(19)日本国特許庁(JP)　(12)公表特許公報(A)　(11)特許出願公表番号
特表2009-508410
(P2009-508410A)
(43)公表日　平成21年2月26日(2009.2.26)

(51)Int.Cl.	F I	テーマコード（参考）
H04L 12/56 (2006.01)	H04L 12/56 100Z	5K030

審査請求 未請求　予備審査請求 未請求　（全 16 頁）

(21)出願番号	特願2008-530233 (P2008-530233)
(86)(22)出願日	平成18年9月8日 (2006.9.8)
(85)翻訳文提出日	平成20年4月23日 (2008.4.23)
(86)国際出願番号	PCT/US2006/035116
(87)国際公開番号	WO2007/030742
(87)国際公開日	平成19年3月15日 (2007.3.15)
(31)優先権主張番号	60/715,388
(32)優先日	平成17年9月8日 (2005.9.8)
(33)優先権主張国	米国 (US)
(31)優先権主張番号	60/716,383
(32)優先日	平成17年9月12日 (2005.9.12)
(33)優先権主張国	米国 (US)

(71)出願人　000005821
パナソニック株式会社
大阪府門真市大字門真１００６番地
(74)代理人　100105050
弁理士　鷲田　公一
(72)発明者　ビュフォード　ジョン
アメリカ合衆国　０８６４８　ニュージャージー州　ローレンスヴィル　リチャーズ　ロード　６

Fターム(参考)　5K030 HA08　LB01　LB05　LD02

最終頁に続く

(54)【発明の名称】マルチデスティネーション・ルーティングを利用したピアツーピア・オーバーレイ通信の並列実行

(57)【要約】
オーバーレイ・ネットワークにおいてオーバーレイ通信動作を並列実行するための方法が提供される。当該方法は、並列メッセージング方式を有するオーバーレイ通信動作を特定することと、前記メッセージング方式による各並列メッセージの送信先アドレスを決定することと、各送信先アドレスを一つのデータ・パケット中に符号化することと、マルチデスティネーション、マルチキャスト・ルーティング・プロトコルを使用して、前記データ・パケットを前記オーバーレイ・ネットワークを介して送信することを含む。

【特許請求の範囲】
【請求項１】
オーバーレイ・ネットワークにおいてオーバーレイ通信動作を並列実行するための方法であり、
並列メッセージング方式を有するオーバーレイ通信動作を特定することと、
前記メッセージング方式による各メッセージの送信先アドレスを決定することと、
送信先アドレスの各々をもつデータ・パケットをフォーマット化することと、
マルチデスティネーション、マルチキャスト・ルーティング・プロトコルを使用して、前記データ・パケットを前記オーバーレイ・ネットワークを介して送信することと
を含む方法。
【請求項２】
明示的マルチキャスト（Ｘｃａｓｔ）プロトコルに従って前記データ・パケットを送信することをさらに含む、請求項１の方法。
【請求項３】
ルーティング装置にて前記データ・パケットを受信することと、マルチデスティネーション、マルチキャスト・ルーティング・プロトコルを使用して、前記データ・パケットを転送することとをさらに含む、請求項１の方法。
【請求項４】
前記データ・パケットを転送することは、
前記データ・パケット中の前記送信先アドレスの各々の次のホップを特定することと、
特定された次のホップのそれぞれに対して前記データ・パケットを複製することと、
各データ・パケットが当該データ・パケットに関連付けられた次のホップを通じてルーティングされるべき送信先アドレスだけを含むように各データ・パケット中にリストされた前記送信先アドレスを修正することと、
修正された各データ・パケットを該当する次のホップへ転送することと
をさらに含む、請求項３の方法。
【請求項５】
前記オーバーレイ・ネットワーク内のノードにて出メッセージ・キューを定義することと、
一つのオーバーレイ通信動作に関連付けられるメッセージを前記キューに追加することと、
前記オーバーレイ・ネットワーク内の異なる送信先をもつが、重複する内容を含むメッセージを前記キュー内で特定することと、
前記オーバーレイ・ネットワークを介して前記データ・パケットを送信する前に、前記特定された各メッセージを単一のデータ・パケットに合成することと
をさらに含む請求項１の方法。
【請求項６】
前記特定された各メッセージを合成することが、前記異なる送信先のそれぞれに対する送信先アドレスを前記データ・パケットのヘッダー内にフォーマット化することをさらに含む、請求項５の方法。
【請求項７】
オーバーレイ・ネットワークにおいてオーバーレイ通信動作を並列実行するための方法であり、
前記オーバーレイ・ネットワーク内のノードにて出メッセージ・キューを定義することと、
メッセージをキューに追加することと、
前記オーバーレイ・ネットワーク内の異なる送信先をもつが、重複する内容を含むメッセージを前記キュー内で特定することと、
前記特定された各メッセージを単一のマルチキャスト・データ・パケットに合成することと

