

# 通信衛星を用いたマルチキャストデータ伝送機構の設計と 実装

石井 秀治, 佐野 晋  
日本電気株式会社

## 概要

我々は過去3回、インターネット災害訓練において、生存者情報検索 (IAA) システムを設計・実装し、その評価を行った。IAA システムでは、データベースを複数サイトに分散配置し、内容を同期させ、検索を分散させることで負荷を減らしている。第2回、第3回の訓練では、この同期機構の一部として、通信衛星を用いたマルチキャストデータ伝送機構を設計し、実装した。本稿では、本機構を概説し、第3回訓練の結果を報告、評価するとともに、今後の課題を示す。

## Design and Implementation of Satellite based Multicast transport mechanism

Shuji Ishii, Susumu Sano  
NEC Corporation

## Abstract

We designed, implemented and evaluated a quering survivor information system, called "IAA system" in Internet disaster drills. In the system, we locate database machines on many sites and synchronize its contents. So they can reduce their load. In second and third drill, we designed and implemented database synchronization mechanism which uses satellite multicast transport system. In this paper, we describe the mechanism briefly, report the third drill and discuss the mechanism and future works.

## 1 はじめに

WIDE プロジェクトでは、1995年の阪神淡路大震災を契機に、インターネットの、特に災害時の社会的基盤としての役割を中心に検討を行っている。

我々は、1996年1月、1997年1月、1998年1月と過去3回にわたり、インターネット災害訓練を実施した[1, 2, 3]。この訓練の内容は、生存者情報の登録、検索を行うものである。災害時には、通信網の障害や、災害地外からの問い合わせが殺到することから、データベースを一箇所に設置し、そこに問い合わせる方式では、現実的な動作は期待できない。

そこで、データベースを地理的に分散させ、登録要求や検索要求を分散させることで、高負荷や部分的障害に強いシステムを設計、実装した。このシステムをIAAシステムと呼んでいる。本稿では、IAAシステムにおける通信衛星を用いたマルチキャスト伝送機構を中心としたデータベースの同期機構について述べる。

## 2 設計の方針

今回、IAAシステムは広域分散データベースを中核とするシステムである。本システムが災害時に利用され、その範囲も特定できないことから、

- 十分なスケーラビリティをもっていること、
- システムの分断や部分的な障害、局所的な負荷の集中に対して、その影響がシステム全体に影響しないこと(頑強性)、

が要求される。一般的に分散データベースは、個々のデータベースの同期に関する時間的制約が強いほど、また排他制御やアトミックオペレーションなどの要求が強いほど、プロトコルは複雑になり、集中化するため、上述の要件を満たすことは困難となる。そこで、このようなコンシステンシに関する制御をできるだけ簡略化し、十分なスケーラビリティと頑強性を確保することとした。

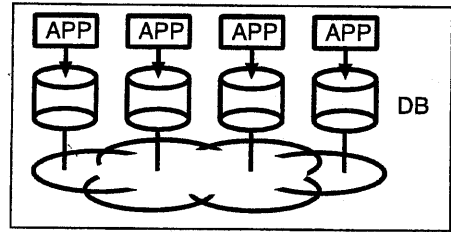


図 1: IAA システム概念図

図 1は本システムの概念図である。

分散されたデータベースがネットワーク上に配置され、それぞれにアプリケーションが動いている。アプリケーションはデータベースと直接接続する。このデータベースは、

- 個々のデータベースはすべて同じ内容である。
- データベースに対する操作は参照と追加のみである。変更と削除はない。
- あるデータベースにデータの追加があると、その内容を放送型の通信によって他のデータベースに伝える

という方法でコンシステンシの維持を行う。放送型の通信機構の信頼性が確保できれば、データベースの整合性も確保できることになる。

以下では、今回用いた放送型の通信の機構を中心に述べることにする。

## 3 データ多重放送との比較

衛星を用いたデータ伝送機構の一つに、テレビ放送におけるデータ多重放送[4, 5]がある。ここでは、今回提案した機構とデータ多重放送では、以下の二点が異なると考えられる。

