

人気度に依存した放送配信スケジューリング方式のシミュレーション

白田由香利^{†1} 飯沢篤志^{†1} 橋本隆子^{†1} 李韻^{†2}

近年、放送のデジタル化が世界的規模で急速に進展している。我々は、現在の放送の発展型として、高度に構造化されたデジタル情報をサービスとして提供する次世代の「情報放送」の研究を行なっている。本稿では、情報放送サービスにおいて、現在のTV視聴者数と同程度の数百万人の利用者がTV受信端末を通じてリクエストを送信し、放送局側でリクエストの人気度に応じた放送配信スケジュールを行う番組配信システムを想定する。番組は有料であることを仮定した場合、設定環境や配信スケジューリング方式の違いにより、どのように利用者(視聴者)への応答時間、及び、サービス提供者の収益性が影響を受けるかをシミュレーションにより分析する。

Simulation Studies on Popularity-based TV Program Scheduling by Broadcasting Stations

YUKARI SHIROTA,^{†1} ATSUSHI IIZAWA,^{†1} TAKAKO HASHIMOTO,^{†1} and JIE LI^{†2}

Digital broadcasting has rapidly spread throughout the world. Information Broadcasting is a next-generation information service, provided by digital broadcasting systems, with highly structured information. This paper describes a popularity-based TV program scheduling system at the Information Broadcasting service. In this system, as many as several million current TV viewers send requests from their TV terminals; the broadcasting stations then decide the TV program broadcasting schedules. Assuming that the Information Broadcasting is a pay service, using simulation studies, we study what program scheduling policies and parameters are effective to increase the satisfaction of Information Broadcasting providers and viewers alike.

1. はじめに

近年、放送のデジタル化が世界的規模で急速に進展している。従来データベースへのアクセスはコンピュータから通信網を通じて行う方式が主流であったが、放送のデジタル化とともに、データベースの情報を放送で配信し、利用者は家庭用TV受信端末でそれを受信するというデータアクセス方式も今後広く普及していくと予測される。

我々は、現在の放送の発展型として、高度に構造化されたデジタル情報をサービスとして提供する次世代の「情報放送」の研究を行なっている。本稿では情報放送サービスにおいて、現在のTVの視聴者数と同程度の数百万人の利用者が家庭用TV受信端末を通じて

データベースにアクセスする環境を対象とする。我々の想定する番組配信システムでは、データアクセスの際、利用者はリクエストを送信する。放送局側ではそのリクエストを集計し、その人気度に応じて放送配信スケジュールを行う¹⁾。スケジュールリング方式に対する要望事項は、データアクセスに要する平均応答時間が短いこと、及び、有料番組放送での収益性が高いことである。本稿では、広く使われている3種類のスケジュールリング方式に対するシミュレーション評価を行った。

2. 関連研究

近年、データベース分野においても放送あるいは放送を用いた情報配信システムに関する研究が盛んであり、利用者のデータアクセスに対する応答時間期待値を短縮するため、各種の放送データ配信スケジュールリング方式が提案されている^{2)~6)}。従来の研究ではその応用として、移動モバイル端末からのデータベースアクセスや、HTMLファイルなど、関連性のあるデータアイテムへのアクセスなどを対象としてきた。一方、

^{†1} (株)次世代情報放送システム研究所、(株)リコーソフトウェア研究所。[(株)リコーから、(株)次世代情報放送システム研究所へ兼任出向中。The authors are partly on loan from Software Research Center, Ricoh Company, Ltd. to Information Broadcasting Laboratories, Inc.]

^{†2} 筑波大学 電子・情報工学系 Institute of Information Sciences and Electronics, University of Tsukuba

我々の対象とする応用は従来、電話や葉書、FAXなどでリクエストしていた利用者がTV受信端末を通してリクエストを送信するようなシステムであり、その規模は現在のTV視聴者数並みの、数百万人の利用者からのアクセスを想定している。また、データ間の相関性はないものと仮定し、リクエストの多さに応じて放送頻度を高める方針をとる。

放送機構を用いたデータアクセス方式として、ビデオ・オン・デマンドシステム(以下VODと略す)及びニア・ビデオ・オン・デマンド(以下NVODと略す)⁷⁾が考えられる。これらのシステムにおいても人気度の高い順に多数のストリームを割り当て応答時間の短縮を実現しようと、各種リソースのスケジューリング手法が研究されてきた²⁾。どのコンテンツにどの程度の程度のストリーム数を確保すればよいかは、過去の人気度情報などから計算される。その結果を元に、RAID(Redundant Array of Disks)を用いた複数ディスク分割方式及びそのクラスタ内のデータ配置方式を制御したり、ビデオサーバの階層的記憶管理の構成制御等を行う^{8),9)}。