マルチデスティネーション、マルチキャスト・ルーティング・プロトコルを使用して、前記マルチキャスト・データ・パケットをそのノードから送信することと
を含む方法。
【請求項８】
前記特定された各メッセージを合成することが、前記異なる送信先のそれぞれに対する送信先アドレスを前記データ・パケットのヘッダー内に符号化することをさらに含む請求項７の方法。
【請求項９】
送信アドレスのリストがサイズ限度を超えていなければ前記特定された各メッセージを合成することをさらに含む、請求項８の方法。
【請求項１０】
前記データ・パケットのペイロードがサイズ限度を超えていなければ前記特定された各メッセージを合成することをさらに含む、請求項７の方法。
【請求項１１】
前記キュー内のメッセージを、当該メッセージに関連した最大キューイング遅延が距離計量を超えているときには、ユニキャスト・ルーティング・プロトコルを使用して送信することをさらに含む、請求項７の方法。
【請求項１２】
重複する内容を含まないメッセージをユニキャスト・ルーティング・プロトコルを使用して送信することをさらに含む、請求項７の方法。
【請求項１３】
明示的マルチキャスト（Ｘｃａｓｔ）プロトコルに従って前記データ・パケットを送信することをさらに含む、請求項７の方法。
【請求項１４】
並列のメッセージを有する少なくとも一つのオーバーレイ通信動作を実行するように動作可能なオーバーレイ・ネットワーク中のホスト・ノードであり、並列のメッセージのそれぞれの送信先アドレスを決定し、各送信先アドレスを単一のデータ・パケット中に符号化し、マルチデスティネーション、マルチキャスト・ルーティング・プロトコルを使用して前記データ・パケットを送信するホスト・ノードと、
マルチデスティネーション、マルチキャスト・ルーティング・プロトコルに従って前記データ・パケットを各送信先アドレスへ転送するように動作可能な下層ネットワーク中に存在する複数のルータとを
含むオーバーレイ・ネットワーク向けのメッセージング方式。
【請求項１５】
ルータの各々は前記データ・パケットを受信するようになされ、前記データ・パケット中の前記送信先アドレスの各々に対する次のホップを特定し、特定された次のホップのそれぞれに対して前記データ・パケットを複製し、各データ・パケットが当該データ・パケットに関連付けられた次のホップを通じてルーティングされるべき送信先アドレスだけを含むように各データ・パケット中にリストされた前記送信先アドレスを修正し、修正された各データ・パケットを該当する次のホップへ転送するように動作可能である、請求項１４のメッセージング方式。
【請求項１６】
前記マルチデスティネーション、マルチキャスト・ルーティング・プロトコルは、明示的マルチキャスト（Ｘｃａｓｔ）プロトコルとしてさらに定義される、請求項１４のメッセージング方式。
【請求項１７】
複数のノードを有するオーバーレイ・ネットワーク向けのメッセージング方式であり、
前記オーバーレイ・ネットワーク内のノード間のデータ・パスを定義する階層ツリー構造を保守することと、
マルチデスティネーション、マルチキャスト・ルーティング・プロトコルに従ってデー

タ・パケットを転送するように前記オーバーレイ・ネットワーク内のノードを構成することと、
前記マルチデスティネーション、マルチキャスト・ルーティング・プロトコルを使用した前記階層ツリー構造に従ってノード間でデータ・パケットを送信することと
を含むメッセージング方式。
【請求項１８】
前記マルチデスティネーション、マルチキャスト・ルーティング・プロトコルは、明示的マルチキャスト（Ｘｃａｓｔ）プロトコルとしてさらに定義される、請求項１７のメッセージング方式。

【発明の詳細な説明】
【関連出願の相互参照】
【０００１】
本願は、２００５年９月８日付出願の米国仮特許出願No. 60/715,388及び２００５年９月１２日付出願の米国仮出願No. 60/716,383の利益を主張する。上記の出願における開示は、参照することにより本文書に援用される。
【技術分野】
【０００２】
本開示は、ピアツーピア・オーバーレイ・ネットワークに関係し、より詳しくは、オーバーレイ・ネットワークにおけるオーバーレイ通信動作を並列実行するための方法に関係する。
【背景技術】
【０００３】
オーバーレイ・ネットワークは、他のネットワークの上に構築されるネットワークである。オーバーレイ・ネットワーク内のノードは、下層のネットワーク内のパスに各々対応する論理リンクによって接続されていると考えることができる。多くのピアツーピア・ネットワークは、インターネットの上に張り巡らされているオーバーレイ・ネットワークとして実現される。従来、オーバーレイ・ネットワークは、ノード間での通信にユニキャスト・メッセージングを頼りにしてきた。
【０００４】
最近になって、ホスト・グループ・マルチキャストが、オーバーレイ・メッセージング運用のために提案された。簡単に述べれば、ホスト・グループ・マルチキャスト・プロトコルは、グループ・アドレスを生成し、各ルータはアクティブであるグループ・アドレスごとにステート(state)を記憶する。ルータに保持されるステートは、同時発生するマルチキャスト・グループの数だけ増える。グループを作るための遅延が存在し、ネットワークがもち得るグループ・アドレスの数には制限がある。
【０００５】
大きなオーバーレイ・ネットワークの場合には、各ノードがメッセージの送信先のその他のノードの各セットに対するグループ・アドレスを保持することは非現実的である。オーバーレイ・ネットワークの全部のまたは大部分のサブセットに対するマルチキャスト・アドレスを各ノードが保持したとするならば、関わりをもつノード数は多大である故、過度の量のトラヒックとルータ・オーバヘッドが生じることになろう。
【０００６】
さらに、一つのセットにまとめられた複数のノードに対してあるピア・ノードが並列のクエリー（queries）を発行するために、元来のホスト・グループ・マルチキャストを使用したい場合、ピア・ノードは先ず各ルータにステートを作成し、受信ノードをマルチキャストに取り込む必要がある。この設定は遅延を加算するし、マルチキャスト・パスがしばらくの間は再使用されることになる場合にのみ妥当である。しかし、ピアツーピア・ネットワークではノードをまとめたセットはかなり動的であり、 ノード間の一連の要求は予測できないので、このようなマルチキャスト・グループの再使用は制限的である。

