情報処理学会論文誌 Vol.54 No.3 1061–1070 (Mar. 2013)

移動透過IPマルチキャストに対応するグローバルライブマイグレーションの設計と性能評価

鎌田 恵介[1,†1] 近堂 徹[2,a)] 西村 浩二[2] 相原 玲二[2]

受付日 2012年6月29日, 採録日 2012年12月7日

概要：仮想化技術の発展を背景に，継続的なシステム運用を目的として，規模や負荷に応じて仮想計算機の構成を動的に変更できるライブマイグレーションが注目されている．また，同一ネットワーク内に限定されているライブマイグレーションを拡張し，IP 移動透過性を持つグローバルライブマイグレーションも提案されている．しかし既存方式では，その対象が IP ユニキャスト通信のみであり，IP マルチキャストの移動透過性まで実現できていない．そこで本研究では，仮想計算機上でマルチキャストとユニキャストを組み合わせて利用する移動透過 IP マルチキャスト方式について提案し，プロトタイプシステムを用いた基本性能評価について述べる．評価結果より，ネットワークセグメントが異なるマイグレーションにおいても，同一ネットワーク内のマイグレーションと同等の途絶時間に短縮できることを確認した．

キーワード：仮想計算機，IP 移動透過性，ライブマイグレーション

Design and Evaluation of Global Live Migration with Mobility Support for IP Multicast

Keisuke Kamada[1,†1] Tohru Kondo[2,a)] Kouji Nishimura[2] Reiji Aibara[2]

Received: June 29, 2012, Accepted: December 7, 2012

Abstract: Virtualization technologies are widely used, and include Live Migration which changes resource allocation for virtual machines according to scale or load. In addition, Global Live Migration with IP mobility is also proposed. It enables migration among distributed sites and provides continuation even if the network of the virtual machines was changed. However, it supports only IP unicast communication, not IP multicast communication. In this paper, we propose a mobility support mechanism for IP multicast on virtual machines, and evaluate its basic performance using prototype system. As a result, the proposed method provides migration function continuously receiving multicast stream stopped same as traditional Live Migration.

Keywords: virtual machine, IP mobility, Live Migration

1. はじめに

インターネットユーザの増加にともない，ネットワークトラフィックは日々増加している．アクセスが集中するサーバに対しては，並列化・冗長化による処理速度の向上が行われてきた．しかし，サーバが物理的に遠距離にある場合など，伝送速度や途中経路のルータ処理による遅延がレスポンスを悪化させている．

これを解消する方法として，近年 CDN（Contents Delivery Network）と呼ばれる配信網が利用されている．CDN は，あらかじめアクセスが集中しやすい大容量コンテンツを世界中のキャッシュサーバに分散配置しておき，リクエストしたユーザに直近のキャッシュサーバからデータを

[1] 広島大学大学院工学研究科
Graduate School of Engineering, Hiroshima University, Higashi-hiroshima, Hiroshima 739–8527, Japan
[2] 広島大学情報メディア教育研究センター
Information Media Center, Hiroshima University, Higashi-hiroshima, Hiroshima 739–8511, Japan
[†1] 現在，日本アイ・ビー・エム株式会社
Presently with IBM Japan
[a)] tkondo@hiroshima-u.ac.jp

送信することによって，負荷の集中を解消しレスポンスを向上させている．これらの通信方式としてはユニキャストが使用されることが多いが，CDN のような管理されたネットワークにおいてはマルチキャストを利用することで，網内の帯域効率化を図ることができる．さらに，マルチキャストアドレスによってコンテンツを識別し，IGMP [1]，MLD [2]，PIM [3] などのプロトコルにより IP 層で配送ツリーを構築してパケットが中継されるため，動的なキャッシュノードの配置に対してもスケーラビリティを確保できる．

一方で近年の仮想化技術の発展により，継続的なシステム運用を目的として，各種サーバを仮想計算機上に構築する仮想化技術が注目されている．仮想化技術の特徴として，セッション状態などアプリケーションの内部状態を保持したまま仮想計算機を任意の実計算機に移動させることが可能である．しかしながら，ネットワークセグメントが異なる広域ネットワーク環境でこれを展開しようとする場合，IP 層での移動透過性をサポートしなければ通信の継続ができない [4]．この問題に対してグローバルライブマイグレーション [5] が提案されており，セッションを引き継いだまま動的な配置転換ができるものの，ユニキャスト通信のみ（**ユニキャストモビリティ**）に限定されておりマルチキャストに対する移動透過性（**マルチキャストモビリティ**）は維持されない．マルチキャストの移動透過性を実現するには，移行先ネットワークにおけるマルチキャストツリーの再構築を早期に完了させることと，移行先ネットワークがユニキャストに限定されたエリアの場合への対応が必要となる．

そこで本論文では，継続的なシステム運用を目的として，仮想計算機（以下，VM; Virtual Machine）がグローバルライブマイグレーションを行う際，VM を稼働させる実計算機（以下，VMS; Virtual Machine Server）でマルチキャスト受信が不可能なネットワーク上に設置されているケースを考慮した，IP ユニキャスト/マルチキャスト両対応のライブマイグレーション手法を設計する．具体的にはユニキャストについて，既存の IP モビリティのうち経路最適化に優れた MAT [6] を用い，マルチキャストについては，VMS 単位に設置するエージェントノードが MLD の先行送信と IP マルチキャスト非対応ネットワークにおけるユニキャスト中継処理を自律的に行う．これにより，VM の移行先ネットワークのマルチキャスト対応状況によらず，VM 上のアプリケーションは継続的なマルチキャストストリームの受信が可能となる．

本研究が想定するアプリケーションの 1 つとして，先に示したような，CDN における配信ノードの動的配置によるリアルタイムストリームの広域配信が考えられる．たとえば，CDN 網内は VM で動作する配信ノードが，マルチキャスト通信を行いながら広域ネットワーク間でグローバルライブマイグレーションを行うことで，配送の継続性を確保しながらネットワーク帯域の効率化を実現できる．

