ホットレプリケーション：三次記憶システムにおける高アクセス頻度データの複製クラスタリング手法

根本利弘† 喜連川優†

本論文では，ホットレプリケーションと名付けた，磁気テープドライブ装置を用いた三次記憶システムにおける高速化手法を提案するとともに，その性能評価結果を示し，有効性を明らかにする．ホットレプリケーションは，あらかじめテープ上に一定の領域を確保しておき，当該領域上に高アクセス頻度データの複製を作成し，クラスタリングを行う．ホットレプリケーションでは，複製へのアクセスが不利にならないよう，テープを巻き戻すことなくロード/イジェクトが可能なテープドライブ装置を用いることを想定している．オリジナルデータが記録されているテープが使用されている場合においても，異なるテープ上に存在する複製をアクセスすることでリクエストがブロックされることがなくなり，また，高アクセス頻度データの複製をクラスタリングすることで，高アクセス頻度データが連続してアクセスされる場合のシーク長が短縮されるため，応答性能が向上する．本論文では，ホットレプリケーション手法について述べるとともに，シミュレーションによりその有効性を明らかにする．ホットレプリケーションを導入することで，大幅な性能向上が得られることを示す．さらに，衛星画像データベースに対するアクセス履歴を用いてシミュレーションを実行し，実アプリケーションに対してもホットレプリケーションが有効であることを示す．

Hot Replication: Clustering Replicas of Frequently Accessed Data on a Tertiary Storage System

Toshihiro Nemoto† and Masaru Kitsuregawa†

In this paper we propse a replicating and clustering method named *Hot Replication* and evaluate its performance. It dynamically replicates frequently accessed data (hot data) on cache disks and clusters them together onto reserved areas on tapes in a tertiary storage system. Tape drives which can load/eject a tape without rewinding are used for Hot Replication in order not to put a replica at a disadvantage. Hot Replication improves the performance of tertiary storage system by both incleasing accessibility of hot data and reduceing seek length between hot data. In this paper we explain Hot Replication precisely and also clarify its performance by simulations using synthetic data. In addition we show its effectiveness for real applications by simulations using access traces of our satellite image database.

1. はじめに

近年，マルチメディアデータベースや地球環境情報システムなどの膨大なデータを扱うシステムがさかんに構築されるようになり[1)]，大規模三次記憶システムの需要が高まってきた．大規模三次記憶システムとしては，容量，コストの点で優れる磁気テープが用いられることが多いが，テープ上のデータへアクセスする場合にはシークが必要であり，そのために要する時間が，リクエストが発行されてからデータの読み書きが終了するまでの時間に対して大きな割合を占める場合も少くない．データを記録する際にアクセス頻度の高いデータをシークが短くなる位置に配置することで応答時間を短縮することが可能であるが[2)]，一般にデータが生成された時点においてそのデータのアクセス頻度を予測することは困難である．一方，一度テープ上へ記録されたデータの再配置のためには時間的，空間的に大きなコストを要するため，各データのアクセス頻度が判明した後に再配列を行うことは必ずしも得策とはいえない．

本論文では，テープ上にあらかじめ確保した領域へ高アクセス頻度データを複製し，クラスタリングをすることで応答性能の向上を図るホットレプリケーションと名付けた手法を提案する．ホットレプリケーションでは，複製されたデータへのアクセスが不利になることがないよう，テープ途中でロード/イジェクト可

† 東京大学生産技術研究所概念情報工学研究センター
Center for Conceptual Processing of Multimedia Information, Institute of Industrial Science, University of Tokyo

能なテープドライブ装置を用いる．ホットレプリケーションでは，データが多重化されていることによるアクセシビリティの向上，および高アクセス頻度データがクラスタリングされていることによるシーク時間の短縮により応答時間の短縮を図る．

ディスクアレイにおけるアクセス頻度に基づいたファイル配置手法に関しては，Copeland らがファイルを格納するディスクを決定するために熱と温度の概念を用いたファイルの配置法を提案した[3]．Weikum らは，さらにこの研究を発展させ，データが作製された場合や拡張された場合の動的なデータ配置法を提案した[4]．また，Mogi らはホットミラーリングと呼ばれる，ディスクアレイ内にミラーリングと RAID5 の 2 つの領域を設け，アクセス履歴に応じてデータを配置する手法を提案した[5]．しかしながら，ディスクアレイとテープアーカイバでは，アクセス方法や転送速度など異なる点が多い．一般のテープドライブ装置では追記のみが可能であって，すでに記録されたデータを破壊することなくテープ上の任意の位置へ新たに書き込むことはできない．また，データの書き込みにはテープのマウント，シークなどに要する時間的なコストが大きい．したがって，ディスクアレイでの結果をテープアーカイバにそのまま適用することはきわめて困難である．

我々は，三次記憶システムの高速化手法に関し，小規模テープアーカイバを複数組み合わせることによって構成されるスケーラブルテープアーカイバを提案するとともに，スケーラブルテープアーカイバ上の性能向上手法であるホットデクラスタリング手法を提案した[6]．本論文では，新たにホットレプリケーションと呼ぶ応答性能向上手法を提案し，その性能を評価する．ホットデクラスタリングはテープを物理的に移動させることにより応答性能の向上を図るのに対し，ホットレプリケーションはテープ内のデータの複製の作製，クラスタリングにより性能向上を図るものである．これらは独立した手法であり，個別に適用することも，同時に適用することも可能である．以降，2 章において，ホットレプリケーション手法について説明を行い，3 章においてシミュレーションによりホットレプリケーションの基本性能について述べる．さらに，4 章では衛星画像データベースに対するアクセス履歴を用いた性能評価結果を示し，実アプリケーションに対するホットレプリケーションの有効性を明かにする．加えて，5 章では，ホットデクラスタリングを用いている場合において，さらにホットレプリケーションを利用したときの効果について示す．

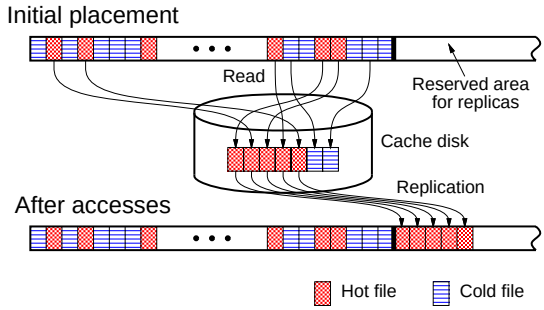

図 1 テープ途中でロード / イジェクト可能なテープドライブ装置を用いたホットレプリケーション
Fig. 1 Hot Replication using a drive which can load and eject without rewinding.

2. ホットレプリケーション

2.1 ホットレプリケーション手法

ホットレプリケーションとは，あらかじめテープ上に確保しておいた領域へ高アクセス頻度データ（ホットデータ）の複製を作成してクラスタリングを行う手法である．ホットレプリケーションは，オリジナルデータ記録時に，テープ全体にオリジナルデータを記録せず，高アクセス頻度データの複製のための領域を確保する．各データに対してある程度のアクセスが行われ，各データのアクセス頻度が判明した時点で，当該領域に高アクセス頻度データの複製を作成し，クラスタリングする（**図 1**）．多くの商用テープドライブ装置では，テープ上の既存のデータを破壊せずに新たなデータを書き込むためには，新たに書き込むデータを既存データの後に追記する以外の方法がないため，複製のための領域をテープの終端部に確保する．

あらかじめ確保しておいた領域にホットデータの複製がクラスタリングされることにより，ホットデータが連続してアクセスされる場合のシーク長が短縮され，応答時間が短縮される．また，データが複製を持たない場合，アクセス要求されたデータが記録されているテープが使用されているときにはそのリクエストをただちにサービスすることはできないが，複製を作成することによりオリジナルデータにアクセスできない場合においても複製をアクセスすることが可能となるため，応答時間が短縮される．すなわち，ホットレプリケーションでは，ホットデータのクラスタリングによるシーク長の短縮，および複製の作成によるアクセシビリティの向上☆により，応答時間の短縮が期待

☆ アクセシビリティが向上するとは，複製を生成することにより原データおよび複製データ両者に対してアクセスが可能となり，アクセス多重度が向上することを意味する．

される．

また，ホットレプリケーションでは少数のホットデータのみを複製の作成対象とする．全データの複製を作成すれば，データのアクセシビリティはさらに向上すると考えられるが，それだけオリジナルデータの読み込み，複製の書き込みにより長い時間を要し，複製を格納するための空間もより多く必要となる．また，アクセス頻度の低いデータの複製を作成してもそのデータがアクセスされることは少なく，応答時間短縮の効果は小さいことが予想される．

さらに，一般にテープドライブ装置ではテープのイジェクト/ロード，シークに要する時間的コストが大きいため，ホットレプリケーションでは，通常のアクセスリクエストへの影響を最小限に抑えるべく，複製作成時には適宜，以下の方針を採用する．

