

(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2009-510953
(P2009-510953A)

(43) 公表日 平成21年3月12日(2009.3.12)

(51) Int.Cl. F I テーマコード(参考)
H04L 12/46 (2006.01) H04L 12/46 Z 5K033

審査請求 有 予備審査請求 未請求 (全 28 頁)

(21) 出願番号 特願2008-533832 (P2008-533832)
 (86) (22) 出願日 平成18年10月2日(2006.10.2)
 (85) 翻訳文提出日 平成20年5月28日(2008.5.28)
 (86) 国際出願番号 PCT/CA2006/001614
 (87) 国際公開番号 W02007/038856
 (87) 国際公開日 平成19年4月12日(2007.4.12)
 (31) 優先権主張番号 60/723,479
 (32) 優先日 平成17年10月5日(2005.10.5)
 (33) 優先権主張国 米国(US)
 (31) 優先権主張番号 60/735,884
 (32) 優先日 平成17年11月14日(2005.11.14)
 (33) 優先権主張国 米国(US)
 (31) 優先権主張番号 60/745,330
 (32) 優先日 平成18年4月21日(2006.4.21)
 (33) 優先権主張国 米国(US)

(71) 出願人 301027052
 ノーテル・ネットワークス・リミテッド
 カナダ国・エイチ4エス 2エイ9・ケベック州・サン ローラン・ブルーバード
 アルフレッド・ノーベル・2351
 (74) 代理人 110000246
 特許業務法人オカダ・フシミ・ヒラノ
 (72) 発明者 アラン, デーヴィッド
 カナダ国ケー・2・ビー 5・ピー・9、
 オンタリオ州オタワ、フォレスト・ストリート 852

最終頁に続く

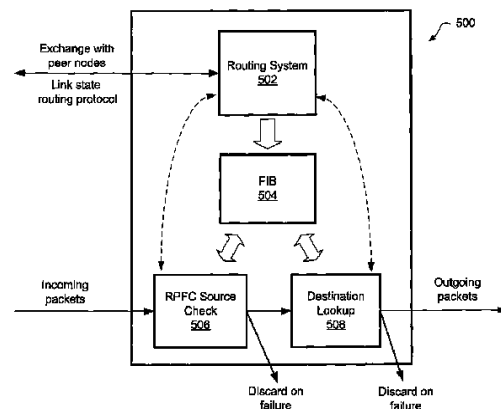
(54) 【発明の名称】 プロバイダ・リンク状態ブリッジング

(57) 【要約】

【課題】 LAN空間からWAN又はプロバイダ・ネットワーク空間にイーサネットを拡大するには、効率的なループ・フリーのイーサネットを生成することが必要とされる。

【解決手段】 ブリッジへのユニキャスト転送用に少なくとも1つのMACアドレスが、またブリッジからのマルチキャスト転送用に少なくとも1つのMACアドレスが割り当てられる。ブリッジは、ネットワークの同期の取れた構成された見取図がノード間で共有されるように、リンク状態ブリッジング・プロトコルによって状態情報を交換する。各ノードは、同位ブリッジング・ノード間を接続する最短パスを計算し、それを適切な転送テーブルに入力する。逆方向パス転送チェックが、ループ抑制を提供するため、着信パケット上で実行される。ネットワークが不安定な間、宛先MACアドレスによって識別されるユニキャスト・パケットについてのループ抑制は使用不能にされる。

【選択図】 図5



【特許請求の範囲】**【請求項 1】**

プロバイダ・リンク状態ブリッジング・イーサネット・ノードであって、
少なくとも1つの関連するユニキャストMACアドレスと、
少なくとも1つの関連するマルチキャストMACアドレスと、
それぞれのユニキャストMACアドレスと同位ノードの複数のマルチキャストMACアドレスとに基づいてノード間でリンク状態ルーティング情報を交換するため、および同位ノード間の最短パス接続性を決定するためのルーティング・モジュールであって、多数の等コスト・パスが利用可能である場合、選択された最短パスが、前記ルーティング情報の交換に関係するすべてのブリッジに対して整合性のあるものとなるように決められる、ルーティング・モジュールと、

前記ノードから同位ブリッジ・ノードへの接続性を識別するための、前記ルーティング・モジュールから受け取った転送情報が入力される転送情報ベース(FIB)であって、前記ユニキャストMACアドレスは同位ノードを指し示し、前記マルチキャスト・アドレスは同位ノードから指し示すものである、転送情報ベース(FIB)と、

着信パケットを検査し、前記パケットが、送信元MACアドレスと等しい宛先MACアドレスをもつパケットを転送するために、前記FIBによって決定された出口ポートとして使用されるポートと同じ入口ポートに到着したかどうかを判定するための逆方向パス転送チェック(RPFC)モジュールと、

前記ノードの出口ポートが同位ブリッジの宛先MACアドレスに関連付けられているかどうかを前記FIBから決定し、前記パケットを転送するための転送モジュールと、
を含むイーサネット・ノード。

【請求項 2】

前記パケットが正しい入口ポートに到着しなかったと決定された場合、前記RPFCモジュールが、前記パケットを廃棄する、請求項1に記載のイーサネット・ノード。

【請求項 3】

ネットワーク・トポロジおよび前記ルーティング・モジュールが収束していない場合、前記RPFCモジュールが、ユニキャスト・トラフィックについて使用不能にされる、請求項1に記載のイーサネット・ノード。

【請求項 4】

パケットの前記廃棄を検出することによって永続的なループが存在するかどうかをチェックするために、前記RPFCモジュールが、ユニキャスト・パケットについて定期的に使用可能にされる、請求項3に記載のイーサネット・ノード。

【請求項 5】

前記パケットの宛先MACアドレス用の出口ポートが前記FIB内で識別されない場合、前記転送モジュールが、前記パケットを廃棄する、請求項1に記載のイーサネット・ノード。

【請求項 6】

前記少なくとも1つのユニキャストMACアドレスが、ライン・カード、仮想スイッチ・インスタンス(VSI)、もしくはUNIポート、またはブリッジにおける終端のその他の任意の命名の1つに割り当てられ、あるいは前記ブリッジの背後のMAC終端を表す、請求項1に記載のイーサネット・ノード。

【請求項 7】

前記ルーティング・モジュールが、フロイドのアルゴリズムまたはダイクストラのアルゴリズムによって前記最短パスを決定する、請求項1に記載のイーサネット・ノード。

【請求項 8】

前記FIBモジュールが、前記最短パス内のノードにのみ関係する情報に基づいてデータ入力される、請求項1に記載のイーサネット・ノード。

【請求項 9】

各ブリッジが、2つのノード間の前記最短パス上にある場合、前記2つのノードが関心

に基づくコミュニティの共通部分を有するかどうかを決定することができ、前記ブリッジが、転送テーブル内に入力されるデータをしかるべく変更するように、前記リンク状態情報が、ルーティング公示内に関心に基づくコミュニティ情報を含む、請求項 1 に記載のイーサネット・ノード。

【請求項 1 0】

前記リンク状態情報が、複数の仮想 LAN (VLAN) 識別子を含むことをさらに含み、各識別子が、コントロール・プレーンでメッシュの 1 つのインスタンスを定義するために使用される、請求項 1 に記載のイーサネット・ノード。

【請求項 1 1】

前記複数の VLAN 識別子が、多数の等コスト・パスが利用され得るように、ネットワークを区分化して、前記メッシュ内におけるトラフィックの拡散を容易にするために使用される、請求項 1 0 に記載のイーサネット・ノード。

10

【請求項 1 2】

前記ルーティング・モジュールが、多数の等コスト・パス間の均衡を破り、VLAN の集まりの間で負荷分散をもたらすために、異なるランク付けアルゴリズムを各 VLAN に割り当てながら、各 VLAN についての前記最短パスを決定する、請求項 1 1 に記載のイーサネット・ノード。

【請求項 1 3】

前記リンク状態情報が、一意のマルチキャスト・グループを識別するために、前記リンク状態ルーティング公示内に含まれる拡張サービス ID (E-SID) を含み、2 つのブリッジ間の前記最短パス上にあるブリッジが、各ブリッジに関連付けられた前記ユニキャスト MAC アドレスと、前記 2 つのブリッジに共通なすべての E-SID に関する前記マルチキャスト MAC アドレスとをインストールする、請求項 1 に記載のイーサネット・ノード。