以下に，本論文の残りの構成を示す．2 章は関連技術について概説し，3 章は提案手法の設計について，4 章は実装について述べる．そして，5 章ではマルチキャストマイグレーションにおける性能評価と考察を行う．最後に，6 章でまとめと今後の課題について述べる．

2. 関連研究

2.1 仮想計算機を用いた CDN の構築

実計算機を用いて構成された既存の CDN では，各サーバの負荷や帯域に応じた動的な構成変更は困難であった．この問題を，VM を用いることによって解決した手法 [7], [8] が提案されており，VM の構成を動的に変更しながらノードを追加・削除することで柔軟な運用が可能である．しかし，VM 上で動作するアプリケーションは実計算機と同様に独立しているため，たとえば高負荷時に VM が追加されたとしても，アプリケーションがセッションを切り替えない限り負荷を軽減できないという問題点がある．

2.2 仮想計算機とグローバルライブマイグレーション

2.1 節で述べた問題を解決する方法として，ライブマイグレーションがあげられる．ライブマイグレーションは，VM のメモリ情報を VMS 間で移行させることにより，VM 内のアプリケーションセッションを保持したまま，最小限の停止時間で動的な計算資源の変更が可能となる．しかし，移行後のネットワークアドレス設定についてはサポートされていない．したがって，異なるサブネットへマイグレーションを行うと到達性がなくなり，アプリケーション層で確立されたセッションが切断されてしまうため，マイグレーションは同一サブネット内のみという制限があった．

この問題に対し，IP モビリティと複数インタフェースを用いたグローバルライブマイグレーション [5] が提案されているが，提案手法ではユニキャスト通信に限定される．VM がマルチキャストを受信している際にグローバルライブマイグレーションを行ったとき，移行先ネットワークに同一のマルチキャストストリームが配送されていない場合には VM 内のマルチキャストアプリケーションに通信途絶が発生してしまう．途絶時間はネットワークによって異なるが，マルチキャストストリームを受信するためには，希望するマルチキャストの Multicast Listener Discovery（以下，MLD）を直近のルータと交換し，そのルータはさらに上位のルータと PIM [3] などを使ってマルチキャストツリーを構築する必要がある．マルチキャストアプリケーションが受信状態を変更する場合以外にも，ノードのマルチキャスト受信状態を報告する Multicast Listener Report（以下，MLD Report）が送信されるが，その間隔はルータが定期的に送信する Multicast Listener Query（以下，

MLD Query）と等しい．MLD Query の送信間隔はデフォルト値が 125 秒と定義 [2] されており，マルチキャストツリーの構築には一般的に数秒から数分程度必要である．すなわち，継続的なマルチキャスト受信のためには，VM がマイグレーションを行う前に，VM が受信中のマルチキャストに関するツリーを移行先ネットワークで構築しなければならない．

3. 移動透過マルチキャストに対応するグローバルライブマイグレーション

3.1 提案手法の考え方

マルチキャストモビリティに対応するグローバルライブマイグレーションを実現するには，次の 2 点が重要である．

1 つ目は，グローバルライブマイグレーションが完了するまでに移行先ネットワークでマルチキャストが受信可能となっていることである．したがって，マイグレーション時点で該当 VM が参加しているマルチキャストを把握し，移行先ネットワークのルータに対しても MLD Report を通知しておく必要がある．MLD はリンクローカルなプロトコルであり，該当 VM はマイグレーションが完了するまで移行先ネットワークに接続されないことから，他のノードが代理で参加要求を行う．

2 つ目は，移行先ネットワークにおいてマルチキャストが受信できない場合に，ユニキャストを用いた伝送方式に切り替えたうえで，VM にそれを隠蔽することである．ユニキャスト中継を VM に直接行った場合，VM ではマルチキャストを直接受信する場合とは異なる宛先アドレスとして受信されるため，その対応関係をアプリケーションが管理しなければならない．さらに，多数の VM が中継のために同一のユニキャストセッションを複数確立して輻輳を発生させないよう，ユニキャスト中継を VMS 単位で管理し，セッション数を受信ノード数に比例させないように考慮する必要がある．

これらの要素技術として，筆者らが提案しているユニキャストを併用する移動透過 IP マルチキャスト [9] を用い，それを仮想化環境に拡張することにより実現する．

3.2 システムの全体構成

提案手法では，**図 1** に示すように VMS と異なるネットワークにマルチキャストの送信ノードが存在し，VMS 上のゲスト VM（以下，VM）がマルチキャストに参加するというモデルを想定する．そして，各 VMS に管理エージェント（以下，Agent）を収容する Agent VM を設置する．Agent は，同一 VMS 内のすべての VM のマルチキャストモビリティを維持するために 3.4 節で示す動作を行う．本論文では，SrcVMS（Source VMS; マイグレーション前の VMS）上の Agent を SrcAgent，DstVMS（Destination VMS; マイグレーション後の VMS）上の Agent を DstAgent と定義する．

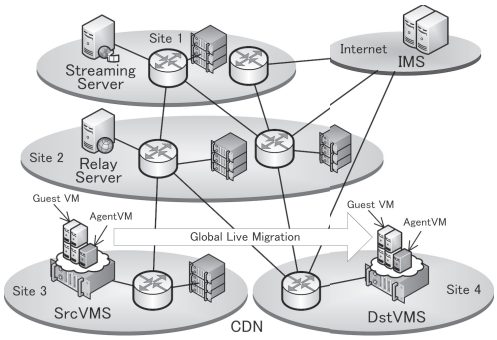

図 1 ネットワーク構成例
Fig. 1 An example of network structure.

