

データ間の相関性を考慮した放送データの スケジューリング法およびキャッシング法

矢島悦子[†] 原隆浩^{††}
塚本昌彦^{††} 西尾章治郎^{††}

近年、有線および無線通信環境において、放送型通信を用いてデータを配送する情報システムに関する研究がさかんに行われている。サーバがクライアントにデータを繰り返し放送する手法は、クライアント数が非常に多い分散システムにおいて、スループットの向上をもたらす可能性がある。これまでに、このようなシステムにおいて、クライアントのデータアクセスに対する応答時間の短縮を図るための様々な戦略が研究されている。本論文では、放送されている各データ間の関連性を考慮した新たな放送プログラムのスケジューリング法を提案する。提案する手法では、相関性の強いデータを隣接して放送することでデータアクセスの応答時間の短縮を図っている。また、クライアント側でのキャッシング法として、従来の手法を相関性を考慮するように拡張した手法を提案する。さらに、提案した手法の有効性を、シミュレーション評価によって検証する。

Scheduling and Caching Methods for Broadcast Data Considering the Correlation among Data

ETSUKO YAJIMA,[†] TAKAHIRO HARA,^{††} MASAHIKO TSUKAMOTO^{††}
and SHOJIRO NISHIO^{††}

Recently, there has been increasing interest in research of the information systems that deliver data using broadcast in both wired and wireless environments. The method in which a server repeatedly broadcasts data to clients can result in a larger throughput. In such an environment, various strategies have been studied to reduce the average response time for data accesses. In this paper, we propose the scheduling strategies for broadcast program considering the correlation among data. These strategies put data items with strong correlation side by side in a broadcast program to reduce the average response time. We also propose a caching strategy which extends a conventional caching strategy so that it can handle efficiently correlation among broadcast data. Moreover, we evaluate the performance of our proposed strategies by simulation studies.

1. はじめに

近年、有線および無線通信環境において、放送型通信を用いてクライアントにデータを配送するプッシュ型情報システムに関する研究がさかんに行われている。クライアントからの要求に応じて個別にデータ配送を行うプル型の配送方式とは異なり、プッシュ型の情報システムでは、サーバはクライアントへの広い帯域幅を持つ通信チャネルを利用して多種のデータを周期的に放送し、クライアントは自分に必要なデータの

みを選択して取得する(図1)。放送型通信を用いてデータを繰り返し散布することにより、各クライアントから離散的に発生する同じデータに対するアクセス要求を、1周期放送するたびにまとめて満たすことができる。したがって、クライアント数が非常に多い分散データベースシステムにおいて、通信のための帯域確保の競合が起りにくくなり、データアクセスに対する応答時間の短縮が期待できる。プッシュ型通信を用いた情報システムは、クライアント数が増加してもシステム全体の負担コストはほとんど変わらないといった特徴があり、サーバからクライアントへの通信チャネルの帯域幅が大きく、その逆が小さい非対称型のネットワークにおいてよく用いられる。

このようなシステムの例としては、放送型情報提供システム^{6),9)}と放送ディスクシステム^{1),2),4),8)}がある。

[†] 株式会社エフエム大阪東京支社営業部
Sales Department, Tokyo Office, FM Osaka Co., Ltd.
^{††} 大阪大学大学院工学研究科情報システム工学専攻
Department of Information Systems Engineering,
Graduate School of Engineering, Osaka University

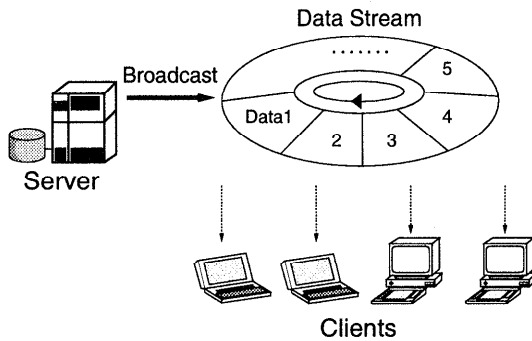


図1 放送型通信に基づく情報システム

Fig. 1 Information system based on message broadcast.

前者は、携帯端末やパソコン、家電機器などをクライアントとして、それらに対して情報を提供するもので、後者は、データベースサーバのディスクの全体または一部の内容を定期的に放送するものである。前者の具体的な例としては、次のようなものが考えられる。

- ショッピングセンターで各店舗のチラシが放送されている。客は携帯端末を持ち歩きながらそれらのチラシを見て、気に入った店に行く。
- 駅周辺で、付近の公共交通機関の時刻表が、その日の臨時増発便の時刻も含めて放送されている。ユーザは携帯端末を用いてそれらの情報を得る。
- 各家庭において、衛星放送や地上放送、有線テレビ放送などのデジタル化された様々なメディアの情報（画像、音声、テキスト）を家電機器やパソコンを用いて取得する。放送される情報には、ニュース、スポーツ、為替、通信販売などが含まれる。

