

論文 / 著書情報
Article / Book Information

論題(和文)	仕様が異なるディスク混在ストレージの帯域・容量同時有効利用に向けた領域割当とデータ配置の評価
Title(English)	Evaluation of disk segmentation and data location policy for efficient use of storage capacity and bandwidth on the heterogeneous systems
著者(和文)	片居木誠, 横田治夫
Authors(English)	Makoto KATAIGI, Haruo YOKOTA
掲載誌(和文)	データ工学ワークショップ論文集
Citation(English)	Proceedings of Data Engineering Workshop
Vol, no, pages	, ,
発行日 / Pub. date	2008, 3
URL	http://search.ieice.org/
権利情報 / Copyright	本著作物の著作権は電子情報通信学会に帰属します。 Copyright (c) 2008 Institute of Electronics, Information and Communication Engineers.

仕様が異なるディスク混在ストレージの帯域・容量同時有効利用に向けた領域割当とデータ配置の評価

片居木 誠[†] 横田 治夫^{†,††}

[†] 東京工業大学 大学院 情報理工学研究科 計算工学専攻 〒152-8552 東京都目黒区大岡山 2-12-1

^{††} 東京工業大学 学術国際情報センター 〒152-8552 東京都目黒区大岡山 2-12-1

E-mail: [†]mkataigi@de.cs.titech.ac.jp, ^{††}yokota@cs.titech.ac.jp

あらまし 近年、帯域や容量等の仕様の異なるディスクが混在するストレージシステム構築の要求が増している。しかし、このようなシステムでは、帯域及び容量が不均衡であるため、速度性能を引き出すことが難しい。本稿では、データに対してプライマリ・バックアップ構造によって冗長性を付与しているシステムにおいて、そのプライマリ・バックアップ間のアクセス頻度の違いに着目し、均一環境において負荷とデータ量を均衡化させるデータ配置法 AOD 法の混在環境への拡張を行う。AOD 法においてはバックアップはプライマリの隣接ディスクに配置しなければならないという配置関係が存在することから、混在環境において容量割り当てが不可能となる領域を減らし、容量利用率を均衡化させるようにデータを配置するため、隣接ディスクのプライマリ・バックアップ対領域の割り当ての調整を行う。同時に、プライマリ・バックアップ対領域内でのプライマリ・バックアップの役割変更により、混在環境において帯域利用率を均衡化させる。また、本手法をシミュレーションで実装し、同様の目的で用いられる BSR を用いた手法と比較し、BSR を用いた手法と同等の帯域利用率でかつより均衡化した容量利用率であるという結果を得た。
キーワード ストレージ, 性能評価, データ構造

Evaluation of disk segmentation and data location policy for efficient use of storage capacity and bandwidth on the heterogeneous systems

Makoto KATAIGI[†] and Haruo YOKOTA^{†,††}

[†] Department of Computer Science, Graduated School of Information Science and Engineering,
Tokyo Institute of Technology, Ookayama 2-12-1, Meguro-ku, Tokyo, Japan

^{††} Global Scientific Information & Computing Center, Tokyo Institute of Technology,
Ookayama 2-12-1, Meguro-ku, Tokyo, Japan

E-mail: [†]mkataigi@de.cs.titech.ac.jp, ^{††}yokota@cs.titech.ac.jp

Abstract Recently, the requirement of building of heterogeneous storage systems, which are constructed by the disks with various bandwidth and capacity is growing. However it is difficult to efficiently use performance on these systems because of the imbalance of both bandwidth and capacity. On the other hand, the data location policy AOD, which uses the difference of access frequency between primary data and backup data on replication systems, can balance both access load and data size of each disks. Thus, we aim to extend the AOD method for heterogeneous storage systems. Since the data location policy of AOD method has to put the backup data to the neighbor disk of disk which primary data locate in, we extend this method to reduce the area, which backup data cannot be put on and balance the capacity utilization rate by adjusting the location of data. At the same time, we extend this method to balance the bandwidth utilization rate by changing the role of primary and backup. We also simulate the system, and compare with BSR method, which is used for same aim. The result shows that the extended AOD method is better than the BSR method in terms of the bandwidth utilization rate and capacity utilization rate.

Key words Storage system, Performance evaluation, Data structure

1. はじめに

近年のハードディスクの急速な広帯域・大容量化に伴い、様々な帯域及び容量のディスクが混在するストレージシステム(混在環境)構築への要求が高まっている。Grochowskiらの報告によると、1990年から2000年において、ハードディスクの帯域は年率約1.25倍、容量は年率約2倍で向上しているという[1]。そのためストレージシステムの、ディスク故障によるディスク交換や、システム拡張のためのディスク追加において、この新しい広帯域・大容量ディスクを利用したいという要求がある。

しかし、混在環境においてはこの帯域及び容量の不均一さのため、システムの速度性能を引き出すことが難しいという問題点がある。これは、帯域が偏って使用されるため、帯域が足りないディスクにおいて速度性能低下を引き起こしたり、容量が偏って使用されるため、容量利用率の高いディスクにおいてシーク性能が低下するなどがあるためである。そのため現状では、ディスク交換において旧来の狭帯域・小容量のディスクを使用したり、新しい広帯域・大容量ディスクを使用したとしても、その帯域や容量をシステム内の他のディスクに合わせて制限するなどしており、広帯域・大容量を有効に利用できていない。

一方、均一環境においては、各ディスクに対する負荷及び保存データ量を均衡化させる手法として Adaptive Overlapped Declustering (AOD) 法 [2] がある。AOD 法は、冗長化のためのバックアップデータを保持しており、データへのアクセスはプライマリデータに対して行われるという環境で用いられる。そのプライマリデータとバックアップデータのアクセス傾向の違いに着目し、アクセス対象となるプライマリデータの配置で負荷の均衡化、アクセス対象とならないバックアップデータの配置でデータ量の均衡化を図る手法である。

