マルチメディア通信と分散処理　106－40
コンピュータセキュリティ　16－40
（２００２．２．１５）

Adaptive Multicast Protocolにおける輻輳制御方式の提案

細谷 篤　　水野 忠則　　佐藤 文明
静岡大学大学院
情報学研究科
静岡大学情報学部

近年，インターネットの急速な発展と共にグループウェアのような共同で作業を行う分散アプリケーションの開発が非常に活発になっている．分散アプリケーションではその特性から高信頼マルチキャストの利用が有用であり，これまでに多数のプロトコルが研究，提案されてきた．本研究では共有仮想環境のような分散アプリケーションに適用することを目的としたスケーラブルな高信頼マルチキャストプロトコルとして Adaptive Multicast Protocol(AMP) を提案している．本稿では AMP における輻輳制御方式として集約したフィードバック情報を用いる方法と各ノードが自分の周辺のリンクの輻輳状況を監視し送信量を加減する方法の２通りを組み合わせた手法を提案する．

A Congestion Control Algorithm for Adaptive Multicast Protocol

Atsushi Hosoya　　Tadanori Mizuno　　Fumiaki Sato
Graduate School of Information, Shizuoka University
Faculty of Information Shizuoka University

In recent years, development of distributed applications such as groupware systems became very active. Since reliable multicast is useful for a distributed application, many protocols have been researched and proposed. We have proposed the Adaptive Multicast Protocol(AMP) that is scalable and reliable for distributed applications as Shared Virtual Environments. In this paper, we propose a congestion control algorithm which combined with two methods based on feedback information and monitoring the link.

1　はじめに

近年，インターネットの急速な発展と共にグループウェアのような共同で作業を行う分散アプリケーションの開発が非常に活発になっている．分散アプリケーションでは同一の情報を多数の受信ノードに転送する必要性からマルチキャスト通信 [1] の利用が有用である．IP マルチキャストはUDP で行うため，パケットの欠落や到着順については保証しない．そこでパケットの再送や全順序保証の機構を持った高信頼マルチキャストプロトコルが必要となる．

我々は共有仮想環境 [2] のような分散アプリケーションに適用することを目的としたスケーラブルな高信頼マルチキャストプロトコルとして Adaptive Multicast Protocol(AMP)[3][4] を提案してきた．AMP はインターネットに接続した多数のユーザがインタラクティブにデータを交換しあう状況を想定して設計されている．インターネットのような受信ノードの処理能力やリンクの転送能力がさまざまであらかじめ予測できないようなネットワークでマルチキャストを行う場合，従来はフロー制御や輻輳制御は行わず，他のフローを不当に圧迫しないように少量を定レートで転送するのが一般的であった．しかしこの方法では大量のデータ交換を効率的に行うことができない．マルチキャスト通信で効率的なデータ転送を実現するにはフロー制御及び輻輳制御が必要であり，そのための研究も行われている．代表的なフロー・輻輳制御の手法とし

てはエンド-エンド間でフィードバックを用いる方法や中継ノード（ルータ）で過剰なパケットを廃棄するといった方法が挙げられる．しかし前者の場合フィードバックの応答爆発やフィードバック情報の伝播遅延によるオーバーシュート，後者の場合はルータの対応が必要で処理が複雑になるという問題点がある．そこで本稿ではAMPにおける輻輳制御方式として集約したフィードバック情報を用いる方法と各ノードが自分の周辺のリンクの輻輳状況を監視し送信量を加減する方法の2通りを組み合わせた手法を提案する．

2 Adaptive Multicast Protocol(AMP)

2.1 AMPのシステム構成

AMPは分散アプリケーションに適用することを目的としたスケーラブルな全順序マルチキャストプロトコルであり，共有仮想環境のような多対多の通信を想定している．AMPではマルチキャストグループを複数のドメインに分割し，グループのメンバはいずれかのドメインに所属する(図1)．各ドメインには1つのシーケンスサーバと1つ以上の再送サーバが存在する．シーケンスサーバはドメイン内のメンバが送信するパケットに対しユニークなシーケンスナンバーを与える．全順序の保証と再送については以下で詳述する．

