

プッシュ型とプル型通信の動的統合による応答時間の短縮

青 野 正 宏[†] 渡 辺 尚^{††} 水 野 忠 則^{††}

無線データ放送において、サーバの情報をユーザに伝える方法としてプッシュ型通信とプル型通信がある。プッシュ型通信は情報の要求を行う上り通信を省略し、通信帯域を節約することができる。一方プル型通信はユーザの要求を把握し、これに的確に対応することができる。両者の特徴を活かし、両通信方式を併用する方式もこれまでに提案されているが、本論文はさらにそれを進めてプッシュ型通信とプル型通信を動的に統合するモデルについて提案と評価を行う。

Reducing the Response Time by Integrated Push-Pull Method

MASAHIRO AONO,[†] TAKASHI WATANABE^{††} and TADANORI MIZUNO^{††}

There are push-based and pull-based communication ways how to inform the data of a server to users about the radio data broadcast. The upstream communication can be omitted and the band can be saved in the push-based communication. On the other hand, the server can grasp requests of users and can correspond to them accurately in the pull-based communication. Some techniques that combine both communication methods for utilizing the characteristic of both are proposed so far. Moreover, we extend these techniques and we propose and evaluate the model that integrates push-based and pull-based communications dynamically in this paper.

1. はじめに

クライアントサーバ型のシステム構成では、サーバに主要データベースを持つが、クライアント端末の携帯化の普及とモバイル通信の進展により、無線を利用して任意の場所からサーバの情報を利用することが可能となってきた。無線通信においては、その覆域内であれば受信周波数を合わせることで、同時に複数のクライアントが同じサーバからのデータを受信することができる。クライアントの数が増えても必要帯域の大きさは変わらない。複数のクライアント利用者が同じデータを必要としているのであれば、無線を通してデータを放送することにより、帯域を有効に利用することができる。本論文は、無線の覆域内でサーバからデータを直接受信することが可能なクライアントが多く存在するシステムにおいて、どのようなデータの提供方式が帯域の利用効率が良く、敏速にクライアントにデータを提供することが可能かを検討するものである。想定するシステム事例としては、地域内のユー

ザに経済・交通・気象・その他のニュースを提供するデータ放送システム、空港・商店街など限られた環境で特定情報を提供する個別情報システムなどである。クライアントは必要なデータのみを取り込み、自分のファイルに格納してクライアント利用者が参照する。

サーバは、クライアントがあるデータを必要とすることを事前に把握しているならば、個々のクライアントからオンデマンドで要求を受けることなく、一方的にデータを放送出力する。これをプッシュ型通信という。それに対し、クライアント利用者はクライアントを通して、必要に応じてサーバにデータの要求を行う（上り通信）。サーバはクライアントからの要求に対し、要求されたデータをそのクライアントに対して返す（下り通信）。これをプル型通信という。プッシュ型通信では、上り通信の帯域を節約することができる。しかし、サーバのどのデータをどの頻度でクライアントが要求しているということを想定できても、個々の要求は把握できないため木目細かい対応はできない。プル型通信では、クライアントからのデータ要求の時刻と累計数からクライアント利用者の待ちをなるべく小さくする出力スケジュールの設定が可能となる。その代り、上り通信の帯域が必要となる。このプッシュ型通信とプル型通信の両者の特徴を生かした通信方式

[†] 三菱電機
Mitsubishi Electric Corp.
^{††} 静岡大学
Shizuoka University

について検討する。クライアントがサーバのデータを利用する形態として次の2つを想定する。ひとつは、サーバのデータが変化しないかあるいは変化の頻度が小さく、クライアントは目的のデータを受信して自ファイルに取り込めれば受信を完了する場合である。もうひとつは、サーバのデータが刻々変化し、クライアントは常に最新データを監視する場合である。データが更新されたとき、サーバは更新データをプッシュ型通信でクライアントに送信する。クライアント利用者は最初のデータをプッシュ型通信またはプル型通信で得るが、その後の更新データはプッシュ型通信でクライアントが受信し自己のローカルファイルに格納したデータを参照して得る。これらのモデルにおいて、プッシュ型通信のみ、プル型通信のみ、両者を併用する通信方式の評価を行い、さらにプッシュ型通信とプル型通信の応答通信の帯域を統合して利用し、両者の動的な割付を行う方式を提案・評価する。なお、言及する既存論文では、プッシュ型通信を放送型通信、プル型通信をオンデマンド型通信という用語を用いている場合もあるが、本論文においてはプッシュ型通信／プル型通信という用語に統一して用いている。また、本論文ではプッシュ型通信とプル型通信の「併用」という用語と「動的統合」という用語を使い分けている。「併用」はプッシュ型通信とプル型通信の各々の帯域とその方式による通信データを静的に定めておいて両者を併用することを指すのに対し、「動的統合」は動的にプッシュ型通信とプル型通信の帯域と通信データの種類の決定することを指す。さらに単にクライアントと記した場合はクライアント端末そのものまたはクライアント利用者を含めるが、クライアント利用者と記述した場合はクライアント端末を利用する人間またはデータを利用するソフトウェアなどを指す。以下、2章にて検討にあたっての前提条件を、3章にてサーバのデータの内容が固定の場合のモデルを、4章にて出力データが更新される場合のモデルについて記述する。5章に先行研究との関連を述べ、6章にてまとめを示す。