ここで、我々はこの AOD 法のプライマリデータで負荷を均衡化し、バックアップデータでデータ量を均衡化させるという点が、混在環境における各ディスクの帯域の差及び容量の差に対処するために利用できるのではないかと考えた。具体的には、各ディスクの負荷の比を各ディスクの帯域の比と一致させるように調整し、また各ディスク上のデータ量の比を各ディスクの容量の比に一致させるように調整する。これにより各ディスクの帯域・容量がそれぞれ異なる環境でも、それぞれのディスクの帯域・容量を有効利用できるのではないかと考えた。

ここで、我々はまず AOD 法を混在環境へ拡張する上で、無駄な領域が出来てしまう問題を解決した。これにはデータ配置の条件を満たしながら帯域・容量の有効利用を可能にするためのデータ配置領域の決定を行った [3]。これは、プライマリデータとバックアップデータの配置に制約があるためであり、データの配置によってはバックアップデータを配置できない領域が出来てしまう可能性があるためである。そのため、バックアップデータを置けない領域を最小化しつつ、各ディスクのプライマリ領域に対する帯域対容量比 (Bandwidth to Space Ratio [4]) を均衡化させるデータ配置領域決定法を提案してきた。これは、BSR が均衡化している環境においては、帯域及び容量の有効利用が可能であるという結果が出ているためである。

本稿ではこれらのデータ配置領域決定法によって決定したデータ配置領域を元に、各ディスクへの実際のデータ配置の決定を行う。まず、最終的にどのようなデータの配置になれば各ディスクの帯域・容量を有効利用でき、それにより速度性能を引き出せるのかについて考察したのち、そのデータ配置にするために従来の AOD 法におけるデータ再配置法であるディスク内データマイグレーション及びディスク間データマイグレーションを混在環境において適用できるように変更を行う。また、本手法をシミュレーションにより実装し、類似手法である BSR を利用した手法及び従来の AOD 法と比較しその性能を評価する。

本稿の構成を示す。まず 2. 章で帯域と容量の異なるディスクが混在するストレージシステムにおいて、帯域と容量を有効利用するための試みとその問題点を示す。その後 3. 章で本手法の前提である AOD 法について述べる。続いて 4. 章で本稿の主張である AOD 法を混在環境に対して適用するための拡張法を述べる。そして、5. 章において本手法を実装したシミュレーションについて述べ、6. 章において BSR を用いた手法と比較した実験結果及びその考察について述べる。最後に 7. 章において本稿のまとめと今後の課題について述べる。

2. 関連研究

様々な帯域と容量の混在するシステム構築のための、混在環境での帯域と容量の有効利用について、様々な角度から研究されている。たとえば、Striping や RAID5 環境において帯域と容量の同時有効利用を目指す研究や、データのアクセス頻度を利用したもの、容量や帯域が混在しているものを仮想的に均一に見せる方法などがある。特に BSR を用いた手法は我々の提案手法の対抗手法となるため、詳しく説明する。

AdaptRAID0 [5] では Striping 環境において、また AdaptRAID5 [6] では RAID5 環境において、各ディスクに配置するストライプユニットの数を各ディスクの容量に比例させることによって、ディスクの容量利用率の向上を図っている。これらの手法では、どのディスクも帯域と容量の比率はほぼ一定であると仮定することにより、どちらか片方、この場合は容量を均衡化させることによって帯域も均衡化するとしている。しかし、1. で述べたとおり、混在環境では容量と帯域の比率は一定にはなっていない。また、Weight Striping [7] では、各ディスクにウェイトという値をヒューリスティックに定めて、その値に比例させるようにストライプユニットを配置することにより、ウェイトの値によって柔軟に帯域・容量のどちらを重視して均衡化させるか調整する手法である。しかし、どちらも同時に均衡化することはできない。どの手法も、データへのアクセス量はサイズに比例すると考えているため、混在環境での帯域と容量の同時均衡化ができない。

この問題点を解決した手法として Resource Based Striping [8] がある。この手法では、データ毎のアクセス量の違いを利用し、各ディスクへのデータの割り当てを、アクセスの多いデータは帯域の多いディスクに多く配置し、アクセスの少ないデータは帯域の少ないディスクに多く配置することにより、システム全体として帯域及び容量の有効利用を図っている。しか

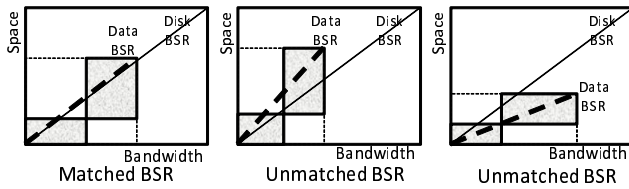


図 1 BSR の一致

し、この手法においてそれぞれのデータ配置を計算するためには、システムに保存される全データ及びその各データへのアクセス量が事前に分かっているとイケない。しかし、実際にはそのようなシステムは一般的ではない。

また、Striping や RAID5 のようなデータを分割して配置する環境だけではなくデータを分割しない環境でも、帯域及び容量が混在する環境でその帯域と容量を有効利用する研究がおこなわれている。例えば、Disk Merging [9] では複数のディスクをまとめて仮想的に一つのディスクとみなす、または逆に一つのディスクを分割して仮想的に複数のディスクにみなす方法である。また、そのそれぞれの仮想ディスクに割り当てる帯域及び容量を調整することにより、システム内での仮想ディスクの帯域対容量比を一定にすることによって、システム全体での帯域と容量の有効利用を行っている。しかし、元のディスクの組み合わせによっては必ずしも仮想ディスクの帯域対容量比を一定にすることができるわけではない。

2.1 BSR を利用した手法

BSR(Bandwidth to Space Ratio: 帯域対容量比) を基にしたデータ配置決定方法 [4] は、ディスク及びデータの BSR を利用し、各ディスクの BSR とそのディスク上のデータの BSR が一致するようにデータを配置する(図 1)。これによって、各ディスクの帯域と容量を有効に利用している。