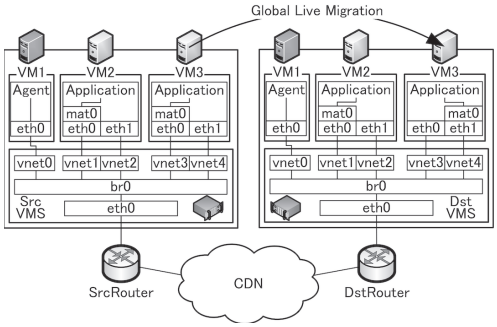

図 2 VMS 内部の構成
Fig. 2 Internal structure of VMS.

VMS 内部のネットワーク構成は，**図 2** に示すようなブリッジ構成とする．一般的な仮想化ソフトウェアでは，ブリッジ接続のほかに NAPT 接続にも対応するが，本提案手法では VM がマルチキャストに参加できることが前提のため，ブリッジを経由して外部ネットワークに接続する．VM 内のユニキャストアプリケーションは，ルーティングテーブルにより後述する移動透過アーキテクチャ MAT の仮想インタフェース `mat0` にいったん集約され，グローバル接続が可能なインタフェースから送受信される．一方，マルチキャストアプリケーションは `mat0` に対して参加処理を行い，ユニキャスト通信に使用されるインタフェースと同一のインタフェースから送受信される．VM 上のインタフェース `eth0` や `eth1` は，VMS 上では `vnet0` といった TAP デバイスと 1 対 1 に対応し，VMS 上のブリッジデバイス `br0` を経由して物理インタフェース `eth0` へと接続される．

3.3 基本動作概要

マルチキャストに対応したグローバルライブマイグレーションでは，マルチキャストとユニキャストそれぞれが独立した手順でモビリティの確保を行う．マルチキャストモビリティに関する対応を 3.3.1 項に，ユニキャストモビリティに関する対応を 3.3.2 項に述べる．

表 1 新しく定義したメッセージプロトコル
Table 1 Defined messages for proposed protocol.

メッセージタイプ	名称	概要
AGENT_SEARCH	Agent 探索	自 VMS の Agent を特定するためのリンクローカル全ノードマルチキャスト.
AGENT_SEARCH_REPLY	Agent 探索応答	自 VMS が送信した VM を格納している場合に応答する Agent の応答.
MIGRATION_REQUEST	マイグレーション要求	VM がマイグレーション先 VMS の DstAgent を SrcAgent に通知する.
MIGRATION_RESPONSE	マイグレーション応答	ACK および，移行先ネットワークのプレフィックス情報を通知する.
AGENT_REQUEST	Agent 代理参加要求	DstAgent が代理参加すべきマルチキャストアドレスを通知する.
AGENT_RESPONSE	Agent 代理参加応答	移行先ネットワークのプレフィックス情報および，マルチキャストアドレスごとに DstAgent がとる受信方法を通知する.
START_MIGRATION	マイグレーション開始	ライブマイグレーションの開始を指示する.（libvirt API）
RELAY_REQUEST	ユニキャスト中継要求	マルチキャストを受信しユニキャストで再送信するユニキャスト中継の要請.
RELAY_RESPONSE	ユニキャスト中継応答	ユニキャスト中継要求への ACK.

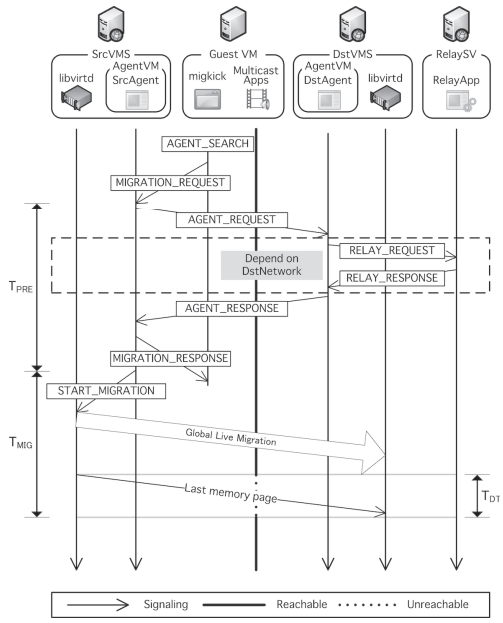

図 3 提案手法によるマルチキャストマイグレーションの流れ
Fig. 3 Sequence on proposed multicast migration method.

3.3.1 マルチキャストモビリティ

提案手法におけるマルチキャストマイグレーションの流れを図 3[*1]に示す．VM には migkick というマイグレーションアプリケーションを用意する．migkick が起動点となり，SrcAgent に対してマイグレーション要求を送信したり，移行に必要な情報を受信する役割を担ったりする．本提案手法のために定義したメッセージを表 1 に示す．

以下に動作手順について説明する．

(1) VM から SrcAgent へのマイグレーション要求

マイグレーション対象の VM が現在受信しているマルチキャストを移行先ネットワークでも継続的に受信できるようにするために，migkick が SrcVMS 上の SrcAgent を検索（AGENT_SEARCH）し，検出された SrcAgent に対してマイグレーション要求（MIGRATION_REQUEST）を送信する．

*1 図中の T_{PRE}，T_{MIG}，T_{DT} については，5 章の評価実験において参照する．

(2) Agent 間でのマルチキャスト代理参加処理

マイグレーション要求を受信した SrcAgent は，DstVMS 上の DstAgent に対してマルチキャストの代理参加と移行先ネットワークのプレフィックス情報（プレフィックスおよびプレフィックス長）を要求（AGENT_REQUEST）する．要求を受信した DstAgent は自ネットワークでマルチキャストストリーム受信の可否を調べる．マルチキャストが受信可能である場合，代理参加要求を受け取ってからストリームが受信できた時点で応答を返すことが可能だが，受信不可を判断するにはタイムアウトを設定して判断するしかない [9]．ゆえに，DstAgent は一定時間内にストリームが受信できた場合は，その可否とプレフィックス情報を応答（AGENT_RESPONSE）する．もし，受信できないと判断した場合には，ユニキャスト中継が可能な中継サーバに対し中継の依頼（RELAY_REQUEST，RELAY_RESPONSE）を行う．中継サーバから送られるユニキャストを DstAgent 自身が代理受信し，自ネットワークにマルチキャストとして再送信する．これによりマルチキャストを継続的に受信できる環境を準備したうえで，SrcAgent に対して応答（AGENT_RESPONSE）する．なお，ユニキャスト中継の詳細については，3.4.3 項で後述する．

