.04

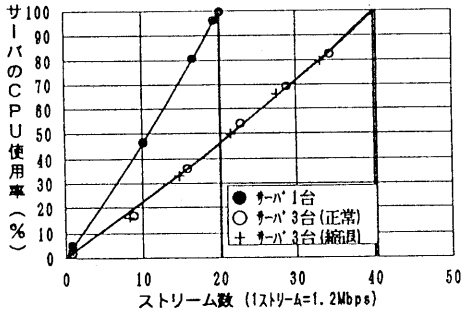

図9　RAID4型の性能評価

表1　コマ落ち率の測定

ケース	ストリーム数	コマ落ち数 (5128フレーム中)	コマ落ち率
RAID4 正常 3サーバ	1.0	0	0.00 %
	9.0	1	0.02 %
	15.9	1	0.02 %
	22.8	0	0.00 %
	28.9	0	0.00 %
	34.3	0	0.00 %
RAID4 縮退 3サーバ	1.0	1	0.02 %
	8.4	0	0.00 %
	14.9	2	0.04 %
	21.5	2	0.04 %
	27.6	0	0.00 %
	33.1	0	0.00 %

%と少なく、良好な映像が転送されていることが判る。又、少なくとも1台のサーバが故障して切り離された縮退運転中に、33ストリームのビデオデータを供給し、コマ落ち率0%で良好な映像であることが事実として確認出来たことも大きな成果である。この時クライアント側CPUはデータ復元のためにExclusive OR演算を実行するが、これは10%強のCPU負荷であり、クライアントCPUにとって大きな負担とはなっていない。

なお疑似故障の発生法は、故障させるサーバの共有ディレクトリサービスを停止させ、突然クライアント側からアクセス出来ないようにすることにより行った。この操作で故障発生させた場合には、放映クライアントの映像には全く影響を与えずに、故障サーバの切り離しと縮退モードへの切替えが行われることを確認した。

3. RAID5型のシミュレーション評価

3.1 正常動作時の性能向上

RAID4型では正常動作時にS台のサーバの内パリ

ティ以外の(S−1)台がオペレーションを実行するが、RAID5型ではパリティが固定されておらずS台のサーバーが動作出来るため、ビデオ供給性能が向上することを発見した。この場合の性能向上率はS/(S−1)と予測出来る。但し、RAID4型では故障発生時にも正常時と比較して性能低下が無かったが、RAID5型では故障発生時には(S−1)台のサーバのみで動くことになり、正常時に比べ(S−1)/Sだけ性能が低下する。

同一のハードウェア量でありながら正常時の性能がRAID4よりも向上する点はRAID5型の大きな魅力である。

3.2 アレイ構成法

RAID5型のアレイ構成法にはいくつかの方法があるが、有望なものは図10、図11の2種類である。

サーバ0 サーバ1 サーバ2

p	2	4
p	8	10
p	14	16

0	p	5
6	p	11
12	p	17

1	3	p
7	9	p
13	15	p

図 10

サーバ0 サーバ1 サーバ2

p	p	p
6	8	10
12	14	16

0	2	4
p	p	p
13	15	17

1	3	5
7	9	11
p	p	p

図 11

図10の方法は、パリティグループ毎にパリティの位置を1つづつずらしていく方法である。図11の方法は、サーバー当たりのディスク台数をDとすると、D回連続して同一サーバーにパリティを持たせる方法である。サーバー数をSとすると、図10の方法ではD＝nSのケース（整数倍）で、パリティだけを持つディスクが出来てしまうという欠点がある。図10はn＝1の場合のこの例であり、各サーバ内で1つのディスクにパリティが集中するため、正常動作時にこれらのディスクは遊んでしまいディスクネックの状況において性能低下を来す。又逆にnD＝Sのケース（整数倍）でもパリティが特定のディスクに片寄るという欠点がある。一方図11の方法はプログラムがディスク台数Dの値を意識しないといけない欠点がある。プログラムの複雑さよりも実利（性能）の方をとって図11を採用することとした。

3.3 ディスク単位の縮退

従来の分散RAID方式[3][4]ではサーバ単位の縮退のみが提案されていた。ここではRAID5型に限って、ディスク単位の縮退も可能とすることによりさらに可用性を増す方式を提案する。可用性を増すとはここでは故障発生による縮退運転時の性能低下を減らすことであり、ディスクネック状況で効果を発揮する。ディスク単位の縮退を実現するには、サーバのディスクを単に市販のRAIDディスクに置き換える方法もあるが、図12のようにパリティブロックが多くなる欠点がある。（図12中＊はRAIDディスクのパリティ、pは分散RAID方式のパリティである）図12では27ブロック中15ブロック＝56％がパリティとなるが、図13の我々の提案する方式では27ブロック中9ブロック＝33％と少なくて済む。一般に、要素サーバ数をS、要素サーバ当たりのディスク数をDとすると、当提案の方式の方がD/(D+S−1)の比率でパリティが少なくて済む。

サーバ0 サーバ1 サーバ2

*	p	p
	*	
		*

*		
p	*	p
		*

*		
	*	
p	p	*

図 12

サーバ0 サーバ1 サーバ2

p	p	p

p	p	p

p	p	p

図 13

次にサーバ縮退に対するディスク縮退の利点を考えてみる。1サーバが縮退した状況ではD(S−1)台のディスクが稼働するが、1ディスクが縮退した状況では(DS−1)台のディスクが稼働出来るため、ディスクネックの状況では(DS−1)／D(S−1)倍の性能向上が期待できる。

