

Source Specific Multicastによる 新たな通信アーキテクチャ

朝枝 仁

INRIA (国立情報科学制御研究所)

Hitoshi.Asaeda@sophia.inria.fr



多対多の通信を実現するインターネット・マルチキャストの研究の中で、そのルーティング・スケーラビリティを解決するための多くの研究がなされ成果も得られてきた。しかし、インターネットで展開される現実的なマルチキャスト・サービスを考えると、多対多よりも一対多で通信を行う形態の方がより一般的であり、そこで、送信元を指定する新たなマルチキャスト通信モデルを定義することにより、今までとは異なるアプローチでスケーラビリティの解決を図ることが可能となる。

本稿では、現在までに行われてきたマルチキャスト・ルーティングの研究成果と既知の問題を簡単に振り返り、送信者を明示して通信を行う Source Specific Multicast (SSM) の解説、さらに、IP バージョン 4 上で SSM 通信を行うために必要となる IGMP バージョン 3 の概念とその仕様について述べる。

インターネット・マルチキャスト・ルーティングの進化

マルチキャストを用いると、一度のデータ送信で複数の受信者がそのデータを受け取ることができる。しかしこれは一方的にデータを送るブロードキャストとは異なり、受信者が存在するネットワークにだけデータが送信され、それを必要とする受信者のみがそのデータを受け取るという仕組みを実現する。これを可能としているのは、「グループ・アドレス」と「マルチキャスト・ルーティング・プロトコル」である。つまり、受信者は必要なデータのグループ・アドレスに対して参加要求、あるいは受信終了を意味する離脱要求を行い、上位ルータはそれらの情報とマルチキャスト・ルーティング・プロトコルにより得られる経路情報を元に適切なマルチキャスト経路木を構成し、これにより通信としてのマルチキャストを実現する。

ここで、インターネットの標準化団体である IETF の活動をもとに、マルチキャスト・ルーティング技術の推移を簡単に振り返ってみよう。

マルチキャストの研究が始まって以来、インターネット上に仮想的なマルチキャスト・バックボーン、いわゆる MBone を構築してその有意性と実現の可能性が

示されてきた。ここで主に用いられていたルーティング・プロトコルはDVMRP¹⁾であり、これは定期的にすべてのサイトに経路情報を流してマルチキャスト経路木を構成し、受信する必要のないサイトは自らその経路木を刈り取るように要求する "Broadcast and Prune Type" と呼ばれるプロトコルである。それ故に、MBoneに参加する組織が増えMBone自体が大きくなっていくと、経路木を構成するためのトラフィックが急増し、経路木の追加と切り離し（フラッピング）も頻繁に生じ、スムーズなデータ配信ができなくなってしまった。

この問題を解決するために、1995年以降、IETFでは、必要なルータのみが経路木に参加する "Explicit Join Type" のCBT²⁾やPIM-SM³⁾などの仕様が活発に討論された。これは「データ転送を必要とするルータのみがCore Router（PIM-SMの場合にはRendezvous Point）と呼ばれるルータを中心とする経路木に参加する」「データの送信者が特定された後、その送信者への最適な経路からなる経路木を再構成する」などの特長を持っている。さまざまな議論の結果、IETFの中ではPIM-SMが標準的なマルチキャスト・ルーティング・プロトコルとして採択されることになり、ルータ・ベンダなどの実装も進み、一部のISPやMBoneを含む実験ネットワークなどで利用され始めた。

しかし、ここで得られた効果的な成果も、インターネットという非常に広範囲なネットワークに対してはさらなる研究の余地が残されていた。そこで、1998年から1999年にかけ、"Inter-domain Multicast" をキーワードとして、Border Gateway Multicast Protocol (BGMP) やSimple Multicastといった新しいデザインのプロトコル、それにPIM-SMと組み合わせて用いるMSDP⁴⁾が提案された。BGMPもSimple Multicastもすばらしい部分は数多く存在したが、完全に新しいMulticast Address Allocationの仕組みを要求する点 (BGMP) やHost Modelの変更を要求する点 (Simple Multicast) などの大々的な変更を避けたいとの意向から、これらは「長期的な解決策」として認識されることになり、最終的にIETFでは、PIM-SM/MSDPにReverse Path Forwarding (RPF) のポリシーを追加するためのMBGP⁵⁾を組み合わせたモデルが選択され、現在も標準化活動が続けられている^{4), 6)}。