(3) SrcAgent から VM へのマイグレーション応答

SrcAgent は DstAgent からの Agent 代理参加応答を受信することで，移行先ネットワークでのストリーム受信の準備が完了したことを把握する．その後，SrcAgent はマイグレーション要求を出した VM に対してプレフィックス情報を応答（MIGRATION_RESPONSE）し，VM は現在通信に使用していないインタフェース（初期状態であれば eth1）に対して，RA [10] の受信を無効化したうえで，新しいプレフィックス情報を用いてユニキャストアドレスを設定する．以上の処理を経て，移行先ネットワークで該当マルチキャスト受信と VM に対する移行先ユニキャストアドレス付与が完了する．

(4) ライブマイグレーションの開始

SrcAgent は VM にマイグレーション応答を返した後，SrcVMS に対してマイグレーションの開始を指示（START_MIGRATION）し，ライブマイグレーションが開始される．本提案手法では Pre-copy 型 [11] のライブマイグレーションを前提とする．本方式は，マイグレーションが開始されると VM を止めずにメモリの内容を移行先に転送し，転送処理の間に VM がさらに更新したメモリページ（ダーティページ）を反復的に再度転送することで，変化するメモリ情報を欠落させることなく転送させる方式である．マイグレーションの最終段階では，最後のメモリページやプロセス，I/O デバイス状態のコピーのために一時的に VM がサスペンドするが，移行先ではあらかじめ準備されたマルチキャストストリームが受信可能であるため，レジューム後はマルチキャストの移動透過性が実現される．

3.3.2 ユニキャストモビリティ

提案手法ではユニキャストの移動透過性を実現するためのアーキテクチャとして MAT [6] を使用し，文献 [5] と同様に複数インタフェース構成を利用し，VM 上でハンドオーバを行うことにより IP 層でのモビリティを確保する．

提案手法におけるユニキャストマイグレーションの流れを図 4 に示す．マルチキャストモビリティ同様に，migkick がマイグレーション起動点となる．マイグレーション前の VM が eth0 経由で通信しているとする[*2]と，マイグレーションを行う前に移行先ネットワークの IP アドレス・ルーティングを eth1 に設定する．マイグレーション操作の最後でサスペンドされるまでは mat0 を通じて eth0 経由でアクセスし，VM が DstVMS 上でレジュームされた後に素早く eth1 経由に切り替えることにより，アドレス取得にかかる時間を短縮できる．最後に IMS に対してマッピングの更新を行うことで，ユニキャストの移動透過性が実現される．

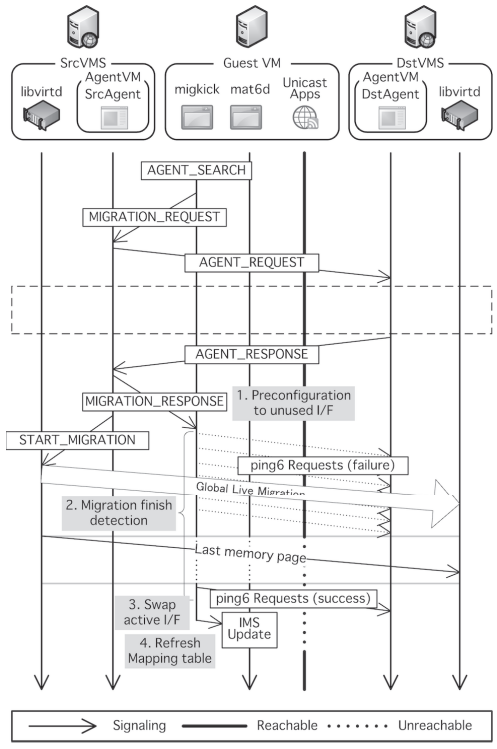

図 4 提案手法によるユニキャストマイグレーションの流れ
Fig. 4 Sequence on proposed unicast migration method.

[*2] 初期状態が eth1 経由の通信の場合，eth0 を eth1 に，eth1 を eth0 に置き換えた同様の手順となる．

ここで問題となるのは，VM 自身がレジュームしたタイミングをどのように検知するかである．提案手法では ICMPv6 を用いて DstAgent との到達性を確認する方法によって解決した．これは eth1 に移行先ネットワークのアドレスを設定すると，そのアドレスと同一リンク上の DstAgent に対するルーティングは eth0 よりも eth1 が上位となり，eth1 経由のルーティングはマイグレーション後のみ有効となる特徴を利用している．

3.4 管理エージェントの動作

VM のマルチキャストモビリティを確保するために，VMS 上で動作する Agent は以下の 3 種類の役割を担う．

3.4.1 マルチキャストアドレスの監視

マルチキャストアドレスへ参加要求を行う MLD Report は，ルータが定期的に送信する MLD Query に対する応答を行う場合と，マルチキャストアプリケーションが受信開始・停止の操作を行った場合に送信される．MLD Report はマルチキャストアドレス宛へ送信されるため，ルータや該当 VM へ機能を追加することなく，リンクローカルの他ノードの受信状態を把握できる．

そのため，Agent は各 VM が送信する MLD Report を監視し，その送信元 MAC アドレスと要請ノードマルチキャスト（ff02::1:ff00:0/104）以外のマルチキャストアドレスを抽出する．これにより，マイグレーション時に該当の VM が受信しているマルチキャストを即時検索することが可能となる．