20

【請求項 1 4】

メッシュ・ネットワーク内でプロバイダ・リンク状態ブリッジング・イーサネット・ノードを構成し、動作させる方法であって、

リンク状態情報を同位ノードと交換するステップであって、各ノードが、少なくとも 1 つの関連するユニキャスト MAC アドレスと、少なくとも 1 つの関連するマルチキャスト MAC アドレスとを有するステップと、

30

前記交換されたリンク状態情報に基づいて、同位ノードに到る最短パスを最短パス・アルゴリズムによって決定するステップであって、多数の等コスト・パスが利用可能である場合、選択された最短パスが、前記ルーティング情報の交換に関係するすべてのブリッジに対して整合性のあるものとなるように決められるステップと、

同位ノードを指し示す関連するユニキャスト MAC アドレスと、同位ノードから指し示すマルチキャスト MAC アドレスとを利用する前記決定された最短パスを転送情報ベース (FIB) に入力するステップと、

着信パケットの送信元 MAC アドレスを検査し、前記パケットが、送信元 MAC アドレスと等しい宛先 MAC アドレスをもつパケットを転送するために、前記ノードの出口ポートとして使用されるポートと同じ、前記ノードの入口ポートに到着したかどうかを判定することによって逆方向パス転送チェック (RPFC) を実行するステップであって、前記 RPFC が失敗した場合、前記パケットが廃棄されるステップと、

40

前記 RPFC が成功した場合、前記 FIB 内で識別される前記パケットの宛先 MAC に関連付けられた前記ノードの出口ポートを介して、前記パケットを同位ブリッジに転送するステップと、

を含む方法。

【請求項 1 5】

ネットワーク・トポロジ、したがって、ルーティング・モジュールが収束していない場合、前記 RPFC を実行する前記ステップが、ユニキャスト・トラフィックについて選択的にバイパスされる、請求項 1 4 に記載の方法。

50

【請求項 16】

最短パスを決定する前記ステップが、フロイドのアルゴリズムまたはダイクストラのアルゴリズムを利用する、請求項 14 に記載の方法。

【請求項 17】

前記最短パス・アルゴリズムが、前記ネットワークを使用するその他のアプリケーションの、キャパシティ、速度、使用量、および正味アベイラビリティを含む群から選択されるコストの尺度をさらに含む、請求項 16 に記載の方法。

【請求項 18】

前記ルーティング・モジュールが、どちらのタイプのパケットの間にも転送の合同性が存在するように、ユニキャストおよびマルチキャスト接続性のための共通メトリックを利用する、請求項 14 に記載の方法。

10

【請求項 19】

前記 FIB に入力する前記ステップが、前記最短パス内のノードにのみ関係する情報にのみ基づく、請求項 14 に記載の方法。

【請求項 20】

前記パケットが、802.1ah に従った MAC-in-MAC カプセル化である、請求項 14 に記載の方法。

【請求項 21】

各ブリッジが、2つのノード間の前記最短パス上にある場合、前記2つのノードが関心に基づくコミュニティの共通部分を有するかどうかを決定することができ、前記ブリッジが、転送テーブル内に入力されるデータをしかるべく変更するように、リンク状態情報を交換する前記ステップが、ルーティング公示内に関心に基づくコミュニティ情報を含むステップをさらに含む、請求項 14 に記載の方法。

20

【請求項 22】

リンク状態情報を交換する前記ステップが、複数の仮想 LAN (VLAN) 識別子を含むステップをさらに含み、各識別子が、コントロール・プレーンでメッシュの1つのインスタンスを定義するために使用される、請求項 14 に記載の方法。

【請求項 23】

前記複数の VLAN 識別子が、多数の等コスト・パスが利用され得るように、前記ネットワークを区分化して、前記メッシュ内におけるトラフィックの拡散を容易にするために使用される、請求項 22 に記載の方法。

30

【請求項 24】

前記最短パスを決定する前記ステップが、多数の等コスト・パス間の均衡を破り、VLAN の集まりの間で負荷分散をもたらすために、異なるランク付けアルゴリズムを各 VLAN に割り当てながら、各 VLAN について繰り返される、請求項 23 に記載の方法。

【請求項 25】

リンク状態情報を交換する前記ステップが、一意のマルチキャスト・グループを識別するために、前記リンク状態ルーティング公示内に含まれる拡張サービス ID (E-SID) を含むステップをさらに含み、2つのブリッジ間の前記最短パス上にあるブリッジが、各ブリッジに関連付けられた前記ユニキャスト MAC アドレスと、前記2つのブリッジに共通なすべての E-SID に関する前記マルチキャスト MAC アドレスとをインストールする、請求項 14 に記載の方法。

40

【請求項 26】

リンク状態情報を交換する前記ステップが、与えられたマルチキャスト・グループの送信元、送信先、または送信元と送信先の両方のどれであることを前記ブリッジが望むかを示すアトリビュートを前記リンク状態公示に追加することによって、非対称 VPN 接続性を提供するステップをさらに含み、2つの他のブリッジ間の前記最短パス上にあることを確定した前記ブリッジが、前記送信元/送信先アトリビュートを使用して、どのマルチキャスト・グループ・アドレスがインストールされるべきかを決定する、請求項 14 に記載の方法。

50

【請求項 27】

イーサネット・ブリッジング・ネットワークであって、

前記ネットワーク内の同位ブリッジに関する転送情報を含む転送情報ベース (F I B) を各ブリッジが有する複数のブリッジであって、着信パケットが、前記着信パケットの送信元 M A C アドレスと等しい宛先 M A C アドレスをもつパケットを転送するための前記ブリッジの出口ポートとして使用されるポートと同じ、前記ブリッジの入口ポートに到着したかどうかを判定するために、逆方向パス転送チェック (R P F C) を各ブリッジが実行することが可能な、複数のブリッジと、

前記ブリッジを相互接続し、メッシュ・ネットワークを形成する複数のパスとを含み、前記 F I B が、前記複数のブリッジ間で交換されたリンク状態情報に基づいてデータ入力され、同位ブリッジ間の最短パスを決定するために使用され、選択されるパスの決定は、多数の等コスト・パスが利用可能である場合、同位ブリッジに対して整合性のあるものとなるように決められる、イーサネット・ブリッジング・ネットワーク。

10

【請求項 28】

1つまたは複数のブリッジが、同位ブリッジを前記 F I B に入力する際に利用される、少なくとも1つの関連するユニキャスト媒体アクセス制御 (M A C) アドレスと、少なくとも1つの関連するマルチキャスト M A C アドレスとを有する、請求項 27 に記載のブリッジング・ネットワーク。

【請求項 29】

ネットワーク・トポロジがブリッジ間で収束していない場合、前記 R P F C が、前記パケットの宛先アドレスによって識別されるユニキャスト・パケットについては選択的に使用不能にされ、マルチキャスト・パケットについては使用可能にされる、請求項 27 に記載のブリッジング・ネットワーク。

20

【請求項 30】

ブリッジのサブセットが、同位 P L S B ドメインを識別する M A C を介して相互接続され得る、個々のプロバイダ・リンク状態ブリッジング (P L S B) ドメインを形成する、請求項 27 に記載のブリッジング・ネットワーク。

【請求項 31】

プロバイダ・バックボーン・トランスポート (P B T) ネットワークが、 P B T M A C 識別子によって、前記ブリッジング・ネットワークのエッジ・ブリッジを用いて相互接続され得る、請求項 27 に記載のブリッジング・ネットワーク。

30

【請求項 32】

8 0 2 . 1 a h プロバイダ・バックボーン・ブリッジが、前記ブリッジング・ネットワーク上にオーバーレイされる、請求項 27 に記載のブリッジング・ネットワーク。

【請求項 33】

接続性方針メカニズムが、各同位ブリッジへのマルチポイント・ツー・ポイント (m p 2 p) ユニキャスト・ツリーのリーフとして、および各関心に基づくコミュニティに関する同位ノードの組への (S , G) ポイント・ツー・マルチポイント (p 2 m p) マルチキャスト・ツリーのルートとして、前記複数のブリッジの1つを利用することによって構成される、請求項 27 に記載のブリッジング・ネットワーク。

40

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、イーサネット・トラフィック・ルーティング・プロトコルに関し、詳細には、メッシュ型イーサネット・ネットワークにおける接続性の構成に関する。

【背景技術】

【0002】

イーサネット・ネットワーク・アーキテクチャでは、ネットワークに接続された装置は、常時、共用遠隔通信パスを使用するための能力を求めて競合する。ネットワーク・セグメントを相互接続するために多数のブリッジまたはノードが使用される場合、メッシュ型