2. 検討の前提条件

以下を仮定する。

- プル型通信において、クライアントからサーバにデータを要求する上り通信とサーバからクライアントにデータを提供する下り通信はひとつの通信チャネルを時分割で利用するものとする。プッシュ型通信のみの場合、下り通信のみとする。

- サーバから出力される下り通信のデータは全て

のクライアントで同時に受信することができる。

- 上り通信と下り通信とではデータの大きさは異なる（上り通信 ≤ 下り通信）が、下り通信の各データの大きさは同じとし、その1単位をページと呼ぶこととする。

- 下り通信で帯域を占有したと仮定したときに、1ページのデータを出力する単位時間をタイムスロット（以下、 t_s で示す。）と呼ぶこととする。

- p_i はページ i の $1t_s$ における要求頻度、 ρ を平均要求頻度 (Request Frequency)、 m をページ数、 θ を分布の偏在係数とすると、各ページに対するクライアントの要求頻度の比率は式 (1) に示す Zipf 分布に従うものとする。

$$p_i = \frac{m \cdot \rho}{i^\theta \cdot \sum_{j=1}^m (1/j)^\theta} \quad (1)$$

この分布は、先行研究のシミュレーションにおいても前提とされているデータ要求頻度の分布を表す一般的な分布のモデルである^{1),5),7),11)}。係数 θ により、要求頻度の偏在を係数として表し易い利点もある。本論文では $\theta=1$ の場合を例として用いる。また、 $m=1000$ を用いる。この分布は、事前に判明してしており、動的に変化しないものとする。また同様に q_i を定義し、ある時点においてページ i を継続して受信しているクライアント数とする。この継続受信者数の分布は本論文では p_i に比例するものとする。クライアントのデータ要求の発生は一定の確率でランダムであるとする。従って、その要求発生の確率分布はポアソン分布に従う。

- プル型通信において、出力待ち時間を除くサーバの処理時間と要求の上り通信に要する時間はそれぞれ $1t_s$ であるとする。

3. 出力データ固定モデル

3.1 出力データのスケジュール方法

本章では、放送されるデータの内容が変わらない場合について検討する。クライアントは必要とするページを1回受信すればそのページの受信は終了して良い。クライアント利用者はできるかぎり早く目的データを手に入れたい。つまり、このモデルにおいては、クライアント利用者がデータ（ページ）を欲してから受信するまでの待ち時間を短くすることが、設計の評価指標となる。この要件におけるデータの提供システムを設計するため、プッシュ型通信とプル型通信をどのように利用すべきか検討する。

プッシュ型通信：クライアント利用者は受信を開始するとサーバからそのデータが出力されるまで待つこ

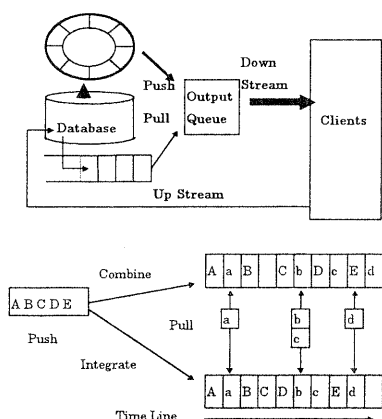


図 1 プッシュ型通信とプル型通信
Fig. 1 Push-based and pull-based communication

となる。平均待ち時間は、出力周期の半分である。当該ページを要求するクライアント数が多くても、上り通信のために帯域を割り当てる必要がない。

プル型通信：クライアント利用者はクライアントを通してサーバへ目的ページを要求する。サーバは要求のあったページを出力する。ページ要求頻度が小さい場合は、サーバはただちに目的ページを出力できる確率が大いので、クライアント利用者の平均待ち時間が小さいが、ページ要求頻度が大きい場合は目的ページの出力に待ち行列ができ、平均待ち時間が大きくなる。また、要求をクライアントからサーバに伝えるための上り通信の帯域が必要である。(図 1 参照)