**時間的制約:** データ多重放送では、データの送信スケジュールがあらかじめ決定していることが多い。つまり、テレビ番組のように、決まった時刻に決まったデータが配布されることが要求されており、厳密な時間的な

制約がある。それに対して、本機構では、データは非同期に発生し、あらかじめ送信スケジュールを決定できない。比較的緩やかな時間的制約のもとで配送されることが要求される。

配布形態：データ多重放送では、時刻と同様に内容もあらかじめ送信スケジュールが決定されていることが多い。つまり、受信者は送信データを選択的に受信する。また、受信者からの（地上網を用いた）要求によって、送信される場合もある。一方、本機構は、全ての受信者に対して均一な配送が要求される。さらに災害時という点を考慮して、地上網による補完が期待できない点が異なる。

## 4 IAA トランスポート

本節では、IAA システムのトランスポート部分の実装について述べる。このトランスポート部では、2 節で述べたように、以下の項目が要求される。

- 回線切断のような障害に強い
- 広範囲にデータを流布させる
  - 緩い整合性

この目的を達成するために、トランスポート部分は、ネットニュースシステム [6] を利用し、データ配送、データベースへの登録を行った。

ネットニュースは、記事を蓄積転送型で広範囲に配布するシステムである。また、サーバ間を冗長に接続することで回線切断のような障害に強いのも特徴である。

これらの要求の他に、セキュリティについての考慮も行った。つまり、ネットニュースでは、平文で記事 (= データ) が送信される。そのため、記事の暗号化を行い、送信経路の盗聴や設定ミスによる誤配送で、データが無関係なサーバへ流出した場合でも、データが暴露されないように配慮した。

衛星リンクは、ニュースシステムにおける記事配送機構の一部として利用した。この機構は、衛星回線の特徴の一つである単方向のリンクにおけるファイル転送機構である。さらに、衛星

の同報性も利用してマルチキャストによるファイル転送を実現した。

### 4.1 単方向ファイル転送機構

単方向ファイル転送機構は、衛星リンクで動作することを目的に設計された。以下の特徴を持つ。

- 単方向リンクで動作 (送達確認が不要)
- UDP を利用 (マルチキャスト環境で動作)
- 簡単な輻輳回避

この機構の基本的な動作は、送信対象のファイルを複数の断片に分割し、これを衛星リンク (一方向) を通じて転送する。TCP のような送達の確認を行わないので、転送中にエラーが生じると、正常にファイルが転送されない。そのため、同一ファイルを繰り返し転送することで、エラーが生じても最終的には受信者に全てのファイルの断片が到達し、ファイルを転送できる (図 2)。また UDP を利用していることからマルチキャスト通信への応用が可能である。

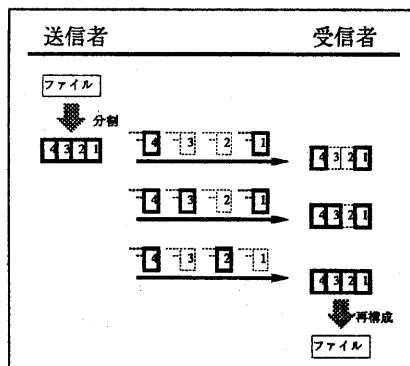


図 2: 転送方法

#### 4.1.1 プロトコル

一連のファイル転送では、複数のコマンドとデータが転送される (図 3)。コマンドには、以下の 4 種類ある。

- FILENAME
- ATTRIBUTE
- DATA
- PAUSE

ATTRIBUTE と FILENAME はファイル転送の開始時に送信され、それぞれファイル属性とファイル名を受信者に通知する。DATA は、ファイルの断片とともに送信される。PAUSE は、全ての断片を送信した後に送信される。

受信側では、上記のコマンド、データが全て受信されると、ファイルが完成する。どれか、一つでも受信できなかった場合、一時ファイルとして保存され、欠けているコマンドあるいはデータの受信を待つ。