50

ネットワーク・アーキテクチャでは、同じ宛先に到る多数の潜在的なパスが存在する。このアーキテクチャの利点は、それがブリッジ間にパス冗長性を提供し、追加リンクの形態でネットワークにキャパシティが追加されることを可能にすることである。しかし、フラディングおよび学習 (flooding and learning) というイーサネットのパラダイムは、折り返しパス (replicating path) がループしないことを保証する要件のために、メッシュ型ネットワーク内に存在する接続性の広さを、一般には任意の瞬間に利用しないことを意味する。ブリッジは、単純なリンクによって、または共用 LAN セグメントもしくは共用仮想 LAN セグメントによって相互接続され得ることに留意されたい。本文書の観点からは、セグメント、仮想セグメント、およびリンクは、実際上交換可能である。

10

【0003】

イーサネット・ネットワーク内の各ブリッジ・ノードは、与えられた装置からのパケットがどのセグメント上で到着したかを観察することによって、どの局所イーサネット・セグメントを経由してどの装置またはステーションに到達し得るかを学習する。未知の装置 (転送情報ベース (FIB: forwarding information base) 内に転送情報が存在しない装置) に送信する場合、ブリッジは、すべての接続セグメント上にメッセージを複製する (これはフラディングとして知られている)。これは2つの効果を有し、すなわち、フラディングされたパケットを観察したブリッジは、発信装置に到るための到達可能性情報を学習し、また将来のある時点で、意図された宛先装置からの、メッセージによって求められた応答またはその後の求められたものではないメッセージが観察され、それから、どのセグメントを経由して装置がメッセージに応答したかが分かることが期待される。次第にブリッジは、ネットワーク内の特定の装置に到達するための次のセグメントとしてどのセグメントが適切であるかについての独自の絵図を作り上げる。その後メッセージが送信されて来た場合、ブリッジは、その FIB を使用して、どの局所接続セグメントにメッセージを転送するかを決定することができる。ブリッジが経験を通してネットワークを学習することを可能にする手法は、透過ブリッジング (transparent bridging) として知られている。この技法の1つの主要な利点は、ブリッジングが管理者によるセットアップを必要としないことである。

20

【0004】

メッシュでは、パケットが同位装置間で転送される場合、パケットは、複数のパスにより、多数のブリッジを通過して横断することが可能であり得る。しかし、パケットが多数の経路上で同時に送信された場合、ネットワーク輻輳が増大し、可能なループ形成条件が生じることがある。このシナリオは、メッシュにおいて、マルチキャストおよび/またはフラディングされたトラフィックに関して生じるが、それは、与えられたパケットのコピーの数が、抑制されなければ指数関数的に増加するためである。参加ノードの組を論理的に接続するスパニング・ツリーを反復的に選出することによって、パケット重複を除去するために、スパニング・ツリー・プロトコル (STP: Spanning Tree Protocol) アルゴリズムが利用される。このアルゴリズムは、特にブリッジ・ループ (1つのセグメントを別のセグメントにリンクして無限ループ状況を引き起こす多数のパス) を回避するように構成される。このアルゴリズムは一般に、すべてのセグメントから選出ルート・ブリッジ (root bridge) への最短パスを計算する。最も良いパスが障害を起こした場合、アルゴリズムは、ネットワークを再計算し、次に良い経路を見つける。ルートが障害を起こした場合、新しいルートが選出され、アルゴリズムが再実行される。同時に、装置到達可能性に関するすべての学習された MAC 情報は廃棄され、新しいスパニング・ツリーが確立されると、FIB には、フラディングおよび学習を通して、徐々にデータが再入力されていく。

30

40

【0005】

STP は、接続性を提供する一方で、ノードまたは装置間に多数のアクティブ・パスがある場合に存在するネットワーク内の望ましくないループを防止する。ループが形成を許された場合、ループに入ったパケットは、ノードがそのパケットを削除するアクションを

50

取るまで、またはループが壊されるまで、巡回し続ける。ループ内のブリッジがフラディングを行った場合、またはパケットがマルチキャスト・パケットである場合、ループを巡るたびにパケットの追加コピーが生成され、これは非常に望ましくない結果である。ループ・フリー (loop-free) な接続性を確立するため、STPは、展開されたネットワーク内のすべてのブリッジに枝を伸ばすツリーを作成し、冗長パスを強制的にスタンバイまたはブロック状態にする。STPは、ネットワーク内の任意の2点間に、1度に1つのアクティブ・パスのみを許可し、これがループを防止するが、物理メッシュ接続性に関連する追加のキャパシティは利用されない。STPは、ブリッジの簡潔性のために状態メモリを持たないノード用に設計された。しかし、ループ・フリーな接続性に対する単純過ぎる手法は、物理トポロジを刈り込んで単一のループ・フリーなスパニング・ツリーにする要件のため、過大な未使用ネットワーク・キャパシティをもたらし得る。

10

【0006】

STP収束 (STP convergence) の最中のルート選出は、どのブリッジが最低IDを有するかをブリッジの組が決定すること、その後、最低IDを有するブリッジ、すなわち「ルート」に到る最低コストのパス上にあるネクスト・ホップ (next hop) を各ブリッジが決定することによって決定される。ブリッジは最初、スパニング・ツリー・プロトコル交換において、自らのIDとヌル・コストとを公示する。より低いIDを有する公示を受信した場合、ブリッジは、自らのIDを公示することを止め、そのより低いIDとコストとを公示する。同様に、現在の最低IDに到るより低いコストを有するセグメントに遭遇した場合、ブリッジは、その再放出される公示と内部状態とをしかるべく修正する。最終的に、ネットワークは、最低IDの共通見解に収束し、各ブリッジは、ルートに到る最低コスト・パス上の次のセグメントを知る。コストが変化した場合、またはスパニング・ツリー内の1つのネットワーク・セグメントが到達不能になった場合、STPアルゴリズムは、セグメントの変更された組を利用する接続性を提供するために、反復的に新しいスパニング・ツリー・トポロジを計算する。ネットワークのブリッジが収束していない間、ループ形成および複製を防止するために、ポートがブロックされて、ネットワーク・トラフィックを停止させる。

20

【0007】

図1は、メッシュ・ネットワークにおけるスパニング・ツリー・プロトコル・アルゴリズムが、ネットワークを介したパケットの転送にどのように影響するかについての一例を示している。送信元装置Aから、STPアルゴリズムは、(マルチキャスト経路として示された) 最低コストを達成する各ノード間のリンクを決定することによって、ネットワークを介したパスを緻密に計画している。この経路は、ネットワークを横断するすべてのトラフィックが利用可能なパスを決定する。ブリッジ110に接続された装置Aが、ネットワーク上において新しい装置Bにパケットを送信する。ネットワークのブリッジは、この装置用のデータが入力された転送テーブルを有していない。パケットがネットワーク内にフラディングされ、スパニング・ツリーによって定義されたパスによって各ブリッジ・ノードに転送される。スパニング・ツリーに従って、パケットは、ブリッジ110からブリッジ112まで転送される。ブリッジ110から隣接ブリッジ116、120へのリンクは、使用不能またはブロックされており、スパニング・ツリーの一部ではないので、利用されない。ブリッジ112から、パケットは、ネットワークのその他のノードであるブリッジ120、122、124、114に転送される。ブリッジ114は、パケットをさらにブリッジ116、118に転送する。装置Bは、パケットのフラディングされたコピーをブリッジ122を介して受信する。すべての中間ブリッジは、送信元「A」に到達するためのセグメントを知り、FIBをしかるべく更新する。その後、装置Bによって送信される応答は、戻り経路がネットワーク内の中間ブリッジに知られているので、フラディングされることなく、ネットワークを横断して装置Aに直接転送される。基本的に、パケットのフラディングがネットワークに入り、装置Bから装置Aに1つの応答が戻され、初期メッセージを観察して「A」への最適転送を学習した中間ブリッジは、それと同様に、応答を観察して「B」への最適転送を学習する。

