

## 広く分散した動的な情報のための情報配信手法の構想

八木 哲†

NTT サイバースペース研究所†

広く分散した多数の情報源において頻繁に変化する情報を観測し、その最新の情報を広く分散した多数の利用者に提供するための、我々が提案している情報共有手法について、これを補完する情報配信手法を提案する。情報共有手法の課題は次のとおりである。(a) 広く分散した多数の情報源に起因する多量の更新処理。(b) 広く分散した多数の利用者に起因する多量の参照処理。(c) 頻繁に変化する情報に起因する高い頻度の更新処理。(d) 最新情報の共有。このような情報共有手法を補完する情報配信手法の課題は、拡張性・信頼性・即時性である。本稿では、拡張性・信頼性・即時性を備えた情報配信手法として、階層構造と冗長な経路とを特徴とするアプリケーション・レイヤ・マルチキャスト方式を提案する。

## A Method of Delivering Dynamic and Globally Distributed Information

Satoru Yagi†

NTT Cyber Space Laboratories†

This paper proposes a method of distributing information, which complements our method of sharing dynamic information observed on globally distributed information sources among globally distributed users in real time. There are four issues concerning our method of sharing information : (a) the volume of operations for updating by a large number of globally distributed information sources; (b) the volume of operations for retrieval by a large number of globally distributed users; (c) the frequency of operations for updating by dynamic information and (d) information sharing in real time. Accordingly, issues concerning a method of distributing information are scalability, reliability and real-timeness. For such a method of distributing information, this paper describes the application layer multicast which has hierarchical topology and redundant path.

### 1 はじめに

広く分散した多数の情報源において頻繁に変化する情報を観測し、その最新の情報を広く分散した多数の利用者に提供できる仕組みがあれば、より利便性の高い広域情報案内サービスを構築できる。情報源は、現実世界において発生する事象や g コンテンツ<sup>1)</sup>を配信するサーバなどである。たとえば、道路交通情報・地域情報・気象環境情報を提供できる仕組みがあれば、デジタルシティ<sup>2)</sup>・カーテレマティクス・歩行者 ITS を融合させたような、計画の立案からドア・ツー・ドアの行動までを支援するサービスや防災支援サービスを構築できる。ネットワーク情報・計算機資源情報を提供できる仕組みがあれば、広域ネットワーク

監視サービス<sup>3)</sup>やグローバルコンピューティング向けの計算機資源管理サービス<sup>4)</sup>を構築できる。このような仕組みを情報共有基盤と呼ぶことにする。情報共有基盤を実現するには、次の課題に対処できる情報共有手法が求められる。

- (a) 広く分散した多数の情報源に起因する多量の更新処理。
- (b) 広く分散した多数の利用者に起因する多量の参照処理。
- (c) 頻繁に変化する情報に起因する高い頻度の更新処理。
- (d) 最新情報の共有。

これに対して我々は、次の特徴を持つ情報共有手法を提案している<sup>5), 6)</sup>。「利用者は情報の詳

細を段階的に参照する」という参照パターンを前提条件として、(1) 自律型モジュールを分散配置し、配置した地域で観測された情報の管理を担当させる。(2) 自律型モジュールは、観測された情報から詳しさを異る複数の情報を動的に生成する。(3) 自律型モジュールは、利用頻度の低い情報を相互に参照して利用者に提供し、利用頻度の高い情報を相互に複製して利用者に提供する。これらにより、参照処理と更新処理を各自律型モジュールに分散・局所化し、課題(a), (b), (c), (d)に対処する。

このような情報共有手法では、自律型モジュールの間で情報を複製するための情報配信手法が重要になる。情報共有手法の課題から、次の課題に対処できる情報配信手法が求められる。

- 課題(a), (b)のために拡張性。
- 課題(c), (d)のために信頼性と即時性。

本稿では、拡張性・信頼性・即時性を備えた情報配信手法として、階層構造と冗長な経路とを特徴とするアプリケーション・レイヤ・マルチキャスト方式を提案する。まず、我々の情報共有手法の概要を説明する(2章)。次に、情報配信手法を提案する(3章)。最後に、本稿の内容をまとめ、今後の課題を示す(4章)。

## 2 情報共有手法の概要

前提条件とする参照パターンの詳細を次に示す。

- 各利用者は、第1段階では大多数の情報の概観を参照する。第2段階では興味のある情報の詳細を参照する。
- 各情報は、第1段階では大多数の利用者から概観を参照される。第2段階では興味を抱いた利用者から詳細を参照される。