我々の行う番組配信システムのスケジューリング方式の研究はこうしたVODにおけるスケジューリング方式の研究と以下の点で異なっている。

- ・VODでは主に映画ビデオ配信を想定している¹⁰⁾が、我々の想定するデータでは、サイズはさらに小さく、データ数はより膨大であることを前提としている。
- ・VODが即時型視聴を想定しているのに対し、蓄積型視聴を想定する。
- ・VODがストリームの多重化のスケジューリング方式を対象としている¹¹⁾のに対し、配信用帯域の時分割多重化のスケジューリング方式を対象とする。

こうした違いがあるので、番組システムの振舞いはVODシステムのそれとは大きく異なってくる。

3. 想定する応用

本節では、番組配信システムの応用として我々が想定する情報放送サービスを説明する。本番組配信シス

テムでは、リクエストによる人気投票制に基づいてデータベース(DB)サイトで配信スケジューリングが行われる。以下では、データアクセス要求を「リクエスト」と、データアイテムを「番組」と呼ぶこととする。

番組の内容は、ビデオクリップ(カタログ/CM、歌曲)、マルチメディアデータベースからの検索結果、ソフトウェアなど多種多様である。以下では、我々が典型的な応用として想定する、ヒットチャートと連動した音楽ソフト配信システムを説明する。

音楽ソフト配信システムとは、現在CDやDVD媒体としてレコード店などで販売されている音楽ソフトを、デジタルコンテンツとして放送配信により販売するシステムである。1曲あたり3~4分のコンテンツを圧縮するが、映像と音楽データの比率によりデータサイズが異なってくる。全編MPEG2とした場合約150MB³⁾、音楽データだけの場合約1.7MBですむ⁴⁾。

上述したように番組サイズは動画の比率に大きく依存するが、一般的な番組では、メニューなどのブラウジングプログラムも含めて50MB~100MB程度が適切ではないかと、我々は考えている。

4. 番組配信システムの概要

本番組配信システムはリクエストによる人気投票制に基づいてデータベース(DB)サイトで配信スケジューリングが行われる。家庭サイトの利用者はDBサイトに対して通信ネットワークを通じて以下のようなリクエストメッセージを送る。その内容は、TV情報端末内のDBに記憶されている。

```
request( userID, contentID, currentDate,
         expireDate )
```

リクエストは有効期限をもち、その期間を過ぎるとシステム(DBサイト及びTV情報端末)上から消される。その期限を示すのがexpireDateである。リクエストが発行された時刻(currentDate)からリクエスト有効期間を加算した時刻がexpireDateである。

³⁾ NTSCの標準テレビジョン用データでMPEG2で圧縮すると3~8Mbpsとなる¹²⁾。1曲あたり3.5分、MPEG2圧縮後平均6Mbpsとすると、3.5min×60sec/min×6Mb/sec=157.5MB。但し、¹³⁾にも指摘のあるように、MPEG2による圧縮率の他、8相トレリスPSKなどの高効率伝送技術という要素があるので、利用する伝送技術により同じ放送帯域幅でもより高い情報レートが得られる可能性がある。

⁴⁾ MPEG2 Audio AAC符号化方式を利用すると64kbpsで原音と殆ど変わらぬ音質が得られる。この方式はBSデジタル放送でも採用が決定された。この方式で計算すると、3.5min×60sec/min×64kb/sec=1.68MB。

¹⁾ NVODとは、例えば2時間の映画を、開始時刻を10分ずつずらして12ストリームで繰り返し放送するサービス。視聴者は最大10分待てば初めから見られる。但しVODがサーバと視聴者が1対1で接続されていないので、ビデオの早送り、逆再生などの制御はできない。

²⁾ ストリームとはビデオサーバの独立した映像チャンネルである。人気の高いコンテンツにはアクセスが集中するので、より多くのストリーム数を確保する必要がある。

DB サイトに到着したリクエストは、リクエスト待ち行列に入れられる。この待ち行列モデルは単純な first in first out 方式ではなく人気度に応じた待ち行列であることを特長とする。配信スケジューリングは一定期間分まとめて行う。この期間を「配信スケジューリング周期」(以下、周期と略す)と呼ぶ。DB サイトは周期ごとに各番組の有効リクエストの人気度を合計し人気度の高い順に配信する。