2.2 全順序保証

シーケンスサーバはドメイン内のメンバのパケット送信要求に対しシーケンスナンバーを発行する．送信ノードはパケットにシーケンスナンバーとドメインIDを付与してマルチキャストする．受信ノードでは受信したパケット送信元のドメインごとにバッファに振り分け，一定

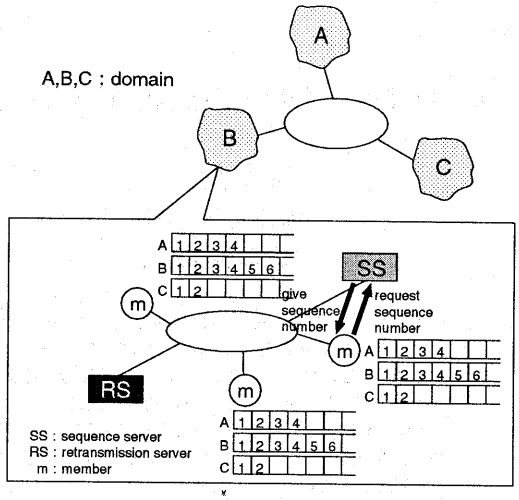

図1: AMPのシステム構成

のアルゴリズムで送信元ドメインを選択して，シーケンスナンバーに従ってアプリケーションに配送する．このとき送信元ドメインを選択するアルゴリズムをすべての受信ノードで同期させることで全順序を保証する．

ひとつのシーケンスサーバがマルチキャストグループすべてのパケットにシーケンスナンバーを与え全体の全順序保証を行うような方式[5]ではサーバに負荷が集中してしまい，スケーラビリティに問題があったがAMPでは複数のシーケンスサーバを導入することでこの問題を解決している．

2.3 再送メカニズム

AMPにおける再送方式[6]の信頼性保証はACK,NACKのユニキャストで行う．再送サーバはそれぞれ再送を受け持つ範囲があり，これを各サーバの再送リージョンと呼ぶ(図2)．すべてのノードはいずれかのサーバの再送リージョンに属しており，どのサーバから再送を受けるかは基本的にはマルチキャストグループに参加するときに決定する．また再送サーバは再送木と呼ばれる仮想的な木構造を構成する．受信ノードは

自身が属する再送リージョンの再送サーバに対し,一定数パケットを受信するとACKを,パケットの欠落を検知するとNACKを送信する.再送サーバはACKや自身が保持していないパケットに対する再送要求のNACKを受信した場合,再送木によって繋がれている隣接サーバに対しACK,NACKを送信する.NACKで要求されたパケットを持っている場合は受信ノードまたは隣接するサーバに対し再送を行う.このとき再送サーバ及び送信ノードの受信するACK,NACKは再送木で繋がれた隣接サーバからのみであるので応答爆発は起こらない.