- キャッシュディスク上に存在するホットデータを複製する．
- 使用されていないテープドライブ装置を用い，そのドライブ内に存在するテープに複製を作成する．
- 上記方針によって複製を作成することができない場合には，新たなデータの読み込みやテープのマウントは行わず，複製の作成そのものを行わない．

キャッシュディスク上のホットデータのみを複製の作成対象とすることにより，複製の作成対象となるデータを新たに読み込む必要がなく，また，すでにテープドライブ内に存在するテープを用いることにより，新たにテープをロードする必要がなくなるため，複製の作成に要するコストを最小限に抑えることができる．

2.2 テープ途中でロード/イジェクト可能なテープドライブ装置

一般的な商用テープドライブ装置では，テープイジェクト時には先頭まで巻き戻す必要があり，テープをマウントした直後はテープ後方に存在するデータへのアクセスには長いシークを必要とする．すなわち，テープ終端部に確保された領域のホットデータの複製へのアクセスは時間的なコストの面で不利となってしまう．ホットレプリケーションでは，複製へのアクセスが不利となることがないようにするため，テープを巻き戻すことなく，テープの途中でロード/イジェクトすることが可能なテープドライブ装置，テープメディアを用いる．

多くの商用テープドライブ装置では，テープメディア先頭にディレクトリ情報やメディアに関する情報などを記録しており，ロード時にはこの情報を読み込むようになっている．このため，イジェクト時にはテープを巻き戻すようになっており，少数のテープドライブ装置によって多数のテープを扱うテープアーカイブシステムでは，テープの交換を繰り返すためにロード/イジェクト時のシークに多くの時間が費やされる．このロード/イジェクト時のシークを削減するために，テープ先頭のみではなくテープ上の複数の位置，あるいは任意の位置にディレクトリ情報やメディア情報などを記録できるようにしたり，テープカセットに不揮発性メモリ素子を設け，これにディレクトリ情報やメディア情報などを記録したりすることで，テープを巻き戻すことなく，ヘッドがテープの途中にあるときでもロード/イジェクトが可能であるドライブ装置が開発され，商用化されている．たとえば，AMPEX 社製 DST312 や SONY 社製 GY-2120[7] ではテープ先頭以外の位置にディレクトリ情報やメディア情報などを記録できるようにすることにより，また，SONY 社製 AIT-S100[8] では，テープカセットに 64 KB の EEPROM を搭載することにより，イジェクト/ロード時にテープを完全に巻き戻す必要がなく，ヘッドがテープの先頭以外の位置に存在するときでもロード/イジェクトが可能となっている．

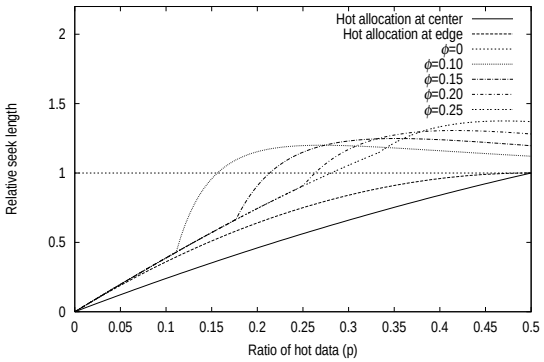

図 2 高アクセス頻度データをテープ終端部に複製したホットレプリケーションにおけるシーク長短縮効果
Fig. 2 Reduction of seek length using Hot Replication.

テープを巻き戻すことなくロード/イジェクト可能なテープドライブ装置はすでに商用のものとなっており，アーカイブシステムでの利用において機能面，性能面で大きな利点を持つため，今後の普及が期待される．

2.3 ホットレプリケーションによるシーク長の短縮

図 2 は，複製が存在せず，複製用の領域を除くオリジナルデータ用の領域上にランダムにデータを配置した場合の平均シーク長を 1 としたときの，複製が存在する場合の相対シーク長を示している．詳しい導出については付録 A.1 に示す．ϕ は全テープ長に対する複製用領域の割合である．解析結果が複雑になるのを避けるため，ここでは，データは，全体の p（$p < 0.5$）

の割合を占め，$1-p$ の割合のリクエストを受けるホットデータと，$1-p$ の割合を占め，p の割合のリクエストを受けるコールドデータ（低アクセス頻度データ）の 2 種類のデータで構成され，それらがテープ上にランダムに配置されているものとする．複製が存在するデータに関しては必ず複製をアクセスするものとしている．また，複製用領域よりもホットデータの量が多い場合には，複製用領域に収まる分の複製を作成するものとしている．複製を作製せずにオリジナルデータをクラスタリングした場合に関して，全ホットデータをテープ中央に配置し，両側に均等にコールドデータを配置した場合，および端部に全ホットデータを配置した場合の相対シーク長も示している．

オリジナル領域が全体に占める割合 $1-\phi$ は，テープ 1 本に記録できるオリジナルデータ量の割合であり，したがって，単位データあたりに要するメディアのコストは $\frac{1}{1-\phi}$ となる．一方，複製が存在しない場合の平均シーク長は $1-\phi$ と比例関係にある．したがって，複製が存在しない場合の平均シーク長を 1 とした図 2 は，メディアコストで正規化したシーク長の短縮効果ということもできる．

図 2 によると，複製用の領域が小さく，その領域以上にホットデータが存在する場合，すなわちすべてのホットデータの複製を作成できない場合には，平均シーク長は急激に長くなる．クラスタリングの効果を保つためには，ホットデータの量よりも複製用の領域を大きくとる必要がある．一方で，アクセス頻度が 70/30 則（全体の 70%のデータが全リクエストの 30%を受ける）よりもアクセスの偏りが緩やかになると，たとえ複製用領域が十分でも，ホットデータのクラスタリングの効果は薄れ，コールドデータをアクセスするためにオリジナルデータ領域へシークする回数が増えるために平均シーク長は延びる．すなわち，シーク短縮の観点のみからいえば，70/30 則よりも緩い分布をなすデータのために，ホットデータのすべての複製を作成できるように複製用領域をテープ全体の 20〜25%以上としても，その効果は望めないことが分かる．複製を作製せずにオリジナルデータをクラスタリングした場合には，複製をクラスタリングした場合と比べ，よりいっそう平均シーク長を短縮することが可能である．しかしながら，オリジナルデータをクラスタリングするためには，一度テープ上の全データをディスク上へ読み込み，その後，目的の配置となるようにデータをテープへ書き戻す必要がある．すなわち，複製を用いずにオリジナルデータをクラスタリングするためには，今日の典型的なテープ装置ではテープ 1 本あたり数時間を要することとなる．一方，ホットデクラスタリングでは少数のホットデータのみをテープへ書き込むだけであるため，複製を作製せずにクラスタリングする場合に比べて，書き込みに要する時間はきわめて小さい．また，ホットレプリケーションは，複製のための領域を必要とするが，今日の一般的なテープメディアの容量あたりのコストは小さく，したがって複製のための領域に要するコストはわずかである．ホットレプリケーションはアクセスの偏りが大きい場合にはクラスタリングを行わない場合と比べ十分にシーク長を短縮することが可能であり，現実的な手法であるといえる．

3. 基本性能評価

本章ではシミュレーションによりホットレプリケーションの基本性能を評価する．ホットレプリケーションは，平均シーク長の短縮および複製によるアクセシビリティ向上の 2 つの効果により応答時間の短縮を図るが，本章ではシミュレーションにより個々の効果を明確にすることを目的とする．まず，3.2 節におけるシミュレーションでは，各テープのヒートがほぼ均一となるようなデータ配置を初期状態とした．このため，複数のリクエストが少数のテープに集中する可能性は低く，したがって，複製によるアクセシビリティの向上による性能向上は期待できず，シーク長の短縮による応答性能の向上の効果が明確になる．また，3.3 節においては，2 つのシミュレーション結果を示す．一方はテープ間のアクセス頻度に偏りがある場合であり，もう一方はテープ間でのアクセス頻度に偏りがない場合である．両者ともテープ内にはデータをランダムに配置し，また，全データについてのアクセス頻度分布はまったく同様としている．すなわち，この 2 つのシミュレーションの相違はテープ間のアクセス頻度の偏りの有無であり，テープ間にアクセス頻度の偏りがある場合には少数のテープにアクセスが集中することとなるため，これら 2 つのシミュレーションの結果を比較することで，複製によるアクセシビリティ向上の効果が明らかになる．

3.1 シミュレーション条件

16 台のエレメントアーカイバにより構成されるスケーラブルテープアーカイバ[6] において，テープの途中でロード/イジェクトが可能なテープドライブ装置を用いてホットデータの複製をテープ終端部に作成した場合のシミュレーションを行う．スケーラブルテープアーカイバとは，小規模のテープアーカイブ装置を 1 つのエレメントとし，複数のエレメントアーカイバを，

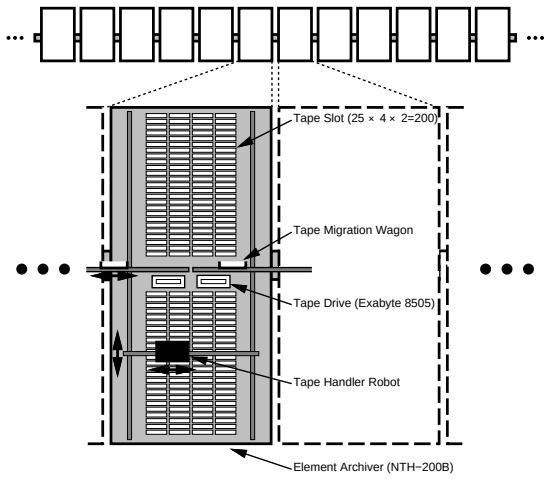

図 3 スケーラブルテープアーカイバ
Fig. 3 Scalable tape archiver.