周期は固定で、利用者にもその長さや開始のタイミングは知られていると仮定する。k 番目の周期に放送される番組は、(k-1)番目の周期中に決定される(図 1 参照)。(k-1)番目の周期に実行されたスケジューリングを k 番目のスケジューリングと呼び、その結果作成された番組表を k 番目の番組表と呼ぶ。k 番目の番組表は、(k-1)番目の周期中に繰り返し放送され、利用者は番組表に関しては受信に失敗することはない、つまり利用者への番組表の周知徹底が行われている、と仮定する。以下では、リクエストした番組がスケジューリングの結果、番組表に選ばれることを「当選」、そうでないことを「落選」と呼ぶことにする。

DB サイト上で稼動するスケジューリングプロセスはスケジューリングに際し、始めに有効期間を過ぎたリクエストを廃棄する。つまりスケジュール開始時点で有効であるリクエストのみを元データとして、番組表を作成する。当選した番組へのリクエストは、DB サイト上の集計対象のリクエスト群から抹消される。スケジュール方式の内容は次節で説明する。

利用者が「リクエストした番組受信に成功する」という場合は以下の 2 条件を満たした場合である。

- **条件 1:** (A)そのリクエストの有効期間中にリクエストした番組が放送される、あるいは、(B)リクエストの有効期間中にリクエストした番組が当選する。

自分のリクエストが集計されスケジューリングされる前に、場合によっては他の利用者のリクエストにより当選し放送される場合がある。前者(A)の条件はそのような場合を含む。後者(B)の場合は、当選が k 番目の周期であった場合、受信するのは(k+1)番目の周期である。

- **条件 2:** 放送された番組を受信端末が正しく受信する。

但しリクエストしたコンテンツが当選し放送されたとしても、正しく受信されるとは限らない。以下では受信端末上での受信エラー確率を $P_f(0 \leq P_f \leq 1)$ で表わす。また、本番組配信システムでは、受信機上でエラー無く蓄積されたことを確認するまで課金は行われないと、

仮定する。

次に、方式評価の指標である、(1)利用者のリクエストへの応答時間、及び、(2)サービス提供者側の収益度を本番組配信システムモデルではどのように定義したかを説明する。

<応答時間>

図 1 に示すように、利用者が番組 x に対してリクエストを送信したとする。そして、x の有効期間中に図に示すように 2 回のスケジュールが実施されたとする。1 回目のスケジュールで x は落選、2 回目のスケジュールでは当選した場合、x は(k+1)番目の番組表に含まれ、(k+1)番目の番組配信で放送される。その場合、応答時間は図に示すように、[リクエスト x の送信時]~[x の受信成功が確認された時点]までとなる。1 回目のスケジュールで当選し、k 番目の番組配信で受信に成功した場合は、その分、応答時間は短くなる。

最終的に利用者が x の受信に成功しなかった場合、利用者は送信される可能性のある (k+1)番目の周期が終了するまで待つとして、そこまでを応答時間と定義する。実際には、利用者にも配信スケジューリング周期とそのタイミングは知られているので、利用者は k 番目の番組表を受信した時点で、自分の出したリクエストは放送されないことが分かる。

DB サイトにおいて当選したリクエストは集計対象のリクエスト群から抹消される。その際シミュレーションにおいては、想定する受信エラー確率 P_f を考慮し、[当選したリクエスト総数 $\times (1 - P_f)$] 個のリクエストを抹消する。

<収益性>

次にサービス提供者の収益性をどのように評価するか説明する。本システムでは、リクエストした後受信端末にエラーなく蓄積されるに至った場合、つまり、受信に成功した場合、1 度数の売上げがあったとする。単純化のため、今回の評価モデルでは番組内容による売上げ度数の違いはないものとし、一律に「1 番組 1 度数」とした。売上げた配信番組の総数を「売上げ総度数」と呼ぶ。

番組配信システムの評価モデルで仮定したことを以下で述べる。まず評価モデルでは、(a)番組人気度は正規分布に従うことを仮定した。つまり人気の高い順に番組を並べ、その人気度をみると、ピークをもつ山形を 1/2 に切断した形状を成すようにした。正規分布採用の理由は、本システムでは対象となるリクエスト数も膨大なので、正規分布が妥当であると考えたからである。

また、問題の単純化のため、評価の 1 バッチ区間に

において、(b)番組及びその人気度の動的変更はない、と仮定した。

5. 配信スケジューリング方式

本節では、リクエスト数に依存して放送頻度を決定するスケジューリング方式として評価を行った数種類の方式を説明する。本稿では以下の A, B, C, 3種類の配信スケジューリング方式について比較評価を行った。以下に各方式の内容を示す。