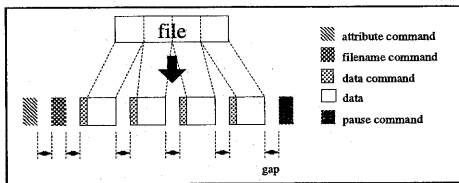


図 3: コマンドパケット

#### 4.1.2 再送機構

ファイルは常に再送されるが、送信待ちの行列を一つにすると、行列の長さが長くなると、新規に待ち行列に入るファイルは、送信開始までの時間が長くなる問題がある。そこで、優先度付の待ち行列を3つ用意し、新規に到着したファイルを優先的に送信する機構を実現した(図4)。

#### 4.1.3 輻輳

本プロトコルでは、送信者は、送信経路上の輻輳の発生が分からない。そこで、ファイルの断片の送信間隔を、断片の大きさ、衛星回線速度、予備実験の結果から決定し、輻輳が起こらないように実装した。今回の実験では100 ミリ秒に設定した。

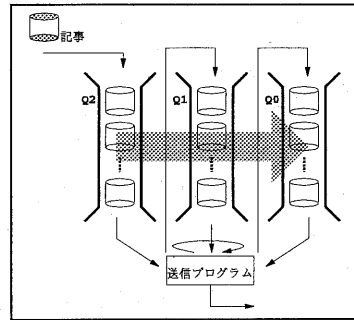


図 4: 再送機構

## 4.2 実験環境

今回の訓練では、データベースを以下の3サイトに設置した。

- 東京大学
- 北陸先端科学技術大学院大学 (以降、北陸先端大)
- 奈良先端科学技術大学院大学 (以降、奈良先端大)

衛星配送は、送信者を東京大学とし、受信者を北陸先端大と奈良先端大とした。そして、北陸先端大と奈良先端大から東京大学へニュースリンクを設定した。このリンクでは、無遅延で記事の配送を行った。つまり、北陸先端大と奈良先端大に登録されたデータは、即座に東京大学に送信され、衛星で配送される(図5)。

このリンク以外にも地上網でのニュースリンクを設定した。このリンクは、衛星配送の障害時にデータを配送するためである。全て、地上網での配送されないように、このリンクでは、遅延配送を行っている。つまり、データが到着して、一定時間経過したのちに配送される。この設定によって、衛星による配送を優先した。

- 奈良先端大 → 東京大学
- 東京大学 → 北陸先端大
- 北陸先端大 → 奈良先端大

ネットニュースシステムは、INN Version 1.5.1 を利用した。各サーバは BSD/OS 3.1 が動作している IBM-PC 互換機である。

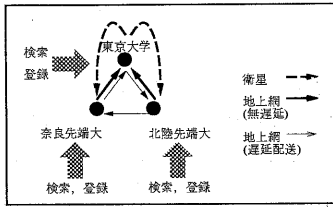


図 5: 実験環境

## 5 結果

実験は、1998年1月17日午前9時から同年1月18日の午後5時までの32時間行われた。ただし、奈良先端大においてデータベースマシンが30分程、運用不能となった。

登録件数は全部で838件であった。最初に登録されたサイトで分類した登録件数の内訳は以下の通り。

- 北陸先端大: 181件
- 奈良先端大: 400件
- 東京大学: 257件

また、衛星経由で登録された件数は、以下の通りである。

- 北陸先端大 654件
- 奈良先端大 420件

奈良先端大と東京大学から北陸先端大への配送経路は衛星リンクであるため、本来は、それぞれ奈良先端大に最初に登録されたデータ数400件と東京大学に最初に登録された件数257件との和657件が衛星経由の件数である。

しかし、実際は654件であるため3件(0.1%)が地上経由になった。同様に奈良先端大へは438件(=400+181)のうち18件(4.1%)が地上経由となった。

北陸先端大の場合、問題の3件の記事は全て、奈良先端大に最初に登録された記事であった。これは、該当記事の奈良先端から東京大学への送信が失敗してしまったため、と考えられる。送信が失敗した理由は不明であるが、奈良-東京間の回線が混雑していて失敗したと推測される。