表 1 シミュレーションパラメータ
Table 1 Simulation parameter.

エレメントアーカイバ	
全エレメントアーカイバ数	16 台
最大テープ数	200 本/台
テープドライブ数	2 台/台
テープドライブ	
ロード時間	35 秒
シーク速度	25 MB/秒
リード/ライト速度	0.5 MB/秒
イジェクト時間	20 秒
テープハンドラロボット	
移動時間（テープの操作なし）	2 秒
移動時間（テープの操作あり）	14 秒
テープマイグレーション装置	
ワゴンの移動時間	9 秒

物理的にテープを移動することが可能な移送装置を用いて接続することにより構成する大規模テープアーカイブ装置である（**図 3**）．ただし，本シミュレーションでは，ホットレプリケーション単独での効果を明らかにするため，移送装置は用いていない．スケーラブルテープアーカイバのパラメータを**表 1** に示す．各テープの容量は 7 GB とし，先頭より 5.5 GB までをオリジナルデータ用，終端部の 1.5 GB の領域をホットデータの複製用領域とする．リクエスト到着間隔は負の指数分布に従う．リクエストサービススケジューリングに関しては，ホットデータのクラスタリングによるシーク長の短縮の効果と複製によるアクセシビリティの向上による効果の両者が得られるよう，次のスケジューリングを採用する．

(1) リクエストキュー内のリクエストを発行順にソートする．
(2) 複製が存在するデータに関しては，複製が優先的にアクセスされるようにするため，まずクラスタリングされた複製のみが存在すると仮定し，リクエストキュー内のサービス可能なもののうちで最も先に発行されたリクエストを選択し，サービスを行うテープを決定する．
(3) (2) において，リクエストされているテープが使用されている，あるいはリクエストされているテープをサービスするためのドライブ装置が，他のリクエストのサービスのために使用されているためにサービス可能なリクエストがない場合には，複製の存在するデータに関し，複製だけではなくオリジナルデータが存在するものとしてサービス可能なリクエストを選択し，サービスを行うテープを決定する．
(4) 選択されたテープ上に記録されたデータに対するすべてのリクエストをテープの先頭方向から順にまとめてサービスする．

なお，詳細なスケジューリング手順は付録 A.2 に示す．

3.2 シーク時間の短縮による効果

本シミュレーションでは，各データのアクセス頻度分布は Zipf 分布☆に従うとし，それらをテープ上のオリジナル領域にランダムに配置した状態をオリジナルデータの初期配置とする．ホットデータの複製もまた，テープ終端部の複製用領域にランダムに配置した状態を初期状態とする．本シミュレーションにおいては，動的な複製の作成は行わない．各データをランダムに配置しているため，各テープのヒートはほぼ均一となり，したがって複数のリクエストが 1 本のテープに集中する可能性は低く，複製によるアクセシビリティの向上による性能向上は期待できず，シーク長の短縮による応答性能の向上の効果が明確になる．本シミュレーションにおいてはアクセス頻度分布がほぼ 90/10 則に従う（全体の 10%のデータが全リクエストの 90%を受ける）ように Zipf 分布のパラメータ z を設定している．

図 4 は，各データのサイズを 100 MB とし，データのアクセス頻度が 90/10 則に従うように $z = 2.0$ の

☆ 全要素数を N としたとき，i 番目の要素の確率 p_i が

$$p_i = \frac{1}{\sum_{n=1}^{N} \frac{1}{n^z}} \cdot \frac{1}{i^z}$$

により表される分布．Zipf 分布において，$N = 167200$ の場合には $z \approx 2.0$，$N = 1672000$ の場合には $z \approx 1.4$ とすることにより 90/10 則に従う分布が得られる．

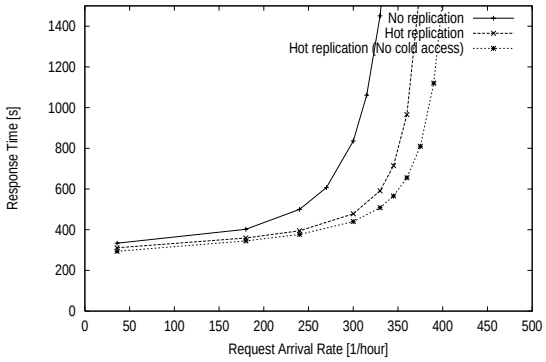

図 4 ホットレプリケーションによる応答時間（ファイルサイズ 100 MB）
Fig. 4 Response time using Hot Replication (file size is 100 MB).

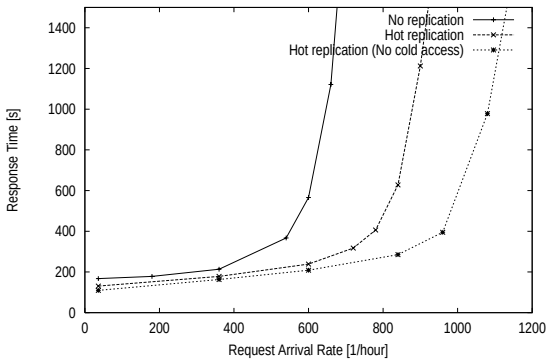

図 5 ホットレプリケーションによる応答時間（ファイルサイズ 10 MB）
Fig. 5 Response time using Hot Replication (file size is 10 MB).

Zipf 分布としたときの初期状態から 50,000 アクセスまでの平均応答時間である．このとき，アクセス頻度の高い順に 10%のデータをホットデータと見なし，複製を作成している．図 4 には，ホットデータの複製が存在しない場合，テープ終端部にホットデータの複製が存在する場合，およびテープ終端部にホットデータの複製が存在し，かつホットデータにのみに対してアクセスリクエストが発行される場合の平均応答時間を示している．応答時間はリクエストが発行されてからデータの読み込みが完了するまでの時間である．ホットデータのみに対してリクエストが発行される場合は，ホットレプリケーションによる応答性能向上の限界値に相当する．図 4 より，テープ終端部にホットデータの複製を作成することにより，平均応答時間が短縮されていることが分かる．

図 5 は，各データのサイズを 10 MB とし，データのアクセス頻度分布が 90/10 則に従うように $z = 1.4$ の Zipf 分布とした場合の初期状態から 50,000 アクセスまでの平均応答時間である．アクセス頻度の高い順に 10%のデータをホットデータと見なし，複製を作成している．ファイルサイズが 100 MB の場合と比較し，ホットデータの複製の作成がより効果的であることが分かる．これは，ファイルサイズが小さくなったためにデータの読み込み時間が短縮され，そのためリクエストが発行されてから読み込みが終了するまでの全応答時間に対してシーク時間の占める割合が大きくなったことによる．ホットレプリケーションは，ホットデータの複製をクラスタリングすることでシーク時間が短縮されるため，ファイルサイズが小さくなるに従いその有効性が向上する．

3.3 複製によるアクセシビリティの向上による効果

本節では，各テープに対するアクセス頻度が Zipf 分布に従い，さらに各テープ内のデータのアクセス頻度分布もまた Zipf 分布に従うとしてデータをランダムに配置したものをオリジナルデータの初期配置とした場合，および全体としては同じアクセス頻度分布を持たせたデータを全テープ上にランダムに配置したものをオリジナルデータの初期配置とした場合のシミュレーションを行う．両者の違いはテープ間のアクセス頻度の偏りの有無であり，テープ間にアクセス頻度の偏りがある場合には少数のテープにアクセスが集中することとなるため，これらの結果を比較することで複製によるアクセシビリティの向上による効果を見ることができる．本シミュレーションにおいてもアクセス頻度分布が 90/10 則に従うように Zipf 分布のパラメータ z を設定している．

図 6 はデータ全体のアクセス頻度分布が 90/10 則に従うように，各テープに対するアクセス頻度が $z = 1.15$ の Zipf 分布に従い，さらに各テープ上のデータそれぞれに対するアクセス頻度もまた $z = 1.15$ の Zipf 分布に従うとした場合の初期状態から 50,000 アクセスまでの平均応答時間である．また，**図 7** は，データ全体としては図 6 のデータと同じ分布を持たせ，それらを全テープ上にランダムに配置した場合の初期状態から 50,000 アクセスまでの平均応答時間である．各データのサイズは 100 MB であり，アクセス頻度の高い順に 10%のデータをホットデータとしている．また，複製は全テープ上の複製用領域にランダムに配置したものを初期状態とし，動的な複製の作成は行っていない．

テープ間のアクセス頻度の偏りがある場合には，テープ間のアクセス頻度の偏りがない場合に比べ，ホット

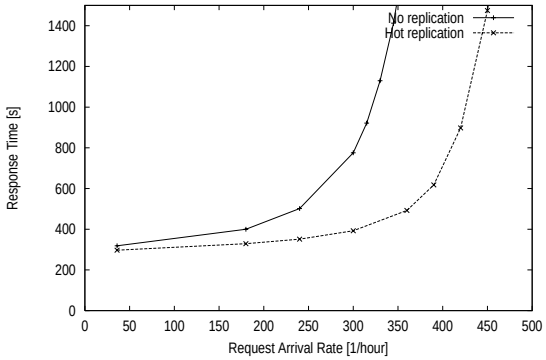

図 6 ホットレプリケーションによる応答時間（テープ間のアクセス頻度の偏りあり，ファイルサイズ 100 MB）
Fig. 6 Response time using Hot Replication (access frequencies are skewed among tapes, file size is 100 MB).