これまでに、プッシュ型情報システムの性能向上のための研究がいくつか報告されている。これらの研究はいくつかに分類でき、主なものとして、サーバ側の放送データのスケジューリング戦略^{1),5),7)~12)}、クライアント側のキャッシュ戦略^{1),3)}、プッシュ型とプル型の融合戦略^{4),6)}などがあげられる。

一方、放送されている各データの間には、あるデータとあるデータにまとめてアクセス要求を起こす確率が高いといったように、相関関係が存在する場合が多い。クライアントが相関性のあるデータの集合全体に対して頻繁にアクセスする場合には、放送データのスケジューリングおよびキャッシングの際に、それらの関係を考慮することにより、クライアントのデータアクセスに対する平均応答時間を短縮することができる。これまでに、多数のデータが特定の1つのデータを参照するような関係が存在する場合に、アクセスの平均応答時間を短縮するためのスケジューリング法が提案

されている⁹⁾。しかし、実際の環境ではさらに複雑な相関関係が存在するものと考えられる。

そこで本論文では、データ間の複雑な相関性を考慮した放送データのスケジューリング法およびキャッシング法を提案する。さらに、シミュレーション評価によって提案した手法の有効性を検証する。

本論文では次のようなシステム環境を想定する。

- サーバは1つとする。
- サーバは、IDが1から M までのデータアイテムから成る1周期分のプログラムを作成し、隙間なく繰り返し放送する。各データアイテムは、1周期中で1回だけ放送される。
- 各データアイテムのサイズはすべて同じものとする。
- クライアントは放送プログラムを知っている。これは、たとえば、プログラム情報を定期的に放送することによって実現できる。
- 各データアイテムへのアクセス確率およびデータアイテム間の相関性は各クライアントで異なるものとし、それらは各クライアントで既知とする。
- データおよび放送プログラムの更新はない。
- プル型配送は行わない。

以下、2章では放送データの相関性について述べる。3章で相関性を考慮したスケジューリング法、4章でキャッシング法を提案する。5章で提案した手法の性能評価を行い、最後に、6章で本論文のまとめを行う。

2. データアイテム間の相関性

一般に、サーバから放送されている様々なデータアイテムは、互いに関連性を持つことが多い。たとえば、サーバが様々なサイトのホームページのHTMLファイルを放送している場合には、各ホームページからリンクしている複数のページに対して、まとめてアクセス要求を出す確率が高いものと考えられる。この確率は、相関関係の強さを表し、各データアイテム間で相関性の強さは異なる。すなわち、クライアントが特定のデータアイテム集合にまとめてアクセス要求を発生する傾向が強いほど、それらのデータアイテムの相関性は強いものと考えられる。

相関性のあるアイテム集合に対するアクセス要求は、それらのアイテムに対して同時に起こる場合と、ある程度の時間差を持って連続して起こる場合が考えられる。本論文では、連続するアクセス要求の時間差が1周期分のプログラムの放送時間に対して無視できるほど小さい場合も含め、相関性のあるアイテム集合に対するアクセス要求は同時に発生するものと想定する。

ここで、各データアイテム間の相関関係は非常に複雑で、放送プログラム中の個々のデータアイテムを表す複数の頂点と相関関係のある頂点どうしを結ぶ辺でグラフを構成すると、木構造ではなくネットワーク構造を示す場合が一般的であると考えられる。このように、データアイテム間に複雑な相関性が存在するときは、相関性を考慮したスケジューリングおよびキャッシングを行うことで、データアクセスに対する平均応答時間を短縮できる。

実環境において、各クライアントにおけるアイテム間の相関性は、クライアントのアイテムへのアクセス履歴を調べることで決定できる。分散システム全体のアイテム間の相関性は、定期的もしくは不定期に、サーバがクライアント各自のアクセス履歴を収集することで決定できる。また、HTML ファイルの転送において、ページを構成するテキストや画像がそれぞれ1つのデータアイテムとして放送される場合では、アイテム間の相関性は、ページ構成や画像データの共有関係などで明示的に決定する。このように、データアイテムの内容によっては、解析的にアイテム間の相関性を決定できる場合もある。

3. 放送プログラムのスケジューリング

本章では、まず、従来の放送プログラムのスケジューリング法について説明し、次に、データアイテム間の複雑な相関性を考慮したスケジューリング法を提案する。

3.1 従来のスケジューリング法

一般に、放送しているデータのうちクライアントが頻繁にアクセスするアイテムは、個々の嗜好によって異なり、放送しているデータアイテムすべてにアクセスするとは限らない。個々のクライアントから発生するアクセスのうち80%が、放送されている全データの20%という限られたデータアイテムに対するものであるとの報告もある¹¹⁾。システム全体としてのアクセスに関しても、あるアイテムに対するアクセスが多いなど、アクセスの偏りがあることが一般的である。このようなクライアントからのアクセスの偏りに対し、従来のスケジューリング手法では、クライアントから頻繁にアクセスされるデータアイテムを高い頻度で放送し、ほとんどアクセスされないアイテムの放送頻度を低くするといった方法がとられている^{1),7),11)}。