3.4.2 マルチキャストへの代理参加要求処理

移行先ネットワークでのマルチキャスト参加をできる限り早く完了させるために，グローバルライブマイグレーションが開始された時点で DstAgent が代理で参加要求処理を行う．

DstVMS 上にマルチキャストを受信しているノードが DstAgent だけであった場合，DstAgent が Leave するとマルチキャストツリーの再構築が中断されたり，配送が停止してしまったりする恐れがある．したがって，Agent はマイグレーション中も参加状態を維持する必要があり，マイグレーション後の VM が MLD Query に応答する MLD Report を送信するまでそのマルチキャストから離脱してはならない．

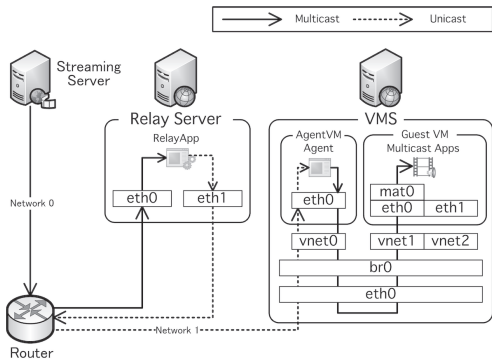

図 5 ユニキャスト中継を用いる際のストリームの流れ
Fig. 5 Stream flow using unicast relaying.

3.4.3 ユニキャスト中継の管理と代理受信

Agent は管理するマルチキャストアドレスすべてに対し，そのアドレスを宛先とするストリームが継続して VMS に受信されているかを確認する．一定時間以上受信が確認できない場合，マルチキャストが受信できないネットワークの可能性があると判断しユニキャスト中継に切り替える．

ユニキャスト中継を使用する際のストリームの流れを**図 5** に示す．`Network 1` はマルチキャストが直接受信できないネットワークであるため，マルチキャストが受信可能な中継ノードでいったん受信し，それを Agent に向けてユニキャストで再送信する．そして Agent は，受信したユニキャストストリームを，再度 TTL を 1 としたリンクローカルなマルチキャストストリームとして再送信することにより，VM 上のマルチキャストアプリケーションへ到達する．なお，ユニキャスト中継を行うサーバのアドレスは，文献 [5] で提案されている MCS（Migration Control Server）や，エニーキャストアドレス，DNS ラウンドロビン，SDP [12] などの手法を用いて解決する．

このような手順により，VMS のネットワークで直接マルチキャストを受信できない場合でも，最終的に VM ではマルチキャストとして受信され，中継による帯域コストを最小限に抑えることが可能となる．

4. 実装

本章では，提案手法の実現のために使用した仮想化環境について述べる．

VMS のハイパーバイザとして Linux Kernel に統合された KVM [*3] を使用し，完全仮想化環境を提供するエミュレータ QEMU [*4] と組み合わせた QEMU/KVM 構成上に実装を行った．さらに仮想化 API として各種仮想化環境をサポートしている libvirt [*5] を用い，Agent は libvirt API を通して QEMU/KVM の制御を行う．したがって，本論文で示した提案手法は特定の仮想化環境に依存するものではなく，他のハイパーバイザに対しても適用可能である．なお，QEMU/KVM，libvirt ともに，現時点で IPv6 によるマイグレーションが未サポートであったため，マイグレーションソケットに IPv6 が利用できるよう修正を行った．

[*3] http://www.linux-kvm.org/
[*4] http://qemu.org/
[*5] http://libvirt.org/

5. 基本性能評価

本章では，提案手法における効果を定量的に示すため，グローバルライブマイグレーションにおけるマルチキャスト受信の通信継続性と移動透過 IP マルチキャストのために追加した操作に要する時間について評価実験を行った．

検証環境を構成するノード群を**表 2** に示す．VM を稼働させる 2 台の VMS および，マルチキャストストリームを送信する Sender は，すべて同一の L3 スイッチに接続されているが，ネットワーク構成については各測定により異なるため後述する．Sender は，Iperf [*6] を用いて任意ビットレートのマルチキャストストリームを送信する．

VM の構成は，CPU は仮想 2 コアとし，メモリは 128 MB，256 MB，512 MB，1 GB の 4 種類においてそれぞれ実験した．VM のシステムを格納するストレージは，外部ストレージへのアクセスが与える影響を取り除くため，事前に双方の VMS 上のローカルディスクに共通のストレージファイルを配置し，スワッピングを無効化して測定を行った．

5.1 グローバルライブマイグレーションによるマルチキャストストリームの途絶時間

本測定では，提案手法を用いて異なるネットワーク上の別 VMS へ移行するグローバルライブマイグレーション（以下，GLM Multicast）と，同一ネットワーク内の別 VMS へ移行する従来のライブマイグレーション（以下，LM Multicast）について，ライブマイグレーション中に発生する通信途絶時間を比較することで，マルチキャストアプリケーションに与える影響と各方式による差異について明らかにする．なおここでは，途絶時間を VM がサスペンド状態になることでパケットが連続的に欠落した時間と定義し，その時間は図 3 の T_{DT} の区間に該当する．

実験構成を**図 6** および**図 7** に示す．測定方法は，Sender からマルチキャストストリームを送信し，VM 内の測定アプリケーションで受信パケットのシーケンス番号を記録する．測定では 20 Mbps（パケット長 1,450 Byte，1 秒あたりの送信パケット数 1,725 パケット）のストリームを利用し，途絶時間は連続して欠落したパケット数と送信パケット間隔から求めた．このビットレートは，地上デジタル放送で利用される ISDB-T 規格の MPEG2-TS ストリームが伝送可能な帯域を想定したものである．

GLM Multicast の結果を**図 8**，LM Multicast の結果を**図 9** に示す．各測定値は，5 回測定した平均値，最小値，

[*6] http://sourceforge.net/projects/iperf/

表 2 実験機材
Table 2 Experiment machines.