これに基づいて、我々の情報共有手法の枠組みを説明する。参照パターンの第1段階では、情報の概観が参照の対象である。各利用者が網羅的に参照するために、各情報の参照頻度は比較的高い。そこで、情報源において観測された頻に変化する情報から、次の特徴を持つ情報を動的に生成し、これを利用者に提供する。

1. 観測した情報の特徴だけを反映する。
2. 観測した情報と比較して時間的変化が少ない。

この情報を「抽象化した情報」と呼ぶ。利用者のいる地域ごとに「抽象化した情報」の複製を配置することにより、「抽象化した情報」に対する参照処理を分散・局所化する。「抽象化した情報」の前記1, 2の特徴により、「抽象化した情報」の複製に対する更新処理の頻度を削減する。参照パターンの第2段階では、情報の詳細が参照の対象である。各利用者が選択的に参照するために、各情報の参照頻度は比較的低い。そこで、情報源において観測された頻に変化する情報を、そのまま利用者に提供する。この情報を「素な情報」と呼ぶ。情報源のある地域ごとに「素な情報」を個別管理することにより、「素な情報」に対する更新処理を分散・局所化する。「抽象化した情報」を索引として利用者に提供し、参照する「素な情報」を利用者に選択させることにより、「素な情報」に対する参照処理を分散させる。すなわち、課題(a)のために、自律型モジュールを分散配置し、配置した地域の「素な情報」の管理を担当させる。これにより、「素な情報」に対する更新処理を分散・局所化する。課題(b)のために、自律型モジュールは、「素な情報」から「抽象化した情報」を動的に生成し、これを相互に複製して、各々が保有する「素な情報」への索引として利用する。これにより、「抽象化した情報」に対する参照処理を分散・局所化し、「素な情報」に対する参照処理を分散させる。課題(c), (d)のために、前記1, 2の特徴を持つ「抽象化した情報」を複製の対象とする。これにより、複製に対する更新処理の頻度を削減したうえで、複製が最新情報を保持できるようにする。

このような情報共有手法を実現するための仕

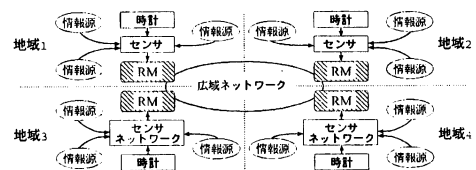


図1: WISEの構成

組みとして、WISE (Wide-area Information Sharing Engine) <sup>5)</sup> の概要を説明する。WISE の構成を図 1 に示す。地域ごとに、センサ・センサネットワーク・RM (Resource Manager) がある。センサ・センサネットワークは、情報源において情報を定期的に観測する。RM は、前記の自律型モジュールである。利用者は、近傍の RM を介して情報を参照する。RM の中核は、データベース・抽象化機能・学習機能である。データベースは、‘素な情報’と‘抽象化した情報’を格納する。‘素な情報’と‘抽象化した情報’は、次に示す単位データ  $d$  の集合として表す。

- $d = (g, t, a, s, v)$ 
  - $g$ : 地域。  $d$  が生成された地域を示す。識別名 <sup>7)</sup> を用いる。
  - $t$ : 時刻。  $d$  の生成時刻を示す。単位時間ごとに単調増加するシーケンス番号を用いる。
  - $a$ : 抽象度。値が大きいほど抽象化されており、よりいっそう、観測した情報の特徴だけを反映し、観測した情報と比べて時間的な変化が少ないことを示す。‘素な情報’は 0。‘抽象化した情報’は正の整数。
  - $s$ : 情報源。  $d$  を生成した情報源を示す。  $g$  に対する相対識別名 <sup>7)</sup> を用いる。
  - $v$ : 値。

$d$  の集合には、前提条件とする参照パターンに即して利用者が単位データを参照するように、WISE 情報モデル (IMW: Information Model for WISE) と呼ぶグラフ構造を与える。IMW は、次に示す 1 種類のノードと 2 種類のアークから構成される。

- ノード: 地域と抽象度が等しい  $d$  の集合。
- ノードの抽象度の高低関係を表すアーク: 抽象度が異なる 2 つのノードの組み。
- ノードの地域の隣接関係を表すアーク: 抽象度が等しい 2 つのノードの組み。

グラフ構造を与えられた  $d$  の集合を、IMW のインスタンスと呼ぶ。図 2 に例示する。抽象度が大きいノードほど広い地域の概観情報を持つ。抽象度が小さいノードほど狭い地域の詳細情報を持つ。利用者は、2 種類のアークをたどって、抽象