さらに、新しいデータの挿入やアクセス傾向の変化に対しては、データの BSR がディスクの BSR より大きいディスク上のデータのレプリカを作成し、他のディスク例えばデータの BSR がディスクの BSR より小さいディスクに配置し、各レプリカへのアクセス割合を調整することにより、再びディスクの BSR とデータの BSR を一致させている。また、レプリカ作成は、システム内の各ディスクの状況により動的に行われるため、あらかじめアクセス傾向を知っている必要はなく、またアクセス傾向が変わった場合にも対応できる。

[4] における実験の結果を見ると、各ディスクの BSR が同じである環境においては良い帯域利用率・容量利用率が得られている。しかし、各ディスクの BSR が均質化していない状況、帯域が最大 20% 違う環境においては結果が芳しくない。しかし前章で述べたとおり、混在環境を構築する際にはシステム内でのディスク間の帯域の違いは 20% よりはるかに大きくなるため、この手法による帯域及び容量の有効利用は難しいと考える。

3. Adaptive Overlapped Declustering 法

我々はディスクの容量及び帯域が混在する環境で、容量及び帯域を有効利用するために、AOD 法を拡張することを考える。

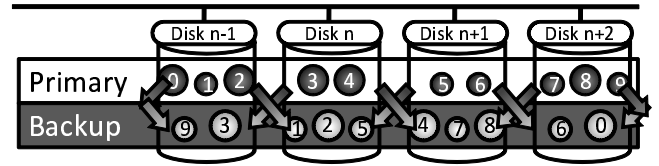


図 2 AOD のデータ配置

そのため、本章ではまず AOD 法 [2] の概略について説明する。

AOD 法は値域分散におけるディスク間での負荷とデータ量を同時に均質化することが困難であるという問題点を解決するための手法である。従来の値域分散を用いたインデックスにおいては、データの配置が値によって一意に決まってしまうため、各ディスクの負荷及びデータ量を同時に均質化させることは非常に難しかった。AOD 法は、冗長化のためにバックアップを保持しており、そのバックアップの配置として Chained Declustering を採用している環境において用いられる。まず、プライマリデータの配置で負荷の均質化、さらにバックアップデータをプライマリデータの左右のディスクへ分散させ、その分散割合を調整することによりデータ量の均質化を行う。また新しいデータの挿入やアクセス傾向の変化に対して、2 種類のデータ再配置により動的に均質化状態を維持する手法である。

なお、各ディスクに論理的な番号が付けてあり、また番号の一つ小さいディスクを左のディスク、番号の一つ大きいディスクを右のディスクと呼ぶことにする。また、以降本稿では負荷とは各データを読み出すために使われているディスク帯域を表す。

3.1 AOD 法におけるデータ配置

AOD 法では、プライマリデータとバックアップデータのアクセス頻度の違いを利用する。アクセスされるプライマリデータの配置によって負荷の均質化、アクセスされないバックアップデータをプライマリデータの左右に分散して配置し、その分散割合を変更させることによってデータ量の均質化を行い、各ディスクの負荷とデータ量を均質化させる(図 2)。

また、新しいデータの挿入やアクセス傾向の変化が発生した際に、負荷とデータ量の均質化状態が崩れてしまう。これに対してデータの動的再配置を用いることによって再均質化を行う。ここでは 2 種類のデータ再配置法、負荷を均質化するためのディスク内データマイグレーションと容量を均質化するためのディスク間データマイグレーションを併用し、均質化を行う。AOD 法でも BSR を用いた手法と同じく、システム内の各ディスクの状況により、動的にデータの再配置が行われるため、あらかじめアクセス傾向が分かっている必要はなく、またアクセス傾向の変化にも対応できる。

3.2 ディスク内データマイグレーション

ディスク内データマイグレーションは隣接ディスク間において、プライマリデータとそのバックアップデータのお互いの役割を入れ替えることによって、負荷を移動させる方法である(図 3)。このとき、データの役割を入れ替える操作だけでよいので、実際のデータの移動を伴うことなく、ディスク間での負荷の移動を行うことができる。AOD 法では、配置条件から、必ずプライマリデータとバックアップデータは隣接ディスクに保存され

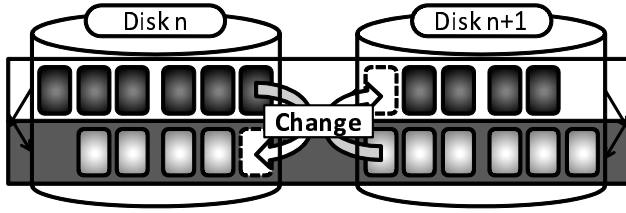


図3 ディスク内データマイグレーション

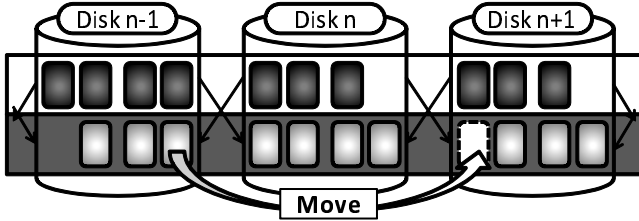


図4 ディスク間データマイグレーション

ているため、このような役割変更が行える。なお、実際のデータ移動を伴わないため、データ量の移動は起こらない。

システム全体としては、隣接するディスク n 及び $n+1$ において、ディスク n 上の各データ $Data\ i$ の負荷 ($A(i)$) の合計とディスク $n+1$ 上の各データ $Data\ j$ の負荷の合計の差の絶対値 δ_{band} を最小化するようにデータの役割変更を行う。順に操作対象となるディスクをずらしていくことにより、漸近的にシステム全体で負荷を均衡化していく手法である。なお、ディスク n 上の全データ集合を $Data_in(n)$ と定義する。負荷の差 δ_{band} は以下のとおりである。

$$\delta_{band}(n) = \left| \sum_{i \in Data_in(n)} A(i) - \sum_{j \in Data_in(n+1)} A(j) \right| \quad (1)$$