【０００７】
ホスト・グループ・マルチキャストは、多くの受信者をまとめた非常に大きなセットの、比較的少数のセットを対象にして設計されている。それゆえ、ホスト・グループ・マルチキャストは、メッセージに関わりあうピアの数が少ない小グループが同時に多数存在するネットワーク・オーバーレイ通信動作の並列実行に使用するためにはよい選択ではない。したがって、オーバーレイ・ネットワークにおけるオーバーレイ通信動作の並列処理の手立てが必要とされている。
【０００８】
この節で述べた事柄は、本開示に関連した背景情報を単に提供するものであり、先行技術を構成するものではない。
【発明の開示】
【０００９】
オーバーレイ・ネットワークにおいてオーバーレイ通信動作を並列実行するための方法が提供される。当該方法は、並列メッセージング方式を有するオーバーレイ通信動作を特定することと、前記メッセージング方式による各並列メッセージの送信先アドレスを決定することと、　各送信先アドレスを一つのデータ・パケット中に符号化することと、マルチデスティネーション、マルチキャスト・ルーティング・プロトコルを使用して、前記データ・パケットを前記オーバーレイ・ネットワークを介して送信することを含む。
【００１０】
さらなる適用分野は、ここで提供した記述から明らかに理解できる。その記述と特定の例は例示の目的で意図されたものであり、本発明の権利範囲を制限することを意図するものではないと理解されるべきものである。
【発明を実施するための最良の形態】
【００１１】
図１は、 オーバーレイ・ネットワークを有するネットワーク構成例の図である。簡単に述べれば、下層ネットワーク１０は、一般に、複数のネットワーク・ルーティング装置１４（すなわち、ルータ）によって相互接続された複数のネットワーク装置１２から構成される。これらの装置間の物理ネットワーク・リンクは、図内の実線で示される。オーバーレイ・ネットワーク２０は、下層ネットワーク１０の上に構築される。オーバーレイ・ネットワーク２０は、装置間に張られた論理リンクの連なりであり、図１で点線で示される。例示的なオーバーレイ・ネットワーク・アーキテクチャは、コンテント・アドレサブル・ネットワーク（Content Addressable Network（ＣＡＮ））、Chord、Tapestry、Freenet、Gnutella、及びFast Trackを含む。本開示は、その他のタイプのオーバーレイ・ネットワーク・アーキテクチャにも関係することは容易に理解される。
【００１２】
図２は、オーバーレイ・ネットワークにおけるオーバーレイ通信動作を並列実行するための方法を説明するフローチャートである。先ず、適切なオーバーレイ通信動作が２２で特定される。典型的なオーバーレイ通信動作は、これらに限定されないが、オーバーレイに加入するノード、オーバーレイから退去するノード、ルーティング・テーブル更新、ルーティング・テーブルまたはルーティング・テーブルの抜粋をその他のノードへ転送するノード、ノード・ステート及び／またはオーバーレイ測定値をほかのノードとの間で交換するノード、数個のその他のノードに要求を送信するノード、及び数個の加入者ノードにイベントを発布するノードを含む。これらの動作の一部を以下にさらに説明する。本方法が並列メッセージング方式（すなわち、一つの送信元ノードから複数の送信先ノードに送信される少なくとも二つのユニキャスト・メッセージ）を有するその他のオーバーレイ通信動作にも適用されることは容易に理解される。
【００１３】
オーバーレイ・ネットワーク上で適用可能なメッセージを送信するために、次に、マルチデスティネーション、マルチキャスト・ルーティングが使用される。一般に、送信元ノードがメッセージの送信先のリストを決定し（２４）、単一のデータ・パケットのヘッダ

ー内に各送信先アドレスを符号化する（２６）。オーバーレイ・ネットワークでは、このようなメッセージの送信先アドレスは、通常、送信元ノードに知られている。図３を参照して、ここでノードＡがノードＢ、Ｃ、及びＤへメッセージを送信しようとしていると仮定する。ノードＡは、データ・パケット・ヘッダーを次のように符号化する。[src = A| dest = Ｂ Ｃ Ｄ | payload]。次に、データ・パケットは送信元ノードから送信される（２８）。

【００１４】

送信パス上にあるマルチキャスト対応の各ルータは、順に、データ・パケットをその送信先まで転送する。データ・パケットを受信すると、マルチキャスト対応ルータはデータ・パケットを次のように処理する。データ・パケット中の各送信先アドレスに対して、ルータは、ルート・テーブル検索を行って次のホップを決定する。異なる次のホップごとに対応してデータ・パケットは複製され、そこで各データ・パケットがそのデータ・パケットに関連付けられた次のホップを通じてルーティングされるべき送信先アドレスだけを含むように送信先のリストが修正される。最後に、修正された各データ・パケットは、当該ルータによって該当する次のホップへ転送される。

【００１５】

図３で、ルータＲ１は、[Ｂ Ｃ Ｄ]の送信先リストをもつ１個のデータ・パケットをルータＲ２へ送信する。ルータＲ２がこのデータ・パケットを受信すると、当該データ・パケットの一つのコピーをルータＲ４へ、当該データ・パケットの一つのコピーをＲ５へ送信する。ルータＲ４へ送信されたデータ・パケットは、修正されたＢの送信先リストをもつ。他方、ルータＲ５へ送信されたデータ・パケットは、修正された、[Ｃ Ｄ]の送信先リストをもつ。このデータ・パケットは、ルータＲ７へ到達する前にルータＲ５とＲ６によって転送される。ルータＲ７では、データ・パケットはＣとＤの送信先をそれぞれにもつ２個のデータ・パケットに再び分離される。単一の送信先をもつデータ・パケットは、これらのルートの残りの区間をユニキャストされ得ることは容易に理解される。

【００１６】

明示的マルチキャスト（Ｘｃａｓｔ）プロトコルは、マルチデスティネーション、マルチキャスト・ルーティング・プロトコルの一例である。Ｘｃａｓｔプロトコルに関する詳細は、参照することにより本文書に援用される、インターネット・エンジニアリング・タスク・フォースによって公開されている明示的マルチキャストの基本仕様に見出すことができる。しかし、その他のマルチデスティネーション、マルチキャスト・ルーティング・プロトコルも本開示の範囲内であることは容易に理解される。