文献1)にあるように、データ間に相関関係が存在しない場合には、各データアイテムの放送周期が一定で、各アイテムの放送頻度が適当なものでさえあれば、各放送データの放送順はアクセスに対する応答時間に

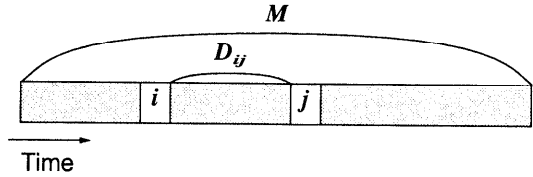


図2 放送プログラム
Fig.2 Broadcast program.

影響しない(後に示すプリフェッチ型キャッシング法を用いる場合を除く)。しかし、先に述べたように、放送されているデータアイテム間には、相関関係が存在する場合が一般的である。スケジューリングの際にこれらの関係を考慮すると、データアクセスに対する平均応答時間をさらに短縮できるものと考えられる。

3.2 アイテム間の相関性を考慮したスケジューリング法

本節では、データアクセスに対する平均応答時間を短縮するための放送プログラムのスケジューリングアルゴリズムとして、CBS (Correlation-Based Scheduling) 法と呼ぶ手法を提案する。本論文では、簡単のために、クライアントからは必ず2つのアイテムに対して同時にアクセス要求が生じるものとする。ただし、3つ以上のアイテムに同時にアクセス要求を出す場合も、同様に議論が可能である。

本論文では、各アイテムが1周期分の放送プログラム中で1回ずつ放送される。したがって、周期的な放送プログラムのスケジューリングは、各アイテムを節点とした完全グラフにおいて、1つのハミルトン閉路を選択し、その閉路に沿う節点のアイテムを順に放送することと等価である。

ここで、図2に示す放送プログラムを繰り返し放送しているとき、任意の時刻においてアイテム*i*と*j*に対して同時にアクセス要求を起こしたときの平均応答時間 avg_{ij} は次式で表される。ただし、ここでは、クライアントがキャッシュを持たない場合を想定している。

$$\begin{aligned}
 avg_{ij} &= \frac{D_{ij}}{M} \cdot \left\{ \frac{D_{ij}}{2} + (M - D_{ij} - 1) \right\} L \\
 &\quad + \frac{M - D_{ij} - 2}{M} \\
 &\quad \left\{ \frac{M - D_{ij} - 2}{2} + (D_{ij} + 1) \right\} L \\
 &\quad + \frac{2}{M} \cdot (M - \frac{1}{2}) L \\
 &= \frac{L}{M} \{ D_{ij}(M - D_{ij} - 2) \} \\
 &\quad + \frac{L(M^2 + 2M - 2)}{2M}
 \end{aligned} \tag{1}$$

M は放送されているアイテムの個数、 L は1アイテム

ムを放送するのにかかる時間、 D_{ij} は放送プログラム内で先に放送されるアイテム i (j) と後に放送されるアイテム j (i) の間に配置されるアイテムの個数とする ($0 \leq D_{ij} \leq M-2$ かつ $D_{ij} = D_{ji}$)。

avg_{ij} の中辺第1項は、図2に示す放送プログラムにおけるアイテム i の放送からアイテム j の放送までの間にアクセス要求が起こった場合の平均応答時間である。同様に、第2項、第3項はそれぞれ、アイテム j の放送から i の放送まで、アイテム i または j の放送中においてアクセス要求が生じた場合の平均応答時間を示している。

式(1)から、 avg_{ij} は D_{ij} の関数であり、これを最小にするには、右辺第1項を最小にすればよいことが分かる。 D_{ij} は0から $M-2$ までの整数値をとるので、 avg_{ij} を最小する D_{ij} は0または $M-2$ であり、いずれの場合もアイテム i と j は隣接して放送される。

しかし、放送プログラム内で隣接させて放送することができるアイテム数は限られているため、相関性のあるアイテムの組合せの数が増えた場合には、相関性の強さを考慮して、放送プログラム内で隣接させるアイテムを最適に決定することは困難である。

そこで、本論文では、クライアントからまとめてアクセス要求を出される確率が高いアイテムを放送プログラム内に隣接して配置することで、アクセスの平均応答時間の短縮を図るヒューリスティックな手法(CBS法)を提案する。

CBS法

- (1) 各データアイテムを節点、節点 ij 間の辺の重みを $W_{ij} = 1 - P_{ij}$ とした完全グラフを用意する。ここで、 P_{ij} はクライアントがアイテム i と j に同時にアクセス要求を起こす確率とする ($\sum_{j=i+1}^M \sum_{i=1}^M P_{ij} = 1$ かつ $P_{ij} = P_{ji}$)。
- (2) すべての節点を1度ずつ通り、かつ、その道に含まれる辺の重みの総和が最小となる閉路を見つける。
- (3) 閉路上のアイテムを順に配置した放送プログラムを作成する。