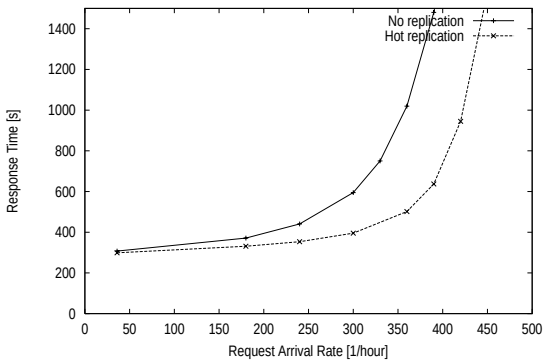

図 7 ホットレプリケーションによる応答時間（テープ間のアクセス頻度の偏りなし，ファイルサイズ 100 MB）
Fig. 7 Response time using Hot Replication (access frequencies are uniform among tapes, file size is 100 MB).

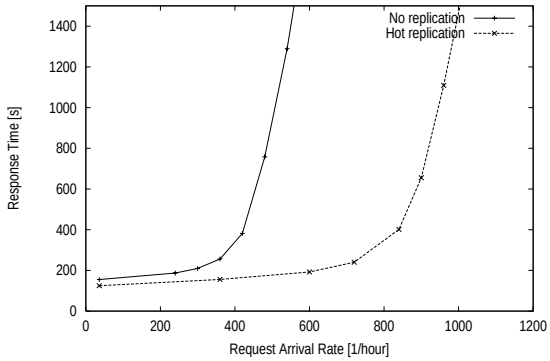

図 8 ホットレプリケーションによる応答時間（テープ間のアクセス頻度の偏りあり，ファイルサイズ 10 MB）
Fig. 8 Response time using Hot Replication (access frequencies are skewed among tapes, file size is 10 MB).

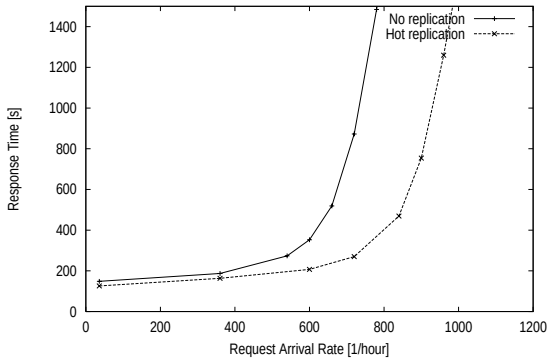

図 9 ホットレプリケーションによる応答時間（テープ間のアクセス頻度の偏りなし，ファイルサイズ 10 MB）
Fig. 9 Response time using Hot Replication (access frequencies are uniform among tapes, file size is 10 MB).

レプリケーションを用いないと大幅に応答性能は劣化している．これは，少数のテープにリクエストが集中してしまうため，あるリクエストによってアクセス要求されたテープが別のリクエストによって使用されているためにブロックされてしまったり，あるいはリクエストされたテープが存在するエレメントアーカイバ内のテープドライブが別のリクエストによりつねに使用中となったりしてしまうためにサービスを行うことができず，テープドライブを効率的に使用することができないためである．このような場合，ホットレプリケーションでは，オリジナルデータと複製の 2 つが存在するため，どちらか一方の存在するテープが使用されていたり，そのテープが存在するエレメントアーカイバのテープドライブが使用されたりしていても，他の異なるエレメントアーカイバ内に存在するもう一方のデータへアクセスすることによりテープドライブを効率的に使用し，応答性能の劣化を防ぐ．

図 8 はデータ全体のアクセス頻度分布が 90/10 則に従うように，各テープに対するアクセス頻度が $z = 1.05$ の Zipf 分布に従い，さらに各テープ上のデータそれぞれに対するアクセス頻度もまた $z = 1.05$ の Zipf 分布に従うとした場合の初期状態から 50,000 アクセスまでの平均応答時間である．また，**図 9** は，データ全体としては図 8 のデータと同じ分布を持たせ，それらを全テープ上にランダムに配置した場合の初期状態から 50,000 アクセスまでの平均応答時間である．各データのサイズは 10 MB であり，アクセス頻度の高い順に 10%のデータをホットデータとしている．

図 6，図 7 に示したファイルサイズが 100 MB の場合と比較し，図 8，図 9 に示すファイルサイズが 10 MB の場合には，応答性能の悪化がより顕著であ

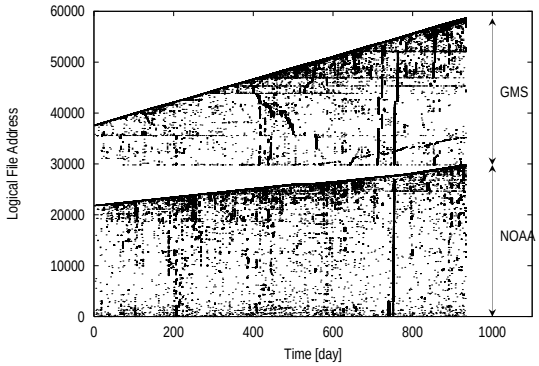

図 10 リクエスト分布
Fig. 10 Distribution of requests.

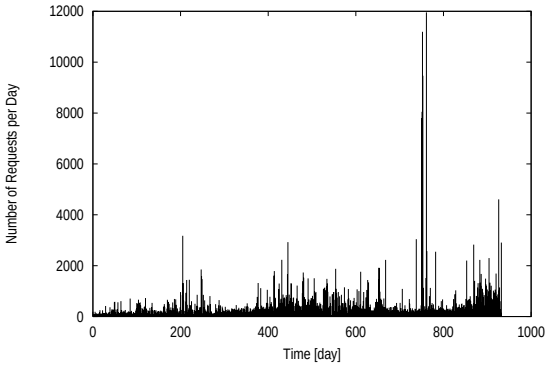

図 11 1 日ごとのリクエスト数
Fig. 11 Number of requests per day.

る．これは，ファイルサイズが小さくなったためにデータの読み込み時間が短縮され，ホットレプリケーションの効果が際立たせられたためである．アクセシビリティの向上による応答時間短縮効果においても，ファイルサイズが小さくなるに従って，ホットレプリケーションの有効性は向上する．

4. 衛星画像データベースのアクセス履歴を用いた性能評価

本章では，東京大学生産技術研究所において World Wide Web（WWW），gopher，ftp により公開している衛星画像データベースのアクセス履歴を用い，ホットレプリケーションの性能を評価する．本シミュレーションにより，ホットレプリケーションが実システムにおいても応答性能向上に対して効果的であることを示す．

4.1 アクセス履歴

図 10 は，1996 年 4 月から 1998 年 10 月半ばまでの衛星画像データベースシステム上のクイックルック画像に対するアクセスの分布である．リクエスト数は ftp によるアクセス約 49,000 件，gopher によるアクセス約 215,000 件，WWW によるアクセス約 197,000 件の合計 461,000 件である．横軸は 1996 年 4 月 1 日からの経過日数，縦軸はデータ番号を表し，グラフの各点は当該データに対し当該日にアクセスがあったことを示している．データ番号は，NOAA 衛星画像データ，GMS 衛星画像データに分けて観測日順に付けられており，0〜29,799 が NOAA 衛星画像データ，29,800〜58,636 が GMS 衛星画像データである．図 10 において，NOAA 衛星画像データ，GMS 衛星画像データとも最上部に斜めの線状の分布が存在しており，最新画像にアクセスが集中していることが分かる．また，縦方向の線状の分布が見られ，短期間に多くのデータがまとめてアクセスされたことが分かる．これらは，特定の利用者がある一定期間の画像データを一括して転送したことによる．

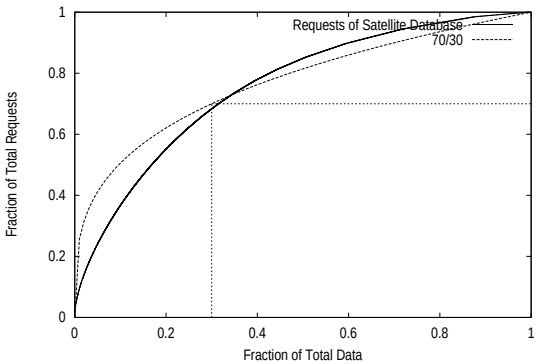

図 12 アクセスローカリティ
Fig. 12 Access locality.