(A) 単純法: 番組枠数分、人気度の高い順に1回ずつ選択。

(B) 比例代表制: 人気度に応じて枠数を比例配分する。これは n 個の番組のリクエスト数が $x[1] \sim x[n]$ で表わされる時に、リクエスト数に応じて配信番組枠数の総数 w を $y[1] \sim y[n]$ と配分する方式で、 $x[1]/y[1] \sim x[n]/y[n]$ ができるだけ近いことを目標とする配分法である¹³⁾。比例配分方式にも各種があるが、我々はそのうちドント方式と呼ばれるものを使った。

(C) 局所最適法: 1回の配信スケジューリングで売上げ総度数を最大にする。この売上げ総度数は受信機側の受信エラー確率 Pf に依存する。ある番組を1回の配信スケジューリング周期の中で $n(>0)$ 回繰り返し放送した場合の延べ受信番組数は、受信失敗確率 0 のときを 1 とすると以下の計算により、 $(1 - (Pf)^n)$ となる。

$$(1 - Pf) + Pf(1 - Pf) + (Pf)^2(1 - Pf) + (Pf)^3(1 - Pf) + \dots + (Pf)^{n-1}(1 - Pf) = (1 - (Pf)^n)$$

C 方式では、この値が最大になるように順に番組選択していく。実際の C 方式の運用では Pf を前もって予測することは不可能なので、予測 Pf 値と実際の Pf 値は異なる。以下の評価では、実際の Pf 値が予測 Pf 値に等しいとした場合の C 方式を「純粋 C 方式」と呼び、予測と実際値が異なるとする C 方式を「予測 C 方式」と呼ぶことにする。

上記 3 種類の方式を比較検討方式とした理由としては、A 及び C 方式は明確なスケジューリング方針で比較が容易であること、及び B の比例代表制は実際に VOD などの資源分割で用いられていることがあげられる。方針の違いを考えると、A 方式が、1 回のスケジュールで、より広い範囲の番組を配信する方式であるのに対し、B 及び C 方式では、少ない番組数になってもよいから、人気の高い番組を繰り返し配信する。B 方式が人気の高さに応じて資源を比例配分しようとする方針をとるのに対し、C 方式は収益性のみを最適化の対象とする。

6. 評価

評価におけるデフォルトの環境設定パラメータ値は以下のようにした。

- ・利用者数: 100 万人
- ・平均リクエスト数: 2 個/[週・人]
- ・番組数: 100 万
- ・配信番組数: 2160 本/日
[高能率伝送技術を考慮せず、単純に放送帯域 10Mbps を圧縮後の番組サイズ 50MB で割り算した値である。]
- ・配信スケジューリング周期: 1 日
- ・リクエスト有効期間: 1 日
- ・シミュレーションバッチサイズ: 2 週間
- ・リクエスト当りの支払料金: 一律 1 度数

リクエストの DB サイトへの到着はポアソン分布を仮定する。シミュレーションでは、上述した A, B, C の 3 つのスケジューリング方式が各種の環境パラメータ値の変化によってどのような影響を受けるかを評価した。 Pf 値は予測困難なため、其々の評価において 10%, 50%, 90% のように典型的な値、数種類をシミュレーションした。 $Pf=10\%$ の場合の A 方式での値を“A10”という記号で表現するとする。

6.1 人気度ピークの鋭さの影響

上述したように、我々の評価モデルでは、番組人気度は正規分布に従うと仮定した。つまり、シミュレーションにおいてリクエスト中の contentID を乱数を用いて生成する際に、そのリクエスト数が正規分布になるように発生させている。一般に、人気度のピークが鋭いほど少数の人気番組にリクエストが集中するため収益性が高い、と予想されるので、そのピークの鋭さと収益性等の関係を分析した。ピークの鋭さを示すパラメータとして正規分布の標準偏差 σ を用いた。 σ が小さいほどピークが鋭いことを表わす。

図 2、図 3 に σ と指標(応答時間、総売上げ数)の関係の評価結果を示した。B 方式は Pf が高くなると C 方式とほぼ同じになるが、 Pf が低いと C に比べてよくないことが分かる。A 方式は Pf が高い場合、B 及び C 方式に比べて悪く、 Pf が低い場合でも、ピークが鋭い範囲では C 方式よりも悪いことが分かる。