アプローチの異なるものとして、最近では、マルチキャストのサービスを別の形で表現するXcast⁷⁾などの研究も始まっている。これは上記にある「グループ・アドレス」と「マルチキャスト・ルーティング・プロトコル」を用いずにマルチキャストがもたらす「サービス」を実現する手法で、主に受信者の数があまり多くなく、かつ特定できるケースをそのターゲットとしている。

PIM-SM/MSDP

現時点では最も標準的なInter-domain Multicastと認識されているPIM-SM/MSDPの概要を述べる。PIM-SMではグループ・アドレス、あるいはグループ・プレフィックス単位にRendezvous Point（以下RP）を中心とした "Rendezvous Point Tree (RPT)" と呼ばれる経路木を構成し、必要なルータのみがその経路木に参加する。その後、各ルータは、ある一定量のデータ受信の後、各々の送信者のアドレスを中心とする "Shortest Path Tree (SPT)" と呼ばれる最適な経路木を構成し直す。このようなプロセスの中で、広域なネットワークでの使用で最も問題になるのはRPTに関する部分である。すなわち、

1. Third Party Dependency
2. Traffic Concentration
3. Bootstrap Scalability

といわれるもので、具体的には、

1. 経路木の頂点（ルート）となるRPが他のISPや遠方の組織にあった場合、そこまでのネットワーク到達性が悪いとSPTへのスイッチが適切なタイミングで行われなかったり、RPTから受け取るデータの品質も悪くなる。
2. データや経路交換のトラフィックがRPおよびその周囲のネットワークに集中してしまうため、局所的な高負荷状態が続く。
3. RPの可用性を高めるためRendezvous Point候補ルータ（以下CRP）を複数の個所（サイト）に配置することができるが、その中から実際に使用される1つのRPを選びそれをネットワーク全体に告知するプロセスがトラフィック量を高めてしまう。

というものである。

このような問題を解決するためにMSDPが提案された。MSDPでは、まずPIM-SMネットワーク（PIM Domain）を図-1のように論理的に複数に分割する。そして各々のPIM Domainとの境目に境界ルータ（Border Router）を設置し、PIM Domain内にはそのDomain専用のCRPを設置する。これにより、上記の問題が生じない程度の大きさの「島」ができることになる。しかしこれだけだとPIM Domainを越えた通信ができない。つまり、自サイトのRPTは異なるDomainで利用可能なデータの存在が分からないためそのデータはRPTを流

