(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第5684110号
(P5684110)
(45)発行日　平成27年3月11日(2015.3.11)　(24)登録日　平成27年1月23日(2015.1.23)

(51)Int.Cl.　　F I
H04L 12/701 (2013.01)　　H04L 12/701
H04L 12/751 (2013.01)　　H04L 12/751

請求項の数 12　(全 14 頁)

(21)出願番号　特願2011-509861 (P2011-509861)
(86)(22)出願日　平成20年5月23日 (2008.5.23)
(65)公表番号　特表2011-521573 (P2011-521573A)
(43)公表日　平成23年7月21日 (2011.7.21)
(86)国際出願番号　PCT/EP2008/056376
(87)国際公開番号　WO2009/141013
(87)国際公開日　平成21年11月26日 (2009.11.26)
審査請求日　平成23年4月22日 (2011.4.22)

(73)特許権者　598036300
テレフオンアクチーボラゲット　エル　エム　エリクソン（パブル）
スウェーデン国　ストックホルム　エス－１６４　８３
(74)代理人　100076428
弁理士　大塚　康徳
(74)代理人　100112508
弁理士　高柳　司郎
(74)代理人　100115071
弁理士　大塚　康弘
(74)代理人　100116894
弁理士　木村　秀二
(74)代理人　100130409
弁理士　下山　治

最終頁に続く

(54)【発明の名称】ルーティングテーブルを維持する方法およびオーバレイネットワーク内で使用するためのノード

(57)【特許請求の範囲】
【請求項１】
Ｃｈｏｒｄ分散型ハッシュテーブル（ＤＨＴ）ベースのオーバレイネットワークのノードにおけるルーティングテーブルを維持する方法であって、所与のノードが有するルーティングテーブルは、近隣のサクセッサノード及びプレディセッサノードのセットの各々に対して、ノードのオーバレイネットワークアドレスと該ノードの物理的ロケータとの間のマッピングを含み、
前記ノードが前記オーバレイネットワークに新たに参加した他のノードを知ることができるように、前記オーバレイネットワークのノード間でＤＨＴ保守メッセージを周期的に交換し、
前記オーバレイネットワークからのノードの撤退に際して、前記撤退するノードの近隣ノードのうちの前記撤退を検知した一つのノードから、前記撤退するノードの他の近隣ノードの各々に、前記撤退を示すとともに受信ノードのルーティングテーブルに含まれていないノードに対する１つ以上のマッピングを含む離脱要求を送信し、
前記他の近隣ノードの各々において離脱要求を受信し、前記受信ノードにおいて前記ルーティングテーブルを更新するために前記マッピングを利用することを特徴とする方法。
【請求項２】
前記離脱要求を送信する前記ノードは、周期的に送信されるキープアライブメッセージに対する応答が前記撤退するノードから送信されないことの結果として前記撤退するノードの撤退を検出し、前記離脱要求を送信することにより前記離脱要求に対して反応するこ

とを特徴とする請求項１に記載の方法。
【請求項３】
前記受信ノードの前記ルーティングテーブル内に含まれていないノードに対する前記１つ以上のマッピングは、前記離脱要求を送信する前記ノードの前記ルーティングテーブル内に含まれているマッピングであることを特徴とする請求項１または２に記載の方法。
【請求項４】
離脱要求を受信するノードにおいて、前記受信ノードが前記撤退するノードの近隣ノードのうち前記離脱要求を送信したノードにとって未知の近隣ノードを知っているかどうかを判断し、知っていると判断された場合には離脱要求を１つ以上の前記近隣ノードに送信し、前記離脱要求は、受信ノードにおけるルーティングテーブル内に含まれていない少なくとも１つのノードに対する物理的ロケータマッピングへの１つ以上のオーバレイネットワークアドレスを含み、前記受信ノード或いはその他の受信ノードにおいて前記離脱要求を受信すると、前記ノードの前記ルーティングテーブルを更新することを特徴とする請求項１乃至３のいずれか１項に記載の方法。
【請求項５】
前記オーバレイネットワークのノードにおいて、ルーティングテーブルに含まれていないノードに対する１つ以上のマッピングをキャッシュし、前記オーバレイネットワークからのノードの撤退が発生した場合に、１つ以上の前記キャッシュされたマッピングを含めるように、前記撤退するノードの少なくとも一つの近隣ノードのルーティングテーブルを更新することを特徴とする請求項１乃至４のいずれか１項に記載の方法。
【請求項６】
前記離脱要求は、前記受信ノードに対する新たなサクセッサおよびプレディセッサルーティングテーブルの一つを含むことを特徴とする請求項１乃至５のいずれか１項に記載の方法。
【請求項７】
前記オーバレイネットワークは、分散型ハッシュテーブルネットワークであり、前記オーバレイネットワークアドレスはハッシュ値であることを特徴とする請求項１乃至６のいずれか１項に記載の方法。
【請求項８】
前記受信ノードにおいて前記ルーティングテーブルを更新するステップは、前記撤退するノードに対応するマッピングを削除することと、前記受信された離脱要求に含まれている新たなマッピングを前記ルーティングテーブルに加えることとを含むことを特徴とする請求項１乃至７のいずれか１項に記載の方法。
【請求項９】
Ｃｈｏｒｄ分散型ハッシュテーブル（ＤＨＴ）ベースのオーバレイネットワーク内で使用するためのノードであって、
近隣のサクセッサノード及びプレディセッサノードのセットの各々に対する、前記ノードのオーバレイネットワークアドレスと前記ノードの物理的ロケータとの間のマッピングを含む、ルーティングテーブルを格納するメモリと、
前記ノードが前記オーバレイネットワークに新たに参加した他のノードを知ることができるように、前記オーバレイネットワークの他のノードとの間でＤＨＴ保守メッセージを周期的に交換する処理ユニットを備え、
前記処理ユニットは、前記オーバレイネットワークからの近隣ノードの撤退に際して、前記撤退するノードを特定するとともに受信ノードのルーティングテーブル内に含まれていない少なくとも１つのノードに対するマッピングを含む離脱要求を前記ノードの１つ以上の近隣ノードに送信する処理ユニットと、を備えることを特徴とするノード。
【請求項１０】
近隣ノードから離脱要求が受信された場合に、前記メモリに格納されている前記ルーティングテーブルから前記離脱要求において特定されている撤退するノードに対応するマッピングを削除し、前記離脱要求に含まれている１つ以上の新しいマッピングを前記ルーテ