図3は、CBS法を適用した様子を表している。IDが a から f のアイテムを左図のような完全グラフで表し、グラフ内のハミルトン閉路のうちで、閉路上の辺の重み $W_{ij} = 1 - P_{ij}$ の総和が最小となる閉路を選択する。その閉路に沿って右図のように各アイテムを配置して、放送プログラムを構成する。

ここで、データアクセスに対する平均応答時間を最短にする放送プログラムを作成するためには、放送す

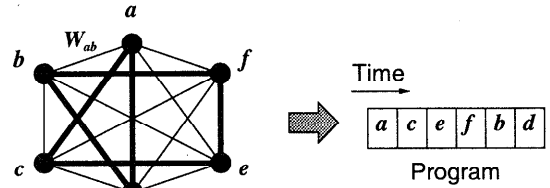


図3 CBS法
Fig.3 CBS method.

るアイテム間に存在する相関性をすべて考慮しなければならない。CBS法では、放送プログラム内で隣接されるデータアイテムの相関性の強さのみを考慮しているため、作成される放送プログラムは必ずしも最適とはならない。しかし、相関性の強いアイテムを隣接させることにより、応答時間の大幅な短縮が期待できる。

CBS法は、最適解を求める問題よりは大幅に簡単になっているとはいえ、その第(2)ステップは、巡回セールスマン問題と等価である。巡回セールスマン問題はNP完全であるため、放送されるアイテム数が大きくなると実時間で解くことができない。しかし、高速に巡回セールスマン問題の近似解を求めるアルゴリズムはこれまでに多く提案されており、工学の様々な分野において利用されている。CBS法においても、アイテム数が多い場合は、高速近似アルゴリズムを利用して近似的に巡回セールスマン問題を解くものとする。

4. 放送データアイテムのキャッシング

頻繁にアクセスするアイテムやアイテム間の相関性などがクライアントによってまったく異なる場合、クライアント全体の嗜好を考慮したスケジューリングによって、あるクライアントでは平均応答時間が非常に大きくなってしまふ。このとき、クライアントがキャッシュを所有していれば、個々のクライアントの嗜好に応じたキャッシングが可能となり、このような問題を緩和できる。

本章では、まず、これまでに提案されているプリフェッチ型キャッシング法のPT法について説明する。次に、PT法をアイテム間の相関性を考慮するように拡張した手法を提案する。

4.1 従来のプリフェッチ型キャッシング法

プッシュ型通信を用いた情報システムにおいて性能向上を目指す研究の一環として、これまでに様々なキャッシング法が提案されている。これらのキャッシング法は、アイテムの置き換えを行うタイミングによって、プリフェッチ型と非プリフェッチ型の2つに分

類できる。ここでは、そのうちのプリフェッチ型キャッシング法に着目する。

プリフェッチ型キャッシング法では、重要度が高いと判断したアイテムをあらかじめキャッシュに格納しておく。フィルタリングの手間以外のコストの増加をとまなうことなく、重要なアイテムに対するヒット率を高め、平均応答時間を短縮できることから、プリフェッチ型キャッシング法はプッシュ型通信に適したキャッシング法と考えられている³⁾。代表的なプリフェッチ型キャッシング法として、文献³⁾においてPT法と呼ばれる手法が提案されている。PT法の概要を次に示す。

PT法

- (1) 各アイテムの放送開始時に、キャッシュ内のアイテムおよび放送されるアイテムに対応付けてPT値と呼ぶ値を与える。アイテム*i*に与えるPT値 L_i は次式で表される。

$$L_i = P_i \cdot T_i - P_i \cdot \lambda_i (= P_i \cdot \tau_i) \quad (2)$$

ここで、 P_i はアイテム*i*に対するそのクライアントのアクセス確率、 T_i はアイテム*i*が放送される周期、 λ_i はアイテム*i*がキャッシュに格納されてから経過した時間、 τ_i は次にアイテム*i*が放送されるまでの時間とする。

- (2) キャッシュに空きがないとき、放送されるアイテム*i*のPT値 L_i が、キャッシュ内でPT値が最小となるアイテム*j*のPT値 L_j より大きい場合、アイテム*i*と*j*を置き換える。

L_i は、アイテム*i*がキャッシュに格納されるときに最大値 $P_i \cdot T_i$ となり、次回のアイテム*i*の放送開始時には0となる。

PT値は、あるアイテムに対してアクセス要求を起こしたとき、そのアイテムをキャッシュに保持していない場合に生じる待ち時間の見積り値である。すなわち、PT法では、各アイテムの放送開始時に、各々のアイテムをキャッシュから追い出すことにより増加する待ち時間を比較し、その値の大きなアイテムをキャッシュに残す。したがって、PT法は、与えられたキャッシュ内のアイテム集合に対して、アイテム放送時の置き換えを最適にし、その瞬間のアクセス要求に対する平均応答時間を最短にする。なお、PT法におけるデータアイテムの置き換えは長期的に見た場合に必ずしも最適とはならないが、従来の手法の中では特に良い性能を示す。