即ちディスク故障が発生した時にサーバー全体を切り離すのでなく、そのディスクだけを切り離すことで上記の性能上の利点が得られる。なおこの機構はRAID5型縮退時の性能低下を縮小するのに有効なものであるため、縮退時性能低下の無いRAID4型には適用しなかった。

3.4 シミュレーションによる性能評価

RAID5型分散RAID方式ビデオサーバーをシミュレーションにより性能評価した。このシミュレータはC言語で1332ステップで構成され、UNIX上で動作する。シミュレーションはディスクについては、シーク時間、回転待ち（確率はランダム）ディスクリード時間等詳細に行われるが、CPU処理時間、ネットワーク転送時間については大まかに近似的に行っている。そのためディスクネック状況のシミュレーションに適したものとなっている。本シミュレータの特徴はコマ

落ち率をシミュレーションしていることである。クライアントが必要時点よりも遅れてデータブロックを受け取った場合、最低次の1ブロックはスキップして（それ以上遅れている場合はその分、数ブロックスキップして）次のブロックに対してリクエストを出す。スキップした画面数をこの場合のコマ落ち画面数として計算する。このようにしてコマ落ちのシミュレーションが行われる。シミュレーションに使用したパラメータは次の通り。

サーバー数＝3、サーバー当たりのディスク数＝3、ディスクの回転待ち＝0～11msec、データブロックサイズ＝64KB、ビデオデータレイト＝1.2Mbps、ビデオデータサイズ＝1000ブロック、ビデオプログラム本数＝15

図14と図15にシミュレーション結果を示す。これらのグラフで横軸はビデオストリーム供給数、縦軸はコマ落ち率を示す。図14はRAID4型とRAID5型の通常動作時を比較したグラフである。ビデオサーバーのビデオ供給能力はコマ落ち率が0から正の値に変わる点と見ることにすると、図14においてRAID5型のビデオ供給能力は53ストリーム、RAID4型のそれは35ストリームとなる。この比53/35=1.51は、この時

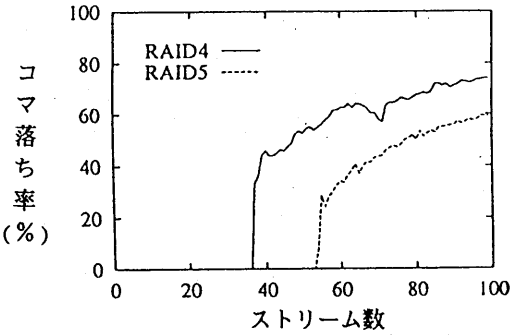

図１４　シミュレーション結果（１）

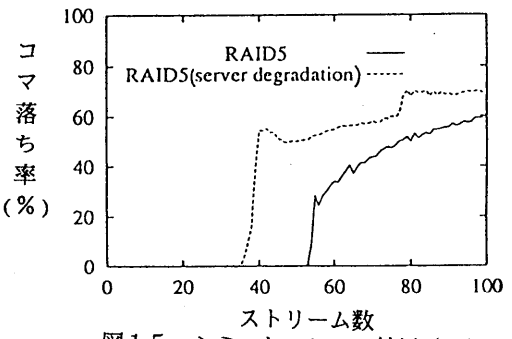

図１５　シミュレーション結果（２）

のそれぞれの動作サーバー数の比3/2=1.5とほぼ一致することが確認出来る。一方、図15はRAID5型の、通常動作時と縮退時を比較したものである。この場合の両者のビデオ供給能力は通常動作時が53ストリーム、縮退時が35ストリームと読みとられる。この比53/35=1.51も同様に動作サーバー数の比3/2=1.5にほぼ一致することが確認出来る。以上の結果から、通常動作時にはRAID5型の方がRAID4型よりビデオ供給能力が高いこと、縮退時には両者共同じになること、またこれらの値はその時の動作サーバー数に比例すること（計算と良く一致すること）、がシミュレーションにより確認出来た。

４．むすび

前回報告したRAID0型に引き続き、RAID4型の試作を行った。RAID0型ではC言語で700ステップであったが、RAID4型ではC言語で1200ステップのプログラムを開発した。又RAID5型についてはシミュレーションによる評価を実施した。これらの結果から、本分散RAID方式では、サーバー数に比例したスケーラブルな性能向上がほぼ実現出来ることが確認された。これは理論通りと言えばそれまでだが、サーバーのH/W（CPU、ディスク）、S/W（OS，通信ドライバー他）の処理性能、LAN上の振る舞い等の種々の要因がからむ中で、実際に検証されたことの意義は大きいと考える。試験は数十ビデオストリームという実際に近い環境で行われれ、映像品質も目視やコマ落ち率測定プログラムにより確認した。又RAID4型については、１サーバーの故障挿入試験も行い、その時にクライアント側の映像の乱れ無しに、縮退運転に移行することも確認した。

以上の結果から、分散RAID方式ビデオサーバーの実現性が大きく前進したと考える。

参考文献

[1] 中村，山口，峯村，渡辺，水野，“ビデオストリーム配信性能の一検証”，情報処理学会研究報告 95－DPS－72, p.37～42, Sep.1995.

[2] 中村，峯村，山口，清水，渡辺，水野，”分散ＲＡＩＤ方式ビデオサーバー”情報処理学会研究報告 95－DPS－72, p123～128,Dec.1995

[3] L.F. Cabrera, and D.D.E. Long “Swift: Using Distributed Disk Striping to Provide High I/O Data Rates,” Computing Systems, vol.4, no.4, Fall, 1991.

[4] 内川，横田，”ＶＯＤにおける耐故障並列ディスクの利用とバケット不配／遅配への対応”，情報処理学会研究報告 94－AVM－6, p1～7, Oct.1994.