

放送型配信機構上での並行処理制御方式：山彦

Concurrency Control Schemes by Broadcasting Mechanism : ECHO

白田 由香利* *** 飯沢 篤志* ** 矢野 隆志**

Yukari SHIROTA Atsushi IIZAWA Takashi YANO

要 旨

大規模なデータベースシステムを構築する場合、利用者からのアクセスを効率良く行うためには超多地点にデータ複製を作ることが有効であるが、従来の通信方式だけでは複製間の一貫性管理コストが大き過ぎてしまい実現が困難である。そこで放送型配信機構を用いた複製間の並行処理制御方式「山彦」を提案する。山彦方式は、サイト数が増えてても配信コストが変わらないという利点に加え、降雨等の影響で放送受信に失敗しても通信で補完する機能を備えている。

ABSTRACT

When constructing a large-scale database system, data replication is useful to give users efficient access to the data. However, using traditional networking concurrency control protocols makes the communication too costly for practical use. Therefore, we have proposed the concurrency control schemes using broadcasting mechanism called ECHO for the concurrency control of these replica data. The advantage of the ECHO mechanism is that the communication cost is same, regardless of the number of sites. The broadcast can be received in an ordinary way. However, weather hazard such as a heavy rain, may disturb to receive the broadcast. The ECHO schemes have a mechanism to ensure stable reception through communication network.

* (株)次世代情報放送システム研究所

Information Broadcasting Laboratories, Inc.

(株)リコーより(株)次世代情報放送システム研究所へ兼任出向中。

** 研究開発本部 ソフトウェア研究所

Software Research Center, Research and Development Group

* (株)次世代情報放送システム研究所は、基盤技術研究促進センターの事業のひとつとして、基盤技術研究促進センター、ソニー(株)、松下電器産業(株)、(株)リコー、日本放送協会、日本テレビ放送網(株)、(株)フジテレビジョンの出資により平成9年2月に設立されたもので、次世代の放送情報放送(IB: Information Broadcasting)の研究開発を行っている。

1. 背景と目的

放送のデジタル化により、従来のような番組のみの放送に加え、コンピュータデータのようなデジタル情報を多重化して送ろうとする動きが高まりつつある。放送という公共性の高いメディアを利用して国民規模の利用者に提供するためには、個々の情報を相互に関連づけ、構造化することが必要となる。我々は高度に構造化された情報により提供される次世代の放送サービスを「情報放送(IB:Information Broadcasting)」として位置づけ、これを実現するための研究開発を行なっている¹⁾。

従来、放送はストリーム状の情報をリアルタイムで視聴するという形態が殆どであったが、これから情報放送システムでは、放送配信されたコンテンツをデータベース(以下ではDBと略す)に一度蓄積してから閲覧する、という利用が増加してくると予想される。こうした蓄積型の閲覧形態としては、(1)利用者の受信機に蓄積された情報を閲覧する方式、(2)外部DBデータベースサーバに蓄積された情報を閲覧する方式の2種類が想定される。

Fig.1に情報放送システムのイメージを示す。図中の外部DBは情報放送のインフラの一部として管理される。このようなDBを用意することで、家庭での受信機による蓄積を補完するとともに、情報の共有化が可能となる²⁾。図の中の中継局にあるDBは放送配信されたコンテンツをキャッシュし、システム全体のアクセス効率向上のために使われる。各種DBの数は図中に示すように非常に大きいことを想定している。

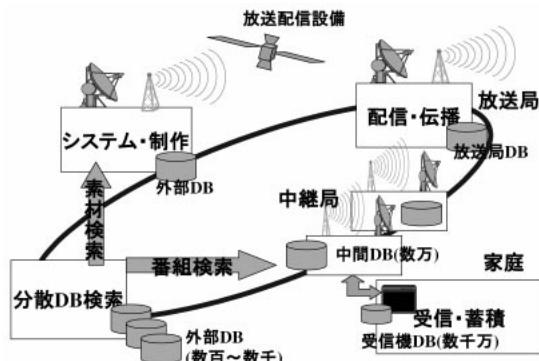


Fig.1 An Image of the Information Broadcasting System

外部DBは受信機からのアクセス要求に応じ、他の外部DB上及び中継局DB上にコンテンツの複製を作り、その最適配置を行なうことで、システム全体としての効率的処理につとめる。こうしたデータ複製により、(a)信頼性の向上、(b)アクセス可能性の向上、(c)性能の向上、というメリットが得られる。しかしデータ複製が数千万サイト以上の超多地点にわたる場合、従来