サーバのデータ出力スケジュール手法を挙げる。

- サーバはプッシュ型通信で各ページを平等に順次出力する。(プッシュ型 FLAT 法)

- サーバはプッシュ型通信でデータを出力する。各ページの出力配分の比率は各ページの要求頻度の平方根に比例させるものとする。この方法は、Wong らにより待ち時間を最も小さくするプッシュ型スケジューリング手法として証明されている^{1),5),6)}。(プッシュ型 SQ [Square root] 法)

- サーバはプル型通信でデータを提供する。クライアントからページの要求があれば、サーバは待ち行列を作って最も古い要求ページから順次処理する。クライアント利用者は要求したページに対する応答を待つ。(プル型個別処理法：QUEUE 法)

- サーバはプル型通信でデータを提供する。クライアントからページの要求があったとき、サーバが先

着順に要求ページを処理するのは QUEUE 法と同じであるが、既にそのページが出力待ちであれば、要求を無視する。クライアント利用者は要求の有無にかかわらず、目的とするページのデータを受信すれば待ちを完了する。(プル型先着順法：FIFO [First In First Out] 法)

- サーバはプル型通信でデータを提供する。当該ページに関し、個々の要求の合計待ち時間を計算する。合計待ち時間値が最も大きいページを選択して出力する。クライアント利用者は目的とするページのデータを受信すれば待ちを完了する。(合計待ち時間累計法：LWF [Longest Wait First] 法)

3.2 各スケジューリング法比較

まず、プル型通信における上り通信の帯域を無視する場合を評価する。各ページの要求頻度は偏りがあるものとする (Zipf 分布において $\theta = 1.0$)。各方法におけるクライアントが目的ページを得るまでの平均待ち時間を、シミュレーションにより求める。 $\rho = 0.2 \sim 5.0$ まで変化させた結果を図 2 に示す。プッシュ型の FLAT 法は平均ページ要求頻度にかかわらず平均待ち時間は一定である。プッシュ型 SQ 法は各ページの要求頻度に応じて出力比率を定めるため、FLAT 法より平均待ち時間が大きく下がる。プル型 QUEUE 法の場合は、M/D/1 型の待ち行列に帰着し、待ち時間は $\frac{\rho}{2(1-\rho)}ts$ となるので、 $\rho \geq 1$ の場合は待ち時間が発散してしまう。QUEUE 法は個別のクライアントと 1 対 1 の通信を行うときのみ必要な考え方であるが、同時に全クライアントに同じ応答を返せる無線環境では、 $\rho \geq 1$ の場合でも有限時間で処理可能である。FIFO 法は、ページ要求頻度が極限まで増大すれば全てのページを順次出力することになるので、FLAT 法と同じとなる。LWF 法は、Wong より、プル型通信の他の手法より待ち時間が小さいことが紹介されている¹⁾。図 2 においても FIFO 法よりも、良い結果を示している。この結果から、LWF 法が、平均ページ要求頻度にかかわらず、各手法のなかでもっとも待ち時間が小さい手法であることを示している。あるタイミングにおいて、ページ i における前回の出力からの時間を T_i とすると、LWF 法における平均の合計待ち時間は、発生要求数と個々の要求の平均待ち時間はそれぞれ T_i に比例するため、 $p_i \cdot T_i^2$ に比例する。総待ち時間累計の最大値の選択を進めると、出力配分は p_i の平方根に比例に近似することになり、プッシュ型通信における SQ 法に収束する。従って、プル型通信の LWF 方式においてページ要求頻度が極限に増大すれば、プッシュ型通信の SQ 方式と同じとなる。小さい

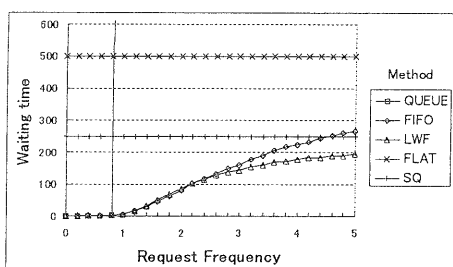


図2 各方式による平均待ち時間 ($\theta = 1.0$)
Fig. 2 Mean waiting time with each method ($\theta = 1.0$)