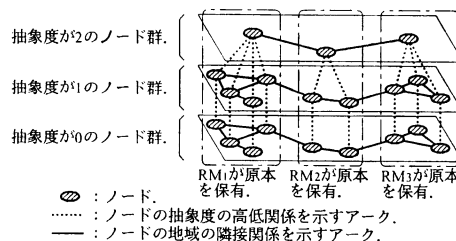


図 2: IMW のインスタンス

度が大きいノードから抽象度が小さいノードへ興味のある地域を絞り込みながら、単位データを参照する。各 RM は、IMW のインスタンスのトポロジ情報を保有する。さらに、配置された地域のセンサ・センサネットワークから直接受け取った抽象度 0 の単位データと、その抽象度 0 の単位データから生成した抽象度 1 以上の単位データとを、単位データの原本として保有する。他の RM が保有する単位データの原本に関しては、その RM への参照リンクと部分的に複製を保有する。抽象化機能は、抽象度 0 の単位データの更新を契機として、抽象度 0 の単位データから抽象度 1 以上の単位データを動的に生成する。また、生成した単位データを他の RM に複製する。学習機能は、単位データの利用頻度に基づいて、RM が保有する単位データの原本のなかから他の RM に複製すべき単位データを動的に選定する。WISE は、これらの機能を実現するためにアクティブデータベース <sup>8)</sup> の枠組を利用する。

### 3 情報配信手法

#### 3.1 方針

情報配信手法の目的は、単位データの原本を持つ RM からその単位データの複製を持つ複数の RM へ単位データを配信することである。情報共有手法の課題から、情報配信手法の課題は次のとおりである。

- 課題 (a), (b) のために拡張性。
- 課題 (c), (d) のために信頼性と即時性。

拡張性を得るには、配信負荷を分散できるために、単位データをマルチキャストする方式がよ

い、これに信頼性を付加するには、単位データを再送する方式と単位データを冗長に送信する方式がある<sup>9), 10), 11)</sup>。さらに即時性を付加するには、信頼性を得るための処理をできるだけ単純化・局所化する必要がある。また、ネットワーク・インフラに依存せずに実現できることが望ましい。そこで、ネットワーク上に配信ノード (DN: Delivery Node) を配置して<sup>12)</sup>、階層構造と冗長な経路を持つオーバーレイネットワークを構築し、その上で単位データを冗長にマルチキャストすることにする。すなわち、階層構造と冗長な経路とを特徴とするアプリケーション・レイヤ・マルチキャスト<sup>11), 12), 13), 14), 16), 15)</sup>方式である。次節以降では、オーバーレイネットワーク (3.2 節) と、その上でのマルチキャスト配信 (3.3 節) について述べる。

### 3.2 オーバーレイネットワーク

まず、オーバーレイネットワークのトポロジについて述べる。拡張性を重視すれば、配信負荷を分散できるために、配信元木 (Source Based Tree) を用いて単位データをマルチキャストする方式<sup>11), 15)</sup>が良い。ただし、DN を冗長化し、DN の間の通信に TCP を用いても、物理的な経路の故障には対処し難い。信頼性及び即時性を重視すれば、冗長な経路を利用して単位データを冗長に送信する処理を単純化・局所化できるために、メッシュを用いて単位データをフラディングする方式<sup>16)</sup>が良い。ただし、メッシュを構成するリンクと物理的な経路とが 1 対 1 に対応することが要件であり、配信負荷も高い。そこで、配信負荷の分散と、冗長な経路を利用して単位データを冗長に送信する

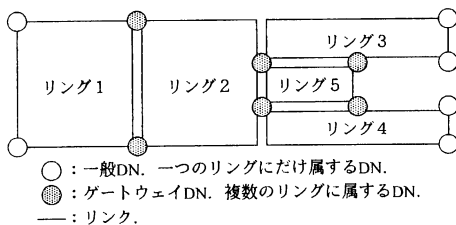


図 3: オーバーレイネットワーク

処理の単純化・局所化とが可能なトポロジとして、リングとリングを複数点で接続したトポロジを用いることにする。リングを構成するリンクと物理的な経路とを 1 対 1 に対応させるには DN を意図的に配置する必要があるが、DN の配置の自由度は比較的高い。このようなオーバーレイネットワークを図 3 に例示する。DN のグループが 5 つあり、グループごとにリングを構成している。単一のリングに属する DN を一般 DN と呼ぶことにする。複数のリングに属しリングを接続する DN をゲートウェイ DN と呼ぶことにする。例えば、リング 1, 2 は 1 次 ISP のネットワークの上のリング、リング 3, 4 は 2 次 ISP のネットワークの上のリング、リング 5 は ISP の接続点と地域 IX の接続点を結んだリングである。