3.3 ディスク間データマイグレーション

ディスク間データマイグレーションはバックアップデータをプライマリデータに対して反対側の隣接ディスクに移動させることにより、データ量を移動させる方法である(図4)。AOD法ではバックアップデータはプライマリデータの左右のどちらかのディスクに配置すればよく、反対の隣接ディスクへ移動することができる。この方法ではプライマリデータは動かないため、データアクセスへの影響は少なく、負荷の移動は起らない。

システム全体としては、隣接ディスク n 及び $n+1$ において、ディスク n 上の各データ $Data\ i$ のデータ量 ($S(i)$) の合計とディスク $n+1$ 上の各データ $Data\ j$ のデータ量の合計の差の絶対値 δ_{cap} を最小化するようにデータ移動を行う。順に操作対象となるディスクをずらしていくことにより、漸的にシステム全体でデータ量を均衡化していく手法である。データ量の差 δ_{cap} は以下のとおりである。

$$\delta_{cap}(n) = \left| \sum_{i \in Data_in(n-1)} S(i) - \sum_{j \in Data_in(n+1)} S(j) \right| \quad (2)$$

上記のディスク内データマイグレーションとディスク間データマイグレーションを合わせることによって、AOD法ではディスク毎の負荷の均衡化、及びデータ量の均衡化を行っている。

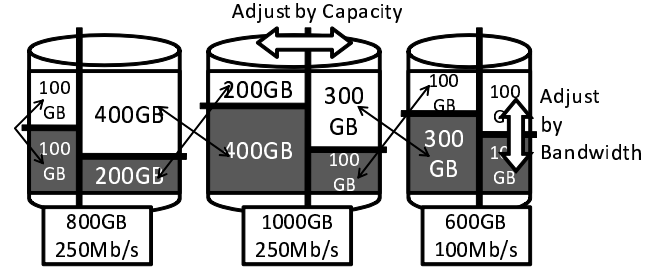


図5 ディスクの領域分割

4. AODの混在環境への適用

本章ではAOD法を混在環境に拡張することによって、帯域及び容量を有効利用し、速度性能を引き出す方法を考える。まず、3.で述べたとおり、AOD法ではデータ配置において、バックアップデータをプライマリデータの隣接ディスクに配置しなければならないという条件が存在する。AOD法を混在環境に適用するためには、この配置の条件を満たさなくてはならない。しかし、無計画にデータを配置してしまうと、バックアップデータが配置できない領域ができてしまう。

そのために、配置条件を満たしつつ利用可能容量が最大となり、プライマリ領域のBSRが均衡化するようなディスクの領域分割を行う。その後、各ディスクの容量利用率及び帯域利用率が均衡化するようにデータの配置を行う。また、データ追加やアクセス傾向の変化に対して、データ再配置によって、各ディスクの帯域利用率および容量利用率を再均衡化させる。

4.1 ディスクの領域分割

我々は[3]において、データ配置条件を満たしつつ、帯域及び容量を有効利用するため、利用可能容量を最大化し、各ディスクのプライマリ領域のBSRを均衡化させるディスク領域分割手法を提案してきた。この手法は、まず利用可能領域を最大化するため、左右どちらのディスクにもバックアップを配置できなくなる領域を最小化させるように、各ディスクのバックアップを左右に振り分けて配置する領域の決定を行った(図5中の縦棒、以降この領域をそれぞれディスク内の左の領域および右の領域と呼ぶ)。その後、各ディスクの帯域を有効利用するため、各ディスクにおけるプライマリデータを配置する領域の大きさを各ディスクのプライマリ領域のBSRが均衡化するように調整した(図5中の横棒、以降この領域をそれぞれプライマリ領域およびバックアップ領域と呼ぶ)。

ディスク領域を分割したのち、どのようにデータを配置すると良いのかを考察する。さらに、その配置にするためのデータ再配置戦略を考察する。また、各ディスクのデータ領域をそれぞれ左右の領域及びプライマリ・バックアップ領域へと分割したことによるデータ配置戦略への影響や均衡化への効果を測定するため、それぞれ領域を分割する再配置戦略(AOD-W)と領域を分割しない再配置戦略(AOD-S)を提案する。まず、混在環境においては、何を均衡化すべきかを考察する。

4.2 混在環境におけるデータ配置

混在環境において、その速度性能を引き出すためには、帯域

利用率及び容量利用率の均衡化が必要である。まず、データを読み出す場合、そのデータにとって必要な分だけの帯域を占有し、また必要な帯域が確保できない場合は読み出しが行えず、速度性能低下につながる。そのため、速度性能を引き出すためには、帯域に対して負荷を偏らせない、つまり帯域利用率を均衡化する必要があるためである。また、ディスクのシーク性能はそのディスク上の容量使用率が大きく影響する。そのため、各ディスクのシーク性能を落とさないためには、容量利用率を均衡化させることが必要である。

しかし、AOD法ではデータの配置は値域分散で決定されるため、初期配置のままでは均衡化が行えない。そのため、データマイグレーションを利用することによって、この帯域利用率及び容量利用率が均衡化した状況を作り出す。以下ではそのデータマイグレーションを各ディスクの帯域利用率・容量利用率を均衡化させられるように拡張する方法を示す。

4.3 ディスク内データマイグレーションの拡張

従来のディスク内データマイグレーションでは、各ディスクへの負荷を均衡化していた。しかし、様々な帯域のディスクが混在する環境においては、単純に負荷を均衡化するだけでは帯域の有効利用はできない。そのためには、各ディスクの帯域利用率の均衡化が必要である。