【００１７】

実施形態の一例では、マルチデスティネーション、マルチキャスト・ルーティング・プロトコルは、送信元ノードのアプリケーション・レベルで実現される。つまり、オーバーレイ通信動作を実行するアプリケーションが、並列メッセージング方式を有するこれらの通信動作を特定し、適宜にメッセージを送信する。

【００１８】

実施形態の別の例では、 各ピアＰｊが、送信待ちメッセージを保持するキューＱを備えている。キュー中のメッセージは、ユニキャスト・メッセージであってもマルチキャスト・メッセージであってよい。マルチキャスト・メッセージは、ピアで実現されるオーバーレイ通信動作によって直接追加されてもよいし、またはＱの内容の前処理中にいくつかのメッセージを合成して作成されてもよい。

【００１９】

ユニキャスト・メッセージをＱに追加した後、ピアはＱを調べて、いくつかのユニキャスト・メッセージの集合ｕをグループg_kに属する一つのマルチキャスト・メッセージm_kに合成することができる。ここでm_kは各ユニキャスト・メッセージの内容を含み、$p_j \in g_k$であり、$|g_k| = |u| + 1$、及び$g_k \in F_i$である、ここでp_jは所与のピアであり、F_iはサイズｉ = 2, 3, ..., nのオーバーレイ内の複数ピアからなる各セットのすべての組み合わせの集合である。ピアは、キューから一つ以上のメッセージを消去

したり、その他のユニキャスト／マルチキャスト・メッセージを合成したり、及び／またはさらに追加されるメッセージを待つことができる。ピアは、各メッセージの最大キューイング遅延をしきい値d_q以下に維持するように動作する。メッセージのマルチキャスティングを抑制するその他の判断基準は次のものを含む。すなわち、パケットがそのペイロードのサイズ限度に達したか、パケットがその送信先アドレスのリストのサイズ限度に達したか、パケットを作成し、保存し、受信し、処理するために必要となる時間またはピアのリソースに関係した処理限度にパケットが達したか、メッセージが送信前にキューに留まることができる時間長に関係した時間遅延にパケットが達したか、またはマルチキャスト・メッセージに合成される各メッセージの内容が完全に重複するか、一部重複するか、または全く重複しないかである（重複部分がより多いほど、マルチキャストを使用する効率ゲインはより大きくなる）。

【００２０】

ユニキャスト・メッセージをマルチキャスト・メッセージに合成すること、並びにユニキャスト・メッセージをマルチキャスト・メッセージから取り出すことに関する規則に各ピアが合意していると仮定する。Ｑにおいてメッセージを合成する際に使用される決定基準は、マルチキャストによるネットワーク効率上のゲインが、合成されたユニキャスト・メッセージの内容重複の量に比例することを考慮するものと仮定する。

【００２１】

マルチキャスト・ルーティングは、オーバーレイ設計者に効率性と並行性を提供する。しかし、第一に、グループ数に関するマルチキャスト・アルゴリズムのスケーラビリティが、オーバーレイのスケーラビリティ要件を満たす必要がある。このオーバーレイで同時に存在するマルチキャスト・グループ・ステートをサポートするためのネットワークの容量をＣとすると、$N_G \leqq C$となる。また、最大グループ・サイズをｖとすると、$|g_{max}| < v$となる。第二に、グループ形成及びグループ構成メンバー変更のオーバーレイの速度ｒは、マルチキャスト・メカニズムによって達成可能である。新しいマルチキャスト・グループを作成するための時間は、$t_c < d_q$となる。

【００２２】

この手法は、下層ネットワークがマルチキャスト対応ルータを使用していることを前提とする。多くの場合において、これは妥当な前提である。ほかの場合には、下層ネットワーク中の一部のルータだけがマルチキャスト対応型である。この場合、マルチキャスト対応ルータは、それら相互間でデータ・パケットを送信するために特別なトンネル接続を使用する。

【００２３】

またほかの場合には、下層ネットワークはマルチキャスト対応ルータを全然提供しない。この場合、専用のコンピュータが下層ネットワーク中のほかのルータのそばに配備され得る。これらのコンピュータは、前述のルーティング・プロトコルを実現することによって論理マルチキャスト・バックボーンを形成するように構成される。マルチキャスト・パケットを送信したい送信元ノードは、論理マルチキャスト・バックボーン内の最も近いコンピュータへパケットを送信し、そのコンピュータが論理マルチキャスト・バックボーンを介して送信先へ順次パケットを転送する。

【００２４】

この手法が特定のオーバーレイ通信動作にどのように適用され得るかをさらに以下に説明する。図４Ａは、オーバーレイ・ネットワークの一例の現在のステートを示す。しかし、ピアツーピア環境は動的になりがちなので、図４Ｂに示すように、ある場合、ノード４２がネットワークに加入することがあり、別のノード４４がネットワークから退去する。こうするためには、加入ノードまたは退去ノードは、そのステートの変化をネットワーク中のその他のノードに通知する必要がある。例えば、図４Ｃに示すように、加入ノードは要求メッセージを複数のノードへユニキャストし得る。複数のユニキャスト・メッセージを送信するのではなく、図４Ｄに示すように、加入ノードは、マルチデスティネーション、マルチキャスト・ルーティング・プロトコルを使用した単一のパケットを送信し得る。

種々のオーバーレイ・ネットワークがノード間の通信のために種々のメッセージング方式を採用することは理解される。とはいえ、このような種々のオーバーレイ通信動作は、以下に述べる方式により並列実行に特に適合するようにされる。

【００２５】

Kademliaは、その対称距離計量（ＸＯＲ関数）に基づいて、そのルーティング・テーブルの保守、検索及び設定のために並列の要求を発行することができるマルチホップ・オーバーレイである。ノード探索を行うとき、ピアはノードまでのＸＯＲ距離を算出し、該当するｋのバケット内を探し、既知であるα個の最も近いノードを選び出し、これらのピアに並列の要求を送信する。応答はより近いノードを返す。Kademliaは、発見したｋ個の最も近いノードから応答を受信するまで、追加の並列の要求をα個の最も近いノードに繰り返し送信する。αの標準値は３である。図５は、１１０ｋのバケット内のノードについてのノード探索を示す。１６０ビットのアドレス空間の場合、最大１６０のバケットが存在する。