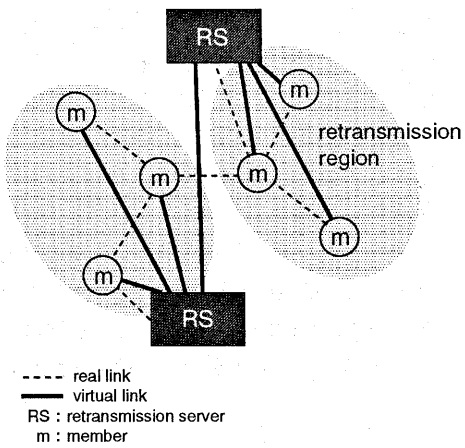

図 2: 再送木の構成

3 高信頼マルチキャストにおけるフロー・輻輳制御

フロー・輻輳制御を行わずにマルチキャストを行うと,次のような問題が生じる.送信ノードはデータを定レートで送信し続けるが,一部はルータのバッファオーバーフローや衝突による損失によって受信アプリケーションまで到達しない.損失してしまったパケットはアプリケーションに届かないという意味では無駄なパケットということになる.しかし送信ノードからパケットの損失を起こしている点までの間に分岐している経路では元のレートで転送ができるので転送全体では必ずしも無駄なパケットとはいえない.また定レートのフローがTCPフローと混在する場合,TCPのフローは輻輳制御により送信レートを絞るため,定レートのフローがリンクの帯域を占有してしまうという現象が起こる.一般に受信ノードの処理能力やリンクの転送容量,輻輳状況はそれぞれ異なること,他のトラフィックと公平に容量を分け合うべきであること,などを考えるとフロー・輻輳制御を行うべきなのは明白である.

マルチキャストにおけるフロー・輻輳制御の代表的な手法としてはエンド-エンド制御で送信ノードにフロー制御情報・輻輳制御情報をフィードバックし,その情報を元に送信ノードで送出量を加減する方法や,ネットワーク内の中継ノード(ルータ)で過剰なパケットの廃棄などを行う方法などが挙げられる[7].

フィードバック情報を用いて送出量を制御する方法の利点は既存のネットワークを変更する必要がなく,ルータの処理が単純で早いことなどがある.しかし送信ノードにフィードバック情報が多数集まってくる応答爆発や,フィードバック情報が伝播しフローが絞られるまでの遅延が比較的大きく,その期間は過大負荷が続く,などの問題がある.これまで提案されてきたフィードバックを用いる方法の多くは最もスループットの低い受信ノードに転送レートを合わせたものである[8].この方法は無駄なトラフィックを減少させることが可能であるが,一方で受信ノードによっては自らが受信できる限界より大幅に低いスループットしか得られないという状況も生じる.

ルータによってパケット廃棄を行う方法の利点は,ルータより下流のリンクについてはTCPトラフィックを不当に圧迫する現象がなくなることである.しかし受信ノードのオーバーフロー

には対応できず, 廃棄したパケットに関しては送信ノードから廃棄点までのトラフィックは無駄なものとなる. またルータに新たな機能を付け加える必要があり処理も複雑になる.

輻輳が生じている下流リンク向けのフローを減らす方法として特殊なサーバにより輻輳が生じているリンクより上流からのフローをバッファリングするという技法もある [9]. これには上流のノードには受信レートの低下を意識させずみ済むというメリットがある. しかしこの方式では, サーバとルータとの密接な関係が必要である.

4 AMP における輻輳制御方式

4.1 方針

AMP に求められる主な要件としては以下のようなポイントが挙げられる.

1. 多対多通信への適合
2. 大規模なネットワークへの適合
3. リアルタイム性
4. TCP トラフィックとの公平性

実装するフィールドとしてインターネットを目標とする以上, 輻輳制御システムを実現するのにルータに様々な付加機能を追加するといった方法は難しい. 従ってフィードバック情報を用いることが考えられる. 1 について, 多対多通信では基本的にすべてのノードが送信端かつ受信端であるので, ある地点で輻輳が起こったとき, その原因を特定の送信端に求めることは難しくあまり意味がないためエンド-エンド制御には向いていない. また 2 についても, エンド-エンド制御は大規模なネットワークでは応答爆発, 伝播遅延などの問題がある.

そこで AMP ではフィードバック情報を集約しながらグループ全体に伝播させる方法をとる. フィードバック情報は再送木でやりとりされる ACK,NACK に付加することで余分なパケットを生成することを防ぐ. この方法と並行して, 送信ノードは自分と自分が所属する再送リージョンのサーバ間のトラフィックを監視し, この間で輻輳を検知したら転送レートを下げる制御も行う. すべてのノードが自分と再送サーバ間のリンクについて注意を払うことでネットワーク全体のトラフィックが安定したものになると考えられる. また輻輳の検知からフローを絞るまでの時間が短く済むため, フィードバック情報を用いる方法の欠点を補完する.