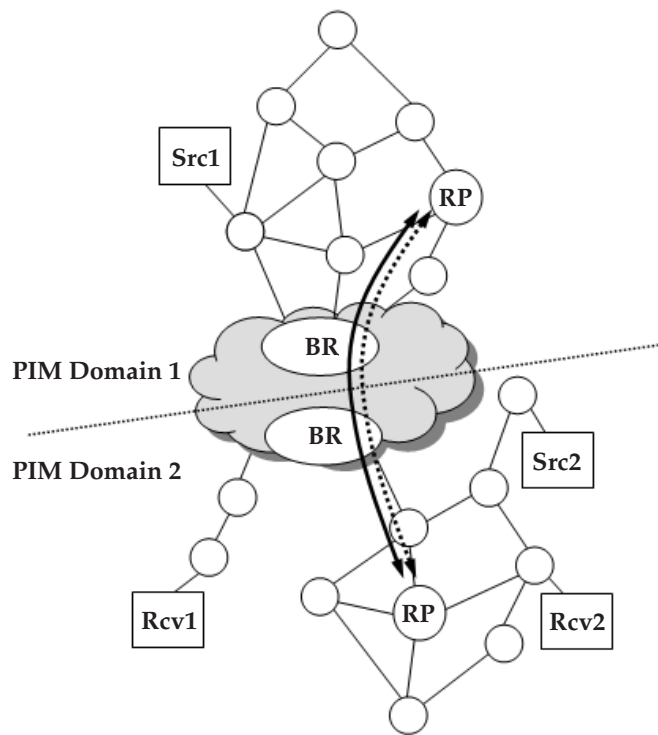


図-1 PIM-SM と MSDP Peering

れて来ない。そこで「架け橋」としてTCPセッション (MSDP Peering) を各々のCRP間で設定し (図-1の実線矢印☆1), そこにデータや利用可能な送信者の情報 (Source Active (SA) Message (図-1の点線矢印)) を流してPIM Domainを越えたマルチキャスト通信も可能とする。

MSDPはPIM-SMの仕様変更を要求せずにそれを非常に広域なインターネットでも使用できる状態にした。しかし、やはりスケーラビリティという点では完全ではない。たとえばSA Messageに関して、これは "Peer-RPF Flooding" と呼ばれるように、すべてのMSDP Peeringルータに配送されなければならない。もちろん、SA Messageの正当性を確認する手法や、そのデータを必要としないルータは自サイト内の配送木にその経路を追加しないなどの機能はプロトコルの一部として組み込まれてはいるが、無駄なトラフィックを完全に防ぐことはできない。さらに、PIM Domain内の各ルータがSPTにスイッチするまでは他サイトのデータはMSDP Peerを通して配送されるため、結局RP自身やその周囲のルータに対する負荷は高いままとなってしまう。これらはRPTを用いる上で避けることのできない問題と考えられる。

☆1 図では2つのRPを結んでいるが、実際は各DomainのCRP間を結ぶ。

Source Specific Multicast (SSM)

ここでもう一度マルチキャスト・ルーティングの問題点を整理してみよう。第1に、ここまで述べてきたルーティング・スケーラビリティに対する問題が最大ののものであろう。そして、そのプロトコルの複雑性ゆえにルーティング障害が生じた時に原因究明が困難であるという点、つまり、オペレーションに対する問題もその1つである。さらに、グループ・アドレスというマルチキャスト特有の識別子を管理する難しさ、そこから派生する認証などを含むセキュリティの問題などもあげられる。

これらの問題を正しく認識した時、マルチキャストが対象とするサービス・モデルを再考することができる。インターネットで展開するためのマルチキャスト、すなわちInter-domain Multicastを考えると、誰もがデータの送信者になることのできる多対多の「ミーティング型」通信形態よりも、ウェブ・アプリケーションのように送信者が固定である一対多の「放送型」通信形態の方がより一般的である。もしマルチキャスト・ルーティングを一対多の通信を前提として考えたならば、異なるアプローチで既存の問題を解決し、マルチキャストを現実的なアプリケーション、あるいはサービスとして展開することができる。1999年の後半以降、