の2相コミットプロトコルなどでは複製間一貫性保持のための通信コストが非常に大きくなり実用上問題がある。そこで我々は放送型配信を活用した、超多地点に分散した複製に対する並行処理制御方式「山彦」を考案した^{3,4)}。本稿では山彦方式について説明する。

2. 山彦方式

2-1 想定する応用

山彦方式が想定する応用は前述した情報放送サービスである。以下に配信されるコンテンツとその更新例をあげる。

- (1) 加入者名簿情報：各種の名簿情報の変化による更新。
- (2) 全国アパート空き部屋情報：仮押さえ、契約などによる状況の更新。
- (3) タウン情報誌：映画館の上映スケジュールの更新、映画観客数の人気ランキングの更新、映画館の混雑度状況(立ち見か)の更新、演劇のキャスティングの変更、各種イベントの内容や開催時間の変更など。

こうした更新を自ら発することができるDBサイトは外部DBサイトに限定されるが、結果の更新情報は全DBで受信可能である。更新頻度はOLTP(Online Transaction Processing)ほど多くはないが、かなり大きいと予想される。また受信機数が非常に大きいので、受信機から通信を用いて外部DBあるいは中継局DBへreadアクセスする頻度は非常に大きくなる、と予想される。

複数のトランザクションが同時に起こると更新が競合する可能性があるので、並行処理制御が必要となる。例えば、複数の支店DBサイトが更新する可能性のある、全国会員情報の更新の場合である。一般に更新では、一つのオブジェクトのサイズを大きくするほど、競合の可能性が高くなる。

情報放送はデジタル放送システムを基盤とする。デジタル放送の形態としては衛星によるものの他、地上波及びCATVを利用したものがある。山彦方式はどのデジタル放送にも利用できる方式であるが、主には衛星放送を想定している。衛星放送の問題点は、その周波数が12GHzと非常に高いため、降雨、降雪、強風等の気象条件により電波が減衰し受信に影響がでることである⁵⁾。従来のデジタル放送では、減衰により映像画面が乱れるだけで済んだが、情報放送ではデータ配信するので1ビットの伝送誤りも許されない。よって、放送配信の誤り訂正符合の他に上位の応用層で誤りを検出し、再送する手段をとらなくてはならない。こうした問題を解決するため、山彦方式には降雨等の影響で放送受信に失敗しても通信で補完する機能をつけた⁴⁾。

2-2 山彦方式の概要

山彦方式の概要を説明する。Fig.2に想定する環境を示した。システムは、唯一の放送局サイト(放送を行なうサイト)と、複数のDBサイトから構成される。図中には更新トランザクションを発することができるDBサイトのみを示したが、この他に受信のみのDBサイト(中継局及び家庭)が多数存在する。代表サイトとは、指定データの最新値をもつことを義務づけられたDBサイトであり、以下の役割をもつ。

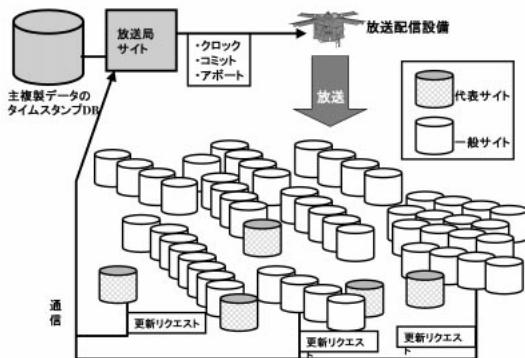


Fig.2 Environment of the ECHO Mechanism

- ・一般サイト(非代表サイト)がデータ最新値をreadしたい場合、readさせてやる。よって受信のみの中継局DBでも、それにより全体の効率が向上すると判断されれば、代表サイトになりうる。
- ・システム全体に、必ずデータの最新値が存在することを保証する。複製が多数あるとお互いに「どこかに最新値が保持されているだろう」と考え、実は最新値が存在しなかった、という可能性がある。それを回避するために代表サイトが役立つ。