ィングテーブルに追加する、さらなる処理ユニットを備えることを特徴とする請求項９に記載のノード。
【請求項１１】
離脱要求が受信された場合に、前記離脱要求が前記撤退するノードから発生したものかどうかを判定し、そうでない場合に、前記ルーティングテーブルを検査して前記撤退するノードの近隣ノードのうち前記離脱要求を送信しているノードが知らないノードを特定し、該特定されたノードに対して、前記離脱するノードを特定するとともに受信ノードのルーティングテーブル内に含まれていない少なくとも１つのノードに対する１つ以上のマッピングを含む離脱要求を送信する、さらなる処理ユニットを備えることを特徴とする請求項９または１０に記載のノード。
【請求項１２】
前記ルーティングテーブル内に含まれていないノードに対する１つ以上のマッピングをキャッシュするさらなるメモリと、前記オーバレイネットワークから隣接ノードの撤退が発生した場合に、１つ以上の前記キャッシュされたマッピングを含むように前記ルーティングテーブルを更新するさらなるプロセッサと、を備えることを特徴とする請求項９乃至１１のいずれか１項に記載のノード。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、オーバレイネットワークにおける分散ハッシュテーブルを維持するためのメカニズムに関する。本発明は特に、オーバレイネットワークからのノードの離脱（leaving）を取り扱うための最適化手順に適用できる。
【背景技術】
【０００２】
ピアツーピアすなわちＰ２Ｐのネットワークは、ファイル共有およびＶｏＩＰ電話を含む多種多様なサービスを容易にするために、処理能力および通信帯域を含めて、加入ノードの共同利用リソースを用いる。セントラルサーバがない場合、特定のＰ２Ｐサービスは、リソース位置を最適化するために“オーバレイネットワーク”を用いる場合がある。オーバレイネットワークは、下位層のネットワーク（たとえば、インターネット）における可能な限り多くの物理リンクにわたるパスを表わす仮想リンクにより接続されるノードを備える。オーバレイネットワークにおける各ノードは、オーバレイネットワーク内の他の特定ノードへの一連のリンクが入っているルーティングテーブルを維持する。リソース要求は、リソース要求が、そのリソースに係るノードに到達するまでノード間を通過する。
【発明の概要】
【発明が解決しようとする課題】
【０００３】
分散ハッシュテーブル（ＤＨＴ）は、リソース名（“キー”）をオーバレイネットワーク内の位置に対応付けるための効率的な手段を提供する。ＤＨＴは、キー、たとえば曲名、ＳＩＰ　ＵＲＩｓ等を有限の値空間、たとえば１２８ビットに対応付けるためにハッシングアルゴリズムを用いる。ハッシングアルゴリズムは、値空間にわたってハッシュ値が比較的一様に広がることを保証するように選ばれる。このようにして、たとえば、１００個の曲名のハッシングは同様に、値空間にわたって比較的等間隔である１００個のハッシュ値をもたらす。オーバレイネットワーク内のノードは、ユーザ名で識別され、ユーザ名はそれ自身がそれぞれのハッシュ値に変換される。各ノードはそれから、自身の値に隣接する、値空間内の一連のハッシュ値に係るようになる。実際、ノードは、自身が“所有する”リソース名に一致するリソースが得られることのできる位置（たとえば、ＩＰアドレス）を記憶するであろう。オーバレイネットワークにおけるノードがリソース要求を受信する場合、ノードは対応するハッシュ値を所有するかどうかを決定する。ハッシュ値を所有すると、ノードはリソースの位置を要求元に（オーバレイネットワークを経て）返す。ハッシュ値を所有していないと、ノードは、リソース要求のハッシュ値に最も近いハッシ

ュ値を有するハッシュテーブル内の該当ノードを識別するために、自身のルーティングテーブルを調べ、そしてリソース要求を当該ノードに送る。受信ノードは、リソース要求に対応するハッシュ値を所有しており、したがってリソース位置を知っているノードにリソース要求が達するまで、同手順を繰り返す。

【０００４】

図１は、リング（リング内に少数のノードのみが図示されている）状に編成されているオーバレイネットワークを説明している。この例では、各ノードはリング中に少数の後続ノードおよび先行ノードばかりでなく、少数のより離れたノードの位置およびハッシュ値が入っているルーティングテーブルを維持する。図示しているネットワークでは、ノードＸは、ルーティングテーブル内に２つのサクセッサ（successor）ノードおよび２つのプレディセッサ（predecessor）ノードに対する位置だけでなく、３つの離れたノードに対する位置を維持する。ルーティングテーブル内の多数のエントリが、ネットワークをルーチングに関してより効率的に、そしてノードの脱退（withdrawal）に対してよりロバストにすることができるが、大きなテーブルは維持するのが難しく、そしてしたがって、ネットワークの不信頼性を増大させる。

【０００５】

オーバレイネットワーク内のノードは、近隣と周期的に連絡をとるように努めることにより、ルーティングテーブルにおける情報が最新であることを保証する。いくつかの異なるメカニズムがこの目的のために用いられる場合がある。

【０００６】

１）ノードは、ルーティングテーブルに挙げられている他のノードがオーバレイネットワークを離脱していないことを調べるために、キープアライブメッセージを周期的に送ることができる。このメカニズムは、ＤＨＴ手法、たとえばＰａｓｔｒｙ［Ａ．ＲｏｗｓｔｒｏｎおよびＰ．Ｄｒｕｓｃｈｅｌ：Ｐａｓｔｒｙ：大規模ピアツーピアシステムのための、拡張可能な分散オブジェクト位置およびルーチング、ミドルウェア、２００１年］、Ｃｈｏｒｄ［Ｉ．Ｓｔｏｉｃａ、Ｒ．Ｍｏｒｒｉｓ、Ｄ．Ｋａｒｇｅｒ、Ｍ．Ｆ．ＫａａｓｈｏｅｋおよびＨ．Ｂａｌａｋｒｉｓｈｎａｎ：Ｃｈｏｒｄ：インターネットアプリケーションのための拡張可能なピアツーピア検索サービス、ＡＣＭ　ＳＩＧＣＯＭＭ’０１会議の会議録、２００１年８月、サンディエゴ、カリフォルニア、ＵＳＡ］および内容でアドレスを指定できるネットワーク（ＣＡＮ）［Ｓ．Ｒａｔｓａｎａｍｙ、Ｐ．Ｆｒａｎｃｉｓ、Ｍ．Ｈａｎｄｌｙ、Ｒ．ＫａｒｐおよびＳ．Ｓｈｅｎｋｅｒ：拡張可能な内容でアドレスを指定できるネットワーク、ＡＣＭ　ＳＩＧＣＯＭＭ２００１、２００１年８月］により用いられる。