要求頻度の条件では、FLAT 法より FIFO 法、SQ 法より LWF 法のプル型通信が優れている。

ところで、サーバからの毎出力タイミングにおけるページ選択の演算負荷を考慮すると、プッシュ型通信のスケジュール法において、FLAT 法は順次出力ページを選択することが可能であるが、SQ 法の場合は候補ページ選択のための計算が必要である。しかし、出力配分比率が動的に変化しないので、効率をあまり落とさずに簡易的に選択計算を行うことができる^{5),11)}。プル型通信における FIFO 法は単に先着順に処理すれば良いのでページ選択が容易である。しかし、LWF 法の場合は毎出力タイミングごとに合計の待ち時間が動的に変化するので、待ち時間累計最大のページを選択するための計算が複雑である。対象となるページが多くページ選択のための処理可能時間が短ければ、大きな負担となる。Aksoy らによる簡易計算法も提案されている¹⁰⁾が、計算負荷や性能面で十分に軽減されているとは言えない。このため、LWF 法の採用を見送ると、プッシュ型通信における SQ 法とプル型通信における FIFO 法が待ち時間減少スケジュール法の候補となる。図3に示すように、平均ページ要求頻度が大きい場合は SQ 法が、小さい場合は FIFO 法が優れている。

3.3 プッシュ型通信とプル型通信の併用方式

要求頻度により、適切なスケジュール方法が異なるのであれば、ページの要求頻度に応じたスケジュール手法を併用することが考えられる。平均要求頻度が大きい順に各ページを並べ、平均要求頻度が大きいページをプッシュ型通信 (SQ 法) で、小さいページをプル型通信 (FIFO 法) で送信する。下り通信の帯域をプッシュ型通信のタイムスロットとプル型通信のタイムスロットに分ける。図4は、平均ページ要求頻度

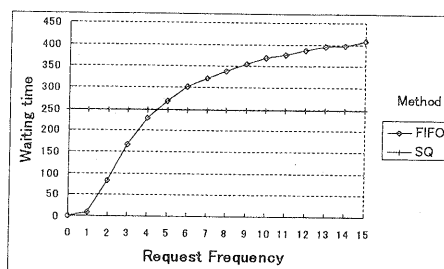


図3 SQ 法と FIFO 法の待ち時間の比較
Fig. 3 Comparison of SQ method with FIFO method

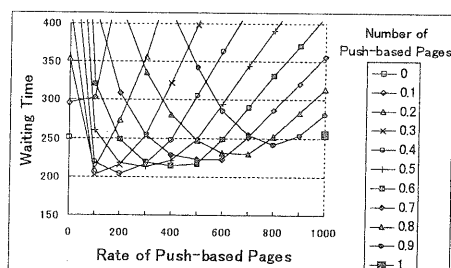


図4 プッシュ型通信 (SQ 法) とプル型通信 (FIFO 法) との併用法による待ち時間

Fig. 4 Waiting time with combined use of SQ method and FIFO method

が 4.5 の場合に、プッシュ型出力ページ数を 100 ページ単位に 0~1000 (プル型出力ページ数は 1000~0) に変化させ、プッシュ型通信とプル型通信の出力配分比率を 0.1 単位に 0~1 に変化した場合の平均待ち時間を示したものである。プッシュ型通信のページ数を 100 ページ (プル型通信は 900 ページ) とし、プッシュ型通信とプル型通信の出力帯域の配分を 3 対 7 とした場合に待ち時間が試験ケースのなかではもっとも小さいことを示している。

3.4 プル型通信における上り通信の負荷の影響

3.2 節の検討ではプル型通信における上り通信の負荷については無視した。しかし、現実には上り通信の通信負荷を無視できない場合がある。また、上り通信の帯域自体を無視することができても、大きい頻度で出力されるデータをクライアントがわざわざ要求するメッセージを発信することは、クライアントにとって余分な負担となり、避けることが望ましい⁹⁾。そのた