まず、プライマリ・バックアップ領域に分割しない手法(AOD-W)を考える。この場合、ディスク内マイグレーションの移動決定のための負荷の差の絶対値 δ_{band} を、以下の式(3)のように、負荷だけではなく、各ディスク毎の帯域 $B(n)$ も用いて帯域利用率(=負荷/帯域)とした手法を提案する。

$$\delta_{band-w}(n) \equiv \left| \frac{\sum_{i \in Data_in(n)} A(i)}{B(n)} - \frac{\sum_{j \in Data_in(n+1)} A(j)}{B(n+1)} \right| \quad (3)$$

また、プライマリ・バックアップ領域に分割した手法(AOD-S)についても考える。この場合、各領域毎の帯域利用率を均衡化することによって、システム全体での帯域利用率の均衡化を考えなくてはならない。しかし、ディスクのデータ配置領域決定時には各領域への帯域は決めておらず、その代りに各領域へ配置されるプライマリデータ量が決められている。このプライマリデータ量は、各領域で使用して欲しい帯域の割合に応じて決められた値である。そのため、各領域について、このプライマリ領域の大きさに応じた帯域を割り当てればよいことになる。よってディスク n の左の領域 L_n 及び右の領域 R_n に割り当てられた帯域 $B(L_n), B(R_n)$ を用いて、ディスク内データマイグレーションの負荷の差の絶対値 δ_{band} を、以下の式(4)のように変更した手法を提案する。

$$\delta_{band-s}(n) \equiv \left| \frac{\sum_{i \in Data_in(R_n)} A(i)}{B(R_n)} - \frac{\sum_{j \in Data_in(L_{n+1})} A(j)}{B(L_{n+1})} \right| \quad (4)$$

これによって、各ディスクの帯域利用率の差が小さくなることにより、システム全体での帯域の有効利用が行える。

4.4 ディスク間データマイグレーションの拡張

従来のディスク間データマイグレーションでは、各ディスクに保存されるデータ量を均衡化していた。しかし、様々な容量

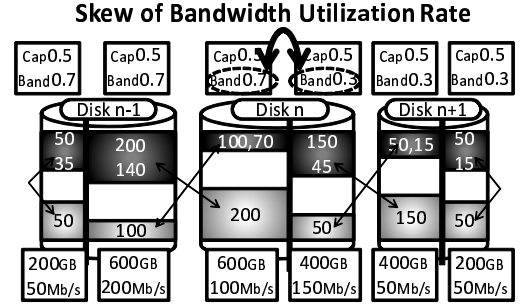


図6 帯域利用率の偏り

のディスクが混在する環境において、容量を有効利用するためには、各ディスクの容量利用率を均衡化させる必要がある。

まず、ディスクを左右の領域に分割しない場合の手法(AOD-W)を考える。この場合、ディスク間マイグレーションの移動決定のためのデータ量の差の絶対値 δ_{cap} を、以下の式(5)のように、データ量だけではなく、ディスク毎の容量 $C(n)$ も用いて容量利用率(=データ量/容量)とした手法を提案する。

$$\delta_{cap-w}(n) \equiv \left| \frac{\sum_{i \in Data_in(n-1)} S(i)}{C(n-1)} - \frac{\sum_{j \in Data_in(n+1)} S(j)}{C(n+1)} \right| \quad (5)$$

また、ディスクを左右の領域に分割した手法(AOD-S)についても考える。この場合、各領域毎の容量利用率を考えることによって、システム全体での容量使用率を考えなくてはならない。よって、ディスク n の左の領域 L_n 及び右の領域 R_n に割り当てられた容量 $C(L_n), C(R_n)$ を用いて、ディスク間データマイグレーションの容量の差の絶対値 δ_{cap} を、以下の式(6)のように変更した手法を提案する。

$$\delta_{cap-s}(n) \equiv \left| \frac{\sum_{i \in Data_in(L_n)} S(i)}{C(L_n)} - \frac{\sum_{j \in Data_in(R_n)} S(j)}{C(R_n)} \right| \quad (6)$$

これによって、各ディスクの容量利用率の差が小さくなることになり、システム全体での容量の有効利用が行える。

4.5 帯域変更

しかし、これだけでは問題がある。実際、図6のように、全領域の容量利用率がすべて0.5で同じになっており、かつDisk $n-1$ の右の領域とDisk n の左の領域の帯域利用率がそれぞれ0.7で等しく、また、Disk n の右の領域とDisk $n+1$ の左の領域の帯域利用率がそれぞれ0.3で等しくなっているような場合がある。この場合、各領域で容量利用率及び帯域利用率は均衡化しているため、ディスク内データマイグレーション・ディスク間データマイグレーションともに行われぬ。しかし、ディスク n の左右の領域の帯域利用率はそれぞれ0.7, 0.3と偏っている。これはディスク内の左右の領域間の帯域利用率を均衡化する方法がないためである。そのため、この領域間での帯域偏りを解消する方法が必要である。

しかし、ディスク内の左右の領域間でのデータ移動を行うことはできない。そのため、この領域間で帯域利用率を均衡化させるためには、各領域に割り当てられた帯域自身を変更する必要がある。そのため、ここではディスク内の左と右の領域に割

り当てられた帯域を再割り当てし、左右の領域の帯域利用率を均衡化させ、左右のディスクとの間の領域を再び不均一にし、再びディスク内マイグレーションを促すことにより、全ディスクにおける各領域の帯域利用率を均衡化させる。

5. シミュレーション実装

提案手法の有用性を確認するために、提案手法のシミュレータを作成した。まず、シミュレーションで対象とするシステムである、マルチメディアデータサーバの構成及びデータに対するアクセス傾向について紹介する。本手法においては、システム内での帯域利用率および容量利用率を均衡化させることにより速度性能を引き出すことを考えるため、本シミュレーションでは特にディスクの帯域及び容量部分についてシミュレートした。また、マルチメディアサーバに対して、想定されるアクセスを与えた時の性能を測定する。特に、帯域に対する性能として、システム全体でのリクエスト待機時間、容量に対する性能として、システム内の各ディスクの容量利用率の分散を測定した。