4.2 アイテム間の相関関係を考慮したプリフェッチ型キャッシング法

相関性を考慮した場合も、PT法と同様に、キャッシュ内および放送中のアイテムに対して、そのアイテム

をキャッシュに格納しなかった場合に増える待ち時間を計算し、その値が小さいアイテムをキャッシュから追い出すことが有効である。そこで、PT法を相関性を考慮するように拡張した、CB-PT法 (Correlation-Based PT) と呼ぶキャッシング法を提案する。

CB-PT法

- (1) 各アイテムの放送開始時に、キャッシュ内および放送中のアイテムに対して、PT値を相関性を考慮するように拡張したCB-PT値 G_i を次式で与える。

$$G_i = \tau_i \cdot \sum_{k \in C} P_{ik} + \sum_{k \in Q_i} (\tau_i - \tau_k) \cdot P_{ik} \quad (3)$$

ここで、 τ_i はアイテム*i*の次回の放送までの時間、 P_{ij} はクライアントがアイテム*i*と*j*に一括してアクセス要求を起こす確率とする。また、 C は放送データのうちキャッシュ内にあるアイテムの集合とし、 Q_i は現時点からアイテム*i*の次回の放送時間までに放送され、かつ、 C に含まれないアイテムの集合とする。

- (2) キャッシュに空きがないとき、放送されるアイテム*i*のCB-PT値 G_i が、キャッシュ内でCB-PT値が最小となるアイテム*j*のCB-PT値 G_j より大きい場合、アイテム*i*と*j*を置き換える。

図4は、 a から h のデータアイテムで構成された放送プログラムを放送しており、放送中のアイテムが a 、キャッシュ内のアイテムが b, e, f の場合、アイテム e のCB-PT値を計算する際の集合 C と Q_e を示している。

式(3)は、アイテム*i*と放送プログラム内の任意のアイテムの2つに同時にアクセス要求を起こしたとき、アイテム*i*をキャッシュに格納していなかった場合に増える待ち時間を示す。したがって、放送中のアイテムおよびキャッシュ内のアイテムのうち、式(3)の値が大きなアイテムをキャッシュに残す置き換えは、その瞬間に生じるアクセス要求に対する平均応答時間を最短にすることができる。CB-PT法における置き換えも、PT法と同様に、長期的に見た場合に必ずしも最適とはならないが、平均応答時間を大幅に短縮できるものと考えられる。

ここで、キャッシュに格納できるデータアイテム数を n とすると、各データアイテムの放送時にPT法は $O(n)$ 、提案したCB-PT法は $O(n^2)$ の計算が必要となり、提案したCB-PT法の方が計算量は大きくなる。

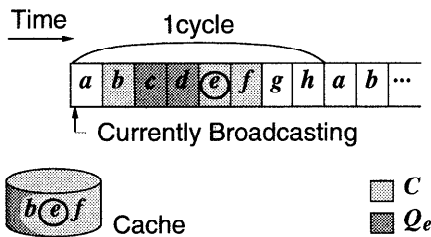


図4 アイテム e に対するアータアイテム集合 C と Q_e .
Fig. 4 Example of item sets C and Q_e for item e .

しかし、実環境における放送プログラム内のデータアイテム数が数十から数百程度で、1アイテムの放送に要する時間が数百ミリ秒から数秒程度であることを考慮すると、CB-PTでもキャッシングの際の計算時間はさほど大きくはならないため、性能上の支障はないものと考えられる。また、放送プログラムやデータアイテム間の相関性に変化が生じない限り、周期的に同じデータアイテムの置き換えが繰り返されるため、その1周期分の置き換え順序を記憶しておけば各データアイテムの放送ごとに計算する必要はない。

なお、式(3)において、 P_{ii} をアイテム i のみに単独でアクセス要求を起こす確率とすると、放送プログラム内に相関性のあるアイテムと相関性のないアイテムが混在する環境においても適用することができる。従来のPT法は、CB-PT法においてアイテム間の相関性がまったくない特殊なケースである。

5. 性能評価

本章では、アイテム間に存在する相関関係が異なるいくつかの環境において、提案したCBS法およびCB-PT法をシミュレーションにより性能評価し、その有効性を検証する。

シミュレーションの条件として、次のものを用いた。

- 放送されるアイテム数を120とし、1アイテムの放送にかかる時間は10(10単位時間)とする。これらの定数パラメータの値は、応答時間に比例的に影響するだけであるから、この値の設定が比較対象と本論文で提案した手法の相対的な差異には影響しない。したがって、シミュレーションでは、特に現実の環境を想定することなく、適当な値を用いた。
- クライアントからは、単位時間あたり10分の1の頻度でアクセス要求が発生する。アクセス頻度は、応答時間に影響しないため、上記のパラメータと同様に適当な値を用いた。
- クライアントは相関性のある2つのデータアイテム