め、上り通信の負担を帯域負荷という形で表し、プル型通信における上り通信の負荷の影響について検討する。単位時間あたりの要求メッセージ数から上り通信に必要な帯域を確保し、残りの帯域を下り通信に配分するものとする。クライアントがプル型通信で要求が発生するとき、上り通信の時間帯を確保してサーバに要求を上げる。厳密には同時に複数のクライアントからの要求の衝突が発生する事象など上り通信についても解析が必要であるが、本論文の検討範囲を絞るため、簡略的に衝突による上り通信帯域の損失も含めて平均的に1つの要求メッセージを送信するのに必要な時間を一定と仮定し、その時間を1サブタイムスロットと呼ぶこととし、 $subts$ で示す。上り通信に必要な帯域を *pullband*、プッシュ型で送信するページ数を $xpush$ となると

$$pullband = \sum_{i=xpush+1}^m p_i \cdot subts < ts \quad (2)$$

でなければ、プル型通信を行うことはできない。また、このとき、1ページの出力可能時間あたりの要求頻度は $pullband / subts$ となる。さらに、プッシュ型のページ i が $period_i$ ページ毎に出力されるとすると、出力周期は $ts \cdot period_i / (ts - pullband)$ となる。図5は ts に対する $subts$ の比が0, 0.2, 0.5, 1の各々についてプル型のFIFO法とプッシュ型のSQ法でデータ提供を行った場合の待ち時間をシミュレーションにより算出した結果を示したものである。 ts と $subts$ との比を除き、実験条件は図2の場合と同じである。 ts に対する $subts$ の比が小さいことは、ページ要求の上り通信メッセージ長に対して下り通信のページ長が大きいことを示し、 ts に対する $subts$ の比が大きいことはその逆を示す。図より、 $subts$ が無視できない場合はプル型通信でスケジュール可能な平均要求頻度には限界があることを示している。また、 ts に対する $subts$ の比が大きくなるにつれて、プッシュ型通信 (SQ法) とプル型通信 (FIFO法) のいずれが有利であるかの分界点が、要求頻度の小さい方向に向かっている。

3.5 プッシュ型通信とプル型通信の動的統合方式

プッシュ型通信とプル型通信の併用方式が有効な場合で、プル型通信出力の平均要求頻度が出力帯域全体のページ出力処理能力よりある程度小さい場合は、プッシュ型通信とプル型通信の動的統合方式が有効である。併用方式がプッシュ型通信とプル型通信の帯域を予め固定的に割り付け、各々の通信はその帯域内で出力スケジュールを定めるのに対し、動的統合方式は、プル型通信に割り付けられたページの出力要求が存在

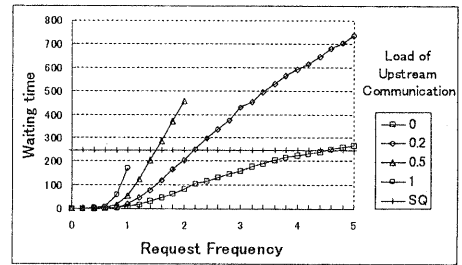


図5 上り通信の負荷を考慮した場合の待ち時間 (FIFO 法 / SQ 法)

Fig. 5 nWaiting time with upstream communication load (FIFO and SQ methods)

する場合はプル型通信のページを出力し、プル型通信の出力ページがなくなればプッシュ型通信のページを出力するものである。図6は ts と $subts$ の比が1:0.5で、 $1ts$ あたりの要求頻度が1.5とした条件下におけるプッシュ型通信とプル型通信の併用方式 (Combine) と動的統合方式 (Integrate) の待ち時間を比較したものである。併用方式の場合は0.1単位でプッシュ型通信の帯域とプル型通信の帯域を割り付けた場合を試算し、プッシュ型通信とプル型通信のページ数の比を100ページ単位で変化させ、そのなかの最適値を選んでいる。図から動的統合方式が併用方式よりも効率が良いことを示している。この理由は、併用方式の場合は、与えられたプル型出力の帯域を利用してプル型通信の出力を行うのに対し、動的統合方式の場合は同じ要求頻度であっても、全帯域を利用してプル型通信の出力を行うため、待ち行列の平均的長さが短くなるためである。一方プッシュ型通信の帯域はプル型通信に必要な帯域の残りであり、併用方式と同様の帯域確保が期待できるため、待ち時間の長さはほとんど変わらない。ただし、プル型通信出力の平均要求頻度が出力帯域全体の能力を越えるとき、併用方式の場合は、全ての要求に個別に対応することをせず、複数の要求に対してまとめて応答する形となるが、動的統合方式の場合は帯域全てをプル型通信出力に使い切ってしまう。このため、動的統合方式はプル型通信方式と同じになってしまう。動的統合方式が併用方式より効果を上げるためには、プル型通信ページの要求がある程度小さくなければならない。