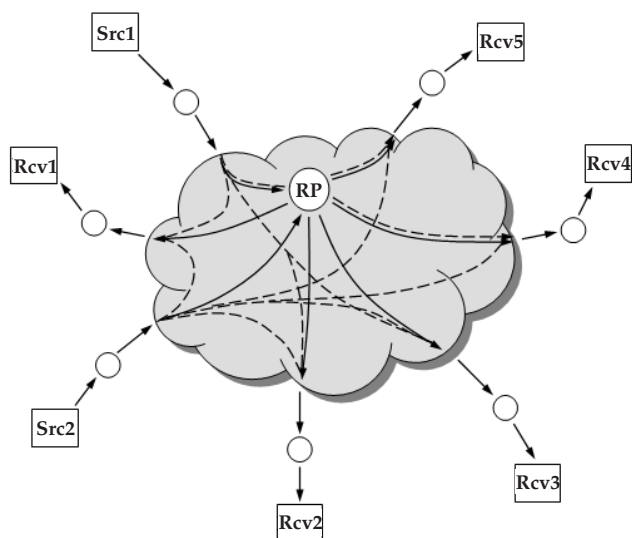


図2 RPTから形成されるマルチキャスト経路木

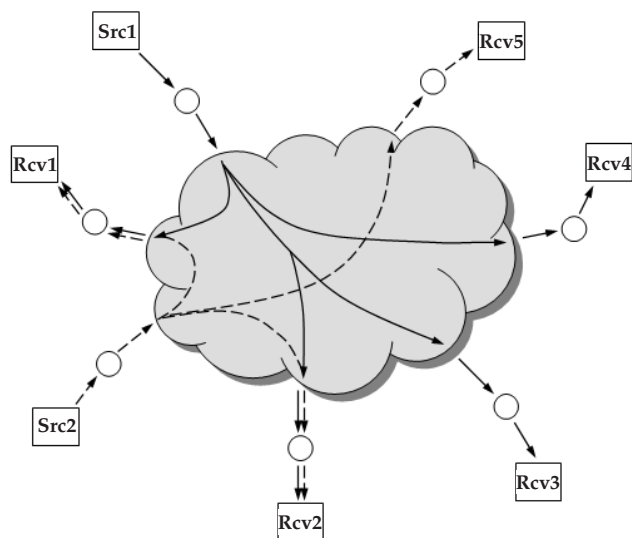


図3 SSMによるマルチキャスト経路木

こういった議論から始まった新しいアーキテクチャが Source Specific Multicast（SSM）といわれるものである。

PCなどのエンド・ノードがマルチキャスト・データの受信を要求する際、IPバージョン4上ではIGMPというプロトコルを用いてそのグループ・アドレス（以下 Group Address）に対してIGMPグループ参加要求（以下Join）を行う。逆に、受信を終了させたい時にはIGMPグループ離脱要求（以下Leave）を行う。これらは自らのアプリケーション（Socket）の登録や解放、ネットワーク・インタフェースに対するフィルタリングのためだけではなく、上位ルータにGroup Addressを通知してデータの転送を促す役目を持っている。多対多のマルチキャスト通信を行うための（*, G）Join/Leaveと呼ばれるこれらのメッセージはGroup Addressだけを指定して交換され、（*, G）Joinメッセージを受け取った上位ルータは、たとえばそのネットワークが従来のPIM-SMで構成されていたならば、同じGroup Addressを扱うRPに向けてPIM（*, G）Joinメッセージを転送してRPT（図2の実線矢印）を形成し、その後、送信者と受信者間のルータで構成される最適なSPT（図2の点線矢印）を形成し直しRP経由の経路を解放する。この時、各エンド・ノードが送信者のIPアドレス（以下Source Address）を知っているか否か、またそれを上位ルータに通知するか否かは必要条件で

はなかった。しかしSSMでは、Group AddressだけではなくSource Addressを明示した（S, G）Join/Leaveを行うことを前提とする。これにより、SSM拡張している上位ルータは、多くの場合、（S, G）Joinを受信してもRPTを形成する必要はない^{☆2}。なぜならば、RPTは同じGroup Addressを用いている送信データを1つに集めて配信する、いわば集約経路の役目を持っていたが、SSMでは直接Source Addressに対するSPTを形成することが可能となるからである。図3では、Rcv1とRcv2はSrc1のデータ（実線矢印）とSrc2のデータ（点線矢印）を受信しており、Rcv3とRcv4はSrc1の、Rcv5はSrc2のデータを受信している状態を示している。図2に比べ経路木がシンプルになっていることが分かる。つまり、エンド・ノードが受け取りたいSource Addressを指定した（S, G）Join/Leaveを行う限りRPTは必要なくなり、同時に上記の経路木の構成は簡素化されて今まで述べたRPTに対する問題も解消できる。

現在、新しいPIM-SM⁶⁾ではSSMの機能追加も含めて研究が継続されている。

☆2 ルータは同一LAN上の各エンド・ノードのFilterを統合して管理するが、その統合したFilterがEXCLUDE（後述）の時はRPTを形成する必要も出てくる。詳細は6)を参照されたい。