放送局サイトのDBでは各データオブジェクトごとに、最終更新タイムスタンプ、最終更新サイトID及びトランザクションID、最新値保有サイトIDの集合、及び代表サイトIDの集合、が格納されている。最新データ自体は蓄積量が大きくなり過ぎるため、放送局DBには蓄積しない。代わりにどこに最新値があるのか、最新値を保有しているサイトIDを管理している。

各DBサイトの更新トランザクションの様子を説明する。山彦方式では、グローバルタイムスタンプを用いて更新の正当性を判定する。各DBサイトで更新トランザクションが発生すると、まずは各DBサイトば他サイトのトランザクションとはデータ競合がないと仮定して、ローカルに処理を進める。その後、コミットの準備ができた時点でコミット・リクエスト・メッセージを通信により放送局サイトへ送る。放送局サイトはリクエスト中のアクセスされたデータをチェックし、全てのデータのタイムスタンプが最新版であった場合、つまり、他のトランザクションにより既に書き換えが行われてい

ない場合にのみ、コミットを許可する。そして、放送局サイトはコミットメッセージを放送によって配信する。コミットメッセージには、更新後のデータの最新値及びタイムスタンプ最新値が含まれる。よって全DBサイトは必要に応じて自分の興味のあるデータの最新値を知り、自サイトの複製を更新することが可能となる。

トランザクションはリクエストを送った後、衛星のような放送配信設備からコミット許可メッセージが配信されるのを待っている。その様子が、山に向かって叫んだ後、返ってくる山彦を待つ様子に似ているので、山彦方式と名付けた。

受信した自分宛のメッセージの内容がアボート、つまりコミット却下の場合を説明する。アボートメッセージには「どのデータが古いためにアボートになったのか」及び「そのデータの最新版はどこのDBサイトに保存されているのか」が含まれているので、アボートを宣告されたトランザクションは、最新版のデータをどこかのDBサイトからコピーすることが可能となる。そしてコピー後、再度、トランザクション処理を行ない、リクエストメッセージを放送局サイトに送る。

放送配信を利用しているので、DBサイトは自分宛のコミットメッセージを待っている間に他のサイト宛のコミットメッセージも受信できる。その時、自分が更新したものと同じデータに対し、他サイトのトランザクションが先に更新を行い、コミットされたことを知ったとする。その場合、自分宛のアボートメッセージを待たずに早めにリトライ処理を開始できる。コミットメッセージには最新値も含まれていることもあり、トランザクションのトータルな応答時間の期待値が短縮できる。

2-3 降雨等による放送受信失敗を補完する機能

山彦方式では、放送と通信の両方を用いている。我々は通信の誤りは、より下位相において回復することが可能なので問題はないと考え、山彦方式におけるDBサイトの放送受信失敗をいかに補完するかのみに注力した。以下この補完機能について説明する。

山彦方式は放送局サイトに負荷が集中する方式なので、その設計ではいかに負荷を分散させるかに注意する必要がある。想定したシステム環境では、更新サイト数が数百～数千に対して、代表サイト数は十～数十、最新値保有サイト数は百～数百というようにかなり限定した。これは放送局サイトへの通信負荷及び蓄積コストを軽減するためである。

DBサイトが自分宛のコミット(アボート)メッセージを受信し損ねた場合を考える。DBサイトはメッセージの到着を待ち、タイムアウトしたとする。DBサイトは放送局サイトに対して、該当するデータの情報を問い合わせる。

合わせる。その結果中に、そのトランザクション自身が更新した痕跡が少なくとも一つでもあればコミットしてよい。

次に、代表サイトが、他サイト宛のコミットメッセージの受信に失敗した場合、つまり、最新値情報を受け損ねた場合を考える(Fig.3参照)。この例では、代表サイトCがサイトA宛のコミットメッセージを受信し損ねたとしてある。次の手順で処理が進む。

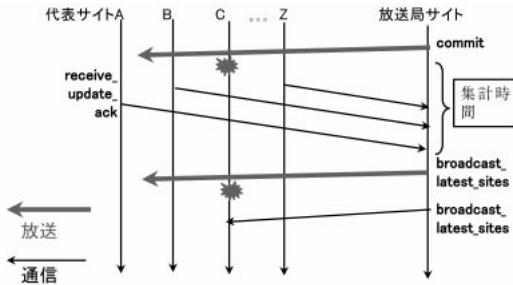


Fig.3 An Example of Collecting a Set of Site IDs that Possess the Latest Version and Broadcasting It.