30

40

50

【0008】

イーサネット・ネットワークにおけるSTPの使用は、ネットワーク内のリンクまたはLANセグメントの利用効率を制限する。STPアルゴリズムによって選択されなかったセグメントおよびリンクは、使用不能であり、ネットワークの物理トポロジに何らかの変化が生じて、そのリンクが結果のスパンニング・ツリーの一部になるまで、基本的にアクティブ・ネットワーク・トポロジの一部ではない。これは、これらのリンクのキャパシティを阻害する。小規模LAN環境では、これは問題にならないかもしれないが、WANまたはプロバイダ・ネットワークでは、未使用キャパシティは、スケーリングに影響を与え、活用が不十分な投資に相当する。加えて、障害が発生した場合、STPを再収束させ、ネットワーク内のすべてのブリッジを安定させるのに要する時間の間、ネットワーク全体のパフォーマンスが影響を受ける。現在のスパンニング・ツリーのルートは、必ずしも物理最短パス上にあるとは限らないので、スパンニング・ツリーを利用する与えられたパケットによって取られるパスは一般に、送信元と宛先の間 shortestパスではない。例えば、装置Cが隣接ブリッジ116に接続された場合、装置Aからのパケットは、必ずしもブリッジ116への最短パスを提供しない現在のスパンニング・ツリーに基づいて、依然としてブリッジ112、114を横断しなければならない。最後に、プロバイダ・ネットワークでは、仮想プライベート・ネットワーク(virtual private network)の形態をとる多くの小規模な関心に基づくコミュニティ(community of interest)が存在することができ、情報のフラッキングをそれらの関心に基づくコミュニティ内に制限するには、一般にコミュニティ毎に異なるスパンニング・ツリーを必要とする。

10

20

【0009】

STPパフォーマンスを改善し、メッシュ型イーサネット・ネットワーク内でのループ形成を緩和するために、様々な方法が提案された。現在の提案のいくつかは、ループ形成の影響を制限するため、MPLSに類似したラベル、またはイーサネット・パケットへの有効期間(TTL: time-to-live)フィールドの追加を必要とする。しかし、これらの解決策を実施するには、基本的なデータ・プレーン(data plane)の修正が必要とされ、それがこれらの解決策の有用性を制限する。加えて、パケットが複製され得る回数を制限するだけでは、ループ形成問題に対する明らかに不完全な解決策にしかない。

30

【発明の開示】

【発明が解決しようとする課題】

【0010】

LAN空間からWANまたはプロバイダ・ネットワーク空間にイーサネットを拡大するには、効率的なループ・フリーのイーサネットを生成するための改善されたメカニズムが必要とされる。これは、メッシュ接続性をより良く利用するメカニズム、障害またはトポロジ変更の影響を制限する(例えば、接続性を中断しない、または接続性の中断を最低限に抑える)メカニズム、ループを抑制する、またはループ形成の影響を緩和するメカニズムである。したがって、ネットワーク利用率を改善しながらも、一時的または永続的ループの落とし穴を回避するイーサネット・ネットワークが必要とされている。

40

【課題を解決するための手段】

【0011】

本発明は、プロバイダ・リンク状態ブリッジング(PLSB: Provider Link State Bridging)を利用することによって、ループ・フリーなイーサネット・ネットワークを生成するためのシステムおよび方法を提供する。PLSBは、コントロール・プレーン(control plane)によってイーサネットMAC転送テーブルの静的構成を拡張し、リンク状態ルーティング・システム(link state routing system)によるイーサネット転送の直接操作を利用する。少なくとも2つのMACアドレスが、各ブリッジに関連付けられ、1つはブリッジへのユニキャスト転送用、1つはブリッジからのマルチキャスト転送用である。ユニキャストM

50

ACは、ブリッジにパケットを送信する場合に、パケットの宛先アドレスとして、同位ブリッジによって使用され、ブリッジからその同位ブリッジへのユニキャストまたはマルチキャスト・トラフィックの送信元としてブリッジを識別する。マルチキャスト・アドレスは、多数の同位ブリッジに同時にトラフィックを送信する場合に、ブリッジによって使用され、ルーティング・システムによって構成された特定のマルチキャスト・ツリーを識別する。

【0012】

ブリッジは、リンク状態プロトコル(link state protocol)によって状態情報を交換し、最短パス・アルゴリズムによって同位ブリッジ間の最短パスを決定するのに利用される状態データベース(state database)を維持する。任意の2つのブリッジ間の等コスト複数パス(Equal Cost Multiple Paths)の条件下では、「行き(go)」および「戻り(return)」ユニキャスト・パス、ならびに2つのマルチキャスト・ツリーの関連するセグメントが、同様にルーティング(co-route)されるように、一貫したパス選択が、すべてのブリッジによって行われる。その後、転送テーブルまたは転送情報ベース(FIB)には、適切なユニキャストおよびマルチキャスト・アドレスが入力される。その後、一意のユニキャストおよびマルチキャスト配送ツリーが、各ブリッジで生成され、より効率的なリンク利用を提供する。

10

【0013】

ループ抑制は、各ブリッジにおいて受信パケット上で実行される逆方向パス転送チェック(reverse path forwarding check、RPFC)によって提供される。RPFCは、パケットがFIB内で識別される正しいポートまたはインターフェースに到着したかどうかを決定し、正しいポートに到着しなかったパケットは廃棄される。これは、(上記のような)任意の2点間のすべてのパスの同ルーティング特性(co-routed property)のために可能になる。ループ形成が潜在的に発生し、ネットワーク・トポロジが収束しない、ブリッジまたはリンク障害の間など、ネットワークが不安定な時間の間、RPFCは、トラフィックへの影響を最低限に抑えるため、ユニキャスト・パケットに対しては使用不能にされることができる。各ブリッジにおいてトポロジが再収束した時、RPFCは、再び使用可能にされることができる。マルチキャスト・トラフィックの場合、RPFCは、無制限な複製の可能性を低減するため、決して使用不能にされることはない。

20

30

【0014】

仮想プライベート・ネットワーク(VPN)は、PLSBネットワーク上にマッピングされることができ、トラフィックが特定の一意の関心に基づくコミュニティにのみ送られるように、一意のマルチキャスト・ツリーが、エッジ・ブリッジ(edge bridge)毎のVPN毎にマッピングされることを可能にする。マルチキャストVPNトラフィックは、共通であるパスに対してインストールされるVPNグループ・マルチキャスト・アドレスによって、VPNに参加しているブリッジにのみ配送される。

【0015】

ネットワークは、(本発明の譲受人に譲渡された出願US 20050220096で説明されるような)プロバイダ・バックボーン・トランスポート(Provider Backbone Transport)、およびPLSBと組み合わせられた802.1ahプロバイダ・バックボーン・ブリッジ(Provider Backbone Bridge)などのその他の技術を利用することによって、拡大されることができる。PBBは、PLSB WANDメインおよびPLSBメトロ・ドメインなどのPLSB領域を互いに結び付けるため、またはPBTを使用して接続性を拡張するために使用されることができる。

40

【0016】

したがって、本発明の一態様は、プロバイダ・リンク状態ブリッジング・イーサネット・ノードを提供し、そのノードは、少なくとも1つの関連するユニキャストMACアドレ

50

スト、少なくとも1つの関連するマルチキャストMACアドレスと、それぞれのユニキャストMACアドレスと同位ノードの複数のマルチキャストMACアドレスとに基づいてノード間でリンク状態ルーティング情報を交換するため、および同位ノード間の最短パス接続性を決定するためのルーティング・モジュールであって、多数の等コスト・パスが利用可能である場合、選択された最短パスが、ルーティング情報の交換に係るすべてのブリッジに対して整合性のあるものとなるように決められる、ルーティング・モジュールと、ノードから同位ブリッジ・ノードへの接続性を識別するための、ルーティング・モジュールから受け取った転送情報が入力される転送情報ベース(FIB)であって、ユニキャストMACアドレスは同位ノードを指し示し、マルチキャスト・アドレスは同位ノードから指し示すものである、転送情報ベース(FIB)と、着信パケットを検査し、そのパケットが、送信元MACアドレスと等しい宛先MACアドレスをもつパケットを転送するために、FIBによって決定された出口ポートとして使用されるポートと同じ入口ポートに到着したかどうかを判定するための逆方向パス転送チェック(RPFC)モジュールと、ノードの出口ポートが同位ブリッジの宛先MACアドレスに関連付けられているかどうかをFIBから決定し、パケットを転送するための転送モジュールとを含む。