【０００７】

２）ノードは、旧エントリに置き換えて、ルーティングテーブルに挿入でき得る新しいノードについて学習するためにクエリを周期的に送ることができる（たとえば、Ｃｈｏｒｄ）。

【０００８】

３）ノードは、近隣のルーティングテーブルにおけるエントリについての情報を要求するクエリを、直通の近隣（direct neighbours）に周期的に送ることができる。この情報は、ノード自身のルーティングテーブルを更新するのに用いられる（たとえば、Ｃｈｏｒｄ）。

【０００９】

４）ノードは、自身のルーティングテーブルを近隣に周期的に送ることができる（たとえば、ＣＡＮ）。

【００１０】

ルーティングテーブルを維持するための別の（付加的な）手法は、リソース要求の発信元がルーティングテーブルに挿入されうるかどうかをノードが調べる工程を伴う（たとえば、Ｋａｄｅｍｌｉａ［Ｐ．ＭａｙｍｏｕｎｋｏｖおよびＤ．Ｍａｚｉｅｒｅｓ：Ｋａｄｅｍｌｉａ：排他的論理和尺度に基づくピアツーピア情報システム、ＩＰＴＰＳ０２の会

議録、ケンブリッジ、ＵＳＡ，２００２年３月］）。

【００１１】

ＤＨＴにおける近隣関係の例を示す図２を考える。図では、リングトポロジが仮定されている。ノードＸは、ルーティングテーブルに３つのサクセッサポインタおよび３つのプレディセッサポインタを維持する。複数のサクセッサポインタおよびプレディセッサポインタを維持する理由がロバスト性を増大させることであることは、既に明瞭であるだろう。単一のサクセッサが停止するようになる確率がｐであると、その場合、すべての３つのサクセッサが同時に停止するようになる確率はp^3である。しかしながら、極めて大きな実世界のＤＨＴに基づくオーバレイネットワークでは、このことはネットワークにおける接続性を維持するには十分でなく、所与のノードのすべての３つのサクセッサ（またはその代わりに、すべての３つのプレディセッサ）が十分短い期間内にネットワークを離脱すると、ネットワークは寸断する。

【００１２】

ノードは、手順を踏んで（gracefully）または手順を踏まずに（ungracefully）ネットワークを離脱できる。ノードは、手順を踏んで離脱する場合、実際に離脱するに先立って、ネットワークを離脱しようとする意思について近隣に知らせる。ノードは、（アプリケーション層で解釈される）離脱メッセージを送ることによりこのことを行う。これは、近隣が自身のルーティングテーブルから離脱するノードを直ちに除去できるようにする。ノードが手順を踏まずにネットワークを離脱する場合、最初に近隣に知らせることなく、ネットワークを去る。したがって、近隣はノードが離脱していることを自ら検出しなければならない。手順を踏まずに離脱する理由には以下がある。（ｉ）ノードが機能停止になった、（ii）Ｐ２Ｐアプリケーションが機能停止になった、または不意に休止した、そして（iii）利己的な振舞い。（iii）の代わりに、ユーザが、手順を踏んだ撤退（departure）につきものの遅延を回避するために手順を踏まずに離脱することを選ぶ場合がある。

【００１３】

オーバレイネットワークから手順を踏まずに撤退する場合、ノードは近隣が２つの異なる方法で離脱していることを知ることができる。

【００１４】

１）下位層のトランスポートプロトコルが信頼できる場合（たとえば、ＴＣＰ）、近隣の撤退は、トランスポート層の接続がダウンするという事実から迅速に検出される。

【００１５】

２）トランスポートプロトコルが信頼できない場合（たとえば、ＵＤＰ）、ノードは、次周期のＤＨＴ保守メッセージを近隣に送ろうと試みるまで、近隣が離脱していることを学習しない。次周期の保守メッセージ送信を待つことに加えて、ノードはまた、近隣が実際に離脱しているということが確実となりうる前に、トランザクションが時間切れするまで待たなければならない。

【００１６】

手順を踏んだ撤退と手順を踏まない撤退との双方の場合、最終結果は、離脱ノードの直通の近隣の各々がルーティングテーブルで１つ少ないポインタを有することである。たとえば、図２におけるノードＳ１がネットワークを離脱すると、ノードＸは、２つのサクセッサポインタ、すなわちＳ２およびＳ３を残しているだけである。また、ノードＸが追加のサクセッサを見つける機会を得る前に、Ｓ２およびＳ３がまたネットワークを離脱すると、ノードＸはもはやサクセッサを知らないので、オーバレイネットワークは分割されるようになる。同じ状況がノードＸのプレディセッサノードについて生じうる。

【００１７】

このようにして、現存する解決策に関連する問題は、所与のノードのサクセッサまたはプレディセッサのすべてが短期間のうちに停止すると、ネットワークは分割されるようになりうるし、そしてリソース要求はギャップを埋めることができないということである。この“短期間”は、２つの連続するＤＨＴ保守メッセージ間の時間に当てはまる。そのような保守メッセージがたとえば、６０秒ごとに送られるとした場合、単一のノードのすべ

てのサクセッサまたはプレディセッサが、この６０秒の期間内にオーバレイネットワークを離脱すると、オーバレイネットワークは崩壊する。このことは、ネットワークが高い“解約（churn）”率に直面している場合、起こり得そうな事象である。この問題に対する直感的な解決策は、ＤＨＴ保守メッセージをより頻繁に送ることであろうが、周期的な保守メッセージの送信間隔は、もたらされる信号負荷がネットワークを過負荷にするようになるので、任意には小さくできない。この問題は、Ｓ．Ｒｈｅａ、Ｄ．Ｇｅｅｌｓ、Ｔ．ＲｏｓｃｏｅおよびＪ．Ｋｕｂｉａｔｏｗｉｃｚ：ＤＨＴにおける解約の取り扱い、ＵＳＥＮＩＸの会議録、年次技術会議、２００４年６月、により確認されている。