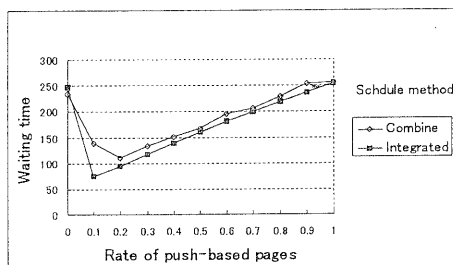


図 6 プッシュ型通信とプル型通信の併用法と動的統合法の比較
Fig. 6 Comparison of combined use with integrated use in push-pull communication

4. 出力データ可変モデル

4.1 モデルの概要

サーバからクライアントに提供するデータが更新される場合を検討する。サーバはデータの中継局の役割を果たす。サーバは外部の情報源からオンラインないしはオフラインで最新データを入手し、そのデータを域内のクライアントに放送する。更新されるデータを受信するクライアントの対応には以下の3種類があり、それが混在することが考えられる。

- クライアント利用者がデータを得ることを欲したときから、クライアントは受信を開始し、利用者は当該のデータを得るまで受信を待ち受ける。目的データを得ればクライアントは受信を打ち切る。この場合、個々のクライアントの待ち時間は、3章で述べた出力データ固定モデルの場合と変わらない。
- 最初の受信開始のとき、クライアント利用者は、当該のデータを得るまで受信を続けて、待ち受ける。目的データを得れば、クライアント利用者は待ち受けを中断する。しかしクライアントはその後も、そのページに関する最新データを入手するため、継続して受信を行ない、自己のローカルファイルに受信したデータを格納しておく。クライアント利用者は、データを利用するときにローカルファイルから読み出して参照する。
- クライアントは常時受信状態にしておいて受信した最新データの更新登録を行う。クライアント利用者は随時登録データの参照を行う。

このモデルにおいて、クライアント利用者に望ましい評価要素として2つある。ひとつは、3章で論じたのと同様に待ち受け開始から目的データ受信までの待ち

時間である。上り通信の帯域に余裕があれば、プル型通信により、クライアント利用者の意向を正確に反映させて出力データを選択するのが望ましい。もうひとつは、クライアント利用者がクライアントのローカルファイルを参照したとき、最新データが記録されているか否かである。サーバが最新データを得ていても、それが出力されておらずクライアントのローカルファイルに古いデータが記録されたままの状態のときに、クライアント利用者がローカルファイルの参照を行えば、クライアント利用者は古い情報しか得られないことになる。更新されたデータは速やかに出力されることが望ましく、全体の利益を考えれば、参照頻度の大きいデータは早く出力することが望ましい。また、クライアントからのデータ要求に拘わらず、データ更新を契機としてデータをクライアントに伝送するのであるから、サーバはプッシュ型通信でデータを伝えれば良い。

この2つの評価要素は同列に論じられない。待ち時間を短くするか遅延時間を短くするかいずれを重視するかは、利用者の価値判断により異なる。常時受信状態のクライアント数と最初の1回めのデータ受信を行おうとしているクライアント数の比によっても異なる。しかし、ここではあえていくつかの評価係数を設定して、両者のバランスをとるプッシュ型通信とプル型通信の動的統合方式を検討する。

プッシュ型通信ページのデータ更新の形態として、更新がランダムに発生する場合と定期的に発生する場合とが考えられる。このうちデータ更新発生がランダムの場合は、データ更新発生によるページ要求をある評価係数で換算したクライアント数からのページ要求とみなせば、ランダムなタイミングでプル型通信によるページ要求に対して、最適ページ出力のスケジュールを設定する課題に帰着する。この課題は、3.2節で述べたように計算負荷を無視すればLWF法が優れていることが判明している。

データ更新が定期的に発生する場合で、最も遅延が少なく効率的なのは、データ更新とプッシュ型通信が同期する場合である。例えば、親局から送られてくるデータを子放送局（サーバ）が中継し、その領域に放送するようなケースである。しかし、場合により更新データが間欠的に発生するケースがある。サーバの複数のページが同時に更新される。サーバが親局から定期的に、放送よりも広い帯域やオフライン媒体でデータを受信したり、サーバが個別に収集したデータを単位時間ごとに編集して、新しい情報としてクライアントに出力したりする場合である。例えば、証券取引所