ムに対して同時にアクセス要求を起こすものとする。2アイテムに同時にアクセス要求を起こす確率、すなわち、アイテム間の相関性は、 120×120 の正方行列として与える。 i 行 j 列目の要素 P_{ij} は、アイテム i と j に同時にアクセス要求を起こす確率とする($P_{ij} = P_{ji}$)。ここでは簡単のために、すべての i において $P_{ii} = 0$ とする。なお、行列の各要素の与え方としては、次の2つを用いた。

行列1: 行列の全要素のうち、ある割合の要素は、一定の大きさの正の値を持つ。それ以外の要素は0とする。0でない要素の割合が増えるにつれて、個々の相関関係は弱くなる。

行列2: 全アイテムをいくつかのグループに分け、同じグループに属するアイテム間の要素に正の値を与える。異なるグループに属するアイテム間の行と列の組合せの要素はすべて0とする。つまり、放送されるアイテムにおいて、同じグループ内のアイテム間には互いに相関性があり、異なるグループのアイテムとの間には相関性はまったくない。与える正の値は、大、中、小の3つがあり、行列の同一グループ内の各要素はそれらの値からランダムに選ばれる。大、中、小の値の比は、8:5:2とする。

シミュレーションでは、この2つの行列を用いて各アイテムの相関性を決定し、クライアントのアクセスに対する平均応答時間を計算した。

5.1 CBS法の性能評価

本節では、CBS法の有効性を検証するために行ったシミュレーション評価の結果を示す。ただし、同時に発生する2アイテムに対するアクセス要求のうち、最初に放送されるアイテムの待ち時間はスケジューリング手法に依存しないため、ここでは1つ目のアクセス要求が満たされた後に2つ目のアクセス要求が満たされるまでの時間(1つ目のアクセスと2つ目のアクセスの時間間隔)を平均応答時間と考える。

行列1と行列2のいずれを用いた場合にも、アイテム間の相関性を考慮しないスケジューリング法、つまり、放送プログラム内にアイテムをランダムに配置する手法をCBS法の比較対象として用いた。ここで、文献1), 7), 11)などの手法を比較対象としないのは、これらの手法は各データアイテムの放送頻度を決定するもので、放送順については考慮していないため、本論文における各データアイテムが1周期で1回ずつ放送される環境では効果がないからである。

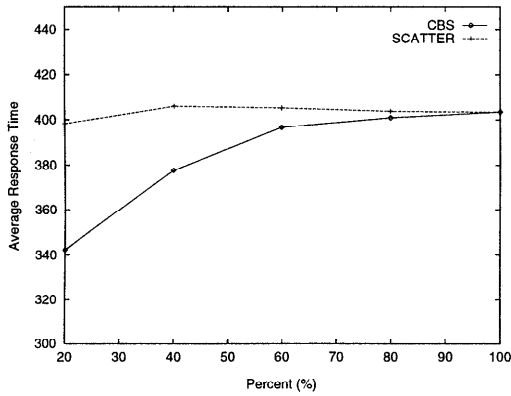


図5 シミュレーション 1における平均応答時間
Fig. 5 Average response time in simulation 1.

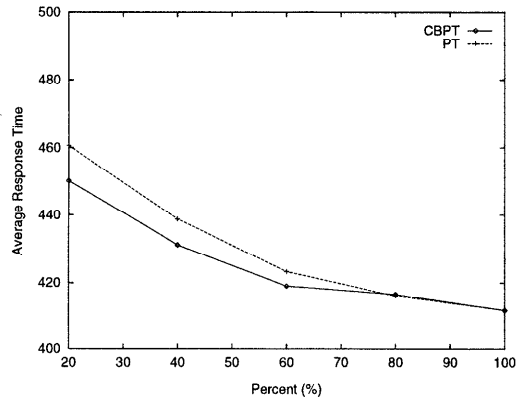


図7 キャッシュがある場合の平均応答時間
Fig. 7 Average response time in case that the client has a cache.

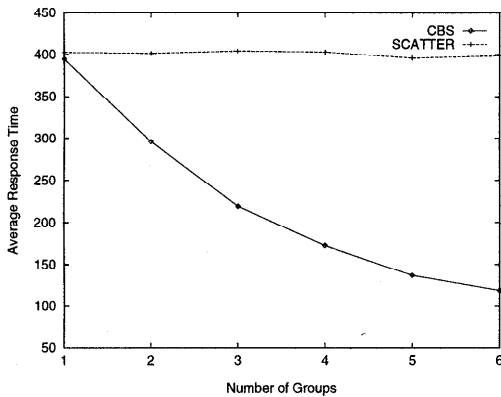


図6 シミュレーション 2における平均応答時間
Fig. 6 Average response time in simulation 2.