【００２６】

ノード探索は、ＤＨＴ保存、ＤＨＴ更新、及びＤＨＴ検索を含むほかのKademlia通信動作によっても使用される。Kademliaのピアは、所与のバケット内でのノード探索を行うにあたり、少なくともｋ／a回の繰り返しを行う。ｋ＝２０、α＝３であれば、３方向へのクエリーを７個のマルチキャスト・グループに対して行う。160のバケットでは、各ピアはそのアドレス空間中の少なくとも160個のグループにクエリーを行う必要がある。マルチキャストのクエリーがα方向であったと仮定すれば、Chuang-Sirbuスケーリング法は、マルチデスティネーション、マルチキャスト・ルーティングを使用した場合に１８％の削減があると算定する。また、クエリーがｋ方向、ｋ＝２０であったと仮定すれば、Chuang-Sirbuは、応答はユニキャストであっても、上記の方式によるマルチキャスティングKademlia要求から４２％の削減があると算定する。

【００２７】

Meridianは、オーバーレイ内のその他のノードからの相対距離を使用して、最も近いノードの発見と中心のリーダーの選択という形のオーバーレイ検索を解決するための測定オーバーレイである。各ピアはその隣接するノードを同心リングのまとまりのセットに編成し、各リングはｋ＝Ｏ(log N)個の一次エントリーとＩ個の二次エントリーを含む。Ｎ＝2500のシミュレーションでは、ｋ＝16、リング数ｉ＊＝９である。Meridianは、オーバーレイ内の構成メンバー変更を伝搬するために、ゴシップ・プロトコルを使用する。ゴシップ期間中、メッセージは上記の各リング内でランダムに選択されたノードに送信される。メッセージは、上記の各リング内でランダムに選択された１個のノードを含む。ユニキャストのゴシップ・メッセージを、単一のｉ＊方向へのメッセージを使用して、ｉ＊個の送信先へ前述の方式でマルチキャストすることができる。

【００２８】

EpiChordでは、ピアは完全なルーティング・テーブルを保守し、ルーティング・テーブル更新及び保存の回数の増加という犠牲を払って、マルチホップ・オーバーレイのＯ(log N)ホップ性能と同等にＤＨＴ動作での１ホップ性能を近づけようとする。EpiChordのピアのルーティング・テーブルは、当該ピアがオーバーレイに加入する時点で、後続及び先行ピアのルーティング・テーブルのコピーを取得することにより初期化される。その後、ルーティング・テーブル内に存在しないピアからの要求が着信するとピアは新しいエントリーを追加し、無効とみなされるエントリーを削除する。探索がルーティング・テーブルに新しいエントリーを追加する速度に比べて、変動速度が十分に高い場合、ピアはスライスと呼ばれるアドレス空間のセグメントにプローブ・メッセージを送信する。現在のピアの位置からアドレス範囲が遠くへ行くにつれて、スライスは指数関数的に増加する規模で編成される。これにより、当該ピアを中心にルーティング・テーブルのエントリーは集中する形になり、ルーティングのコンバージェンスがよくなる。

【００２９】

探索の成功を向上させるために、EpiChordは、当該ノードに最も近いピアへ向けたｐ方

向への要求を使用する。変動速度が高い期間中は、ピアはそのルーティング・テーブルの各スライスにおいて少なくとも２個のアクティブ・エントリーを保守する。スライスにおけるエントリー数が２を下回ると、ピアは並列の探索メッセージをスライスに属する各ＩＤに発行する。これらの並列の探索メッセージは、前述の方式によるマルチデスティネーション、マルチキャスト・ルーティングを用いて送信され得る。これらの探索に対する応答は、ルーティング・テーブル中の当該スライスへエントリーを追加するために使用される。

【００３０】

Accordionは、保守用のトラヒックが各ピアで使用可能な帯域幅に基づいて割り当てられることを除き、EpiChordと同様である。Accordionは、オーバーレイ上のその近隣における最新のルーティング・テーブル・エントリーを保守し、タイムアウトの可能性を減少させるように、反復的な並列探索を使用する。探索の要求元のピアは、そのルーティング・テーブル内のキーと共に空白(gap)に基づいて送信先を選択する。転送された探索に対する応答は、これらのルーティング・テーブル内の空白に入るエントリーを含む。ピアの帯域幅割り当てを超過する帯域幅が、ピアのルーティング・テーブル内の最大の規模の空白を埋めるルーティング・テーブル・エントリーを得るための並列の探査的探索のために使用される。並列性の程度は、探索トラヒックと帯域幅割り当てに基づいて、例えば、６方向などの最大設定内で動的に調整される。Accordionのｐ方向へ転送する探査的探索をマルチデスティネーション探索に置き換えた場合、エッジ・トラヒックを(ｐ-1)/2ｐ削減する。例えば、ｐ= 5 とすれば、エッジで４０％の削減となる。固定的な帯域幅割り当ての場合、これはピアが探査速度を2.5倍増加でき、その結果ルーティング・テーブルの正確度を十分に高められることを意味する。あるいは、より低い帯域幅割り当てにおいては、ピアは同一レベルのルーティング・テーブルの正確度（探索当りのホップ数）で動作できる。