図 11 は 1 日ごとのリクエスト件数を示している．1 日に 10,000 件以上のリクエストを受けている日もあるが，多くのリクエストを受けている日は図 10 の縦方向の線状の分布と対応するものが多く，特定の利用者が一定期間の画像データを一括して転送したことによるものが多い．

図 12 は，衛星画像データベース内の全データに対するリクエストに関して，アクセス頻度上位のデータに対するリクエストが全リクエストに対して占める割合を表したグラフである．比較のために論文 3) において示された 70/30 則に従う曲線☆も示している．衛星画像データベースに対するリクエストは全体の約 30%のデータが 70%のリクエストを受けているが，論文 3) による分布とはやや異なり，アクセス頻度が極

☆ Fraction of Total Requests = Fraction of Total Data$^{\frac{\log \beta}{\log \alpha}}$ により表される分布．

端に高いデータはなく，より緩やかな分布をしていることが分かる．

4.2 シミュレーション条件

シミュレーションには，28,000 件の新たに受信されたデータの書き込みリクエストを加えた 489,000 のリクエストを用いる．シミュレーションは 489,000 リクエスト終了時まで実行するが，性能評価には初期状態から 450,000 までのリクエストを使用する．WWW，gopher，ftp を通じた読み込みリクエストはクイックルック画像に対するものであるが，シミュレーションではこれらのリクエストは対応する衛星原画像へのリクエストであると仮定する．クイックルック画像のアクセス分布と原画像へのアクセス分布とは必ずしも一致しないが，原画像とクイックルック画像は 1 対 1 に対応しており，また，最新画像に対するアクセスが多い，特定ユーザが短期間に一括してデータにアクセスしているなどの特徴は原画像に対するアクセスにも共通すると考えられる．ただし，クイックルック画像のサイズは 50〜100 KB と衛星原画像に比べて小さく，クイックルック画像に対するアクセス系列を原画像に対するアクセス系列とするにはリクエスト間隔が短いため，連続する 2 つのリクエストの時間間隔に対して一定の値（リクエスト遅延率：slow down ratio）を掛け，リクエスト到着率を下げた場合のシミュレーションも行った．アーカイブシステムはなるべく実システムに即した環境を想定するが，実システムでは各テープの全領域にオリジナルデータが記録されており，複製用の領域は確保されていないため，シミュレーションにおいては次のデータ配置を初期配置とする．各テープの容量は 7 GB（非圧縮時）であるとし，そのうち先頭より 5.5 GB の領域をオリジナルデータ用の領域として，衛星原データをデータ番号順に配置する．NOAA 衛星画像データ，GMS 衛星画像データともファイルサイズは約 100 MB であり，テープ上ではいずれもテープドライブ装置に備え付けられている圧縮機能を用いて圧縮されているものとする．実システムにおいても，大部分のデータはテープドライブによって圧縮されているが，個々のデータに対する圧縮率を得ることは困難であるため，非圧縮時のテープ容量と圧縮機能を用いて実際に記録されたデータ量より求めた圧縮率の平均値に基づき，NOAA 衛星画像データは一律に 67%，GMS 衛星画像データは一律 20%に圧縮されているものとする．たとえば GMS 衛星画像データは，テープ上では見かけ上約 20 MB のデータとなり，読み出し時間，書き込み時間とも 20%に短縮される．テープは NOAA 衛星画像データ用 328 本，GMS 衛星画像データ用 108 本の計 436 本で構成される．スケーラブルテープアーカイバは 4 台のエレメントアーカイバ NTH-200B で構成され，初期状態では各エレメントアーカイバに均等に 109 本ずつテープを配置する．また，テープアーカイバ上のデータをキャッシュするためのディスクは 10 MB/s のデータ転送速度を持ち，LRU によりデータを管理する．ディスク容量は，十分な容量である 40 GB，および平均的なファイルサイズの約 3 倍ときわめて小さい容量である 300 MB とする．アクセススケジューリングとしては，3 章と同様のスケジューリングを採用する．その他のパラメータは 3 章の表 1 に従う．本シミュレーションでは，シミュレーション開始時には複製は存在せず，シミュレーション開始後に 10 度以上アクセスされたデータをホットデータと見なし，その複製をテープ終端部の複製用領域に動的に作成する．複製の作成対象となるテープは，複製用の領域を除くオリジナルデータ用の領域がすべて記録されたテープである．シミュレーション開始後にデータが記録されるテープに関しては，オリジナルデータ用の全領域にオリジナルデータが記録された時点で複製作成対象のテープとなる．

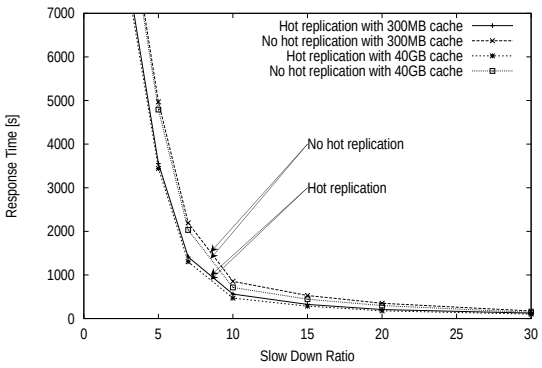

図 13 ホットレプリケーションによる平均応答時間
Fig. 13 Response time using Hot Replication.

4.3 シミュレーション結果

図 13 は初期状態から 450,000 アクセスまでの平均応答時間である．横軸はリクエスト到着率を変化させるための各リクエスト間隔を延ばす際の倍率を表すリクエスト遅延率である．キャッシュディスクサイズが 300 MB，および 40 GB の場合それぞれに対し，ホットレプリケーションを用いた場合，およびホットレプリケーションを用いない場合についての結果を示している．また，**図 14** は各シミュレーション条件におけるホットレプリケーションを用いない場合の応答時間を 1 としたときのホットレプリケーションを用いた場合の相対平均応答時間を表している．

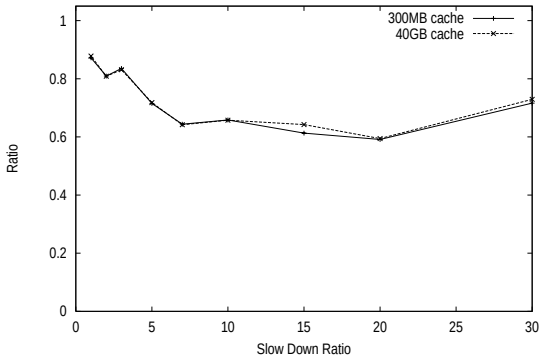

図 14 ホットレプリケーションによる平均応答時間の短縮率
Fig. 14 Relative response time using Hot Replication.

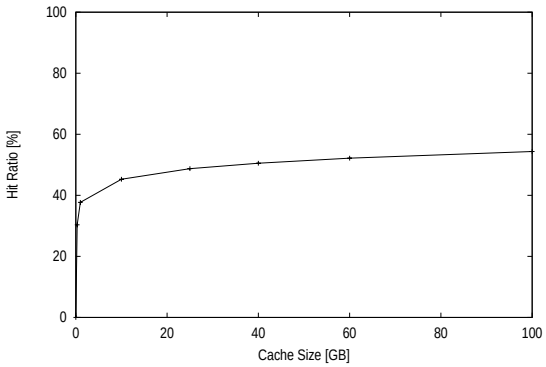

図 15 キャッシュヒット率
Fig. 15 Cache hit ratio.