- (1) 放送局サイトがコミットメッセージを放送する。放送局サイトは「最新値保有サイトID集合」をクリアする。
 - (2) コミットメッセージに含まれる最新値を自分のDBに反映したDBサイトのうち、他のDBサイトにreadさせてもよいと考えるサイト、及び代表サイトはreceive_update_ackを放送局サイトに通信で送る。
 - (3) 放送局サイトはreceive_update_ackを集計し、「最新値保有サイトID集合」に付加する。集計時間は予め決めておくが、締め切りに遅れた場合もDBには付加される。代表サイトのいずれもがackを返さない場合は、集計時間を延長する。これは、システムとして信頼性の高い代表サイトのうち少なくとも一つにはデータ最新値が蓄積されていることを保証するためである。
 - (4) 放送局サイトは締め切りに間に合った最新値保有サイトID集合を放送で配信する(broadcast_latest_sites)。
 - (5) 放送局サイトは、代表サイトのうち最新値保有サイトID集合に含まれないもの、つまり、ackを返さなかつた代表サイトに対して、broadcast_latest_sitesを通信で送る。よって代表サイトが最新情報を受け損ねた場合でも、受信失敗を知ることができ、その代表サイトは最新値を他のサイトに取りにいくことができる。
- 上記のようにして最新値保有サイトを放送あるいは通信により配信することにより、以下の効果が得られる。
- ・DBサイトが最新値を必要としている場合、複数の最新値保有サイトが配信されるので、その中からアクセスコスト最小のDBサイトを選ぶことができる。
 - ・システム中、代表サイトのうち少なくとも一つにはデータ最新値が蓄積されていることが保証できる。
 - ・代表サイトが放送によるコミットメッセージを受信し損なっても、通信で知らせてくれる。

3.まとめ

山彦方式は、以下のような条件の応用で適応すると効果がある。

- ・超多地点サイトで複製を管理
- ・複製へのread-onlyトランザクション頻度が非常に大きい
- ・複製への更新トランザクション頻度はかなりある

山彦方式の利点は次の通り。

- (1) 放送を使っているので、サイト数が増えても実現可能であること(2相コミットのような従来のプロトコルでは通信負荷が大きすぎて実用が困難)。
- (2) 放送を使っているので、コミットまでの応答時間の予測が容易であること。
- (3) 放送を使っているので、他のサイトへのメッセージも受信することが可能。よって、応答時間の期待値が短縮でき、全サイトで必要に応じて最新値を取得可能。
- (4) 降雨等の影響で放送受信に失敗した場合でも、通信補完できると同時に、最新値保有サイトを配信することにより、最新値をコピーするコストが小さくてすむこと。

4.今後の展開

今後はシミュレーションにより本方式の性能評価を行なう予定である。シミュレーションにより、更新頻度及び利用者数がどの程度まで増大しても許容できるか、及び、放送と通信の利用のバランス比をどの程度にした場合全体のシステム効率が最適になるか、を検証したい。また、利用者のアクセスコストを下げるため、複製データの最適配置アルゴリズムについても研究する。

参考文献

- 1) 権野,山岸,木村:情報放送への技術展望,映像情報 メディア学会年次大会予行集(1998).
- 2) 浅田,飯沢:情報放送における情報配信及び検索方式の検討,映像情報メディア学会年次大会予行集(1998).
- 3) 白田,飯沢,矢野:放送型配信機構上の並行処理制御方式:山彦,アドバンスト・データベース・シンポジウム(ADBS)97論文集(1997)pp. 15-22.
- 4) 白田,飯沢,矢野:放送配信機構上の並行処理制御方式「山彦」の補完機構,情報処理学会研究会報告,DBS115-15(1998)pp.111-118,1998年5月.
- 5) 山田:デジタル放送のさらなる発展をめざして,NHK技研R&D, No.4(1997)pp.17-40.