次に、オーバーレイネットワークの構築と維持について述べる。

- DN の参加: 参加するグループを決める。DN に、そのグループに属する DN の IP アドレスを一つ与える。DN は、その DN から、そのグループに属する全ての DN の IP アドレスを取得する。DN は、そのグループに属する全ての DN に参加を広報する。
- DN の脱退: DN は、グループに属する全ての DN に脱退を広報する。また、次項で示す「距離情報の広報」を一定期間行わなかった DN は、自動的に脱退したと見なされる。
- リングの作成と維持: 各 DN は、グループに属する全ての DN に対して Traceroute・Tcptraceroute を実行し、他の DN との距離情報を取得する。得られた距離情報を、グループに属する全ての DN に広報する。これを定期的に行なう。また、グループに属する全ての DN の中で最も大きい IP アドレスを持つ DN を、代表 DN として選出する。代表 DN は、ホップ数や遅延をメトリックとして、広報された距離情報からネットワーク地図を作成する。この地図を用いて、全ての DN を含み一回りの距離が最短であるリングの構成情報を作成する。作成したリングの構成情報を、グループに属する全ての DN に広報する。リングの構成情報を作成する問題は、巡回セールスマン問題である。ここでは速

度のために、欲張りアルゴリズムを用いることにする。例えば、条件付（次数は2以下、閉路は辺の数がDNの数に等しい時だけ可）のクラスカスのアルゴリズム<sup>17)</sup>を用いる。

### 3.3 マルチキャスト配信

まず、オーバレイネットワークの経路情報について述べる。経路情報には、リング内の経路情報とリング間の経路情報がある。リング内の経路情報は、リングの構成情報そのものである。リング間の経路情報は、BGPに類似した方法で作成する。ゲートウェイDNは、属するリングごとに異なるDNがあるかのように振舞う。同じリングに属するDNとは、リング間の経路情報を共有する。異なるリングに属するDNとは、グループに割り振られたリング識別子を用いて、ルーティングプロトコルを動作させる。作成されるリング間の経路情報は、宛先であるグループのリング識別子と、ネクストホップであるグループのリング識別子との組みである。

次に、マルチキャスト配信経路について述べる。前記の経路情報を用いて配信元DNから配信先DNへコマンド・パケットを送信し、コマンド・パケットを中継したゲートウェイDNをプログラムすることにより、マルチキャスト配信経路情報を作成する<sup>11), 15)</sup>。

- 配信元DNは、次のようにコマンド・パケットを送出する。
  1. 配信先DNが属するグループのリング識別子を、配信先DNに問い合わせる。
  2. 自らが属するグループの代表DNから、シーケンス番号を取得する。
  3. 配信する単位データの地域と情報源から、コンテンツ識別子を作成する。
  4. パケット種別・シーケンス番号・配信元DNについて所属グループのリング識別子とIPアドレス・配信先DNについて所属グループのリング識別子とIPアドレスをヘッダ、操作種別・コンテンツ識別子を本体として、コマンド・パケットを作成する。
  5. コマンド・パケットを、リングの両隣のDNに送出する。

- 配信元DNから配信先DNへの経路上にあるDNは、次のようにコマンド・パケットを中継する。
    1. 一般DNは、リングの隣のDNにコマンド・パケットを中継する。
    2. ゲートウェイDNは、ネクストホップであるグループに属していない場合は、一般DNと同様である。ネクストホップであるグループに属している場合は、ネクストホップであるグループが構成するリングの両隣のDNにもコマンド・パケットを中継する。
  - 操作種別が配信先DNの追加であるとき、グループからグループへコマンド・パケットを中継したゲートウェイDNは、次のようにプログラムされる。
    1. ネクストホップであるグループのリング識別子とコンテンツ識別子との組みを、マルチキャスト配信経路情報に追加する。既に組みが存在していれば、その組みに付属するカウンタを1増やす。
  - 操作種別が配信先DNの削除であるとき、グループからグループへコマンド・パケットを中継したゲートウェイDNは、次のようにプログラムされる。
    1. マルチキャスト配信経路情報において、ネクストホップであるグループのリング識別子とコンテンツ識別子との組みに付属するカウンタを1減らす。もしカウンタが0になれば、その組みを削除する。
  - 配信先DNは、次のようにプログラムされる。
    1. マルチキャスト配信受信情報に、コンテンツ識別子を追加あるいは削除する。
- 最後に、単位データのマルチキャスト配信について述べる。
- 配信元DNは、次のように情報パケットを送出する。
    1. 自らが属するグループの代表DNから、シーケンス番号を取得する。
    2. 配信する単位データの地域と情報源から、コンテンツ識別子を作成する。