ディスクによるキャッシュのサイズにかかわらず，ホットデータの複製を作成することにより平均応答時間が短縮されることが分かる．ホットレプリケーションは 40 GB のディスクによるキャッシュ以上に平均応答時間を短縮している．**図 15** はキャッシュサイズとヒット率の関係を示しているが，キャッシュサイズを 40 GB 以上にしてもヒット率はほとんど向上せず，40 GB が十分なサイズであることが分かる．すなわち，ホットレプリケーションは十分なサイズのキャッシュ以上に応答性能を向上させることが分かる．また，ホットレプリケーションはキャッシュサイズが 300 MB の場合においても平均応答時間を短縮し，キャッシュディスクがきわめて小さくても十分に複製が作成され，平均応答時間が短縮されることが分かる．シミュレーション開始から 450,000 アクセスまでに作成された複製データ数は，キャッシュサイズが 300 MB の場合では 9,337〜9,762，40 GB の場合では 9,675〜10,066 であり，これは全データの約 16〜17%にあたる．この結果からも，キャッシュディスクがきわめて小さい場合においても，十分にキャッシュディスクが存在する

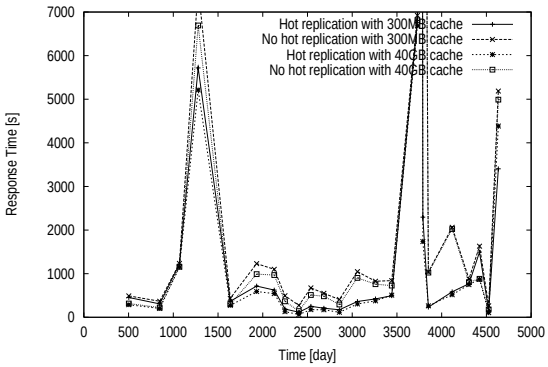

図 16 ホットレプリケーションによる平均応答時間の変化
Fig. 16 Response time transition using Hot Replication.

場合とほぼ同程度に複製が作成されており，ホットレプリケーションは有効であることが示されている．リクエスト遅延率が小さい場合には，ホットレプリケーションの効果が低下しているが，これはリクエスト遅延率が小さくなるとアクセスリクエストへの対応のためにテープドライブ装置の使用率が上昇し，複製の作成が効果的に行われなくなるためである．

図 16 はリクエスト遅延率が 5 のときの 20,000 アクセスごとの平均応答時間を表している．1,600 日（リクエスト遅延率 1 では 320 日）以前では，まだ十分に複製が作成されておらず，キャッシュがホットレプリケーション以上に効果を示しているが，1,600 日以降はホットレプリケーションはキャッシュ以上に応答時間を短縮している．また，図 10 において垂直方向の線状の分布が見られ，過去のある一定期間の連続する多数のデータへのアクセスが行われる場合においてもホットレプリケーションは効果を示している．これは，ホットレプリケーションを用いていない場合，ほぼ同時に同一テープ上の連続する一連のデータに対してリクエストが発行されることが多いが，それらを同時に読み込むことはできず，順次読み込まなければならないため，多数のリクエストが待たされることとなり応答性能が悪化する．また，アクセスされるデータは最新画像データではないためにその直前にアクセスされていることはほとんどなく，キャッシュも有効ではない．一方，ホットレプリケーションでは，このような場合，異なるテープ上の複製にアクセスすることが可能であるため，応答時間が短縮される．大規模アーカイブシステムにおいては，過去のある一定期間の連続する多数のデータへのアクセスのように，少数のメディア上の複数のデータをまとめてアクセスするということは多々あるが，ホットレプリケーションではこのようなアクセスに対しては，異なるテープ上の複製

を用いて並列アクセスが行われるため，きわめて効果的である．

5. ホットデクラスタリング環境下での性能評価

我々は論文 6) においてホットデクラスタリングと名付けた負荷分散手法を提案している．ホットデクラスタリングとは，スケーラブルテープアーカイバにおいて，テープ移送装置を用いて物理的にテープを移動させることにより各エレメントアーカイバ間の負荷を分散させ，各エレメントアーカイバ内のテープドライブ装置を効率的に利用することによって応答性能を向上させる手法である．本章では，本論文で提案するホットレプリケーション手法とすでに提案したホットデクラスタリング手法の 2 つの手法を併用することにより，さらに性能向上できることを示す．ただし，ホットデクラスタリングを利用した場合には，用いなかった場合と比べ，ホットレプリケーションの効果は低減することとなる．これらの点に関し以下評価結果をまとめる．

5.1 基本性能評価

5.1.1 シーク時間の短縮による効果

本シミュレーションでは，データのアクセス頻度分布が Zipf 分布に従うとし，それらをランダムに配置した状態を初期配置としている．本シミュレーションは，ホットデクラスタリングを用いていることを除き，シミュレーション条件は 3.2 節におけるファイルサイズが 100 MB の場合のシミュレーションとまったく同じである．

図 17 は，ホットデクラスタリングを用いている場合の，初期状態から 50,000 アクセスまでの平均応答時間であり，ホットデクラスタリングを用いていない場合の図 4 と対応する．図 17 より，ホットデクラスタリング環境下においてもテープ終端部にホットデータの複製を作製することにより，平均応答時間が短縮されることが分かる．

5.1.2 複製によるアクセシビリティの向上による効果

本項では，各テープに対するアクセス頻度が Zipf 分布に従い，さらに各テープ内のデータのアクセス頻度分布もまた Zipf 分布に従うとしてデータをランダムに配置したものをオリジナルデータの初期配置とした場合，および全体としては同じアクセス頻度分布を持たせたデータを全テープ上にランダムに配置したものをオリジナルデータの初期配置とした場合のシミュレーションを行う．ホットデクラスタリングを用いている

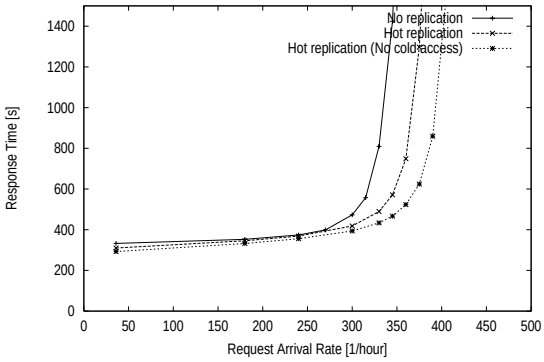

図 17 ホットデクラスタリング環境下でのホットレプリケーションによる応答時間（ファイルサイズ 100 MB）
Fig. 17 Response time using Hot Replication with Hot Declustering (file size is 100 MB).

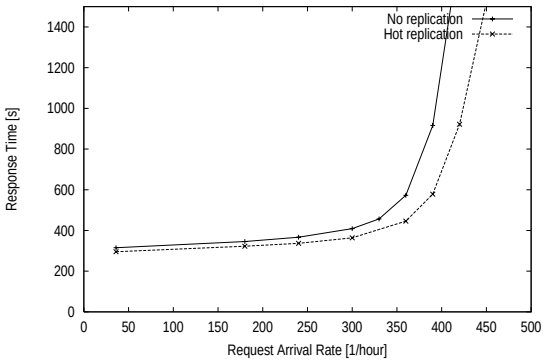

図 18 ホットデクラスタリング環境下でのホットレプリケーションによる応答時間（テープ間のアクセス頻度の偏りあり，ファイルサイズ 100 MB）
Fig. 18 Response time using Hot Replication with Hot Declustering (access frequencies are skewed among tapes, file size is 100 MB).

ことを除き，シミュレーション条件は 3.3 節におけるファイルサイズが 100 MB の場合のシミュレーションとまったく同じである．**図 18** は各テープに対するアクセス頻度が $z = 1.15$ の Zipf 分布に従い，さらに各テープ上のデータそれぞれに対するアクセス頻度もまた，$z = 1.15$ の Zipf 分布に従うとした場合の初期状態から 50,000 アクセスまでの平均応答時間である．また，**図 19** は，データ全体としては図 18 のデータと同じ分布を持たせ，それらを全テープ上にランダムに配置した場合の初期状態から 50,000 アクセスまでの平均応答時間である．それぞれ，ホットデクラスタリングを用いていない場合の図 6，図 7 に対応する．

図 18，図 19 より，ホットデクラスタリング環境下においても，テープ間のアクセス頻度の偏りがある場合には，テープ間のアクセス頻度の偏りがない場合に

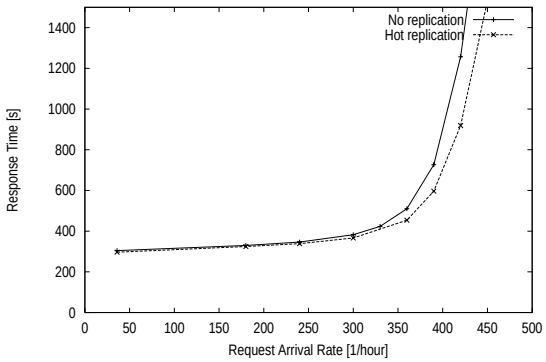

図 19 ホットデクラスタリング環境下でのホットレプリケーションによる応答時間（テープ間のアクセス頻度の偏りなし，ファイルサイズ 100 MB）

Fig. 19 Response time using Hot Replication with Hot Declustering (access frequencies are uniform among tapes, file size is 100 MB).

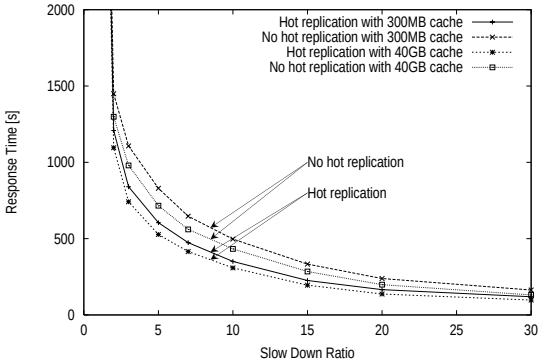

図 20 ホットデクラスタリング環境下でのホットレプリケーションによる平均応答時間

Fig. 20 Response time using Hot Replication with Hot Declustering.