から株価情報を高速回線でバッチ処理として定期的にサーバに送付してくる。サーバは低速の無線放送にて域内のクライアントに株価データを放送するといったケースである。本論文では典型例として、定期的に全ページが一斉に更新される場合を想定する。更新データのみを考えた場合、当該ページを受信しているクライアントの数が多きページから出力すれば、クライアント利用者が最新データを参照する確率が高くなる。定期的に各ページの更新データが外部システムよりサーバに送られてくる間隔を *cycle* とする。クライアントが自己のローカルファイルを参照したときに最新データが得られる確率を *newrate* とすると以下の式 (3) で得られる。

$$\text{newrate} = 1 - \frac{\sum_{i=1}^m k_i \cdot q_k}{\text{cycle} \cdot \sum_{i=1}^m q_k} \quad (3)$$

出力帯域はできるだけ広いほど、クライアントは最新データを得られる確率が高くなる。

4.2 モデルの評価

クライアントは最初に求めるページを受信するために、そのページの出力要求をサーバに対して行い、サーバは要求に対応してページ出力をプル型通信で行う。サーバは、データ更新されたページを出力する場合はプッシュ型通信で行う。このようなモデルを想定して、出力データが可変な場合のプッシュ型通信とプル型通信の動的統合モデルについて検討する。3章に述べたように最初にデータを受信する場合でもプル型通信のみでなく、意識的にプッシュ型出力を併用した場合が効率的であるケースも存在するが、省略する。全ページ数を 1000 ページとし、各ページの一斉更新間隔を 2000 ts とする。一斉更新のデータを受け取ったサーバは、プッシュ型通信の出力時間帯に受信者数の多いページから順次出力を行う。プル型出力の各ページの平均要求頻度は 0.2 / ts とし、 $\theta = 1$ の Zipf 分布に従うとする。また、各ページの更新データを受信するクライアントの比率はプル型出力の各ページの平均要求頻度に比例するものとする。まず、出力帯域をプッシュ型通信とプル型通信に固定的に分離し、各々の帯域で対応する通信方式を利用する併用型を試算する。プッシュ型の帯域の比率を 0.5 から 0.95、プル型の比率を 0.05 から 0.5 に変化させる。この条件にてシミュレーションを行い、プッシュ型通信の遅延時間とプル型通信の待ち時間を求める。サーバのデータの一斉更新後、あるページについてプル型通信で出力された時刻がプッシュ型通信で出力される時刻よりも前であれば、そのページの遅延時間はデータ更新時からプル型通信のページ出力までの時間とみなす。プッシュ

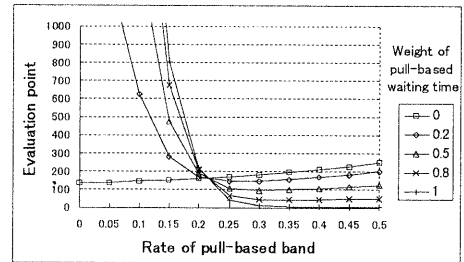


図 7 プッシュ型通信とプル型通信の併用方式による時間評価点
Fig. 7 Evaluation points by combined use of push and pull based communication

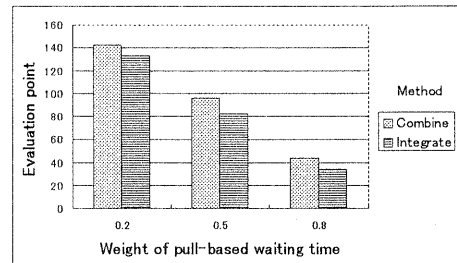


図 8 各方式による時間評価点 ($\rho = 0.2$)
Fig. 8 Evaluation points by each methods ($\rho = 0.2$)

型通信で出力されたページについても、そのページを初期受信のため待ち受けているクライアント利用者にとっては、初期受信が完了したことになる。このプッシュ型通信／プル型通信の遅延時間／待ち時間の結果について、それぞれ重みを与えて加えた時間を評価係数として算出する。重みの与え方により最適配分比率は異なる。(図 7 参照)

出力データ可変モデルにおいても、固定モデルの場合と同様に動的帯域統合方式の適用を考える。図 8 は図 7 における最も良いプッシュ型通信とプル型通信の帯域比の場合と動的統合方式を適用した場合を比較したものである。いずれの場合も動的統合方式が併用方式より評価時間が短く、優れていることを示している。