Pf が低い範囲では、A あるいは C 方式が、 Pf が高い範囲では B あるいは C 方式が優位であり、全般的にみて C 方式が他方式に比較して優位であると言える。C の優位性はピークが鋭いほど、また、 Pf が高いほど顕著になる。ピークが緩やかになると C 方式

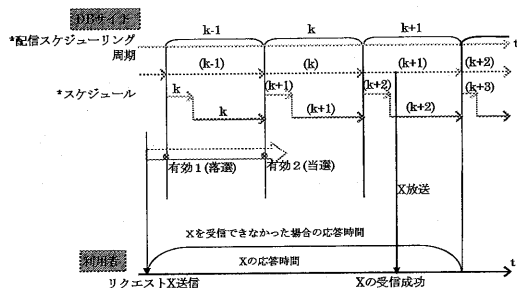


図1 スケジューリング・タイムチャート

Fig.1 Scheduling Time Chart

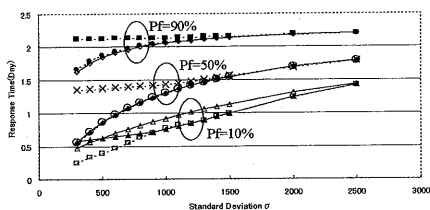


図2 標準偏差に対する応答時間

Fig.2 Response Time with Varying Standard Deviation

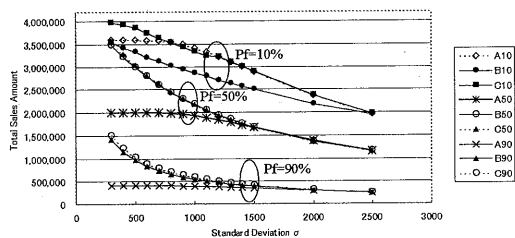


図3 標準偏差に対する総売上げ度数

Fig.3 Total Sales Amount with Varying Standard Deviation

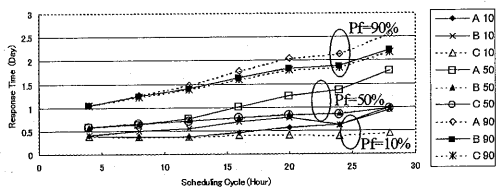


図4 スケジューリング周期に対する応答時間

Fig.4 Response Time with Varying Scheduling Cycle

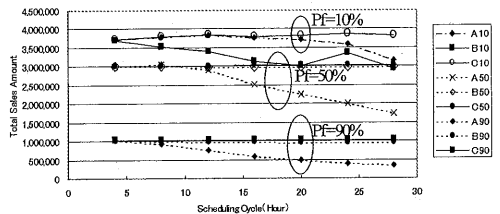


図5 スケジューリング周期に対する総売上げ度数

Fig.5 Total Sales Amount with Varying Scheduling Cycle

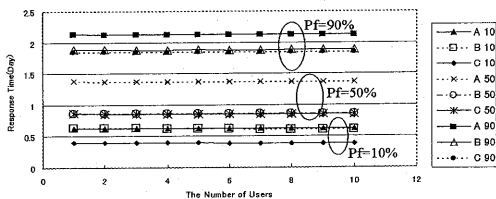


図6 利用者数に対する応答時間

Fig.6 Response Time with Varying the Number of Users

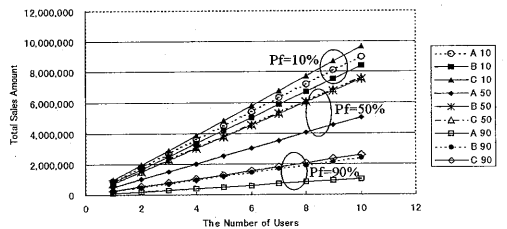


図7 利用者数に対する総売上げ度数

Fig.7 Total Sales Amount with Varying the Number of Users

の優位性が目立たなくなるのは、どの方式でも番組配信は1回という状況になるからである。

6.2 有効期間と周期の比の影響

リクエスト有効期間とスケジューリング周期の関係を考えた場合、周期のほうを長く設定するのは無意味であり、[有効期間>周期]と設定するのが一般的である。それでは、有効期間と周期はどの程度の比にすれば効果的であるかを検討するため、有効期間、周期ともに1日をデフォルト値とし、両者の値を動かして評価を行った。C方式は純粋C方式とした。Pf値は10%,50%,90%の3種類で計算した。 σ は500とする。