10

20

30

40

50

【0017】

本発明のさらなる態様は、メッシュ・ネットワーク内でプロバイダ・リンク状態ブリッジング・イーサネット・ノードを構成し、動作させる方法を提供し、その方法は、リンク状態情報を同位ノードと交換するステップであって、各ノードが、少なくとも1つの関連するユニキャストMACアドレスと、少なくとも1つのマルチキャストMACアドレスとを有するステップと、交換されたリンク状態情報に基づいて、同位ノードに到る最短パスを最短パス・アルゴリズムによって決定するステップであって、多数の等コスト・パスが利用可能である場合、選択された最短パスが、ルーティング情報交換に係るすべてのブリッジに対して整合性のあるものとなるように決められるステップと、同位ノードを指し示す関連するユニキャストMACアドレスと、同位ノードから指し示すマルチキャストMACアドレスとを利用する決定された最短パスを転送情報ベース(FIB)に入力するステップと、着信パケットの送信元MACアドレスを検査し、そのパケットが、送信元MACアドレスと等しい宛先MACアドレスをもつパケットを転送するために、ノードの出口ポートとして使用されるポートと同じ、ノードの入口ポートに到着したかどうかを判定することによって逆方向パス転送チェック(RPFC)を実行するステップであって、RPFCが失敗した場合、パケットが廃棄されるステップと、RPFCが成功した場合、FIB内で識別されるパケットの宛先MACに関連付けられたノードの出口ポートを介して、パケットを同位ブリッジに転送するステップとを含む。

【0018】

また、本発明のさらなる態様は、イーサネット・ブリッジング・ネットワークを提供し、イーサネット・ブリッジング・ネットワークは、ネットワーク内の同位ブリッジに関する転送情報を含む転送情報ベース(FIB)を各ブリッジが有する複数のブリッジであって、着信パケットが、着信パケットの送信元MACアドレスと等しい宛先MACアドレスをもつパケットを転送するためのブリッジの出口ポートとして使用されるポートと同じ、ブリッジの入口ポートに到着したかどうかを判定するために、逆方向パス転送チェック(RPFC)を各ブリッジが実行することが可能な、複数のブリッジと、ブリッジを相互接続し、メッシュ・ネットワークを形成する複数のパスとを含み、FIBは、複数のブリッジ間で交換されたリンク状態情報に基づいてデータ入力され、同位ブリッジ間の最短パスを決定するために使用され、選択されるパスの決定は、多数の等コスト・パスが利用可能である場合、同位ブリッジに対して整合性のあるものとなるように決められる。

【0019】

本発明のその他の態様および特徴は、本発明の特定の実施形態についての以下の説明を添付の図面と併せて検討することで、当業者に明らかになるであろう。

【0020】

本発明のさらなる特徴および利点は、以下の詳細な説明を添付の図面と併せて理解する

ことから明らかとなるであろう。

【発明を実施するための最良の形態】

【0021】

添付の図面の全体で、同様の特徴は同様の参照番号によって識別されることに留意されたい。

【0022】

本発明の実施形態が、図2～図9を参照しながら、以下で例としてのみ説明される。本発明は、ループ・フリーなイーサネット・ネットワーキングのためのシステム、方法、および装置を提供する。

【0023】

プロバイダ・リンク状態ブリッジング(PLSB)は、ループ・フリーな最短パス転送とともに、ネットワーク・キャパシティのより効率的な使用を提供することによって、イーサネット・ネットワークが、LAN空間からWANまたはプロバイダ・ネットワーク空間に拡大されることを可能にする。透過ブリッジングと組み合わされたスパニング・ツリー・プロトコル(STP)を使用することによって各ノードにおいて学習されたネットワーク図を利用する代わりに、PLSBベースのネットワークでは、メッシュ・ネットワークを形成するブリッジは、ネットワーク・トポロジの同期が取られた見取図を有する。これは、リンク状態ルーティング・システムのよく理解されたメカニズムを通して達成される。ネットワーク内のブリッジは、ネットワーク・トポロジの同期が取られた見取図を有し、必要なユニキャストおよびマルチキャスト接続性の知識を有し、ネットワーク内の任意の1対のブリッジ間の最短パス接続性を計算することができ、ネットワークの計算された見取図に従って個別に転送情報ベース(FIB)にデータを入力することができる。すべてのノードが、同期が取られた見取図におけるその役割を計算し、そのFIBにデータを入力した場合、ネットワークは、1組の同位ブリッジからある任意のブリッジに到るループ・フリーなユニキャスト・ツリーと、ある任意のブリッジから同じ1組の同位ブリッジに到る合同かつループ・フリーなポイント・ツー・マルチポイント(p2mp)マルチキャスト・ツリーとを有する。その結果、あるブリッジ対の間のパスは、スパニング・ツリーのルート・ブリッジの移動に制約されず、全体的な結果は、メッシュの接続性の広さをより良く利用することができる。

【0024】

PLSBは、イーサネット・ブリッジ接続性の同等物を提供するが、これをフラッディングおよび学習の代わりにFIBの構成によって達成する。そのようなものとして、PLSBは、B-MAC(バックボーンMAC)の転送を構成し、PBBアダプテーション機能をごくわずが変更した、プロバイダ・バックボーン・ブリッジ(PBB)またはMAC-in-MACと呼ばれるIEEE(電気電子技術者協会)802.1ahドラフト規格など、新たに出現した規格によって使用されることができ、クライアント・イーサネットがPLSBネットワークによって提供される接続性を変更なく利用し得るように、クライアント・ブロードキャスト挙動をPLSBマルチキャストにマッピングする。

【0025】

PLSB動作は、仮想LAN(VLAN)の使用によるネットワークのパーティショニングを介して、その他のコントロール・プレーンまたは透過ブリッジングと組み合わされることができる。VLAN IDが、コントロール・プレーン・レベルでメッシュの1つのインスタンスを定義するために使用されることができ、PLSBの場合、これは、分散リンク状態ルーティング・システムによって駆動される。ネットワーク機能のVLANパーティショニングが使用される場合、PLSBは、本発明の譲受人に譲渡された2004年4月4日出願された米国特許出願第20050220096号で開示されたようなプロバイダ・バックボーン・トランスポート(PBT)など、その他のイーサネット・ネットワーク技術とサイド・バイ・サイド(side-by-side)な互換性がある。

【0026】

PLSBは、任意の2つのブリッジ間の接続性が双方向で同じパスをたどるように、対

10

20

30

40

50

称なメトリックを使用し、マルチキャストされるパケットとユニキャストされるパケットの間に転送の合同性が存在するように、ユニキャストおよびマルチキャスト接続性に対して共通のメトリックを使用する。イーサネット・クライアントをトランスポートするために PLSB が使用される場合、これが望ましい以下のような多くの理由が存在する。

【0027】

初期パケットがマルチキャスト・パス上でフラッディングされるフロー内ではパケットの並べ替えが可能でなく、ユニキャスト・パス上で転送を可能にするために、フローの最中に転送情報が学習される。

【0028】

非対称な障害がループを引き起こし得る場合、スパニング・ツリーを利用するクライアントに関する知られた問題に対処するために、はるかに高い確率の障害が対称的である。

【0029】

クライアント IEEE 802.1ag マルチキャスト接続性障害管理 (CFM: Connectivity Fault Management) パケットの転送と PLSB ネットワークを介した対応するユニキャスト・パスの合同性。

【0030】

PLSB レイヤ CFM パケットの同様の合同性。

【0031】

対称なメトリックは、等しい「行き」および「戻り」トランスポート遅延をもたらし、それが、ネットワーク内でタイミング情報を広めるためのレイヤ 2 クロック分配方式の有効性および堅牢性を大幅に改善する。

【0032】

以下の説明でより詳細に説明されるように、イーサネット・クライアントのサービス要件とは独立に、対称なメトリックは、ループ抑制も容易にする。

【0033】

PLSB は、透過 LAN サービスを使用することができる C-MAC (カスタム MAC) レイヤまたはその他のレイヤ・ネットワークに透過 LAN サービスを提供するために、MAC 構成を使用して、1 組の (わずかに変更された) 802.1ah プロバイダ・バックボーン・ブリッジの間の最短パスのループ・フリーな (ユニキャストおよびマルチキャスト双方の目的のため) 接続性を構成する。これは、関連する VLAN、およびルーティング・システム公示上での MAC 情報のピギーバックングのため、STP の代わりに、プロバイダ・バックボーン・ブリッジ・ネットワーク内でリンク状態ルーティング・プロトコルの動作を必要とする。

【0034】

トランジット・ブリッジは、ユニキャストまたはマルチキャスト接続性を終了させる必要がない場合、MAC 情報を提供せず、その他のブリッジからの MAC アドレス公示を処理することを選択することができる。802.1ah の場合、プロバイダ・バックボーン・ブリッジ (PBB) として知られるエッジ・ブリッジは、ほぼ常に MAC 情報を提供するが、純粋にトランジット・ブリッジは提供しない。