奈良先端大で18件が地上経由になった原因は、データベースマシンの運用不能状態のためと考えられる。この時間帯を除くと全ての記事が衛星経由となっていた。

両サイトとも、マシンのダウンや送信の失敗という障害にもかかわらず、データベースの内容の整合性が取れており、本システムの頑強性を示す結果となった。

## 6 考察

運用という観点から今回の実験について考察する。

**設定・構築:** 今回の実験では、ニュースサーバ、データベースサーバ、衛星などの設定を全て静的に行った。しかし、災害時では、あらかじめネットワークの構成やシステムの設定を行うのは不可能である。また、システムを止めることなく、構成や設定の変更も要求される。災害が起きてから、迅速にシステムを構築、設定する機構が必要と考えられる。

**セキュリティ:** 本システムで、扱うデータは個人情報である。そこで、送信経路での盗聴や設定ミスによるデータの漏洩事故に対処するため、データを暗号化して送信している。この暗号鍵は、あらかじめ各サイトに配付している。しかし、この作業も設定問題と同様に現実的ではない。

現在、設定・構築に関する問題の解決方法の一つとして「定常システムの運用」を検討中である。何も無い状態から大規模なシステムを短時間で設定・構築するのはコストがかかるので、基幹的なトランスポート部分をあらかじめ構築しておき、災害時には、この基幹部分を中心にシステムをスムーズに拡張する方法である。定常運用の、もう一つの意味として常に動作しているシステムで、運用の経験を積んでおかないと、災害時に効率の良い運用が期待できない点があげられる。

鍵配付の問題は、本システムだけでなく、一般的なセキュリティシステムにおける問題であ

る。この問題の解決方法の一つに、公開鍵暗号系を利用した証明書発行局 (CA)[7] を使う方法があり、CA を用いる方法も考えられる。

## 7 まとめ

本稿では、第3回インターネット災害訓練のために設計・実装した、衛星を用いたマルチキャストデータ伝送機構について述べた。そして、災害訓練での本機構の実験結果を報告し、評価を行った。本訓練ではマシンダウンや通信障害が発生したが、地上網による補完でデータベースの整合性が正しく取れ、本システムの有効性を示した。最後に、今後の課題を考察した。この中で、運用の観点から、解決すべき問題点と解決案を示した。

## 8 謝辞

衛星回線を提供して頂いた株式会社日本サテライトシステムズの皆様に感謝致します。また、いつも議論に参加して頂いている WIDE プロジェクト lifeline ワーキンググループの皆様に感謝致します。

## 参考文献

- [1] WIDE プロジェクト, 1995 年度 WIDE プロジェクト研究報告書第2部 ライフラインとしてのインターネットに関する考察, 1996 年
- [2] WIDE プロジェクト, 1996 年度 WIDE プロジェクト研究報告書第2部 ライフラインとしてのインターネットに関する考察, 1997 年
- [3] WIDE プロジェクト lifeline WG ホームページ, <http://www.iaa.wide.ad.jp/>
- [4] IP 対応始まるデータ放送, 日経コミュニケーション 1997 年 9 月 1 日号 p108-p113
- [5] 月額費用無料のデータ放送インターネットとの連携も可能, 日経コミュニケーション 1997 年 7 月 21 日 p122-p125
- [6] Brian Kantor and Phil Lapsley, Network News Transfer Protocol(RFC977), 1986
- [7] S Kent, Privacy Enhancement for Internet Electronic Mail: PartII - Certificate Based Key Management. RFC1422, Feb. 1993
- [8] Yukimitsu Izawa, Shuji Ishii, Nobuhiko Tada and Masaya Nakayama, Implementation and evaluation of Widely Distributed Database System using Satellite based Multicast and NetNews system for the transport mechanism, Proceedings of The International Workshop on Asia-Pacific area advanced research information sharing technology Internet Workshop '98 (IWS'98), 1998