3 について, リアルタイム性を高めるにはできるだけ大きなスループットを確保したい. しかしルータに新たな機能を持たせられないためリンクの転送容量に応じてフローを制御するといったことはマルチキャストでは不可能である. そこで AMP では 4 の TCP トラフィックとの公平性を最低限維持しつつネットワーク全体の送出量をできるだけ高めるアプローチを取る. 送信ノードは任意の転送レートでマルチキャストを行い送信が問題なく行われている間はレートを少しずつ上げてゆき輻輳を検知したらレートを下げる.

4.2 メカニズム

AMP における輻輳制御方式は 2 通りの方法を組み合わせて実現する.

4.2.1 送信ノードがトラフィックを監視する方法

送信ノードは自分と最寄りの再送サーバ間の輻輳状態をチェックしフローを制御する. 前提条件として送信ノードは自分と自分が所属する再送リージョンのサーバ間の RTT を知っている必要がある.RTT は加重平均を取り, 新しいものに更新されるようにする. 送信ノードはデータパケット送信と同時にタイマをセットし, 再

送サーバからのACKを受信するまでの時間を計測する. RTTに基づいて設定されたタイムアウト時間を経過した場合, 輻輳と判断しフローを絞る. フローの増減はTCPのような倍数減少やスロースタートをベースに行う. 図3において, メンバmはマルチキャストをしたが自分とサーバ間の経路上にあるリンクで輻輳が発生したため再送サーバからACKが返ってこない状態である. このような場合,mはACKのタイムアウトによりこの輻輳を検知しすぐにフローを絞る.

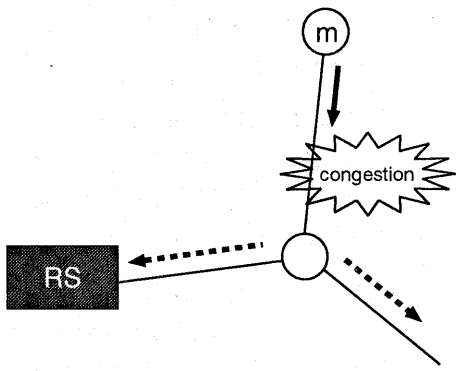

図 3: リンクの監視

4.2.2 フィードバック情報を用いる方法

この方法ではACK,NACKを利用し輻輳制御情報を伝播させ, 複数のノードが輻輳に対し少しずつフローを絞る. 再送サーバはパケットロスを検知したノード及び再送要求を受けたが該当するパケットを保持していなかった隣接サーバからNACKを受信する. 再送サーバは一定数NACKを受信すると輻輳が起きていると判断する. そしてACKまたはNACKに輻輳制御情報として自分の所属するドメインIDと送信元ドメイン別のパケットロス率を付加して隣接サーバに送信し, 同時に再送リージョン内のノードにフローを絞るように通知をする. この通知もACKに付加されて送信される (図4). 輻輳制御情報が付加されているACKまたはNACKを受信した再送サーバはこの輻輳制御情報をそのまま付加したACKまたはNACKの送信と, 再送リージョン内のノードにフロー制御通知を同様に行う.

各ノードフロー制御量は以下のように計算する. ドメイン a,b,c,d,…が送信元であるパケットの送信量を F_a,F_b,F_c,F_d,…, ロス率を L_a,L_b,L_c,L_d,…とすると総フロー量Fとパケットロス量Pは

$$F = F_a + F_b + F_c + F_d + \cdots$$

$$P = F_a \cdot L_a + F_b \cdot L_b + F_c \cdot L_c + F_d \cdot L_d + \cdots$$

と表すことができる. 輻輳制御情報が生成されたときマルチキャストグループは総フロー量を

$$\alpha \cdot F \quad (0 < \alpha < 1)$$

に抑制しようとする. このときドメインxが抑制するパケット量 C_x 及び抑制率 R_x は

$$C_x = \frac{F_x \cdot L_x}{P}$$

$$R_x = \frac{F_x - C_x}{F_x}$$

である. 輻輳制御情報を生成した再送サーバ及びそれを受信した再送サーバは情報をもとにこの計算を行い再送リージョン内のノードにこの抑制率を知らせる. 受け取ったノードは抑制率に従って送信レートを下げる.