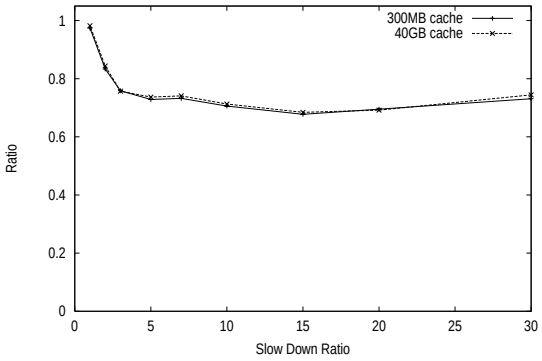

図 21 ホットデクラスタリング環境下でのホットレプリケーションによる平均応答時間の短縮率

Fig. 21 Relative response time using Hot Replication with Hot Declustering.

比べ，ホットレプリケーションを用いないと応答性能は若干劣化しているが，ホットレプリケーションを用いることによりテープ間のアクセス頻度の偏りが存在する場合の応答性能の劣化を防げることが分かる．しかしながら，図 6，図 7 と比べ，ホットデクラスタリングを用いている場合には，応答性能の劣化の度合いは小さい．応答性能の劣化は，少数のテープにリクエストが集中してしまうため，あるリクエストによってアクセス要求されたテープが別のリクエストによって使用されているためにブロックされてしまったり，あるいはリクエストされたテープが存在するエレメントアーカイバ内のテープドライブが別のリクエストによりつねに使用中となってしまうためにサービスを行うことができず，テープドライブを効率的に使用することができなかったりするために生じる．複製が応答時間の短縮効果を示す状況とホットデクラスタリングが効果を示す状況が一部重なっており，ホットデクラスタリングが用いられていない場合において，ホットデータの複製のクラスタリングにより効果が得られる状況の一部は，ホットでクラスタリングによって解消されるため，応答性能の劣化の度合いは小さくなる．たとえば，あるテープに対してリクエストが発行されたとき，そのテープが存在するエレメントアーカイバ内のテープドライブ装置がすべて使用されていることによるリクエストのブロックは，ホットデクラスタリングにより隣接するエレメントアーカイバへそのテープを移送することで解消することも，異なるエレメントアーカイバ内のテープ上の複製をアクセスすることで解消することもできるためである．このため，ホットデクラスタリング環境下においては，テープ間のアクセス頻度の偏りによる応答性能の劣化は小さくなる．

5.2 衛星画像データベースのアクセス履歴を用いた性能評価

本節では，東京大学生産技術研究所において公開している衛星画像データベースのアクセス履歴を用い，ホットデクラスタリングを用いている場合においてホットレプリケーションを併用した場合の評価を行う．本節のシミュレーションは，ホットデクラスタリングを用いていることを除き，4 章におけるシミュレーションとまったく同じ条件で行っている．

図 20 は，ホットデクラスタリングを用いたときの初期状態から 450,000 アクセスまでの平均応答時間である．キャッシュディスクサイズが 300 MB，および 40 GB の場合それぞれに対し，ホットレプリケーションを用いた場合，およびホットレプリケーションを用いない場合についての結果を示している．また，**図 21**

は各シミュレーション条件におけるホットレプリケーションを用いない場合の応答時間を 1 としたときのホットレプリケーションを用いた場合の相対平均応答時間を表している．それぞれ，ホットデクラスタリングを用いていない場合の図 13，図 14 に対応する．

図 20，図 21 より，ホットデクラスタリングを用いている場合においても，キャッシュサイズにかかわらず，ホットレプリケーションは平均応答時間を短縮することが示されている．しかしながら，図 14 と比較すると，ホットデクラスタリングを用いていない場合には最大で約 60%に応答性能が短縮されるのに対し，ホットデクラスタリングを用いている場合においては平均応答性能の短縮率は約 70%であり，ホットデクラスタリングを用いていないときよりも平均応答時間の短縮効果が小さい．これは，テープドライブがすべて使用されているエレメントアーカイバ内のテープに新たにリクエストが発行された場合に，ホットデクラスタリングではそのテープの移動を行い，ホットレプリケーションでは別のエレメントアーカイバ内の複製を参照するというように，リクエストがブロックされることを避ける場合など，ホットデクラスタリングとホットレプリケーションが効果を示す状況が一部重複しているためである．すなわち，ホットデクラスタリング適用時には，ホットレプリケーションが効果を発揮する状況の一部をすでにホットデクラスタリングが解消してしまっているため，その分，ホットレプリケーションの効果は小さくなる．

6. おわりに

本論文では，高アクセス頻度データの複製を作成し，それらをあらかじめ確保しておいたテープ上の領域にクラスタリングするホットレプリケーションの提案を行い，シミュレーションによりその有効性を示した．ホットレプリケーションは，複製によるアクセシビリティの向上，および複製のクラスタリングによるシーク長の短縮により応答性能を向上させる．スケーラブルテープアーカイバ環境を想定したシミュレーションにより基本性能を明らかにするとともに，生産技術研究所において WWW，gopher，ftp により公開している衛星データのクイックルック画像に対するアクセス履歴を用い，これらのアクセスを衛星原画像へのアクセスであると仮定してシミュレーションを行うことで，実システムに対してもホットレプリケーションは応答時間の短縮に有効であることを示した．また，キャッシュ，ホットデクラスタリングを導入している環境においても，さらにホットレプリケーションを導入することで応答時間の短縮可能であることを示した．

本論文で述べたシミュレーションにおいては，複製用領域のガベージコレクションは行っていない．一般のデータアーカイブにおいては，アクセスが頻度が高くなるデータの分布は時間の経過とともに変化する．したがって，複製用領域を効率的に使用するためにはガベージコレクションは必要であると考えらる．今後，ガベージコレクションについて検討を進めていく予定である．

謝辞 衛星画像データの利用に関して貴重なコメントをいただきました東京理科大学高木幹雄教授に感謝いたします．

参考文献

1) Kobler, B., Berbert, J., Caulk, P. and Hariharan, P.C.: Architecture and Desing of Storage and Data Management for the NASA Earth Observing System Data and Information System (EOSDIS), *Proc. 14th IEEE Symposium on Mass Storage Systems*, Monterey, California, pp.65–76 (1995).
2) Christodoulakis, S., Triantafillou, P. and Zioga, F.A.: Principles of Optimally Placing Data in Tertiary Storage Libraries, *Proc. 13rd Very Large Database Conference*, Athenes, Greece, pp.236–245 (1997).
3) Copeland, G., Alexander, W., Boughter, E. and Keller, T.: Data Placement In Bubba, *Proc. 1988 ACM SIGMOD International Conference on Management of Data*, Chicago, Illinois, pp.99–109 (1988).
4) Weikum, G., Zabback, P. and Scheuermann, P.: Dynamic File Allocation in Disk Arrays, *Proc. 1991 ACM SIGMOD International Conference on Management of Data*, Denver, Colorado, pp.406–415 (1991).
5) Mogi, K. and Kitsuregawa, M.: Hot mirroring : A method of hiding parity update penalty and degradation during rebuilds for RAID5, *Proc. 1996 ACM SIGMOD International Conference on Management of Data*, Montreal, Canada, pp.183–194 (1996).
6) 根本利弘，喜連川優：スケーラブルテープアーカイバにおけるテープマイグレーションを用いた負荷分散手法とその性能評価，電子情報通信学会論文誌，Vol.J82-D-I, No.1, pp.53–69 (1999).
7) Sony Corporation: DTF-1 テープドライブ GY-2120. http://www.sony.co.jp/sd/ProductsPark/Professional/DataArchive/BC2/BC2-1/GY2120/index.html
8) Sony Corporation: AIT-2 テープドライブ AIT-S100. http://www.sony.co.jp/sd/ProductsPark

/Professional/DataArchive/BC2/BC2-2/AIT-S100/index.html
9) Nemoto, T. and Kitsuregawa, M.: Scalable Tape Archiver for Satellite Image Database and its Performance Analysis with Access Logs — Hot Declustering and Hot Replication, *Proc. 16th IEEE Symposium on Mass Storage Systems in cooperation with the 7th NASA GSFC Conference on Mass Storage Systems and Technologies*, San Diego, California, pp.59–71 (1999).
10) 根本利弘，喜連川優：テープアーカイブシステムにおけるホットレプリケーションの性能評価，電子情報通信学会技術研究報告，Vol.100, No.226, pp.105–112 (2000).
11) 根本利弘，喜連川優：衛星画像データベースのアクセス履歴を用いたホットレプリケーションの評価，電子情報通信学会 1998 年情報・システムソサイエティ大会講演論文集，D-4-8 (1998).
12) 根本利弘，喜連川優，高木幹雄：スケーラブルテープアーカイバにおけるテープ上での動的データ再配置，情報処理学会第 55 回全国大会講演論文集，4AC-6 (1997).
13) 根本利弘，喜連川優，高木幹雄：大規模テープアーカイバにおけるデータ再配置手法の検討，情報処理学会第 54 回全国大会講演論文集，3R-4 (1997).