	Sender	VMS	VM
CPU	Intel Core Solo U1400 1.20 GHz	Intel Core i7 2600 K 3.40 GHz	Virtual 2-core CPU
RAM	1.5 GB	8 GB	128 MB, 256 MB, 512 MB, 1 GB
OS	Fedora release 14 (Laughlin)		
Kernel	linux-2.6.35.14-96.fc14.i686.PAE SMP		linux-2.6.32.16 SMP
Network Speed	100 Mbps	1 Gbps	
Hypervisor	–	QEMU/KVM 0.13.0-1	–
API	–	libvirt-0.8.3-10	–

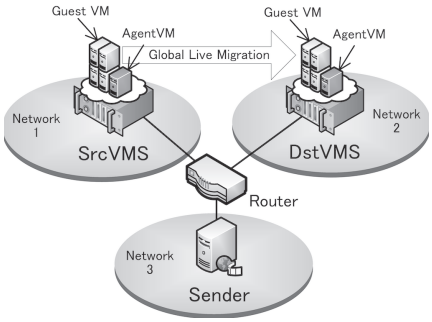

図 6 グローバルライブマイグレーション（GLM）の実験環境
Fig. 6 Experiment network for Global Live Migration.

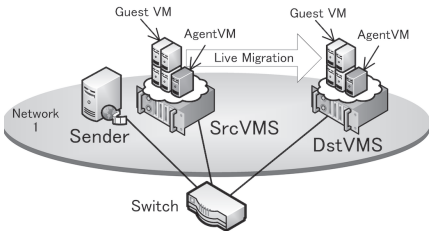

図 7 従来のライブマイグレーション（LM）の実験環境
Fig. 7 Experiment network for traditional Live Migration.

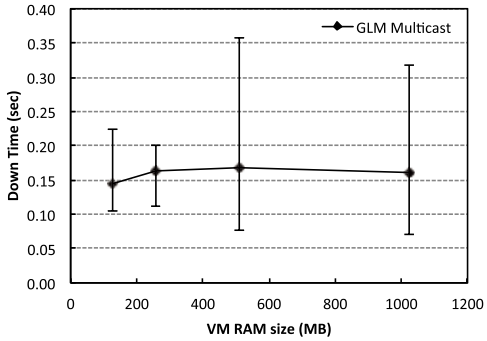

図 8 GLM Multicast の通信途絶時間
Fig. 8 Down time of GLM Multicast.

最大値を示している．この結果から，GLM Multicast においても LM Multicast と同等の途絶時間に抑えられていることが分かる．平均値としては GLM Multicast が LM Multicast を下回る結果となったが，最大途絶時間はメモリサイズ 512 MB の VM で GLM Multicast を行った場合の 0.36 秒であった．各測定結果にばらつきがあるのは，

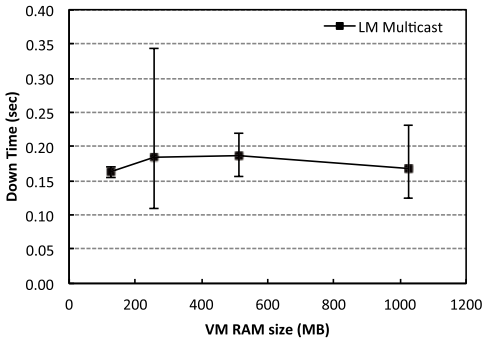

図 9 LM Multicast の通信途絶時間
Fig. 9 Down time of LM Multicast.

Pre-copy 型のライブマイグレーションの特性により，VM のサスペンド後に行われる最後のメモリページ転送量が一定でないため，転送時間に差異が生じていると考えられる．なお，両方式ともに途絶時間はメモリサイズに依存しないことが分かる．以上の結果から，本提案手法により DstAgent があらかじめ移行先ネットワークでマルチキャストストリームを受信可能な状態にしておくことで，GLM Multicast の途絶時間が LM Multicast の途絶時間と同等にできることを示した．

なお，マルチキャストを受信しながらマイグレーションすると，GLM Multicast においても LM Multicast においても重複したパケットが計測された．重複した時間は，計測ごとにばらつきが見られたが平均して 0.3 秒であった．重複が発生する要因としては，VMS でのパケットバッファリングと Pre-copy 型のライブマイグレーションの特性によるものと考えられる．本提案手法は，DstAgent がマルチキャスト代理参加を開始すると，VM がマイグレーション前であっても DstVMS 上のブリッジや TAP デバイスにマルチキャストが流入する．よって，一時的に移行元と移行先ネットワークで同一マルチキャストパケットをバッファリングすることになる．この状態でライブマイグレーションが行われると，サスペンド直後の VM のメモリ状態や I/O デバイス状態のコピーで，バッファリングされているパケットが DstVMS へ移行される．つまり，これらのパケットが DstVMS のネットワークデバイスですでにバッファリングされているパケットとともに VM 内のアプリ

表 3 GLM におけるマルチキャストストリーム途絶時間
Table 3 Down time of multicast stream on GLM.