【0035】

少なくとも 2 つの MAC アドレスが、各ブリッジに関連付けられ、1 つはブリッジへのユニキャスト転送用、少なくとも 1 つはブリッジからのマルチキャスト転送用である。ユニキャスト MAC は、ブリッジにパケットを送信する場合に、パケットの宛先アドレスとして、同位ブリッジによって使用され、ブリッジからその同位ブリッジへのユニキャストまたはマルチキャスト・トラフィックの送信元としてブリッジを識別する。マルチキャスト・アドレスは、多数の同位ブリッジに同時にトラフィックを送信する場合に、ブリッジによって使用され、ルーティング・システムによって構成された特定のマルチキャスト・ツリーを識別する。マルチキャスト・ツリーは、PLSB ドメイン内のすべてのブリッジにパケットを配送するように構成されることができ、したがって、特定の送信元ブリッジのためのブロードキャスト機能を実施する。マルチキャスト・ツリーは、特定の関心に基

10

20

30

40

50

づくコミュニティ、一般にクライアント・レイヤV P Nに属するブリッジにのみ配送を制限するために、任意選択的にブロードキャスト・ツリーの限定的なサブセットを識別することができる。

【0036】

P L S Bは、構成された情報を利用し、与えられたメッシュ・ネットワーク内のすべてのセグメントを使用可能にするので、透過ブリッジングに関連するフラッディングは、必要とされず、望ましくない。したがって、P L S B挙動に割り当てられたV L A Nパーティションはどれも、透過ブリッジングのためのものとは異なる「未知の」宛先M A Cアドレスを有するパケットを扱うために変更された転送挙動も持たなければならない。P L S Bの場合、未知の宛先アドレスを有するパケットは、黙って廃棄される。

10

【0037】

図2 aは、P L S Bを利用するネットワークの概略図である。共用ネットワーク・トポロジから、各ノードは、最短パス・アルゴリズムを使用して、ネットワーク内のその他のプロバイダ・バックボーン・ブリッジ(P B B)またはノードに到る最適最短パスを計算する。ネットワーク全体に対する最短パスの適用の結果、およびブリッジ内のF I Bへの対応するデータ入力が、各ブリッジからネットワークのメンバ・ブリッジに到るメッシュ経路の一意のツリーを提供する。例えば、図2 aの装置Aからの最短パスを使用して計算されたマルチキャスト・ツリーは、特にS T Pのようにルート・スイッチ選出が存在しない点で、図1に示されたスパニング・ツリーとは異なる。図2 aに示されるように、最短パス・アルゴリズムの利用は、装置Aから発信されたパケットが、隣接ブリッジ120、116へのより直接的な経路を進むことを可能にするが、S T Pを使用すると、最短パス経路は、図1に示されたように、使用不能にされることがある。

20

【0038】

フラッディングおよび学習の透過ブリッジング動作は、P L S Bを実施する802.1 a h P B Bによって、P L S B上にマッピングされることができる。例えば、クライアント装置Bのロケーションが、P L S Bネットワークのブリッジ110に知られていない場合、AからBに宛てたパケットは、そのブリッジに割り当てられたグループ・アドレスを使用するブリッジ110によって、マルチキャスト・パケット内にM A C - i n - M A Cカプセル化され、ブリッジ110の送信元アドレスを有する。マルチキャスト・メッセージは、P L S Bを介してネットワークを横断し、コピーは最終的にノード122に到着し、そこでM A C - i n - M A Cカプセル化が解除され、コピーが装置Bに転送される。ブリッジ122のM A C - i n - M A C透過ブリッジング機能は、M A C - i n - M A Cカプセル化内の送信元B - M A Cアドレスを観察し、Aに到達するには、ブリッジ110を経由すべきであることを連想する。装置Bは、その後、メッセージに応答するとき、「A」に宛てたメッセージをブリッジ122に送信する。ブリッジ122は、AのM A C - i n - M A C宛先がブリッジ110であることに気づき、ブリッジ110に宛てたユニキャスト・パケット内にメッセージをラップする。パケットは、ブリッジ112を介してブリッジ110に送信され、その後、ブリッジ110は、M A C - i n - M A Cカプセル化を解除し、装置Aに到達するための正しいポート上にパケットを転送する。同様に、ブリッジ110は、P L S Bネットワーク内でBに到達するには、ブリッジ122を経由することを観察する。装置Aから装置Bおよび装置Bから装置Aに送信される将来のメッセージはどれも、P L S Bネットワークを介した学習されたユニキャスト転送を今では使用することができる。

30

40

【0039】

各ブリッジのユニキャストおよびマルチキャスト・ツリーは、合同であることが必要とされ、これは、対称リンク・メトリックの使用の直接の結果である。等コストの多数のパスに遭遇した場合、パスの共通ランク付けを達成する分散手段が必要とされ、これは多くの潜在的な解決策を有するよく理解された問題である。1つの典型的な解決策は、等コストのパスが分かれ、合流する2つのブリッジを識別し、最高のブリッジ番号を選択し、等コスト・パス・セグメント上に最高の番号の隣接ブリッジを有するそのブリッジからのノ

50

へのパスを選択することである。その他の方法は、当業者には明らかであろう。任意の2つのブリッジ間のランク付けされた最短パスは対称であり、そのため、任意の2つのブリッジ間のユニキャストおよびマルチキャスト転送の合同性が達成される。

【0040】

与えられたブリッジについて、それが与えられたブリッジ対の間の最短パス上にあるかどうかを決定する場合、様々な最短パス・アルゴリズムが、最適パスを計算するために使用されることができる。フロイドのアルゴリズム [R. Floyd: Algorithm 97 (shortest path), Communications of the ACM, 7:345, 1962]、またはダイクストラのアルゴリズム [E. W. Dijkstra: A note on two problems in connexion with graphs, Numerical Mathematics, 1:269-271, 1959] など、グラフ・ベースのアルゴリズムが、同位ノード間の最短パスを計算するために、PLSBブリッジ内で実施されることができる。任意の適切な最短パス・アルゴリズムも利用され得ることを理解されたい。フロイドのアルゴリズムは、コスト行列から距離行列を計算し、一方、ダイクストラのアルゴリズムは、頂点から他のすべての頂点までの最短距離を計算する。ツリーの数は基本的なフロイドのアルゴリズムの計算量に影響を与えないことに留意されたい。アルゴリズムは、パスがルート・ブリッジ・パスから延びる必要がない点で、STPアルゴリズムとは異なるパス図を生成する。STPは、すべての可能なパスについて使用されるより制約的な「ツリー」構造を生成し、一方、最短パス法は、同様に経路選択によって制限されない。

10

20

【0041】

最短パス・アルゴリズムは、トラフィック工学情報を考慮するように変更されることができる。例えば、最短パスは、キャパシティ、速度、使用量、およびアベイラビリティなどのコストの尺度を含むことができる。ブリッジ間での送信元MACの保存は、実際のトラフィック行列が観察され、ルーティング・システムへの入力として使用され、リンク負荷当たりの標準偏差を最小化し、負荷バランスを容易にし得ること意味する。最短パス・アルゴリズムの副作用は、現在ノードよりも宛先に近いことが知られているネクスト・ホップに対する一般に認められた業界用語である、「ループ・フリーな代替 (loop-free alternate)」が、コネクションレス高速経路変更のために、ネットワーク接続性を計算する一部として計算され得ることであることも留意されたい。

30

【0042】

図2bに示されるように、装置Bからブリッジ122を見る場合、最短パス・アルゴリズムから得られるツリーは、図2aに示された装置Aからブリッジ110を見た場合のツリーとは異なる。装置Bからのトラフィックは、接続ブリッジ122からの最短パスによってその宛先に到着することができ、生成されたツリーは、ブリッジから見て一意であることができる。各ブリッジから一意のツリーは、ネットワーク・リンクの効果的な利用を可能にする。しかし、2つの装置 (AおよびB) の間のパスまたは関連するブリッジは対称であり、したがって、各方向で同じである。