【課題を解決するための手段】

【００１８】

本発明の目的は、ノードが停止する、またはそうでなければネットワークから脱退する場合、オーバレイネットワークに対して混乱を最小にすることである。少なくとも、本発明の特定の実施形態は、脱退するノード、または脱退するノードの近隣ノードが他の近隣のルーティングテーブルを更新できるようにすることにより、この目的を達成する。

【００１９】

本発明の第１の態様によれば、オーバレイネットワークのノードでルーティングテーブルを維持する方法が提供され、所与ノードのルーティングテーブルには、一連の近隣サクセッサノードおよび近隣プレディセッサノードのうちの各々に対して、ノードのオーバレイネットワークアドレスとノードの物理的ロケータとの間のマッピングが入っている。

【００２０】

本方法は、ノードがオーバレイネットワークから撤退する時点でまたはその直前に、撤退ノード（または撤退を承知している、撤退ノードの近隣ノードのうちの１つ）から、各近隣ノード（または撤退ノードの他の各近隣ノード）に離脱要求を送る工程、撤退を表示する工程および受信ノードのルーティングテーブルに入っていないノードに対する１つ以上のマッピングを入れる工程を備える。各近隣ノード（または他の各近隣ノード）が離脱要求を受信し、そして前記（複数の）マッピングを用いてルーティングテーブルを更新する。

【００２１】

本発明の実施形態では、脱退ノードの近隣が自身のルーティングテーブルを置換プレディセッサノードまたは置換サクセッサノードで迅速に更新できるようにする。ネットワークにおいて高い解約である場合において、ネットワーク鎖における破断の危険性が大いに減少する。

【００２２】

本発明の第２の態様によれば、オーバレイネットワークのノードでルーティングテーブルを維持する方法が提供され、所与ノードのルーティングテーブルには、一連の近隣サクセッサノードおよび近隣プレディセッサノードのうちの各々に対して、ノードのオーバレイネットワークアドレスとノードの物理的ロケータとの間のマッピングが入っている。

【００２３】

本方法では、ノードがオーバレイネットワークを撤退する直前に、撤退ノードから撤退ノードの各近隣ノードに離脱要求を送信し、撤退を示し、そして、受信ノードのルーティングテーブル内に入っていないノードに対する物理的ロケータマッピングへの１つ以上のオーバレイネットワークアドレスを入れることを備える。各近隣ノードで離脱要求を受信する時点で、ノードは（複数の）マッピングを用いて自身のルーティングテーブルを更新する。

【００２４】

本発明のこの態様の実施形態では、離脱要求内に入っている少なくとも１つのマッピングは、受信ノードには未知の、撤退ノードの近隣ノードに対応する。

【００２５】

本発明の第３の態様によれば、オーバレイネットワークのノードでルーティングテーブルを維持する方法が提供され、所与ノードのルーティングテーブルには、一連の近隣サク

セッサノードおよび近隣プレディセッサノードのうちの各々に対して、ノードのオーバレイネットワークアドレスとノードの物理的ロケータとの間のマッピングが入っている。

【００２６】

本方法では、ノードがオーバレイネットワークから撤退する時点で、撤退を承知している、撤退ノードの近隣ノードのうちの１つから、撤退ノードの他の近隣ノードに離脱要求を送信し、ここで離脱要求は撤退を示すとともに受信ノードのルーティングテーブルに入っていないノードに対する１つ以上のマッピングを含んでいる。前記他の各近隣ノードで離脱要求を受信すると、そのノードはその（複数の）マッピングを用いて自身のルーティングテーブルを更新する。

【００２７】

本発明のこの態様の実施形態では、（複数の）離脱要求を送るノードが、周期的に送信されるキープアライブメッセージに対して撤退ノードが応答しなかったという結果から撤退ノードの撤退を検出する。離脱要求を送るノードは、（複数の）離脱要求を送ることによりその検出への反応を示す。受信ノードのルーティングテーブルに入っていないノードに対する前記１つ以上のマッピングは、離脱要求を送るノードのルーティングテーブルに入っているマッピングであってもよい。

【００２８】

離脱要求を受信するノードは、その受信ノードが撤退ノードの近隣ノードのうちの、該離脱要求を送っているノードが承知していない近隣ノードを承知しているかどうかを判断するようにしてもよい。その受信ノードが、離脱要求を送っているノードが承知していない近隣ノードを承知している場合、離脱要求を送っているノードに対して、離脱要求を送信する。ここで送信される離脱要求は、その受信ノードのルーティングテーブルに入っていないノードに対する物理的ロケータマッピングへのを１つ以上のオーバレイネットワークアドレス含んでいる。受信ノードまたはさらなる各受信ノードにおいて、離脱要求を受信する時点で、ノードは自身のルーティングテーブルが更新される。

【００２９】

当然のことながら、受信ノードでルーティングテーブルを更新するステップは、撤退ノードに対応するマッピングを削除すること、および受信した離脱要求に入っている新しいマッピングをルーティングテーブルに追加することを含んでもよい。

【００３０】

本発明の実施形態に都合よく組み込み可能な機能は、それぞれのルーティングテーブルに含まれていないノードに対する１つ以上のマッピングをキャッシュすることである。ネットワークからノードが撤退する場合に、撤退するノードの少なくとも１つの近隣ノードのルーティングテーブルは１つ以上のキャッシュされたマッピングを含むように更新できる。

【００３１】

本発明の第４の態様によれば、オーバレイネットワーク内で用い、そして一連の近隣サクセッサノードおよび近隣プレディセッサノードの各々に対して、ノードのオーバレイネットワークアドレスとノードの物理的ロケータとの間のマッピングが入っているルーティングテーブルを記憶するための記憶装置を備えるノードが提供される。ノードはまた、ノードまたは近隣ノードがネットワークから撤退する時点で、ノードの１つのまたは複数の近隣ノードに離脱要求を送るように構成された処理部を備え、離脱要求は撤退するノードを識別し、そして受信ノードのルーティングテーブルに入っていないノードに対する物理的ロケータマッピングへの１つ以上のオーバレイネットワークアドレスが入っている。