付　　録

A.1 平均シーク長

全体の p（$p < 0.5$）の割合を占め，$1-p$ の割合のリクエストを受ける高アクセス頻度データと，$1-p$ の割合を占め，p の割合のリクエストを受ける低アクセス頻度データの 2 種類のデータ構成され，それらがテープ上にランダムに配置されている場合の平均シーク長を求める．リクエストは，発行された順に 1 つずつサービスされるものとし，複製が存在するデータに対するリクエストに対してはつねに複製がアクセスされるものとする．

まず，テープ上の長さ l の領域 $[0, l]$ のある任意の点 x から他の任意の点 y までの平均シーク長 $\lambda_1(l)$ を求める．$\lambda_1(l)$ は，

$$\begin{aligned}\lambda_1(l) &= \int_0^l \int_0^l \frac{|x-y|}{l^2} dxdy \\ &= \frac{l}{3} \end{aligned} \quad (1)$$

となる．また，$[0, l_1]$ 内の任意の位置 x から $[l_1 + l_2, l_1 + l_2 + l_3]$ 内の任意の位置 y への平均シーク長を $\lambda_2(l_1, l_2, l_3)$ とすると，

$$\begin{aligned}\lambda_2(l_1, l_2, l_3) &= \int_0^{l_1} \int_{l_1+l_2}^{l_1+l_2+l_3} \frac{y-x}{l_1 l_3} dydx \\ &= \frac{l_1 + 2l_2 + l_3}{2} \end{aligned} \quad (2)$$

となる．

A.1.1 複製が存在しない場合

テープ長 L に対し，ϕ の割合の高アクセス頻度データ用領域を確保した場合，オリジナルデータ用領域のテープ長は $(1-\phi)L$ となる．したがって，オリジナル領域すべてにデータが記録され，複製が存在しない場合の平均シーク長は

$$\lambda_1((1-\phi)L) = \frac{(1-\phi)L}{3} \quad (3)$$

となる．

A.1.2 複製が存在する場合

全テープ長 L に対し，オリジナルデータ用領域すべてにデータを記録した場合，$(1-\phi)L$ に相当するデータ量になり，高アクセス頻度データは，$p(1-\phi)L$ に相当するデータ量になる．

A.1.2.1 高アクセス頻度データが複製用領域にすべて複製される場合

複製容量域が十分に存在し，すべての高アクセス頻度データが複製されている場合，複製用領域へのアクセス確率 P_r は $P_r = 1-p$，オリジナルデータ用領域へのアクセス確率 P_o は $P_o = p$ となる．したがって，平均シーク長は

$$\begin{aligned}&P_o^2 \lambda_1((1-\phi)L) \\ &\quad + 2P_o P_r \lambda_2((1-\phi)L, 0, p(1-\phi)L) \\ &\quad + P_r^2 \lambda_1(p(1-\phi)L) \\ &\quad = (-2p^3 - p^2 + 4p)\frac{(1-\phi)L}{3} \end{aligned} \quad (4)$$

となる．

A.1.2.2 複製用領域より高アクセス頻度データが多く存在する場合

複製領域以上に高アクセス頻度データが存在し，すべての高アクセス頻度データが複製できない場合を考える．複製用領域に可能な限り高アクセス頻度データの複製が作成され，複製が作成されなかった高アクセス頻度データについては，オリジナル領域のデータがアクセスされるものとする．このとき，複製用領域へのアクセス確率は $P_r = (1-p)\frac{\phi}{p(1-\phi)}$，オリジナルデータ用領域へのアクセス確率は $P_o = p + (1-p)\{1 - \frac{\phi}{p(1-\phi)}\}$ である．したがって，平均シーク長は，

$$
\begin{aligned}
&P_o^2\lambda_1((1-\phi)L)\\
&+2P_oP_r\lambda_2((1-\phi)L,0,p(1-\phi)L)\\
&+P_r^2\lambda_1(p(1-\phi)L)\\
&=\{\phi^3-4\phi+1+(-2\phi^3+5\phi^2+\phi)\frac{1}{p}\\
&-2\phi^2\frac{1}{p^2}\}\frac{L}{3(1-\phi)^2} \qquad (5)
\end{aligned}
$$

となる．

A.1.3 全ホットデータをテープ中央に配置し，両側に均等にコールドデータを配置した場合

全データ量が L に相当する場合，平均シーク長は

$$
\begin{aligned}
&(1-p)^2\lambda_1(pL)\\
&+2\left(\frac{p}{2}\right)^2\lambda_1\left(\frac{1-p}{2}L\right)\\
&+4(1-p)\frac{p}{2}\lambda_2\left(pL,0,\frac{1-p}{2}L\right)\\
&+2(\frac{p}{2})^2\lambda_2\left(\frac{1-p}{2}L,pL,\frac{1-p}{2}L\right)\\
&=(5p-2p^2)\frac{L}{6} \qquad (6)
\end{aligned}
$$

となる．

A.1.4 端部に全ホットデータを配置した場合

テープ先頭に全高アクセス頻度データを配置し，その後に低アクセス頻度データを配置した場合の平均シーク長は

$$
\begin{aligned}
&(1-p)^2\lambda_1(pL)\\
&+2p(1-p)\lambda_2(pL,0,(1-p)L)\\
&+p^2\lambda_1((1-p)L)\\
&=(4p-4p^2)\frac{L}{3} \qquad (7)
\end{aligned}
$$

となる．

A.2 アクセススケジューリング

以下に本論文で用いたアクセススケジューリング手順を記述する．

```
Q = sort_by_time(Q)

D = {}
t = NULL

r = first(Q)
while (r != NULL && t == NULL) {
    if (have_replica(r) &&
        is_accessible(tape_id(replica_id(r)))) {
        t = tape_id(replica_id(r))
        D = {replica_id(r)}
        Q = Q - {r}
    }
    else if (!have_replica(r) &&
            is_accessible(tape_id(original_id(r)))) {
        t = tape_id(original_id(r))
        D = {original_id(r)}
        Q = Q - {r}
    }
    r = next(Q)
}

if (t == NULL) {
    r = first(Q)
    while (r != NULL && t == NULL) {
        if (is_accessible(tape_id(original_id(r)))) {
            t = tape_id(original_id(r))
            D = {original_id(r)}
            Q = Q - {r}
        }
        r = next(Q)
    }
}

if (t != NULL) {
    r = first(Q)
    while (r != NULL) {
        if (have_replica(r) &&
            tape_id(replica_id(r))== t) {
            D = D + {replica_id(r)}
            Q = Q - {r}
        }
        else if (tape_id(original_id(r))== t) {
            D = D + {original_id(r)}
            Q = Q - {r}
        }
        r = next(Q)
    }
}
D = sort_by_position(D)
```

ただし，

Q　リクエストキュー

D　アクセスデータリスト

sort_by_time(Q)　Q を発行時間順にソートする．

first(Q)　Q の先頭リクエスト ID を返す．

next(Q)　直前に実行された *first*(Q) または *next*(Q)

によって返されたリクエストの次のリクエスト ID を返す．次のリクエストがない場合は NULL を返す．

have_replica(r) リクエスト r により要求されているデータが複製を持つ場合は真，持たない場合は偽を返す．

original_id(r) リクエスト r により要求されているデータの複製のデータ ID を返す．

replica_id(r) リクエスト r により要求されているデータのオリジナルのデータ ID を返す．

tape_id(d) データ d が記録されているテープ ID を返す．

is_accessible(t) テープ t がアクセス可能な場合は真，テープドライブ装置が使用されている，テープハンドラロボットが使用されている，テープが他のリクエストにより使用されてるなどの理由によりアクセスできない場合は偽を返す．

*sort_by_position(**D**)* ***D*** をテープ上の位置順にソートする．

(平成 13 年 4 月 11 日受付)
(平成 14 年 2 月 13 日採録)

根本 利弘（正会員）

平成 2 年東京大学工学部電気工学科卒業．平成 6 年同大学大学院博士課程退学．同年同大学生産技術研究所概念情報工学研究センター助手，現在に至る．衛星画像データベースシステム，大規模三次記憶システムに関する研究に従事．

喜連川 優（正会員）

昭和 53 年東京大学工学部電子工学科卒業．昭和 58 年同大学大学院工学系研究科情報工学博士課程修了．工学博士．同年同大学生産技術研究所講師．昭和 59 年助教授．現在同教授，概念情報工学研究センター長．並列コンピュータアーキテクチャ，データベース工学に関する研究に従事．ACM SIGMOD Japan Chapter Chair．VLDB-Trustee．IEEE ICDE ステアリングコミッティメンバ．IEEE TKDE エディタ．