RAM size (MB)	Stream bitrate / Down time (sec)					
	2 Mbps			20 Mbps		
	min	avg	max	min	avg	max
128	0.00	0.19	0.37	0.10	0.14	0.22
256	0.15	0.21	0.39	0.11	0.16	0.20
512	0.10	0.22	0.37	0.08	0.17	0.36
1,024	0.11	0.20	0.37	0.07	0.16	0.32

ケーションへ渡されることで，重複として観測されていると考えられる．

途絶時間に対するアプリケーションへの影響について考える．たとえば本研究で想定する動画像のリアルタイムマルチキャスト配信を考えたとき，通信途絶に対する耐性を向上させるためにはバッファリング機構や FEC・再送などの QoS（Quality of Service）制御が必要になる一方で，必要以上の制御はリアルタイム性を低下させる．通信要求特性（伝送遅延やパケット損失許容度，ジッタなど）はアプリケーションに依存するものの，本提案手法により，異なるネットワーク間でマイグレーションを行う場合であっても同一ネットワーク内のマイグレーションと同等の途絶時間として見積もることが可能となる．これはリアルタイムアプリケーションの QoS 制御を行ううえで重要な要素となる．本結果から得られた途絶時間はアプリケーションに対する適切な制御を行う際の指標の 1 つとすることができる．

5.2 受信ビットレートとストリーム途絶時間

本測定では，Sender が送信するマルチキャストストリームの帯域が 20 Mbps と 2 Mbps の 2 種類の場合について，GLM を行う VM 上アプリのマルチキャストストリームに与える影響を，図 6 に示す環境で測定した．測定方法は，5.1 節と同様である．なお，2 Mbps の場合は，パケット長 1,450 Byte，1 秒あたりの送信パケット数 174 パケットのストリームを利用した．

表 3 に各 VM メモリサイズにおける途絶時間の最小値，最大値，平均値を示し，**図 10** に平均値をグラフ化したものを示す．この結果から，平均値では 2 Mbps の場合が 0.04 秒ほど途絶時間が長くなっているが，最小値，最大値で比較するとビットレートの違いで大きな差異はないことが分かる．結果のばらつきの要因として，5.1 節で述べた Pre-copy 型のライブマイグレーションの特性から最終的なメモリページ転送量が一定でないこと，また 2 Mbps の場合は送信パケット間隔が長くなるため，途絶時間の計測粒度が粗くなることにより誤差が生じたと考えられる．なお図 10 より，途絶時間とメモリサイズの関係性はビットレートが異なる場合でも同様の傾向にあることが分かる．今回測定した条件での途絶時間の最大値は 0.39 秒であり，

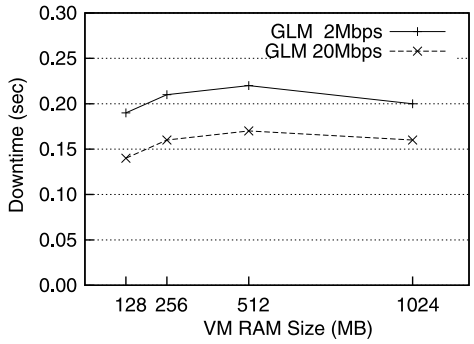

図 10 GLM におけるマルチキャストストリーム途絶時間
Fig. 10 Down time of multicast stream on GLM.

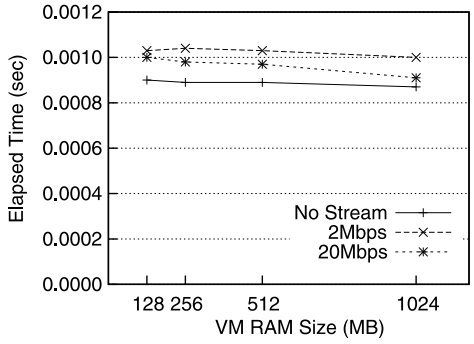

図 11 提案手法のための追加操作に必要な時間
Fig. 11 Necessary time for additional operation of propsed method.

図 9 における同一ネットワーク内のライブマイグレーションと同等の途絶時間であることが確認できた．この結果は，5.1 節でも述べたとおり，通信要求特性はアプリケーションに依存するものの，アプリケーションに対する同一ネットワーク内のライブマイグレーションにおける途絶時間への対策を大幅に変更することなく，グローバルライブマイグレーションへ適用できることを示している．

5.3 Agent の追加操作およびマイグレーションに要する時間

本測定では，マイグレーション開始から終了までの経過時間を，VM のメモリサイズとマルチキャストストリーム別に測定した．図 3 に示す T_{PRE} および T_{MIG} の区間について，図 6 に示す環境の SrcAgent 上で測定した．

(1) 提案手法において追加した代理 Join のためのシグナリング時間（T_{PRE}）
(2) 実際に VM のメモリ情報などを SrcVMS から DstVMS にコピーする時間（T_{MIG}）

図 11 および**図 12** に実験結果を示す．T_{PRE} の区間では，DstAgent において VM が受信しているマルチキャストストリームの代理参加処理が行われているため，マルチキャストを受信していないときよりも受信しているときの方が若干時間がかかっている．また，DstAgent がマルチキャストの代理参加処理を行った後には，ストリームが受

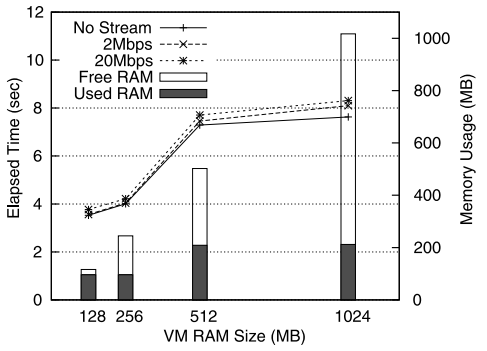

図 12 提案手法におけるマイグレーションに必要な時間
Fig. 12 Necessary time for migrate operation of propsed method.