5 おわりに

本稿では共有仮想環境のような分散アプリケーションに適用することを目的としたスケーラブルな高信頼マルチキャストAMPの輻輳制御方式を提案した.

従来の高信頼マルチキャストプロトコルでの

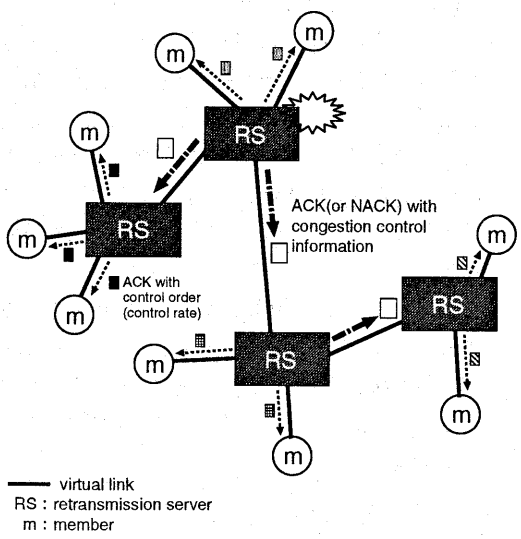

図 4: 輻輳制御情報の伝播

輻輳制御方式にはフィードバック情報を用いる手法やルータで過剰なパケットを廃棄する手法などがあるが, 前者には応答爆発や伝播遅延による輻輳への対応の遅さ, 後者にはルータに新たな機能を追加する必要があるなどの問題点がある. また一般的なマルチキャストの課題としてTCP トラフィックとの公平な共存が挙げられる.

本稿ではAMP の再送システムで用いられるACK,NACK を利用して余分なトラフィックを増やさず簡単なフィードバック情報をやりとりする手法と, 各ノードが自分と最寄りの再送サーバ間のトラフィックを監視して自律的に送信量を調節する手法を組み合わせて,AMP の求める要件に対応できる輻輳制御方式を提案した.

本方式はまだ提案段階であり, 現在はシミュレーションによる評価を行なっている.

参考文献

[1] S.Deering, "Host Extentions for IP Multicasting", Indianapolis,IN:New Riders, (1996).

[2] M.R.Maccedonia et al, "NPSNET: A Network Software Architecture for Large-Scale Virtual Environments", Presence Volume 3,Number 4, (1994)

[3] 細谷篤, 佐藤文明, 水野忠則, "適応的全順序マルチキャストの拡張", 情報処理学会論文誌 Vol.42,No.2,pp.138-146, (2001).

[4] 細谷篤, 佐藤文明, 水野忠則, "適応的全順序マルチキャストの改良", 第 62 回情報処理学会全国大会講演論文集 (3),pp.3-415-3-416, (2001).

[5] 南端邦彦, 佐藤文明, 水野忠則, "共有仮想環境のための高信頼マルチキャスト方式の提案と評価", 情報処理学会論文誌 Vol.41,No.2,pp.254-261, (2000).

[6] 森田悟史, 江崎美仁, 佐藤文明, "高信頼マルチキャストにおける再送木構築方式の提案", マルチメディア, 分散, 協調とモバイル (DICOMO2001) シンポジウム論文集, 情報処理学会シンポジウムシリーズ Vol.2001,No.7,pp429-434, (2000).

[7] 山内長承, 佐野哲央, 城下輝治, 高橋修, "高信頼マルチキャストにおけるフロー・輻輳制御", インターネットコンファレンス'97, (1997).

[8] Luigi Rizzo, "pgmcc: a TCP-friendly single-rate multicast congestion control scheme", ACM SIGCOMM '2000, (2000).

[9] 山口誠, 山本幹, "信頼性マルチキャストにおける Intra-session Fairness を考慮した輻輳制御方式", 信学技報 IN2000-234, (2001).