【0043】

再び図2aを参照すると、いずれかの地点でネットワーク上にリンクの障害が存在する場合、例えば、ブリッジ116とブリッジ118の間のリンクが障害を起こした場合、唯一の影響は、そのリンクを通過するトラフィックに対するものである。障害はルーティング・システムによって公示され、最短パス・アルゴリズムが実行される。この時点で、唯一の変更は、障害リンクを通過していたパスに対するものであり、影響のない最短パスは変化しない。正味の結果は、新しいFIBが古いFIBと大部分は同じであることであり、影響のないパス上でのパケットの転送には実質的な影響がないことを暗示している。図2bを参照すると、ブリッジ116とブリッジ118の間のリンクにおける障害は、最短パスの一部ではないので、ルーティング・ツリーに影響を与えない。本発明の範囲外にある、障害シナリオにおいて実行される計算を最小化する技法が存在することは、当業者であれば理解されよう。上記の説明は、トポロジ変更の通知の受信、新しい転送テーブルの

40

50

計算、および F I B へのデータ入力など、ルーティング・システムのアクションを説明する場合の最も単純なケースを解説したに過ぎない。

【 0 0 4 4 】

ブリッジに関連付けられた M A C アドレス（ユニキャストおよびマルチキャスト）は、リンク状態制御イーサネット・サブネットワークに対してグローバルであり、宛先ベースの転送用を使用される。これは、ルーティング・システム公示において、それらが単純にフラディングされ得ること、ルーティング・システムの局所収束時に、それらがルーティング・システムによって指示されたように局所ブリッジ転送データベース（または F I B ）内でインスタンス化され得ることを意味する。このように、レイヤ 2 接続性の分散計算は、接続性をトポロジと関連付けために異なるシグナリング・システムを必要とせず、イーサネット・ブリッジに適用されることができる。その最も単純な形態では、ブリッジは、それが 2 つの与えられたブリッジ・ノード間の最短パス上にあることを計算した場合、それらのブリッジに関連付けられた M A C アドレスを F I B 内に単にインストールし、ユニキャスト M A C アドレスは、対象ブリッジの各々を指し示し、マルチキャスト M A C アドレスは、対象ブリッジから指し示す。さらなる改良は、ブリッジが 2 つのノード間の最短パス上にある場合、それらが関心に基づくコミュニティの共通部分を有するならば、ブリッジが転送テーブルに入力されるデータをしかるべく変更し得ることをブリッジが決定するように、関心に基づくコミュニティ情報をルーティング公示に追加することを含むことができる。

10

【 0 0 4 5 】

一意のメッシュが、P L S B メカニズムを使用して、V L A N 毎に構成されることができ。通常は単一のメッシュで十分であるが、等しいパスの場合、等コスト・パスが利用され得るようにトラフィックを拡散することが望ましいことがある。接続性の多数の置換が必要とされる場合（例えば、等コスト・パスの効率的な利用）、これは、依然として単一のルーティング・プロトコル・インスタンスを使用しながら、2 つ以上の V L A N を P L S B 動作に委任することによって達成されることができ。最短パスを計算する場合、各 V L A N について動作が繰り返されながら、均衡を破るために異なるランク付けアルゴリズムを各 V L A N に割り当て、このモードの動作に委任された V L A N の集まりの間でエッジにおける負荷分散を行う。追加の M A C アドレスは、それらの機能はエンド・ポイントを明白に識別することなので、必要とされず、V L A N は、経路選択のためのランク付けアルゴリズムを決定する。

20

30

【 0 0 4 6 】

ブリッジ毎に単一のユニキャスト M A C アドレスが説明されたが、何ものもより細かい粒度の使用を排除せず、ユニキャスト M A C アドレスは、ライン・カード、仮想スイッチ・インスタンス（V S I）、または U N I ポートに関係し得ることを理解されたい。これは、宛先ブリッジにおけるフローのデマルチプレキシングを簡単にするために望ましいことがある。

【 0 0 4 7 】

（トポロジ変更から、ルーティング・システムによるネットワーク内のすべてのブリッジへのトポロジ変更の公示、新しいトポロジの共通見取図への再収束および転送情報の対応する更新までの期間である）不安定な期間の間、（潜在的に悪化した形態であっても）接続性を維持するために、ループ抑制が、ネットワークにおいて必要とされる。分散システムの不安定性はしばしば、少なくとも一時的にネットワークの全体的な見取図が同期していないことを意味する。

40

【 0 0 4 8 】

イーサネット・アドレッシングは、フラットおよび非アグリゲータブル（non-aggregateable）であり、これは単一のルーティング・ドメインを暗示し、そのため、一時的なユニキャスト・ループは局所的で、相対的に無害であり、永続的なユニキャスト・ループおよびすべてのマルチキャスト・ループのみが対処される必要がある。メトリックは、ユニキャストおよびマルチキャスト転送用とも対称的で共通であるので、その

50

場合、各方向における最短パスは、ユニキャストおよびマルチキャスト・パスとも任意の2つの装置間で同じである。それが同じである場合、収束したネットワークでは、各ブリッジは、与えられた送信元MACを予期すべきインターフェースを知っており、それは、FIBでは、それを指し示すようにMACが構成されるポートは、戻りパスだからである。パケットの送信元アドレスは、ユニキャストおよびマルチキャスト・パケットで共通であり、それは、送信者のユニキャスト送信元アドレスであることにさらに留意されたい。

【0049】

どの分散システムでも、トポロジ変更に関連して、一時的なループが発生する。まれに、実装エラーまたはハードウェア問題の結果として、永続的なループも発生し得る。これは、構成または方針問題のために多数のルーティング・ドメインを含むネットワークの場合にも当てはまるが、PLSBでは、(ルーティング・ドメインの同位メッシュに対して)単一のドメインまたは限定的な階層のみが考えられる。いくつかのノードはネットワークについての情報を知っており、他のノードは知らない場合に、ループが形成される。これは、必ずしもすべてのノードが同時に収束していないことをしばしば意味する、ネットワーク情報の伝播遅延のために発生する。分散ルーティングが機能するためには、最終的にすべての健全なノードが同期が取られた見取図を有し、共通の結果を計算するという仮定が存在する。

10

【0050】

先に言及されたように、任意の2つのノード間の最短パスが双方向で同じになるように、対称メトリックが使用される。リンク状態ルーティングによって転送データベースの構成と組み合わせられた場合、逆方向パス転送チェック(RPFC)と呼ばれるものを実行することによって、パケットが予期されたインターフェース上で受信されたかどうか(与えられた送信元からの到着のセグメントがその送信元への最短パス上のセグメントに一致するかどうか)の検査となるように、従来のMAC学習手順が変更されることを可能にする、十分な情報が転送データベース内に存在する。これは、ポートをブロックする必要なく、または他にネットワーク接続性を大幅に妨げる必要なく、ループ・フリー性のパケット毎の検査を可能にする。これは、STPを使用せずに、ループ・フリーな任意対任意(any-to-any)のイーサネット接続性を生成する。

20

【0051】

ブリッジは、パケットに含まれる送信元MACアドレスおよびパケットが到着したセグメントと、転送データベース内で宛先としてのその同じMACアドレスのために構成されたものとの比較に基づいて、パケット上で健全性チェックを実行する。RPFC監視(RPFC policing)が使用可能にされている場合、送信元MACアドレスの学習されたセグメントが静的エントリを変更するならば、または静的エントリが存在しないならば、パケットは廃棄される。例えば、図2aを参照すると、ブリッジ110の送信元MACを有するパケットが、ブリッジ114、118、120、122、または124へのその他の接続リンクに関連付けられたポートを介してブリッジ112に到着した場合、ポートはブリッジ110の送信元MACに関連付けられたエントリを持たないので、パケットは廃棄される。RPFCは、サブネットワーク内におけるリンク状態ルーティング・システムのバグのない一貫した実施を想定する。

30

40

【0052】

先に言及されたように、正しく収束したネットワークは、ネットワーク内の任意のブリッジ対の間に1つの双方向最短パスを有する。PLSBルーティングは、定義によって、リンクの両方向に「等しい重み」を割り当てる。使用可能にされたRPFCを用いると、ブリッジAからブリッジBに到達するパケットは、両方向についてのすべての中間ノードの正しい収束の論理「AND」である。正しくないまたは収束しないパスに遭遇したパケットはどれも、ループする可能性を取るのではなく、廃棄される。