【００３２】

ノードは、近隣ノードから離脱要求を受信し、上記記憶装置に入っているルーティングテーブルから離脱要求で識別されている撤退ノードに対応するマッピングを削除し、そして離脱要求に入っている１つ以上の新しいマッピングをルーティングテーブルに追加するように構成されるさらなる処理ユニットを備えるようにしてもよい。また、さらなる処理ユニットが、離脱要求が撤退するノードから発しているかどうかを決定し、撤退ノードか

ら発していない場合には、撤退ノードのすべての近隣ノードのうちの上記離脱要求を送ったノードが承知していないノードを識別するためにルーティングテーブルを調べ、そしてそのような識別されたあらゆつノードに対して離脱要求を送るように構成されてもよい。この離脱要求には、撤退ノードを識別するとともに、受信ノードのルーティングテーブルに入っていないノードに対する１つ以上のマッピングが入っている。

【００３３】

ノードは、近隣ノードがネットワークから撤退する場合に、ルーティングテーブルに含まれていないノードに対する１つ以上のマッピングをキャッシュするためのさらなる記憶装置、およびキャッシュされた１つ以上のマッピングを含むようにルーティングテーブルを更新するためのさらなるプロセッサを備え得る。

【００３４】

本発明の第５の態様によれば、オーバレイネットワークのノードでルーティングテーブルを維持する方法が提供され、所与のノードのルーティングテーブルには、一連の近隣サクセッサノードおよび近隣プレディセッサノードの各々に対して、ノードのオーバレイネットワークアドレスとノードの物理的ロケータとの間のマッピングが入っている。

【００３５】

本方法は、ノードに更新されたアドレス指定情報を提供するためにそれらノード間で保守メッセージを周期的に交換することを含む工程を。アドレス指定情報がピアノードに対して所与ノードで受信され、そしてかかるピアノードが所与ノードのルーティングテーブルに含まれていない場合、情報は所与ノードでキャッシュされる。所与ノードのルーティングテーブルに入っているノードがネットワークから脱退する場合、ピアノードはキャッシュされた情報を用いてルーティングテーブルに追加される。

【図面の簡単な説明】

【００３６】

【図１】いくつかのノードを備えているＤＨＴに基づくリング状のオーバレイネットワークを概略的に説明する図である。

【図２】さらに、ノード間の近隣関係を示す、ＤＨＴに基づくリング状のオーバレイネットワークを説明する図である。

【図３】図２のオーバレイネットワークのノードを概略的に説明する図である。

【図４】図２のオーバレイネットワークからの、手順を踏んだノードの撤退を取り扱うためのメカニズムを説明するフロー図である。

【図５】図２のオーバレイネットワークからの、手順を踏まないノードの撤退を取り扱うためのメカニズムを説明するフロー図である。

【発明を実施するための形態】

【００３７】

本明細書で説明している最適化ＤＨＴ離脱操作は、近隣ノードのルーティングテーブルが、あるノードがネットワークから撤退することにより影響を受ける場合に、近隣ノードが自身のルーティングテーブルを迅速に更新するのを手助けするノードに依存する。撤退するノードは、たとえば手順を踏んだ撤退シナリオでの近隣ノード、または手順を踏まない撤退の場合での別の近隣ノードである場合がある。これら２つのシナリオについてこれから詳細に考察することとする。

【００３８】

リングトポロジを用いるＣｈｏｒｄ　ＤＨＴに基づくオーバレイネットワークの例を示す図２を再度参照する。Ｃｈｏｒｄ　ＤＨＴは、本明細書では例として用いられているが、説明する手順は他のＤＨＴに基づくオーバレイネットワークに同様に適用できる。説明している例では、ＤＨＴに基づくオーバレイネットワークにおける各ノードは、６つの近隣、すなわち、３つのプレディセッサノードおよび３つのサクセッサノードへのポインタを維持するということを仮定している。勿論、提案メカニズムは任意の数のサクセッサポインタおよびプレディセッサポインタとともに機能する。

【００３９】

図２で、ノードＸは３つのサクセッサ、Ｓ１、Ｓ２、およびＳ３を有する。オーバレイネットワークが重大な“解約”に遭遇していると、ノードＸのサクセッサのすべてが短時間フレームのうちにオーバレイネットワークを離脱することを選ぶ可能性がある。ノードが手順を踏んで撤退することを仮定すると、ノードは自身の近隣ノードに離脱要求を送るであろう。しかしながら、上述したように、ノードＸは、すべての３つのサクセッサがネットワークを離脱してしまう前に、他のサクセッサノードのどれをも識別する時間を有しない場合がある。このシナリオで、ネットワークの分割を回避するために、オーバレイを離脱しようとする各ノードは、実際にネットワークを離脱する前に、近隣が自身のルーティングテーブルを代替の近隣ノードで埋めるのに協力する。

【００４０】

以下の本文で、用語“近隣テーブル”が、直接につながる近隣へのポインタが入っているルーティングテーブルの一部を引用するのに用いられ、用語“プレディセッサテーブル”が、プレディセッサポインタが入っている近隣テーブルの一部を引用するのに用いられ、そして用語“サクセッサテーブル”が、サクセッサポインタが入っている近隣テーブルの一部を引用するのに用いられる。

【００４１】

図２のノードＳ１が、手順を踏んでオーバレイネットワークを離脱することを選ぶ場合を考える。ネットワークでのすべての他のノードと同じように、Ｓ１は自身のルーティングテーブルに近隣ノードへのポインタを維持している。図２では、ノードＳ１が３つのプレディセッサポインタおよび３つのサクセッサポインタを維持することを仮定している。ノードＳ１のサクセッサはＳ２、Ｓ３およびＡである。ノードＳ１のプレディセッサは、Ｘ、Ｐ１およびＰ２を含む。ネットワークを離脱する前に、ノードＳ１は近隣のノードＮの各々に対して以下の手順を繰り返す。