シミュレーション 1:

行列 1 において行列の要素のうち 0 以外の値を持つ要素の割合を変化させて、クライアントからのアクセスの平均応答時間を計算した。

図 5 に、シミュレーションの結果を示す。図中の 'SCATTER' は、CBS 法の比較対象である相関性を考慮しないスケジューリング手法を示している。横軸は行列 1 の要素のうち 0 以外の値を持つ要素の割合 (%), 縦軸はアクセスの平均応答時間を表している。この結果から、アイテム間の相関性を考慮しない手法に比べて、提案した CBS 法の応答時間が短くなる事が分かる。また、行列中の 0 でない要素の割合が小さくなる、つまり、相関性が強くなるにつれて、提案した手法の効果が顕著に現れている。

シミュレーション 2:

行列 2 においてグループの数を変化させて、クライアントからのアクセスの平均応答時間を計算した。

シミュレーションの結果を図 6 に示す。横軸は行列 2 におけるグループの数、縦軸はアクセスの平均応答時間を表している。この結果から、相関性を考慮しない手法に比べて、CBS 法の平均応答時間が短くなる事が分かる。また、グループ数の増加にとまらぬ、従来の手法と提案した手法の性能の差が大きくなっている。これは、グループの数が増加するにつれてアイテム間の相関関係が明確になり、相関性を考慮した手法の効果が顕著に現れるためである。

5.2 CB-PT 法の性能評価

次に、CB-PT 法の有効性を検証するために行ったシミュレーション評価の結果を示す。

シミュレーションでは、クライアントのキャッシュサイズが放送プログラムのサイズの 40% の場合に、行列 1 において行列中の 0 でない要素の割合 (%) を変化させて、クライアントからのアクセスの平均応答時間を計算した。筆者らは、40% 以外のキャッシュサイズにおいても同様の評価を行ったが、いずれの場合も CB-PT 法は相対的にはほぼ同様の性能を示した。そこで、本論文では 40% の場合の結果のみを示すものとした。

CB-PT 法の比較対象としては、PT 法を用いた。ただし、PT 法では PT 値の計算の際に、アイテム i へのアクセス確率 p_i を次のように計算している。

$$p_i = \sum_{j=1}^M P_{ij}$$

シミュレーション結果を図 7 に示す。この結果から、CB-PT 法が、従来の手法のうちで最も良い性能を示すと考えられている PT 法よりもさらに良い性能を示す事が分かる。特に、行列内で 0 でない要素の割合が

小さくなる、つまり、アイテム間の相関性が強くなると、PT法との平均応答時間の差が大きくなっている。

ここで、両手法の性能差は最大で2.2%程度であるが、これはシミュレーション条件としてクライアントが2つのデータアイテムに対して同時にアクセス要求を起こす場合を想定していることに起因する。CB-PT法は相関性を考慮したキャッシュの置き換えを行うため、3つ以上のデータアイテムに対して同時にアクセス要求を起こすといった、より実環境に近い場合において、手法の有効性が顕著に現れるものと考えられる。

6. おわりに

本論文では、放送されるデータアイテム間の関連性を考慮した放送プログラムのスケジューリング法およびキャッシング法を提案した。さらに、シミュレーションによって提案した手法の性能評価を行い、有効性を確認した。提案手法は、それぞれヒューリスティックな手法であるため、今後はさらに、提案手法と最適解との性能比について調査する必要がある。

本論文では、簡単のためにクライアントが2つのデータアイテムに対して同時にアクセス要求を起こす場合を想定したが、提案したCBS法およびCB-PT法は同時要求を出すアイテム数が3つ以上の場合にもそのまま適用可能である。CB-PT法については、5.2節で述べたように、3つ以上アイテムに対して同時にアクセス要求を起こす場合の方が、有効性が顕著に現れるものと考えられる。今後は、このような環境において、提案手法の詳細な評価を行う予定である。

また、実際の環境においては、本論文で想定したように相関性のあるアイテムに対して同時にアクセス要求を起こす場合だけでなく、ある程度の時間差を持って連続して起こす場合があるものと考えられる。筆者らは、文献13)において、このような環境におけるアイテム間の最適な放送時間間隔の決定法と、放送プログラムのスケジューリング法を提案している。今後は、このような環境における効果的なキャッシング法について考察する必要がある。さらに、従来のスケジューリング手法のようにアクセス確率の高いアイテムの放送頻度を高くする場合についても、アイテム間の相関性を考慮した効果的なスケジューリング戦略を考える必要がある。

謝辞 本研究は、日本学術振興会未来開拓学術研究推進事業における研究プロジェクト「マルチメディア・コンテンツの高次処理の研究」(Project No.JSPS-RFTF97P00501)の研究助成によるものである。ここに記して謝意を表す。