【００３１】

Ｄ１ＨＴは、オーバーレイ保守アルゴリズムＥＤＲＡ（イベント検出及び報告アルゴリズム）を定義する１ホップのオーバーレイであり、ここでのイベントは加入／退去動作である。ＥＤＲＡは、システム中のすべてのイベントを対数時間内に伝搬する。各加入／退去イベントは、図６に示すように、相対的位置$\log_2(0)$ ～$\log_2(n)$にある$\log_2(x)$の後続ピアへ転送される。従来の表現に従うと、Θはピアがリング内の後続ピアへイベントを伝搬する間隔であり、ｐ = [$\log_2$ n] はその間隔内にピアが送信する最大メッセージ数である。伝搬されたイベントは、最後のイベント・メッセージ後に直接受信されたものおよび先行ピアから受信されたものである。各メッセージは有効期間（ＴＴＬ）をもち、肯定応答される。報告すべきイベントがない場合には、ＴＴＬ = ０のメッセージのみが送信される。

【００３２】

各間隔Θ中にピアは、その現在のイベントを有する最大ｐ = [$\log_2$ n] 個のメッセージを送信する。各メッセージは、同一のイベントのセットを有するが、[０～ｐ]の範囲で異なるＴＴＬをもつ。われわれは、ｐ個のユニキャスト・メッセージを、イベントのセットと［ピア、ＴＴＬ］ペアのリストを含むｐ方向へのマルチデスティネーション・パケットに置き換える。メッセージを受信する各ピアは、リストから自分のＴＴＬを抜き出す。ｎ = $10^{\wedge}6$の規模で、Chuang-Sirbuスケーリング法の算定は41.6％のメッセージ削減（ｐ = 20）を示す。ｎ = $10^{\wedge}3$の規模では、Chuang-Sirbuの算定は34％の削減（ｐ = 10）を示す。

【００３３】

ランダム・ウォークは、べき乗則グラフとして表わされる非構造化トポロジーにおいて最も効率的な探索技法であることがわかった。ランダム・ウォークでは、着信したクエリーはその場で照合することはできず、受信した要求を発した相手以外のランダムに選択した近隣の相手に当該要求を転送する。ランダム・ウォークを使用するシステムは、Ｇｉａ及びＬＭＳを含む。エッジ・トラヒック並びに内部トラヒックを減少させるために、マル

チデスティネーション、マルチキャスト・ルーティングを並列のランダム・ウォークにおける開始ノードで使用することができる。また後続のホップでもこれを使用することができる。

【００３４】

いくつかのピアツーピア・オーバーレイは、オーバーレイ・ネットワーク内のノードがマルチキャスト・ツリー中の子ノードへデータ・パケットを転送するという形式のアプリケーション層マルチキャスティングをサポートする。マルチキャスト・ツリーは、オーバーレイ・ネットワーク内のノード間のデータ・パスを定義する。マルチキャスト・ツリーは、ノードの入次数と出次数の制約を考慮することによって形成される。ノードは、親ノードと子ノードを接続するために、通常、ユニキャスト・リンクを使用するので、各リンクはノードのネットワーク・インタフェース上の帯域幅を使用する。各ノードで許容された限られた分岐数に適応させると、ツリー中のパス長は一般に増加し、エンド・ツー・エンド・レイテンシーがより大きくなる。こうした形式のマルチキャスト・ツリーを構成し、保守するための様々なプロトコルが当技術分野で知られている。

【００３５】

マルチキャスト・ツリー中のノード間でデータ・パケットを送信するためにマルチデスティネーション、マルチキャスト・ルーティング・プロトコルを使用する新しいメッセージング方式が提案される。これを実現するために、オーバーレイ・ネットワーク中のノードは、マルチデスティネーション、マルチキャスト・ルーティング・プロトコルに従ってデータ・パケットを転送するように構成される。そうすることで、マルチデスティネーション、マルチキャスト・ルーティング・プロトコルを使用したマルチキャスト・ツリーに従ってノード間でデータ・パケットを送信することができる。図７Ａと７Ｂは、従来の方式と新たに提案されたメッセージング方式との比較を示す。図７Ａでは、従来のユニキャスト・アプローチを使用してデータ・パケットが送信される。一方、図７Ｂでは、マルチデスティネーション、マルチキャスト・ルーティング・プロトコルを使用してデータ・パケットが送信される。したがって、あるノードで多数の出リンクに振り分けられた内容を、マルチデスティネーションの宛先をもつ一連のパケットで運ぶことができる。一般に、従来のアプローチと比較して、マルチデスティネーション・ルーティング・ノードの出次数はずっと多くすることができるため、それだけレイテンシーを低下させたマルチキャスト・ツリーとなる。

【００３６】

さらに、マルチデスティネーション、マルチキャスト・ルーティングのこの統合は、マルチデスティネーション・ルーティングの規模制限を克服できることを意味する。マルチデスティネーション・パケットが最大５０の送信先に制限され、各ノードが仮にＣとする接続数に制約されるとしよう。それでも、各ノードが最大でＣ*50の出側ノードへ接続する、何百万のノードのオーバーレイ・ツリーをわれわれは形成することができる。単一の入パケットを受信する各ノードは、それに隣接する各ノードに対応するアドレスをまとめたセットを使用してパケットを転送する。当該ツリーのルートは、Ｃ*50の子ノードへ接続可能である。これらのノードの各々は、続いて、別々のマルチデスティネーション・パケットを使用してＣ*50までの子に接続可能である。ツリーの第三のレベルで起こり得る展開は(Ｃ*50)^3である。Ｃ＝2とすると、高さ３のツリーでは10^6のノードがアドレス指定可能である。

【００３７】

さらに別の例では、分散ハッシュ・テーブル（ＤＨＴ）が位置ベースの検索をサポートする。例えば、アプリケーションが、緯度－経度位置などの特定の位置に関連したサービスまたは情報を検索することができる。多くの場合、グリッドが、複数の位置を一つの識別子に相互に関連付けるために使用される。特定の位置について、当該位置に最も近いグリッド点を見つけるためにグリッドが参照される。次に、グリッド点に対応する位置データ（例えば、郵送先住所、郵便番号、緯度－経度位置など）が、ＤＨＴへアクセスするためのキーとして使用される。ある場合には、グリッド上の複数の点が並列にクエリーの対