5.1 前提となるシステム

本シミュレーションで対象とするマルチメディアデータサーバとは、クライアントの要求に応じて動画データを返すようなストレージシステムのことである。このようなシステムに使われるストレージシステムは、大量かつサイズの大きい動画データを保存する必要がある、また動画データは増えていく一方であるため、巨大化する傾向がある。

マルチメディアサーバのデータは、すべてのデータが一律にアクセスされるということではなく、アクセス頻度の高いデータとアクセス頻度の低いデータが混在するため、データ間のアクセス頻度が違ってくる。また、動画データでは、動画を途切れることなく見るためには、常に一定以上のデータ転送速度が要求される。そのため、このようなシステムではいかに一定以上のデータ転送速度で送信できるかが重要になってくる。

5.2 シミュレーション環境

まず、帯域に対する性能として、システム全体のリクエスト待機時間を測定する。これは、各クライアントはそれぞれこのディスクに対してコネクションを張り、動画データの読み出しに必要な帯域を占有することによりデータを読み出す。しかし、必要な帯域を占有できなかった場合は、その読み出しリクエストは待機させられる。これは、占有できないまま帯域を使用してしまうとほかのクライアントの読み出しにも影響を与えてしまうためである。次に、容量に対する性能として、それぞれのディスクの容量利用率を測定する。これは、それぞれ違う容量を持つ複数のディスクにおいて、シーク性能を引き出すために、各ディスクにおいて使用されている容量の割合を一定に保つ必要があるからである。

また、ネットワークについては、各クライアントから全ストレージにそれぞれ接続されており、ユーザは任意のストレージのデータを取得できるとする。またストレージ間もそれぞれお互いにネットワークでつながれており、任意のストレージ同士でデータを転送できるものとする。前提システムのようなマルチメディアサーバにおけるアクセスは、読み出しリクエストが

表 1 パラメータ設定

Storage Number	10
Client Number	300
Initial Storage Capacity	30GB
Capacity Growth Rate	200%/year
Initial Storage Bandwidth	160Mb/s
Bandwidth Growth Rate	125%/year
Total Data Size	50% of Total Capacity Size
Data Size	1 ~ 5GB (uniform distribution)
Data Required Bandwidth	30 ~ 40Mb/s (uniform distribution)
Request Number	100000
Zipf Parameter	0.271

中心であり、書き込みリクエストは非常に少ない。本実験においては、読み出しリクエストのみを行い、その性能を測定する。また、本システムは動画配信のためにだけ用いられるため、クライアント以外からのアクセスはないとする。マルチメディアサーバのような環境では、アクセスはすべてのデータに対して均一になされるのではなく、アクセス頻度の高いデータとアクセス頻度の低いデータが出てくる。なおそのアクセス傾向は Zipf 分布で近似される [10]。

6. 実験

本実験の目的は拡張した AOD 法のデータ配置による帯域利用率及び容量利用率均衡化の効果を測定することにある。そのため、本実験ではデータ再配置が終わった後の均衡化状態の性能について注目する。なお、本実験において使用したパラメータは表 1 に示すとおりである。以下に実験手順を示す。

- (1) 10 台のストレージによりシステムを構築する
 - (2) システムにデータを挿入する、このとき挿入するデータの総サイズをプライマリデータ及びバックアップデータ合わせてシステムの総容量の 50% となるようにする、なお 2 回目以降は古いデータを削除したのち再挿入する
 - (3) クライアントから合計 100,000 回アクセスを行う、システムはこの間にデータへのアクセス傾向を取得する
 - (4) 各手法によりデータの再配置を行う
 - (5) 再度クライアントから合計 100,000 回アクセスを行い、性能測定を行う、なお 4 回繰り返し測定しその平均を取る
 - (6) ディスクを 1 台取り換える、このとき取りのぞかれるディスクはランダムに選ばれる
 - (7) (2) ~ (6) を 20 回繰り返す
- ストレージとしては、現在のモデル (1000GB, 500Mb/s) [11] を基準に、帯域向上率 1.25 倍/年、容量向上率 2 倍/年として、5 年前 (30GB, 160Mb/s) に使用を開始したとして計算を行った。このストレージを 10 台用いて、故障したら故障した時点の最新のディスクに入れ替えるとし、入れ替えを行った合計台数をシステムの混在具合のパラメータとした (以降これを混在度と呼ぶ)。また、データについてはサイズを 1 ~ 5GB、及び要求帯域を地上デジタル放送によるデータ配信をモデルとして 30 ~ 40Mb/s とし、これを現在の基準として、5 年前 (30

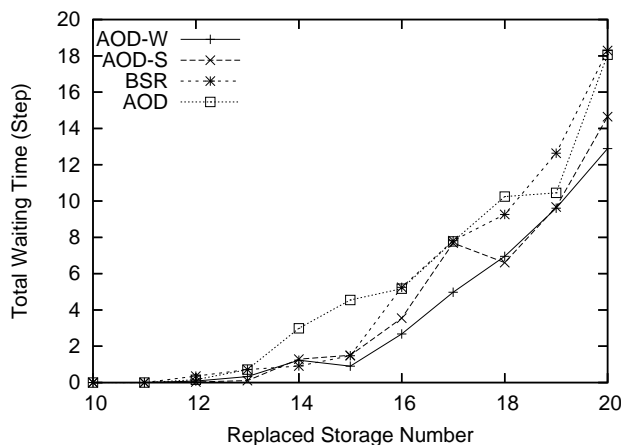


図 7 混在度に対するリクエスト待時間

~150MB, 9.8~13Mb/s) からディスクと同様にサイズが年率 2 倍, 要求帯域が年率 1.25 倍でそれぞれ増加しているとする。また, それぞれ一様分布とした。また, アクセス傾向としては, [10] で述べられている, パラメータ 0.271 の Zipf 分布とした。また, クライアント数を 300 台としている。