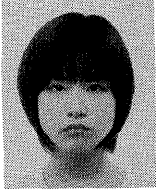
参考文献

- 1) Acharya, S., Alonso, R., Franklin, M. and Zdonik, S.: Broadcast Disks: Data Management for Asymmetric Communication Environments, *Proc. ACM SIGMOD Conference*, pp.199-210 (1995).
- 2) Acharya, S., Franklin, M. and Zdonik, S.: Disseminating Updates on Broadcast Disks, *Proc. VLDB Conference*, pp.354-365 (1996).
- 3) Acharya, S., Franklin, M. and Zdonik, S.: Prefetching from a Broadcast Disks, *Proc. Int'l Conference on Data Engineering*, pp.276-285 (1996).
- 4) Acharya, S., Franklin, M. and Zdonik, S.: Balancing Push and Pull for Data Broadcast, *Proc. ACM SIGMOD Conference*, pp.183-194 (1997).
- 5) Gondhalekar, V., Jain, R. and Werth, J.: Scheduling on Airdisks: Efficient Access to Personalized Information Services via Periodic Wireless Data Broadcast, Technical Report CS-TR-96-25, Univ. Texas at Austin, Dept. of Comp. Sci. (1996).
- 6) 箱守 聡, 田辺雅則, 石川裕治, 井上 潮: 放送型通信/オンデマンド型通信を統合した情報提供システム, 情報処理学会研究報告, Vol.34, No.8, pp.55-60 (1997).
- 7) Hameed, S. and Vaidya, N.H.: Log-time Algorithms for Scheduling Single and Multiple Channel Data Broadcast, *Proc. MOBICOM 97*, pp.90-99 (1997).
- 8) Imielinski, T., Viswanathan, S. and Badrinath, B.R.: Energy Efficient Indexing On Air, *Proc. ACM SIGMOD Conference*, pp.25-36 (1994).
- 9) 石川裕治, 田辺雅則, 箱守 聡, 井上 潮: ハイパーテキスト間のデータ共有を考慮した放送型情報提供方式, 信学技報, Vol.97, No.160, pp.121-126 (1997).
- 10) Jain, R. and Werth, J.: Airdisks and Air-RAID: Modelling and Scheduling Periodic Wireless Data Broadcast (Extended Abstract), DIMACS Tech. Report 95-11, Rutgers University (1995).
- 11) Lin, L. and Xingming, Z.: Heuristic MultiDisk Scheduling for Data Broadcasting, *Proc. Int'l Workshop on Satellite-Based Information Services (WOSBIS'97)*, pp.1-5 (1997).
- 12) Vaidya, N.H. and Hameed, S.: Improved Algorithms for Scheduling Data Broadcast, Tech. Report 96-029, Comp. Sc. Dept., Texas A&M University (1997).
- 13) 矢鳥悦子, 原 隆浩, 塚本昌彦, 西尾章治郎:

相関性を持つデータ間の放送時間間隔について、情報処理学会論文誌, Vol.40, No.1, pp.188-196 (1999).

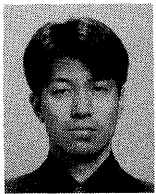
(平成 10 年 8 月 17 日受付)

(平成 11 年 6 月 3 日採録)



矢島 悦子 (正会員)

1998 年大阪大学工学部情報システム工学科卒業。同年, (株) エフエム大阪入社, 現在に至る。データベースシステムおよび放送型通信に興味を持つ。



原 隆浩 (正会員)

1995 年大阪大学工学部情報システム工学科卒業。1997 年同大学院工学研究科博士前期課程修了。同年, 同大学院工学研究科博士後期課程中退後, 同大学院工学研究科情報システム工学専攻助手となり, 現在に至る。1996 年本学会山下記念研究賞受賞。データベースシステム, 分散処理興味を持つ。IEEE, 電子情報通信学会各会員。



塚本 昌彦 (正会員)

1987 年京都大学工学部数理工学科卒業。1989 年同大学院工学研究科修士課程修了。同年, シャープ (株) 入社。1995 年大阪大学大学院工学研究科情報システム工学専攻講師, 1996 年より, 同大学院工学研究科情報システム工学専攻助教授, 現在に至る。工学博士。時空間データベースおよびモバイルコンピューティングに興味を持つ。ACM, IEEE 等 7 学会の会員。



西尾章治郎 (正会員)

1975 年京都大学工学部数理工学科卒業。1980 年同大学院工学研究科博士課程修了。工学博士。京都大学工学部助手, 大阪大学基礎工学部および情報処理教育センター助教授を経て, 1992 年より大阪大学大学院工学研究科情報システム工学専攻教授となり, 現在に至る。この間, カナダ・ウォータールー大学, ピクトリア大学客員。データベース, 知識ベース, 分散システムの研究に従事。現在, Data & Knowledge Engineering, Data Mining and Knowledge Discovery, New Generation Computing 等の論文誌編集委員。ACM, IEEE 等 8 学会の会員。