○Source Address Filtering

以上のように、SSMアーキテクチャへの対応はマルチキャスト・ルータの機能追加だけではなく、エンド・ノードに対しても「必要なSource Addressの明示」を要求する。そしてその機能を満たすのがIGMPの新しいバージョン、IGMPv3⁸⁾であり、これは、Source Addressを明示した(S, G) Join/Leaveを提供する。ここで書かれた(S, G)の"S"は表記上1つのSource Addressを示しているが、IGMPv3では、同時に複数のSource Addressを指定することができる。さらに、Source Addressの識別の方法(Filter)としてはINCLUDEあるいはEXCLUDEのいずれかを指定し、INCLUDEが「受け取りたいSource Address」を要求するのに対し、EXCLUDEでは「受け取りたくないSource Address」を意味するJoin要求を送ることが可能となっている。

ではここでSource Filterの指定の仕方を説明しよう。IGMPv3仕様⁸⁾にあるように、アプリケーションは

```
IPMulticastListen (socket, interface,  
                  multicast-address,  
                  filter-mode, source-list)
```

という形で表される。これはSocketごとに実行されるため、エンド・ノードのOS内部(以下Kernel)でinterfaceとmulticast-addressを単位に全Socketの要求を統合する。今、(i, m)という組に対して、s1というSocketが送信者aとbからのデータを要求するINCLUDE({a, b})のJoin要求を、s2が送信者b, c, dからのデータを要求するINCLUDE({b, c, d})のJoin要求をしたとする。この時、

定義 1. INCLUDE全体のSource Listは各Socketが受信要求をしたSource Addressの和集合により得られる。

から、Kernelが保持するMulticast Source Filterの状態(以下MSF State)はINCLUDE({a, b, c, d})になる。次に、s3とs4というSocketが同じ(i, m)に対して、それぞれc, d, e以外の送信者からのデータを要求するEXCLUDE({c, d, e})とd, e, f以外の送信者からのデータを要求するEXCLUDE({d, e, f})というJoin要求を行うと、

定義 2. EXCLUDE全体のSource Listは各Socketが非受信要求をしたSource Addressの積集合により得られる。

から、EXCLUDE({d, e})のEXCLUDE Source Listが得られる。この時点でKernelはINCLUDEとEXCLUDE両方のSource Listを持っているため、

定義 3. 1つでもSocketがEXCLUDEの要求をしている場合にはMSF StateはEXCLUDEになる。

および

定義 4. MSF StateがEXCLUDEの時には、MSF StateのSource ListはEXCLUDE Source Listの要素からINCLUDE Source Listの要素を差し引いたものになる。

により、最終的にKernel自身の(i, m)に対するMSF Stateは

$$\text{EXCLUDE}(\{d, e\}) - \text{INCLUDE}(\{a, b, c, d\}) = \text{EXCLUDE}(\{e\})$$

となる。

○新しいQueryおよびReport

IGMPv3では、QueryとReportの両方のMessageでSource Addressを明示できるよう、メッセージ・フォーマットに工夫が施された。フォーマットそのものはIGMPv3仕様を見ていただくとして、ここではその種類と特長について述べる。

まずReport Messageであるが、Group Addressに加えSource AddressとMSF Stateを通知できるようになった。その種類は大きく"Current-State Report"と"Filter-Mode-Change Report"の2種類に分けられる。前者はQuery Messageに対する現時点のKernelのMSF Stateを通知するReportになり、後者はMSF Stateの変更が生じた時にその変更内容を通知するReportとなる。これらはすべて複数のGroup Addressに対するレコードを一括して送ることができるようになっている。

Query Messageに関しては、新しく"Group-and-Source-Specific Query"というタイプのMessageがサポートされ、1つのGroup Addressと複数のSource Addressをペアにして「受信を必要とする」Source Listを問い合わせる機能が追加された。たとえば、エンド・ノードがmというGroup Addressとa, bという