【００４２】

●ノードＮがノードＳ１のプレディセッサであると、ノードＳ１は、Ｎに対する新しいサクセッサテーブルを構築し、そしてサクセッサテーブルをノードＮに送られる離脱メッセージに含める。ノードＳ１は、ノードＮに対して創出するサクセッサテーブルに自身を含めるべきではない。サクセッサテーブルにはＳ１のサクセッサばかりでなく、Ｓ１とＮとの間のノードを含めることができる。実際には、ＮがＳ１の直接のプレディセッサでない場合に、そのようになる。これらの介在するノードのいくつかは、それまではノードＮには未知である場合がある。（Ｃｈｏｒｄ　ＤＨＴアルゴリズムに従って、新規に加わったノードのサクセッサのみがノードを承知していて、他のノードは、次周期のＤＨＴ保守メッセージが予定される場合に新規のノードについて学習するであろう）。

【００４３】

●ノードＮがノードＳ１のサクセッサであると、ノードＳ１はノードＮに対する新しいプレディセッサテーブルを創出し、そしてプレディセッサテーブルをノードＮに送られる離脱メッセージに含める。このテーブルには、ノードＳ１とノードＮとの間に位置するＳ１のサクセッサのどれもを含む。ノードＳ１は、勿論、ノードＮに送られるプレディセッサテーブルに自身を含めるべきではない。

【００４４】

●ノードＳ１からのサクセッサ／プレディセッサテーブルが入っている離脱メッセージを受信して、ノードＮは最初にノードＳ１を自身のルーティングテーブルから取り除く。次に、Ｎは離脱メッセージで伝えられるノードのリストを細かく調べる。リストにある各ノードに対して、ノードＮはノードを自身のサクセッサテーブルおよびプレディセッサテーブルにあるエントリと比較し、そしてあるノードが存在していない場合に、そのノードをテーブルの正しい位置に挿入する。

【００４５】

この手順は、オーバレイネットワークが安定状態に留まっており、そしてノードＳ１の撤退後でも完全な接続性を保持することを保証する。すなわち、ノードＳ１の撤退は決してネットワークの運用を妨害しない。

【００４６】
ノード（たとえば、図２のノードＡ）が近隣に知らせずにネットワークを離脱すると、異なる状況が生じる。これは、Ｐ２Ｐアプリケーションの突然のクラッシュまたは何らかの他の異常終了のために、または離脱ノード側での利己的な振舞いのために起こる可能性がある。ノードＡの近隣のノードＢが、オーバレイネットワーク内で、ノードＡが停止していることを（たとえば、ノードＡに宛てた周期的なキープアライブメッセージ、保守メッセージまたは任意の他のメッセージが停止しているという事実から）最初に検出するノードであると仮定する。このとき、ノードＢは、ノードＡの撤退をノードＡの他の近隣に知らせ、そして他の近隣の近隣テーブルの内容を更新させる責務を負わされる。しかしながら、ノードＢはノードＡの近隣テーブルを完全には再現できない（この例では、ノードＢは、ノードＡの近隣であるが、しかしノードＢの近隣ではないＳ１に関する知識を有していない）ので、ノードＢ以外の近隣ノードは、ノードＢが承知していないそれらの近隣ノードの近隣テーブルを更新することによる手順に関与しなければならない。

【００４７】
ノードＡの撤退を最初に検出する近隣ノードが、ノードＡの最も離れたサクセッサまたは最も離れたプレディセッサである例外的な場合では、かかる近隣は、他の近隣ですでに知らないものとなっている、利用できる情報を有しない。このようにして、最も離れた近隣ノードはより近くの近隣ノードの近隣テーブルの内容を更新できない。しかしながら、最も離れた近隣ノードはさらに、ノードＡに代わって、より近くの近隣ノードが他の近隣ノードに近隣テーブルを伴った離脱要求を送らせるようになる空の離脱要求を送ることができる（より近くの近隣ノードが他のノードに恩恵をもたらす情報を有する）。

【００４８】
例として、図２のノードＸが自身の近隣、すなわち、ノードＳ１、Ｓ２、Ｓ３、Ｐ１、Ｐ２およびＰ３に知らせずに、ネットワークを離脱することを仮定する。さらに、たとえばノードＸがノードＳ１からのキープアライブメッセージに応答し損なうために、ノードＳ１が、ノードＸが離脱していることを検出する最初のノードであると仮定する。ノードＸの停止を検出した直後に、ノードＳ１は、ノードＸの近隣テーブルの内容のできる限り正確な表現を創出することになる。この場合、Ｓ１はノードＸの６つの近隣のうちの５つに関する情報を再現できる。これは、Ｓ１がノードＸのすべてのサクセッサおよびノードＸのプレディセッサのうちの２つを知っているからである。次に、ノードＳ１は、Ｓ１が承知している、ノードＸの各近隣Ｎに対して以下の手順を行う。

【００４９】
●ノードＳ１はノードＮの近隣テーブルの内容を構築する。

【００５０】
●ノードＳ１は、Ｓ１がノードＮに対して構築した近隣テーブルからノードＸを取り除き、ノードＸの回復された近隣テーブルから適切なノードを選び、そしてこれらをノードＮに対する新しい近隣テーブルに挿入する。より具体的には、ノードＮがノードＸのプレディセッサであると、ノードＳ１はノードＮに対する新しいサクセッサテーブルを創出する。その代わりに、ノードＮがノードＸのサクセッサであると、ノードＳ１はノードＮに対する新しいプレディセッサテーブルを創出する。

【００５１】
●ノードＳ１は、ノードＸに代わって、ノードＮに離脱要求を送る。離脱要求には、Ｓ１が創出している新しいサクセッサテーブルまたはプレディセッサテーブルを含める。

【００５２】
●ノードＮは、ノードＳ１から受信した離脱要求に基づいて自身の近隣テーブルを更新する。

【００５３】
ノードＸのルーティングテーブルの内容を決定する場合、ノードＳ１はノードＸの最も離れているプレディセッサ、すなわち、Ｐ３のＩＤ（identity）を知らない。しかしながら、Ｐ１とＰ２の双方はＰ３のＩＤを知っている。したがって、ノードＸの近隣Ｎが、ノ

ードＸに代わって別のノードから送られた離脱要求を受信する（離脱要求の発信元のソースアドレスがノードＸのアドレスと一致しないので、近隣Ｎがこれを検出できる）と、近隣ノードＮは以下の処理動作を実行する。