5. 関連研究

Wong は、プッシュ型通信の最適スケジューリング法として、各ページをクライアントの要求頻度の平方根に比例させて出力配分することが望ましいこと、プル型

通信のスケジューリング法として LWF 法が待ち時間を短かくすることを示している。さらに両者を組み合わせたハイブリッド法にも言及している¹⁾。しかし、ハイブリッド法はディスクからのデータ読み出しと出力を意識し過ぎているためか、プッシュ型を重視して柔軟性に欠け、組み合わせによる待ち時間や遅延時間の効率向上に結びついていない。Imielinski もモバイル通信に関する論文²⁾のなかでプッシュ型通信とプル型通信の組み合わせについて言及しているが、定量的な分析や本論文に示した動的統合方法については述べられていない。箱守らはプッシュ型通信とプル型通信を組み合わせたデータの提供方法と最適組み合わせ方法について検討している^{4),8)}が、本論文の用語でいう併用方式に過ぎず動的な統合方式について論じていない。また、プル型通信について全要求について順次処理するプル型個別処理法(3.1 節及び 3.2 節参照)で解析しているため、個々の要求数が出力帯域を超えれば出力不能となり、現実と合わない。本論文においては、同じページの要求が累積すれば、ひとつの出力でまとめて応答することを意識している。また、プッシュ型通信とプル型通信の組み合わせ出力について Achara らの研究⁷⁾があるが、キャッシュに関連するものであり、本論文で述べたデータのバッチ更新と更新データを速やかにクライアントに伝える方法とは視点が異なる。

6. ま と め

サーバからクライアントにデータを提供する方法としてプッシュ型通信とプル型通信がある。無線を用いてサーバから無線領域内のクライアントにデータを提供するとき、プッシュ型通信およびプル型の下り通信のいずれも放送として全クライアントに同じデータを同時に提供できる。サーバから提供データの内容が変化しない場合において、上り通信の負荷を考慮しないときは、理論的にプル型通信はプッシュ型通信よりも待ち時間が小さいスケジューリングを組むことが可能である。しかし、プル型通信における最適スケジューリングを計算するための演算負荷、上り通信の帯域の必要性や上り通信の負荷を無視できないときは、プッシュ型通信が有効となる場合もある。両者の通信方式を最適な割合で組み合わせる併用方式を採用すれば、それぞれ単独の通信方式を採用するよりも全体としての待ち時間の効率が上がる。さらに、クライアントからのページ要求頻度がある範囲に入る場合に限られるが、動的にプッシュ型通信とプル型通信を統合する方式が併用方式よりもさらに効果を上げる。

また、サーバのデータがまとめて更新される場合、

評価要素としてサーバにおける更新から出力までの遅延時間とクライアントにおける待ち受け開始から受信までの待ち時間の両方があり、両者のいずれを重視するかにより結果は一律でないが、一般に動的統合方式が優れていることを示した。

参 考 文 献

- 1) Wong, J. H., Broadcast Delivery, Proceedings of the IEEE, VOL.76, No.12, December 1988 pp1566-1577 (1988).
- 2) Imielinski, T. and Viswanathan, S., "Adaptive wireless information system", Proc. of SIGDBS(Special Interest Group in DataBase Systems) Conference, pp.19-41, Tokyo Japan, October (1994).
- 3) Franklin, M. and Zdonik, S., "Dissemination-based information systems", IEEE Data Engineering Bulletin, Vol.19, No.3, September (1996).
- 4) 田辺雅則, 箱守聡, 井上潮, モバイル環境における放送とオンデマンドを組み合わせた情報提供方式, 情報研報 Vol.96 No.MBL-1 (1996.11).
- 5) Vaidya, N. H. and Hameed, S., "Scheduling data broadcast in asymmetric communication", Proc. of Workshop on Satellite-based Information Services (WOSBIS), New York November (1996).
- 6) Chi-Jiun and Tassiulas, L., "Broadcast Scheduling for information distribution" IEEE INFOCOM Kobe Japan (1997).
- 7) Acharya S., Franklin, M. and Zdonik S., "Balancing Push and Pull for Data Broadcast", Proceedings of ACM SIGMOD conference, Tucson, Arizona, May, (1997).
- 8) 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 情報処理学会 DoCoMo ワークショップ pp. 55-60 July (1997).
- 9) Stathatos, K, Roussopoulos, N. and Baras, J. S., Adaptive Data Broadcast in Hybrid Networks, Proceedings of the 23rd VLDB Conference Athens Greece (1997).
- 10) Aksoy, D. and Franklin, M., Scheduling for Large-Scale On-Demand Data Broadcasting, Source: Proc. IEEE INFOCOM Conf. San Francisco CA, March (1998).
- 11) 青野正宏, 田窪昭夫, 渡辺尚, 水野忠則, "データ放送におけるスケジューリング決定法「二重循環法」の提案と評価, 情報処理学会論文誌 March (1999)