象になる。例えば、単一のグリッド点よりも大きな区域内のサービスを検索したい場合には、その所定の区域内の隣接したグリッド点に対してクエリーを行う。各グリッド点にユニキャスト・メッセージを送信するよりむしろ、 隣接するグリッド点をまとめたセットに対してクエリーを行うために、マルチデスティネーション、マルチキャスト・ルーティング・プロトコルを使用することが提案される。

【００３８】

この技法は、サービス発見メカニズムを探すために特に適合させ得る。タイプを問わずサービス発見メカニズムは、発見、通知及び革新のための特定のプロトコルをサポートし得る。また、特定のサービス記述フォーマット及びセマンティックをサポートし得る。サービス発見メカニズムは、ネットワーク管理ドメイン内で管理することができ、そのプロトコル及びフォーマットを定義するタイプを有する。タイプの例は、ＳＬＰ、ＵＤＤＩ及びＬＤＡＰを含む。ピアツーピア環境内で興味のあるサービス発見メカニズムを探すためにＤＨＴを使用できると想定される。この技法のさらに詳細については、参照することにより本文書に援用される、２００５年９月８日付出願の米国仮特許出願No. 60/715,388に見出すことができる。

【００３９】

識別子をまとめた非空セットを連結して、ＤＨＴへの入力として使用することができる。サービス発見メカニズムへのこのようなキー及び参照がＤＨＴに書き込まれる。ＤＨＴへの参照は、サービス発見メカニズムの記述及びそのアクセス方法、ＵＲＩ、またはサービス発見メカニムズと通信するためのソフトウェアインタフェースがあり得る。任意のサービス発見メカニズムに対して複数のキーをＤＨＴに書き込んでも良く、それによりそのメカニズムを検索するための異なる方法をサポートできる。実務上行われているように、識別子をセグメントに分割してもよく、各セグメントを個別にＤＨＴに書き込んでもよい。これは、ある種のＤＨＴベースのシステムにおいてワイルド・カードと全文読出し検索をサポートする。

【００４０】

サービス発見メカニズムはまた、例えば、ドメインの位置またはドメインによって管理されるサービスの位置などのその他の属性をもつことができる。これらの場合には、適切なサービス発見メカニズムを探すために、位置ベースのＤＨＴの検索を使用することができる。対象の位置の近くの複数のグリッド点に対して、前述のようにマルチデスティネーション、マルチキャスト・ルーティング・プロトコルを使用したクエリーを行うことができる。このようにして、ピアは、位置に基づいて、サービス発見メカニズムを見つけることができる。

【００４１】

再度ことわっておくが、オーバーレイ通信動作のいくつかの例のみを上に説明した。前述のマルチデスティネーション、マルチキャスト・ルーティング・プロトコルは、並列メッセージング方式を有するその他のオーバーレイ通信動作にも適用可能であることは容易に理解される。以下の説明は、本質的に単なる例示にすぎず、本開示、出願、または使用を限定するものではない。

【００４２】

ここに示した図面は、例示の目的でのみ示すものであり、本開示の範囲を決して限定するものではない。

【図面の簡単な説明】

【００４３】

【図１】オーバーレイ・ネットワークを有するネットワーク構成例の図である。

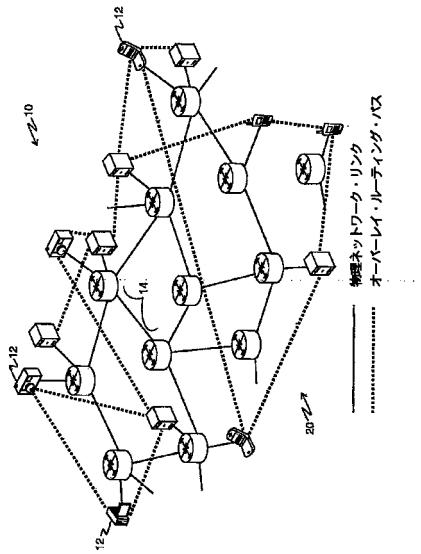
【図２】オーバーレイ・ネットワークにおけるオーバーレイ通信動作を並列実行するための方法の一例を説明するフローチャートである。

【図３】オーバーレイ・ネットワークの一部の図である。

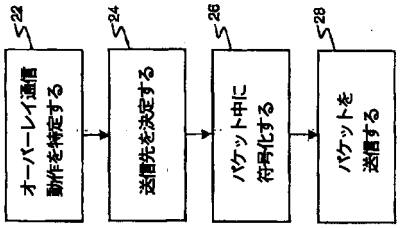
【図４】マルチデスティネーション、マルチキャスト・ルーティング・プロトコルの一部分を記述するためにどのように使用されるかを説明する図である。

【図５】Kademliaオーバーレイ・ネットワークにおけるノード探索を説明する図である。
【図６】イベント検出及び報告アルゴリズムを説明する図である。
【図７】マルチキャスト・ツリーを辿る従来の方式とマルチデスティネーション、マルチキャスト・ルーティング・プロトコルに依存する提案されたメッセージング方式をそれぞれ説明する図である。