3. パケット種別・シーケンス番号・配信元 DN について所属グループのリング識別子と IP アドレス・コンテンツ識別子をヘッダ、単位データを本体として、情報パケットを作成する。
4. 情報パケットを、リングの両隣の DN に送出する。

- 配信元 DN から配信先 DN への経路上にある DN は、次のように情報パケットを中継する。
  1. 一般 DN は、リングの隣の DN に情報パケットを中継する。
  2. ゲートウェイ DN は、コンテンツ識別子をキーとしてマルチキャスト配信経路情報を検索する。適合するエントリがない場合は、一般 DN と同様である。適合するエントリがある場合は、そのエントリのリング識別子が示すリングの両隣の DN に情報パケットを中継する。

- 配信先 DN は、次のように情報パケットを受信する。

1. コンテンツ識別子をキーとしてマルチキャスト配信受信情報を検索し、適合するエントリがあれば情報パケットを取り込む。

なお DN は、コマンド・パケットと情報パケットに対して以下処理を行なう。

- リングの隣の DN に対してパケットの送信を試みたときに、その DN が応答しない場合には、応答しない DN の隣の DN に対してパケットの送信を試みる。
- 受信したパケットのシーケンス番号と配信元 DN の所属グループのリング識別子との組が既知である場合には、パケットを破棄する。

## 4 おわりに

本稿では、拡張性・信頼性・即時性を備えた情報配信手法を提案した。提案手法では、信頼性・即時性のために、階層構造と冗長な経路とを持つオーバーレイネットワークを構築する。拡張性のために、その上で単位データを冗長にマルチキャストする。今後、実験システムを実装し、提案手法を検証、改良する。

## 参考文献

- 1) g コンテンツ流通促進協議会。  
<http://gisclh.dpc.or.jp/gcnt/>
- 2) 石田 亨：デジタルシティの現状，情報処理学会誌， Vol.41, No.2, pp.163-168 (2000).
- 3) 荒野，西郷：大規模商用ネットワークの運用事例，情報処理学会誌， Vol.39, No.10, pp.969-975 (1998).
- 4) Czajkowski, K., Fitzgerald, S., Foster, I. and Kesselman, C.: Grid Information Services for Distributed Resource Sharing, *Proc. 10th IEEE International Symposium on High-Performance Distributed Computing* (2001).
- 5) 八木 哲：情報の抽象化に基づいた広く分散した動的な情報の共有手法，情報処理学会論文誌：データベース， Vol.43, No.SIG8(TOD18), pp.126-135 (2003).
- 6) 八木 哲：広く分散した動的な情報のための情報共有手法の最適化， DEWS2003 (2003).
- 7) 大山，千田，戸部，窪田，田中，空：X.500 ディレクトリ入門，東京電気大学出版局 (1997).
- 8) 石川博：アクティブデータベース，情報処理学会誌， Vol.35, No.2, pp.120-129 (1994).
- 9) 木下真吾：リライアブルマルチキャスト技術の最新動向，電気情報通信学会論文誌 B, Vol.J85-B, No.11, pp.1819-1842 (2002).
- 10) 中内，チェン，森川，青山：ピア・ツー・ピアマルチキャストのためのマルチパスルーティングアルゴリズム， *DICOMO2001* (2001).
- 11) Zhuang, S., Q., Zhao, B., Y., Joseph, A., D., Katz, R., H. and Kubiatowicz, J., D. : Bayeux : An Architecture for Scalable and Fault-tolerant Wide-area Data Dissemination, *Proc. NOSSDAV2001* (2001).
- 12) Chawathe, Y., McCanne, S., Brewer, A., E. : RMX : Reliable Multicast for Heterogeneous Networks, *Proc. IEEE Infocom2000* (2000).
- 13) Francis, P. : Yoid:Extending the Internet Multicast Architecture.  
<http://www.aciri.org/yoid/> (2000).
- 14) Chu, Y., Rao, G., S., Zhang, H. : A case for End System Multicast, *Proc. Sigmetrics2000* (2000).
- 15) Rowstron, A., Kermarrec, A., M., Castro, M. and Druschel, P. : SCRIBE : The design of a large-scale event notification infrastructure, *Proc. NGC2001* (2001).
- 16) Ratnasamy, S., Handley, M., Karp, R. : Application-level Multicast using Content-Addressable Networks, *Proc. NGC2001* (2001).
- 17) Aho, A.V., Hopcroft, J.E., Ullman, J.D. (著)、大野義男 (訳)：データ構造とアルゴリズム，培風社 (1987).