以上の環境において, 帯域及び容量を分割しない方法で AOD 法を拡張した手法 (AOD-W) と帯域及び容量を分割した方法で AOD 法を拡張した手法 (AOD-S) と BSR を用いた手法 (BSR) 及び AOD 法を拡張せずにそのまま適用した手法の 4 手法についてリクエスト待機時間及び平均容量利用率を測定した。なお, BSR を用いた手法では, バックアップを持つことを前提としていないため, AOD 法と同じ条件にするために, 最低 1 つレプリカを持つこととした。また, 各手法 3 回の平均値を取得した。

6.1 帯域比較

リクエスト待機時間の結果は図 7 のようになった。グラフからわかるとおり, 混在度が高い状態では, AOD 法を拡張した 2 手法の待機時間は少なくとも BSR を用いた手法及び従来の AOD 法の待機時間を下回ることはなく, 帯域利用率が均衡化できているという結果になった。これは AOD 法を拡張した手法では, 各ディスクの「帯域利用率」を均衡化しており, 全ディスクの帯域利用率が均衡化し, 一部ディスクへのリクエスト集中がなかったが, BSR を用いた手法ではディスクとデータの BSR のかい離度のみ見ているため, また従来の AOD 法各ディスクのデータ量のみを見ているため, 必ずしも帯域利用率が均衡化せず, 一部ディスクにリクエストが集中したためである。

次に AOD を拡張した 2 手法について比較する。両手法の結果をしてみると, 混在度によってその優劣は変わっており, どちらかの方が優れているとはいえない。これは, 負荷については必ずしもディスク内の左右の領域で均衡化している必要はなく, 左右の領域では偏っていても, ディスク全体として偏っていなければ, 十分帯域の有効利用が可能であるからである。

6.2 容量比較

容量利用率の分散の結果は図 8 のようになった。AOD を拡張した 2 手法と従来の AOD を比較すると, その容量利用率の分散は小さくなっており, 容量利用率が均衡化されたことがわ

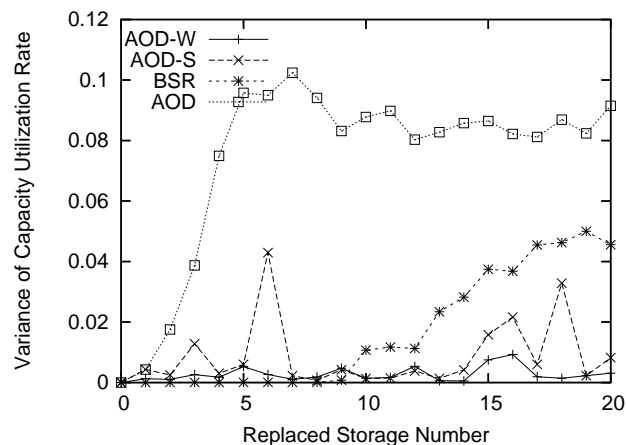


図 8 混在度に対する容量利用率の分散

かる。また, AOD を拡張した 2 手法と BSR を用いた手法を比較すると, 交換台数 10 台以降は AOD 法を拡張した手法の方が容量利用率は均衡化されているという結果となっている。これは, BSR を用いた手法では, レプリカを作成することによって帯域の有効利用を図るため, 帯域利用率の偏りが大きくなってきた時点で大量にレプリカを作成したためである。それに対して, AOD 法を拡張した手法では, データの移動のみを行い帯域利用率を均衡化しているため, 容量利用率は最初から最後までほぼ変わらない。

次に AOD を拡張した 2 手法について比較する。両手法の結果をしてみると, 分割を行った手法の方が容量利用率の偏りが大きくなっている。しかし, 必ずしもすべての点において大きいわけではなく, 同程度の値を示している点と, 大きくなっている点が存在することがわかる。これは分割を行った手法を用いた場合, ディスクの組み合わせによっては, 非常に容量の小さな領域ができてしまい, その領域が壁となり容量分散が進まない状況になるためである。たとえば図 9 における Disk $n-1$ の右の領域及びディスク n の左の領域がそれぞれ 4GB 及び 6GB となり, 他の領域 (240GB や 390GB) に比べて非常に小さい容量になっている。容量の小さい領域においては, 領域へのデータ配置や領域からのデータ移動を行った場合, 容量利用率が急激に変わってしまう。結果, 容量の小さい領域においてデータを配置したり, 逆にデータを移動させることができなくなってしまう。結局, 小さな領域の左右で, データを移動できなくなってしまうため, 容量の小さな領域の隣接ディスクにおいて容量利用率の均衡化がうまくできなくなってしまう。その結果, 確実に分散させられる分割しない手法 (AOD-W) の方が平均するとよくなるためである。

6.3 実験考察のまとめ

以上により, AOD 法を拡張した手法では BSR を用いた手法及び従来の AOD 法に比べて, ほぼ同程度帯域利用率が均衡化しており, 容量利用率は大きく均衡化している。よって, 速度性能について帯域利用率の均衡化による効果及び容量利用率の均衡化によるシーク性能の双方を考慮すれば, AOD 法を拡張した手法は BSR を用いた手法及び従来の AOD 法に比べ向上

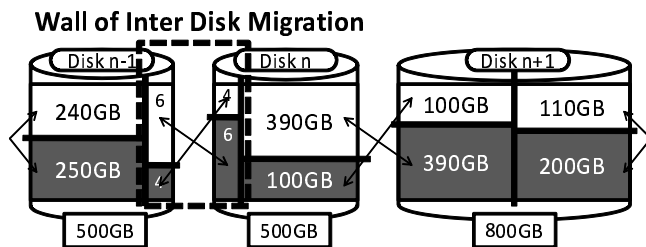


図9 ディスク間データマイグレーションのボトルネック

すると考えられる。

AOD法を拡張した2手法では、分割しない方が若干良いという結果になった。これは、分割した手法ではその分割の仕方によっては逆に戦略上の壁となってしまいうためである。今後分割した手法をよくするためには、データ移動を考慮して、一時的に領域を超える量になったとしてもデータ移動ができるようにするなどの改良が必要である。