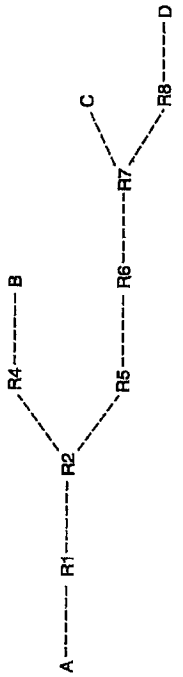
【図１】



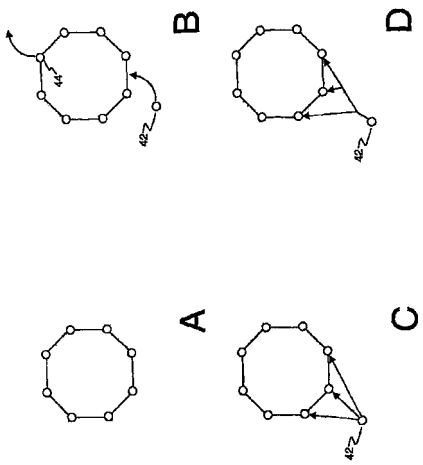
【図２】



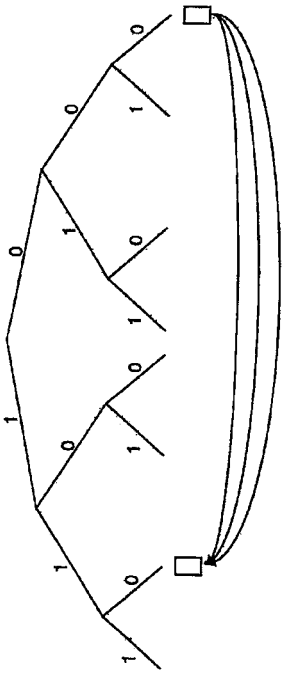
【図３】


A
R1
R2
R4
B
R5
R6
C
R7
R8
D

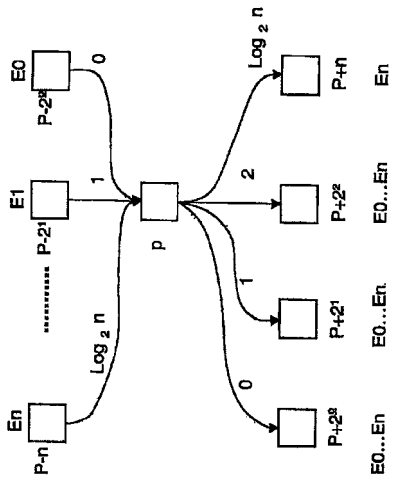
【図４】


A
B
C
D
44
42

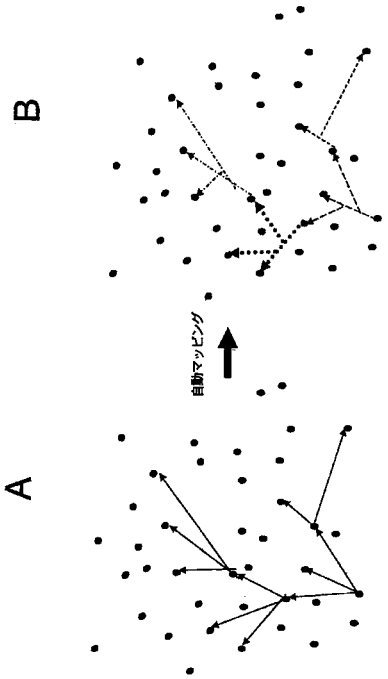
【図５】



【図６】


E0
P-2^n
E1
P-2^1
En
P-n
Log 2 n
p
Log 2 n
P+n
En
P+2^n
E0...En
P+2^1
E0...En
P+2^n
E0...En

【図７】



【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US06/35116

A. CLASSIFICATION OF SUBJECT MATTER
IPC: **G06F 15/173**(2006.01);**G06F 15/16**(2006.01)

USPC: 709/225,229
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 709/225, 229

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Google

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US PUB 2002/0069278 A1 (FORSLOW) 06 June 2002 (06.06.2002(, page 1 [0024], page 2 [0028] [0031], page 3 [0035], page 6 [0094] [0095], page 7 [0102], page 10 [0133].	1-18
X, P	US 6,954,790 B2 (FORSLOW) 11 October 2005 (11.10.2005), column 7 lines 9-35.	1-10
A	US 5,822,608 (DIEFFENDERFER et al) 13 October 1998 (13.10.1998)	
A	US 5,991,271 (JONES et al) 23 November 1999 (23.11.1999)	
A	US 6,195,347 (SEHGAL) 27 February 2001 (27.02.2001)	
A	US PUB 2005/0195774 A1 (CHENNIKARA et al) 08 September 2005 (08.09.2005)	

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
01 May 2007 (01.05.2007)

Date of mailing of the international search report
15 JUN 2007

Name and mailing address of the ISA/US
Mail Stop PCT, Attn: ISA/US
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450
Facsimile No. (571) 273-3201

Authorized officer
Saleh Najjar
DEBORAH A. THOMAS
PARALEGAL SPECIALIST

Telephone No. (703) 305-9000

Form PCT/ISA/210 (second sheet) (April 2005)

フロントページの続き

(81)指定国 AP(BW,GH,GM,KE,LS,MW,MZ,NA,SD,SL,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,MD,RU,TJ,TM),EP(AT,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR,GB,GR,HU,IE,IS,IT,LT,LU,LV,MC,NL,PL,PT,RO,SE,SI,SK,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AT,AU,AZ,BA,BB,BG,BR,BW,BY,BZ,CA,CH,CN,CO,CR,CU,CZ,DE,DK,DM,DZ,EC,EE,EG,ES,FI,GB,GD,GE,GH,GM,HN,HR,HU,ID,IL,IN,IS,JP,KE,KG,KM,KN,KP,KR,KZ,LA,LC,LK,LR,LS,LT,LU,LV,LY,MA,MD,MG,MK,MN,MW,MX,MY,MZ,NA,NG,NI,NO,NZ,OM,PG,PH,PL,PT,RO,RS,RU,SC,SD,SE,SG,SK,SL,SM,SV,SY,TJ,TM,TN,TR,TT,TZ,UA,UG,US,UZ,VC,VN,ZA,ZM,ZW