【００５４】

●ノードＮは、ノードＸの近隣テーブルの内容の表現を再現する。より具体的には、ノードＮがノードＸのプレディセッサであると、ノードＮはＸのプレディセッサテーブルを再現する。他方、ノードＮがノードＸのサクセッサであると、ノードＮはノードＸのサクセッサテーブルの表現を再現する。

【００５５】

●ノードＮが構築した近隣テーブルに基づいて、ノードＮは、ノードＸの任意の近隣のうち、離脱要求の送信元（すなわち、ノードＳ１）が承知していないノードを承知しているかどうかを調べる。ノードＮが図２のノードＰ２であると仮定すると、Ｐ２は離脱要求の送信元、ノードＳ１がノードＰ３（ノードＸの第３のプレディセッサ）を承知していないことを検出する。

【００５６】

●ノードＮは、ノードＸに代わって、離脱要求の送信元が承知していない、ノードＸの近隣の各々に離脱要求を送る。再び、ノードＮが図２のノードＰ２であり、そして送信元がＳ１であると仮定すると、Ｐ２はノードＰ３に離脱要求を送ることになるであろう。

【００５７】

●最後に、ノードＮは、離脱要求の送信元（すなわち、ノードＳ１）が承知していない、ノードＸの近隣のリストを（離脱要求に対してノードＮが生成する応答で）返送する。離脱要求の送信元（すなわち、ノードＳ１）はそれから、これらのノードを自身のプレディセッサテーブルに挿入できる。

【００５８】

上記例では、ノードＸがオーバレイネットワークを離脱していることを検出する最初のノードが、サクセッサ（Ｓ１）であると仮定した。ノードＸがオーバレイネットワークを離脱していることを検出する最初のノードがプレディセッサであると、プレディセッサ（たとえば、Ｐ１）は、ノードＸの最も離れたサクセッサ、すなわちこの例ではＳ３に関する知識を有しない。この場合、ノードＸのサクセッサのうちの１つ、たとえばノードＳ１が上述の工程を実行する。

【００５９】

保守運用は、ＤＨＴネットワークで実行され、そして新しいノードについて学習し、近隣ノードの状態を調べるための、参加ノード間における周期的なメッセージ交換を含む。しかしながら、ＤＨＴに基づくオーバレイネットワークにおけるノードは慣習的に、ノードが自身のルーティングテーブルを更新するために近隣から受信する情報の一部のみを用い、そして情報の残りを廃棄する。たとえば、図２のネットワーク例では、ノードＸは直通の近隣Ｐ１からノードＥの存在について学習するけれども、ノードＸがプレディセッサテーブルに３つのプレディセッサポインタ、プレディセッサＰ１、Ｐ２およびＰ３のみを維持しているので、通常、ノードＸはノードＥの情報を廃棄する。しかしながら、ノードＸがノードＥの連絡情報をキャッシュすると、ノードＸはノードＰ１（およびまたＰ２およびＰ３）の手順を踏まない、考えうる撤退からより迅速に回復できる。言い換えると、ノードＸはノードＥを自身の近隣テーブルに挿入しないけれども、ノードＸは近隣キャッシュを実装する別のデータ構造にノードＥの連絡情報を記憶する。ノードＰ１が突然ネットワークを離脱すると、ノードＸは、ノードＥが活性しているか、そしてルーティングテーブルに追加できるかどうかを確認するために、キャッシュを調べ、そしてたとえばノードＥに連絡する。

【００６０】

以下、上述したメカニズムを実装するのに相応しいＤＨＴに基づくオーバレイネットワークのノード１が概略的に説明されている図３を参照する。ノード１は、ノードのためのルーティングテーブルを記憶するように構成された記憶装置２を備える。ノードには、オ

ーバレイネットワークにおける他のノードとのインタフェース３（典型的には、ＩＰネットワーク、たとえばインターネットとのインタフェース）が備わっている。第１の処理部４は、インタフェース３を通して近隣ノードから離脱要求を受信するように構成されている。離脱要求が受信された場合、第１の処理部はルーティングテーブルを上述のように更新するように構成されている。離脱要求はまた、撤退ノードが要求を送ったノードであるかどうかを決定する第２の処理部５により受信される。撤退ノードが要求を送ったノードでない場合には、第２の処理部５が、要求を送ったノードが承知していない、撤退ノードの近隣ノードを識別するために、ルーティングテーブルを調べる。第２の処理部はそれから、１つ以上のさらなる離脱要求を構築し、そしてこれらを識別されたノードに送る。

【００６１】

第３の処理部６は、ノードのネットワークからの手順を踏んだ撤退を取り扱うように構成されている。第３の処理部６は離脱要求をノードの近隣に送り、各近隣ノードに対する１つ以上の代替マッピングを識別することによりこれを行う。この第３の処理部はまた、近隣が手順を踏まずにオーバレイネットワークを離脱する場合、近隣ノードにキープアライブメッセージを周期的に送ることに関与する、接続性検出器７により通知されるように、そして撤退ノードの近隣に適切な離脱要求を送るように構成されている。

【００６２】

第４の処理部８は、メモリ９内に、その時点のルーティングテーブルに入っていないノードに対するマッピングが入っているキャッシュを維持する。第４の処理部が近隣から離脱要求を受信する場合、第４の処理部はキャッシュメモリからマッピングを抽出し、そしてこれをルーティングテーブルに追加する場合がある。

【００６３】

次に、手順を踏んだ撤退シナリオで適用されるメカニズムを説明している図４を参照する。工程Ｓ１からＳ３までで、撤退ノードは離脱要求を近隣ノード（ここでは、３つのみが示されている）に送る。工程Ｓ４からＳ６までで、近隣ノードの各々はルーティングテーブルを更新する。

【００６４】

図５は、手順を踏まない撤退のシナリオの場合に適用されるメカニズムを説明している。工程Ｔ１で、撤退ノードの第１の近隣ノードが周期的なキープアライブメッセージを撤退ノードに送る。第１の近隣で、キープアライブへの応答がその間に受信されていないタイムアウトＴ２の後、第１の近隣は、撤退ノードに対するエントリを削除するために工程Ｔ３でルーティングテーブルを更新する。工程Ｔ４およびＴ５で、第１の近隣は撤退ノードに代わって、離脱要求を撤退ノードの近隣に送る。離脱要求には、前述のように新しいマッピングが入っている。離脱要求の受信時点で、工程Ｔ６およびＴ７で近隣は自身のルーティングテーブルを更新する。近隣のうちの１つ、この場合では第３の近隣は、第１の近隣が承知していない、撤退ノードのさらなる近隣を承知している。工程Ｔ８で、近隣は、かかるさらなる近隣にさらなる離脱要求を送る。工程Ｔ９で、さらなる近隣が自身のルーティングテーブルを更新する。