7. まとめと今後の課題

本稿では、ハードディスクドライブの帯域及び容量の向上に伴う、様々な帯域や容量のディスクが混在するストレージシステム構築の要求を満たすため、ディスク混在環境においても速度性能を引き出すため、その帯域利用率と容量利用率を同時に均衡化することを目指した。そのために均一環境において負荷及びデータ量を均衡化する手法であるAOD法の拡張を行った。本稿では[3]で提案したディスク領域分割手法を基にデータ配置の決定を行った。まず、プライマリデータの配置で帯域利用率を均衡化、バックアップデータの配置で容量利用率を均衡化することを考えた。このために、ディスク内データマイグレーション、及びディスク間データマイグレーションの拡張を行った。また、シミュレーションにより実験し、BSRを用いた手法及び従来のAOD法と比較することにより、AOD法を拡張した手法の有用性を確認した。特に、データ配置領域を分割しない手法による帯域利用率及び容量利用率の均衡化を確認し、総合的な速度性能を最も引き出せると考えられることを示した。

今後の課題としては、まず、より詳細な調査が必要である。本稿におけるシミュレーションでは容量利用率のみを測定したが、より詳細には容量利用率によるシーク性能への影響をシミュレーションに含める必要がある。次に、データ移動時に容量の小さい領域が壁になってしまうという問題点について考慮して改良する必要がある。このとき、データ移動時に一時的に領域の容量を超えたとしても、最終的にその領域の大きさ以内に収まるならば移動できるようにすることなどにより解決できると考える。さらに、本手法によるデータ配置が、最適なデータ配置と比較して、どの程度の最適性を達成しているか検証を行う必要がある。加えて、マイグレーションそのもののコストの評価や、マイグレーション実行時の性能評価についても行う必要がある。今回は特にReadアクセスが大半と仮定して評価を行ったが、Writeアクセスが多い場合の評価も必要である。

また、実験についても、さらに様々な状況、特に容量が飽和

近くになった場合にも本アルゴリズムは正常に動くか、またシステム規模がさらに大きくなった場合に現実的時間で均衡化処理が終わるのかなどについても検証を行う必要がある。また、本手法を実際の運用環境に適用し、実機でもシミュレーションの通りに動くかどうか検証を行う必要がある。特に実際のストレージの挙動、キャッシュ等の影響や、ストレージ以外の部分、例えばメタデータサーバやキャッシュサーバ等の影響や、ネットワーク構成による違い等の検証が必要である。

謝 辞

本研究の一部は、科学技術振興機構戦略的創造研究推進事業CREST、文部科学省科学研究費補助金特定領域研究(19024028)および東京工業大学21世紀COEプログラム「大規模知識資源の体系化と活用基盤構築」の助成により行なわれた。

文 献

- [1] E. Grochowski and R. D. Halem. Technological impact of magnetic hard disk drives on storage systems. *IBM Systems Journal*, Vol. 42, No. 2, pp. 338–346, 2003.
- [2] Akitsugu Watanabe and Haruo Yokota. Adaptive overlapped declustering: A highly available data-placement method balancing access load and space utilization. pp. 828–839. 21st International Conference on Data Engineering (ICDE'05), 2005.
- [3] Makoto Kataigi, Dai Kobayashi, and Haruo Yokota. Consideration of disk segmentation and disk sequencing for efficient use of disk resource on heterogeneous disk environment. In *Proceedings of the Second IEEE International Conference on Digital Information Management*, Vol. 107, pp. 577–579, 2007.
- [4] Asit Dan and Dinkar Sitaram. An online video placement policy based on bandwidth to space ratio (BSR). *SIGMOD Rec.*, Vol. 24, No. 2, pp. 376–385, 1995.
- [5] T. Cortes and J. Labarta. A case for heterogeneous disk arrays. *IEEE International Conference on Cluster Computing (Cluster'00)*, Vol. 00, pp. 319–325, 2000.
- [6] T. Cortes and J. Labarta. Extending heterogeneity to RAID level 5. In *Proceedings of the 2001 USENIX Technical Conference*, pp. 119–132, Boston, 2001. USENIX Association.
- [7] Yuewei Wang and David Hung-Chang Du. Weighted striping in multimedia servers. In *International Conference on Multimedia Computing and Systems*, pp. 102–109, 1997.
- [8] Jen-Wen Ding and Yueh-Min Huang. Resource-based striping: An efficient striping strategy for video servers using heterogeneous disk-subsystems. *Multimedia Tools and Applications*, Vol. 19, No. 1, pp. 29–51, 2003.
- [9] Roger Zimmermann and Shahram Ghandeharizadeh. Continuous display using heterogeneous disk-subsystems. In *MULTIMEDIA '97: Proceedings of the fifth ACM international conference on Multimedia*, pp. 227–238, New York, NY, USA, 1997. ACM.
- [10] A. Dan, D. Sitaram, and P. Shahabuddin. Scheduling policies for an on-demand video server with batching. In *MULTIMEDIA '94: Proceedings of the second ACM international conference on Multimedia*, pp. 15–23, New York, NY, USA, 1994. ACM.
- [11] Hitachi Ultrastar A7K1000 Datasheet. [http://www.hitachigst.com/tech/techlib.nsf/techdocs/DF2EF568E18716F5862572C20067A757/\\$file/Ultrastar_A7K1000_final_DS.pdf](http://www.hitachigst.com/tech/techlib.nsf/techdocs/DF2EF568E18716F5862572C20067A757/$file/Ultrastar_A7K1000_final_DS.pdf).