Source AddressからなるGroup-and-Source-Specific Queryを受信した時には、mに対するKernelのMSF StateがINCLUDE ({a,c})ならばaだけを、EXCLUDE ({e})ならばa,b共にReport Messageに入れて返答する。

ところで、この新しいタイプを含め、Query Messageは従来通り、1つのGroup Addressしか指定できないようになっている。この理由は、IGMPv3をサポートしていないエンド・ノードもあらゆるIGMPv3 Query Messageを今までの(*,G) Queryと認識して返答させたいという下位互換性への配慮にある。それとは異なり、IGMPv3 Report MessageのCurrent-State ReportとFilter-Mode-Change Reportは完全に新しいフォーマットのReport Messageであり、それらを(IGMPv3エンド・ノードも含めて)ルータ以外が受け取る必要がないため、宛先アドレス(Destination Address)も224.0.0.22という新規のものが割り当てられた。このため、上位ルータがIGMPv3をサポートしていない場合はこれらのReport Messageを解釈できないので、IGMPv3エンド・ノードはルータから送られて来るQuery Messageのフィールドの値や長さから上位ルータが使用しているIGMPのバージョンを読み取り、そのバージョンに適したReport Messageを送るような機能を備えることになっている。

○Multicast Source Filters API

Group AddressへのJoinやLeaveはアプリケーションの命令により実行されるため、Source Addressを指定するApplication Program Interface (API)も新たに必要となる。MSF拡張仕様⁹⁾では、新たなSocketの操作方法としてset/getsockopt(), およびioctl()用にOptionが追加され、これによりKernelに向けてSSMオペレーションが指示される。Optionの種類はIPv4用のIPv4 MSF APIとIPのバージョンに依存しないProtocol-Independent MSF APIに大別され、さらにそれぞれ、要求されるGroup AddressとSource Addressが1対1のDelta-base APIと、複数のSource Addressをリスト形式で渡せるFull-state APIが存在する。Delta-base APIは既存のSocket操作に非常に近いため、SSM対応させたいアプリケーションの変更を最小限にとどめたい時には有効である。

通信と放送の融合に向けて

SSMはすべてのエンド・ノードにも多くの変更を要するため、その環境に移行する時間はそれほど短くはないかもしれない。しかしこれが実現できると、インターネットを技術的にもビジネス的にも新たな方向に導ける可能性がある。この時、サービス・モデルの一例として、インターネット上で通信と放送が融合できる基盤が得られると考えるのは早計であろうか。

将来に向けてインターネット・マルチキャストのさらなる研究や実装が進み、次への新たな展望が見出されることを期待したい。

参考文献

- 1) Waitzman, D., Partridge, C. and Deering, S.: Distance Vector Multicast Routing Protocol, RFC1075 (1988) .
- 2) Ballardie, A.: Core Based Trees (CBT version 2) Multicast Routing, RFC2189 (1997) .
- 3) Estrin, D. et al.: Protocol Independent Multicast - Sparse Mode (PIM-SM) : Protocol Specification, RFC2362 (1998) .
- 4) Meyer, D. and Fenner, B.: Multicast Source Discovery Protocol (MSDP) , Internet Draft - work in progress (2001) .
- 5) Bates, T. et al.: Multiprotocol Extensions for BGP-4, RFC2858 (2000) .
- 6) Fenner, B. et al.: Protocol Independent Multicast - Sparse Mode (PIM-SM) : Protocol Specification (Revised) , Internet Draft - work in progress (2001) .
- 7) Boivie, R. et al.: Explicit Multicast (Xcast) Basic Specification, Internet Draft - work in progress (2001) .
- 8) Cain, B. et al.: Internet Group Management Protocol, Version 3, Internet Draft - work in progress (2001) .
- 9) Thaler, D., Fenner, B. and Quinn, B.: Socket Interface Extensions for Multicast Source Filters, Internet Draft - work in progress (2001). (平成14年2月4日受付)