信できることを確認しているため，パケットの到着間隔が疎である 2 Mbps よりも，密である 20 Mbps の方が時間が短くなる．なお，ここでは VM のメモリサイズに依存する処理はないため，メモリサイズによって経過時間は変化しない．

T_{MIG} の区間では，VM のメモリ内容や CPU・VGA の情報を SrcVMS から DstVMS へと転送しているため，VM のメモリサイズが大きいほど経過時間が長くなっている．しかし仮想化ソフトウェアの特性により，空メモリ部分の転送が高速化されているため，経過時間とメモリ割当て量は比例せず，実際のメモリ使用量および，ストリーム受信によって書き換えられるメモリ量に応じて増加している．

T_{PRE} におけるシグナリング時間は，SrcAgent と DstAgent 間のネットワーク距離に応じて増加する可能性はあるが，従来手法のマイグレーション後に VM が MLD Query を受信して MLD Report を送信するまでの時間よりも十分小さく，マイグレーション全体に必要な時間は T_{MIG} の部分が支配的である．以上の結果から，提案手法を用いることによりマルチキャストの移動透過性が期待できる．

6. おわりに

本論文は，従来のユニキャストに限定されたグローバルライブマイグレーションに対し，マルチキャストに対する IP 移動透過性を実現する手法について述べた．提案手法では，各 VMS 上に分散配置する Agent が，マルチキャストストリームの受信状態を管理するとともに，VM の移動に際してマルチキャストツリーの事前構築のための代理参加要求処理を行う．またマルチキャスト非対応ネットワーク上へマイグレーションする場合は，ユニキャスト中継サーバと連携した配信を可能とする．これにより，マルチキャストアプリケーションの変更を必要とせず，継続的なマルチキャストストリームの受信が可能となる．

評価実験では，マルチキャストマイグレーションにおける性能評価を行い，グローバルライブマイグレーション後のマルチキャスト途絶時間を同一サブネットのライブマイグレーションと同等の途絶時間に短縮することができることを確認した．マイグレーション処理全体に要する時間は VM のメモリサイズやページ書き換え状況，マイグレーション前後のネットワークの遅延などの影響を受けることが予想されるが，アプリケーションに直接的な影響を与える途絶時間を短縮できることは大きな利点であると考えている．

今後の課題として，マルチキャストが利用できないエリアでの動作検証および性能評価を行うとともに，広域の実環境ネットワークのように遅延やジッタが発生する環境下での性能評価を進める予定である．

謝辞 本研究の一部は，日本学術振興会科学研究費補助金基盤研究（B）（課題番号 23300026，24300025）および基盤研究（C）（課題番号 24500083）の助成を受けたものである．

参考文献

[1] Cain, B., Deering, S., Kouvelas, I., Fenner, B. and Thyagarajan, A.: Internet Group Management Protocol, Version 3, RFC 3376, IETF (2002).
[2] Vida, R. and Costa, L. (Eds.): Multicast Listener Discovery Version 2 (MLDv2) for IPv6, RFC 3810, IETF (2004).
[3] Fenner, B., Handley, M., Holbrook, H. and Kouvelas, I.: Protocol Independent Multicast Sparse Mode (PIM-SM): Protocol Specication (Revised), RFC 4601, IETF (2006).
[4] 近堂 徹，西村浩二，相原玲二：移動透過通信機能を持つ仮想計算機によるセッションモビリティの実現，インターネットコンファレンス論文集，Vol.2008, pp.42–48 (2008).
[5] 渡邉英伸，大東俊博，近堂 徹，西村浩二，相原玲二：IP モビリティと複数インタフェースを用いたグローバルライブマイグレーション，電子情報通信学会論文誌，Vol.93-B, No.7, pp.893–901 (2010).
[6] 相原玲二，藤田貴大，前田香織，野村嘉洋：アドレス変換方式による移動透過性インターネットアーキテクチャ，情報処理学会論文誌，Vol.43, No.12, pp.3889–3897 (2002).
[7] 神屋郁子，下川俊彦，吉田紀彦：柔軟な構成変更を可能とする広帯域配信システムの構築，電子情報通信学会技術研究報告 IA，インターネットアーキテクチャ，Vol.109, No.421, pp.13–16 (2010).
[8] 宮城亮太，池部 実，猪俣敦夫：クラウド環境におけるサーバ負荷に応じた動的計算資源割当システムの提案と評価，電子情報通信学会技術研究報告，Vol.110, No.430, pp.17–22 (2011).
[9] 鎌田恵介，近堂 徹，相原玲二：ユニキャストを併用する移動透過 IP マルチキャストの設計，電子情報通信学会技術研究報告 IA，インターネットアーキテクチャ，Vol.110, No.304, pp.13–18 (2010).
[10] Thomson, S., Narten, T. and Jinmei, T.: IPv6 Stateless Address Autoconfiguration, RFC 4862, IETF (2007).
[11] Salfner, F., Tröger, P. and Polze, A.: Downtime Analysis of Virtual Machine Live Migration, *The 4th International Conference on Dependability* (*DEPEND*), pp.100–105 (2011) .
[12] Handley, M., Jacobson, V. and Perkins, C.: SDP: Session Description Protocol, RFC 4566, IETF (2006).

鎌田 恵介 （正会員）

2010 年広島大学工学部第二類（電気・電子・システム・情報系）卒業．2012 年同大学大学院工学研究科博士課程前期修了．IP 移動透過通信に関する研究に従事．現在，日本アイ・ビー・エム株式会社に勤務．

近堂 徹 （正会員）

2001 年広島大学工学部第二類（電気系）卒業．2006 年同大学大学院工学研究科博士課程修了．現在，広島大学情報メディア教育研究センター准教授．博士（工学）．コンピュータネットワーク，リアルタイムマルチメディア通信，QoS 保証技術に関する研究に従事．電子情報通信学会会員．

西村 浩二 （正会員）

1989 年広島大学工学部第二類（電気系）卒業．1991 年同大学大学院工学研究科博士課程前期修了．広島大学総合情報処理センター助手，同大学情報メディア教育研究センター准教授等を経て，2011 年より同教授．博士（工学）．コンピュータネットワークの運用管理，移動透過通信，情報セキュリティに関する研究に従事．電子情報通信学会会員．

相原 玲二 （正会員）

1981 年広島大学工学部第二類（電気系）卒業．1986 年同大学大学院工学研究科博士課程後期修了．同大学助手，同大学集積化システム研究センター助教授を経て，現在，同大学情報メディア教育研究センター教授．工学博士．コンピュータネットワークに関する研究に従事．電子情報通信学会，IEEE Computer Society，IEEE Communications Society 各会員．