【００６５】

本明細書で説明したメカニズムは、ＤＨＴに基づくオーバレイネットワークのロバスト性を向上するので、本メカニズムは、インターネット技術特別委員会（ＩＥＴＦ）のＰ２ＰＳＩＰ作業部会により標準化途上にある、Ｐ２ＰＳＩＰ電話のような重大なＤＨＴに基づくシステムに特に有用である。

【００６６】

当業者には当然のことながら、本発明の範囲を逸脱することなく、様々な変更が上述した実施形態になされる可能性がある。

【図１】

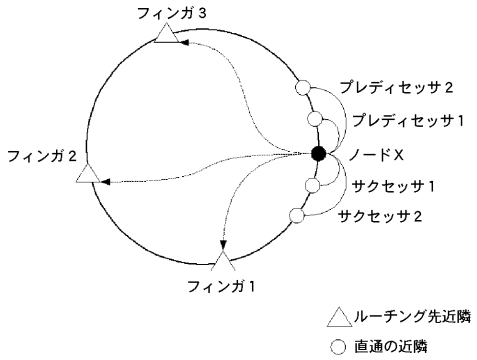

Figure 1

【図２】

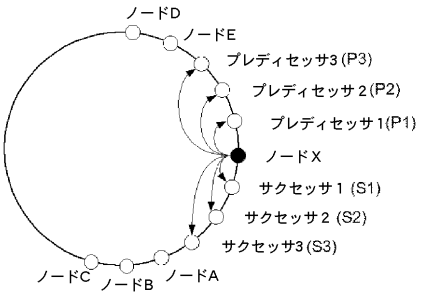

Figure 2

【図３】

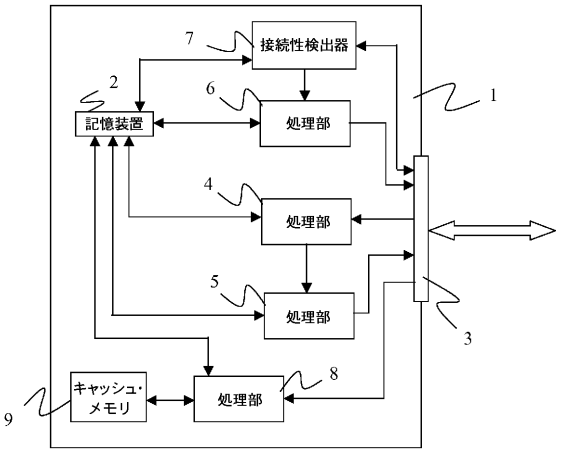

Figure 3

【図４】

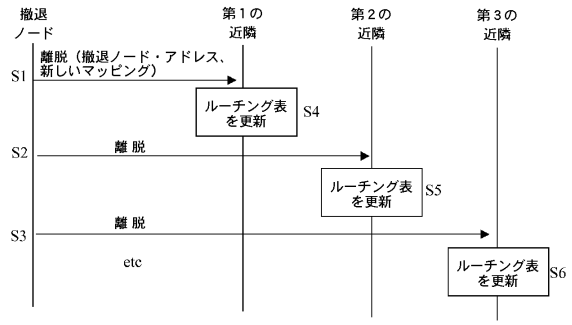

Figure 4

【図５】

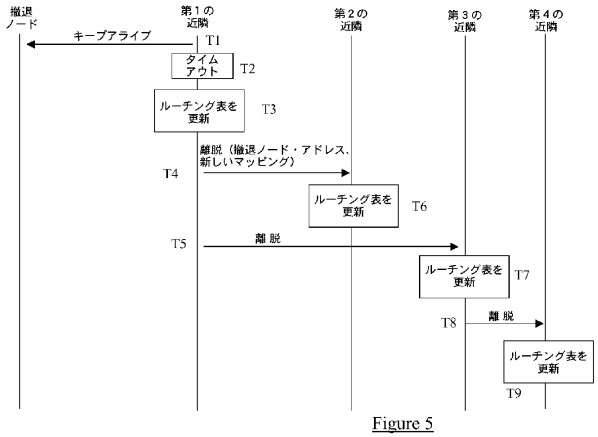

Figure 5

フロントページの続き

(72)発明者　メエンペー，　ヨウニ
フィンランド国　ヌメラ　エフアイ－０３１００，　ナーランパユンティエ　３２　エー５

審査官　永井　啓司

(56)参考文献　Ion Stoica, Chord: A Scalable Peer-to-Peer Lookup Protocolfor Internet Applications, Networking, IEEE/ACM Transactions on (Volume:11 , Issue: １) , IEEE, ２００３年　２月, 17-32頁, ＵＲＬ, http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1180543&tag=1
中村　元紀, ネットワーク分割に対するＤＨＴの可用性向上, 電子情報通信学会２００５年総合大会講演論文集　通信２, 日本, 社団法人電子情報通信学会, ２００５年　３月　７日, S-26,S-27
白石　俊之, Ｃｈｏｒｄネットワークにおけるルーチング情報の効率的な維持管理方法, 電子情報通信学会技術研究報告　Ｖｏｌ．１０６　Ｎｏ．５７７, 日本, 社団法人電子情報通信学会, ２００７年　３月　１日, 第106巻, 137-142頁
西谷　智広, 分散ハッシュテーブル入門, ＵＮＩＸ　ｍａｇａｚｉｎｅ　Ｖｏｌ．２１　Ｎｏ．６, 日本, 株式会社アスキー, ２００６年　９月２１日, 第21巻, 26-33頁
Sean Rhea, Handling Churn in a DHT, Appears in Proceedings of the USENIX Annual Technical Conference, ２００４年　６月

(58)調査した分野(Int.Cl., ＤＢ名)
Ｈ０４Ｌ　１２／００－１２／２６、１２／５０－１２／９５５