有効期間を1日に固定し、周期を4時間から28時間まで変化させた場合のシミュレーション結果を図4、図5に示す。C方式が一貫して優位性を示す。しかも、応答時間及び総売上げ度数の値は周期の値を変えても変わらずほぼ一定を示す。

次に、周期を1日に固定し、有効期間を1日~3日まで変化させた場合でも、同様の結果が得られた。但し、有効期間が長いとその分1個のリクエストがスケジューリング集計対象となる回数が増加するので、その分総売上げ度数は増加する傾向を示した。

まとめると、応答時間を短縮するためには、C方式を、有効期間と周期がほぼ同じ値の範囲で用いるのがスケジューリングコストも低く、最も効果的である、と言える。

6.3 利用者数の影響

次に利用者数が増加した場合の影響を調べた。1週間当たりの、一人の利用者のリクエスト数は固定とする。Cは純粋C方式、 $\sigma=500$ の場合の結果を図6、図7に示す。結果は、総売上げ度数は線形に増加するが、応答時間は一定であった。これは、「利用者数が増加しても配信コストは一定ですむ」という放送方式の利点を示している。

7. 結論

情報放送システムにおける番組配信システムにおいて、どのような番組配信スケジューリング方式をとると、収益性及び応答時間の面で効果的であるかをシミュレーションにより評価した。比較した方式は、単純に人気度の高い順に1回のみ配信するA方式、人気の高さに応じて放送帯域の比例配分を行うアルゴリズムを用いたB方式、1回のスケジューリングで収益性が最も高くなるように同じ人気番組を複数回選択するC方式の、3方式である。

Pfが低い範囲では、AあるいはC方式が、Pfが高

い範囲ではBあるいはC方式が優位であり、全般的にみてC方式が他方式に比較して優位であると言える。Cの優位性はピークが鋭いほど、また、Pfが高いほど顕著になる。ピークが緩やかになるとC方式の優位性が目立たなくなるのは、どの方式でも番組配信は1回という状況になるからである。また、応答時間を短縮するためには、C方式を、有効期間と周期がほぼ同じ値の範囲で用いるのがスケジューリングコストも低く、最も効果的である、ことが分かった。利用者数を増加させたシミュレーションでは、「総売上げ度数は線形に増加するが、応答時間は一定」という放送方式の特長が確認できた。

参考文献

- 1) 白田由香利、飯沢篤志、真野博子、李頌:「放送配スケジューリング方式のシミュレーション」、映像情報メディア学会冬季大会予稿集、p. 67、1998年12月。
- 2) 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.
- 3) Hameed, S. and Vaidya, Nitin H.: Log-time Algorithms for Scheduling Single and Multiple Channel Data Broadcast, Proc. MOBICOM 97, pp. 90-99, 1997.
- 4) Imielinski, T., Viswanathan, S., and Badrinath, B. R.: Energy Efficient Indexing On Air, Proc. ACM SIGMOD Conference, pp. 25-36, 1994.
- 5) 石川裕治、田辺雅則、箱守聰、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」、電子情報通信学会信学技報、Vol.97、No.160、pp.121-126、1997。
- 6) 矢島悦子、原隆浩、塚本昌彦、西尾章治郎:「データ間の相関性を考慮した放送データの送受信方式について」、情報処理学会研究報告 98-DBS-116(1)、pp.95-102、1998年7月。
- 7) 長沢邦彦:「ケーブルテレビとNVOD」、テレビジョン学会誌、Vol. 49、No. 5、pp. 618-624、1995。
- 8) Chang, C. K., Nguyen, T. T., and Mongkolwat, P.: "A Popularity-based Data Allocation Scheme for a VOD Server," Proc. Annual International Computer Software & Application Conference., pp. 62-67.
- 9) Little, T. and Venkatesh, D.: "Popularity-based assignment of movies to storage devices in a video-on-demand system," IEEE Multimedia, Vol. 2, pp. 280-287, Spring, 1995.
- 10) 田中良明、加藤壮太郎:「VODサービスの価格とトラフィック量予測」、電子情報通信学会信学技報SSE95-123、pp. 43-48、1995年12月。
- 11) 小西智夫:「VODシステムにおけるビデオサーバとセットトップの基礎技術 3-1 ビデオサーバ」、テレビジョン学会誌、Vol. 49、No. 5、pp. 605-608、1995。
- 12) 安田靖彦:「放送のデジタル化と情報処理」、情報処理、Vol. 40、No. 4、pp. 409-413、1999年4月。
- 13) 有澤誠:「リクリエショナルプログラミング -遊びのなかの情報処理」、ソフトバンク、1990。