【0053】

図3は、ネットワーク内でループ形成がどのように生じ得るかについての概略図である。例は、ステップ301において、作業システムが装置Bから装置Aおよび装置Aから装

50

置 B に転送を行うことから開始する。ステップ 302 に示されるように、可能な一時的ループを開始するには、例えば、ノード Y から装置 A へのリンクが障害を起こしていなければならない、かつ装置 A がループ上でデュアル・ホーム (d u a l - h o m e d) でなければならない、そのため、ノード Y は、第 2 のリンクを介してパケットを配送しようとし続ける。 P L S B ネットワーク内でのループの生成は、ループ抑制メカニズムが何らかの障害を起こしたことを仮定する。

【 0 0 5 4 】

ステップ 303 に示されるように、ブリッジ Z においてループを進め続けるには、ブリッジ Z が、ブリッジ Y に到る最短パスは、ブリッジ Z からブリッジ Y、またはブリッジ Z から装置 A とだけ矛盾のない、ブリッジ Y からブリッジ Z に直接進むことを決定した、ブリッジ Y への向きとは正反対の、ブリッジ X を経由する時計回りであると信じなければならない。

10

【 0 0 5 5 】

ブリッジ X においてループを進め続けるには、ステップ 304 に示されるように、第 2 の障害が存在しなければならない。しかし、 R P F C が利用される場合、順方向および逆方向におけるループが同時に必要とされる。順方向におけるループを概念的に想像することは容易であるが、両方向での発生を想像することはあまり容易ではない。ループは、2 つの装置間の最短パスに固有である。そのため、特定の M A C アドレスに対するループは、そのアドレスへのすべてのトラフィックがループすることを意味せず、戻りパスも合同なループ内に存在するトラフィックのみがループすることを意味する。

20

【 0 0 5 6 】

保証されるループ・フリー性は、直観的な略式の証明によって確認されることができる。ループ上で 2 つの逆循環方向は、同時に存在することができず、そのような転送構成を生成する条件は、1 つの方向でループを生成する条件がミラーされた場合、本質的に自己矛盾を有する。 R P F C は、ループが順方向および逆方向の両方で合同であることを要求する。直観的な証明の核心は、健全なシステムでは、これが両方向について同時に成り立たないことである。パス上の与えられたノード対の間に現在の作動パスよりも短いパスが存在するためには、リンクまたはリソースが追加される必要があり、その知識は、実際に形成されたループの単一の方向に制限される必要がある。追加されるリソースについての知識は、 R P F C が本質的に自己矛盾となることを覆すためには、同時に (両方向で)

30

【 0 0 5 7 】

障害の組合せまたは不健全な実施が実際にループを形成した場合、ひとたび形成されると、さらなるトラフィックはそのループに入ることができず、その結果、パケットは複製地点に到着しなければ出ることができないという特性を、ループが有することにも留意されたい。図 4 に例示的に示されるように、ループ 402 が多数のブリッジ間で形成された場合、 R P F C のため、ループ内のパケットの唯一の有効な送信元は、ループ自体である。ループするパケットは、ループ内にすでに捕らえられたパケットに制限される。ループの外部のブリッジであるノード 404、406、408、410 からのトラフィックは、ループに入ることができない。これは、形成されるとしても、単純なループのみが形成され得ることを意味する。ループは与えられた送信元アドレスの有効な送信元として 2 つ以上のインターフェースを受け入れることは不可能であるので、重複パケットは元のフローに再合流することができない。

40

【 0 0 5 8 】

したがって、 R P F C は、 (T T L またはポート・ブロッキングに対して) ループ・フリー性のパケット毎の監視を実行するループ防止メカニズムとして利用されることができる。ループは送信元 / 宛先の対に対して存在しなければならないので、ループの粒度が著しく制限されることに留意することも重要である。

【 0 0 5 9 】

R P F C は、イーサネット・パケットには変更を必要とせず、ブリッジの実装には最低

50

限の変更しか必要としないという利点を有する。予期せぬインターフェースに到着するパケットにはその他の原因も存在し得るので、R P F Cは積極的な監視であり、したがって、それを選択的な状況において利用しなくてもよい能力が望ましい。P L S Bマルチキャストは、(例えば、V L A Nにおけるフラディングとは反対に)もっぱらイーサネットM A Cアドレスに基づいている。イーサネット・マルチキャストM A Cアドレスは、独特な監視処理がユニキャストまたはマルチキャスト・パケットに適用されることを可能にする、明示的なマルチキャスト表示を宛先M A Cアドレス上に含む(P B Tと同様に、未知パケットのフラディングは禁止されることに留意されたい)。構成されたユニキャスト転送の場合、ループが解消する時まで、またはバッファリング容量が限界を超え、受信された追加のトラフィックが廃棄されるようになる時まで、ネットワークはループ内のパケットを単にバッファリングするので、一時的なループは破局的ではない。これは、一時的なループは害がなく、複製を行わないので(ネットワークはループが解消されるまで単にバッファリングする)、ネットワーク収束の間、ループ抑制はオフにされることができ、ネットワークが安定したとき、どのような永続的なループも即座に抑制するために、ユニキャスト抑制が再びオンにされることを意味する。

【0060】

(トラフィックの廃棄によって検出される)永続的なループ状態が存在するかどうかをチェックするために、単にユニキャスト用のR P F Cを定期的に変更可能にし、存在しない場合、再びオフにするなど、その他の戦略も考えられ得る。ループ形成の結果は、マルチキャスト・トラフィックの場合は著しく異なり、一時的なループは、無制限な複製をもたらすことができ、したがって、マルチキャスト宛先M A Cアドレスを有するパケットに対しては、R P F Cは決して使用不能にされない。マルチキャストの場合、非収束マルチキャスト転送がパケットを積極的に廃棄し、ループを防止するように、ループ抑制は常にオンにしておくべきである。

【0061】

トポロジ変更によって影響されないツリーは、F I Bに対する変更がなく、またはトポロジ変更に関連しないそれらの関連パスには不安定性がないので、正常に機能し続けるが、変更によって影響されるツリーは、R P F Cが非収束マルチキャスト・パスに関するパケットを廃棄するので、接続性の中断を経験することがある。ネットワーク全体に関して、これは、サービスの一時的な「低下(brown out)」に似ている。学習されたエントリは、構成されたエントリに取って代わることはなく、そのため、監視の使用可能/使用不能が、ルーティング・システムによってインスタンス化された構成の改悪をもたらすことはない。

【0062】

図5は、P L S Bを実施するためのブリッジ・ノード500の可能な実装の概略図である。ルーティング・システム・モジュール502は、リンク状態プロトコル・ルーティングを使用して、ネットワーク・トポロジに関する情報を、ネットワーク内の同位ブリッジと交換する。先に説明されたように、情報の交換は、ブリッジが、ネットワーク・トポロジの同期の取れた見取図を生成することを可能にし、その後、同期の取れた見取図は、ルーティング・システム502モジュールが、収束の最中に(上で説明されたアルゴリズムを使用して)最短パス・ツリーを計算することを可能にする。F I B 504は、決定されたパスに基づいてネットワークを介してトラフィックを送るのに適切なエントリを入力される。R P F C送信元チェック・モジュール506は、着信パケットを処理し、受信ポートが特定の送信元M A C用のF I B 504内で識別されたポートと一致するかどうかを決定するために、F I B 504内で検索を実行する。受信ポート/送信元M A Cが、予想ポート/送信元M A Cと一致しない場合、パケットは廃棄される。同様に、ルーティング・システム502が、送信元チェック506に対して、ネットワークが収束プロセス中であることを識別した場合、ユニキャストに関するループ抑制は使用不能にされる。ネットワークが収束した場合、ユニキャスト・パケットに関するループ抑制は再び使用可能にされる。マルチキャスト宛先アドレスによって識別されるマルチキャスト・パケットの場合、

収束の最中、R P F C送信元チェック506は決して使用不能にされない。パケットがR P F C送信元チェック506モジュールに合格した場合、またはチェックが使用不能にされている場合、宛先検索508モジュールは、どのポートからパケットが宛先ユニキャストまたはマルチキャストM A Cアドレスに転送されるべきかを、F I B 504から決定する。有効なエントリが存在しない場合、パケットは廃棄される。ブリッジがネットワークのエッジにある場合、発信パケットの転送の前に、パケットのM A C - i n - M A Cカプセル化が、一意ユニキャストおよびマルチキャスト・アドレスを利用して行われることがある(図示されず)。説明されたモジュールは、例示的な目的のものに過ぎず、当業者であれば理解されるように、ブリッジ・ノードのモジュールの間で機能を結合または分散することによって実施されてよいことも理解されたい。

10

【0063】

図6は、動作のためにP L S Bブリッジ500を構成する方法の一実施形態を示している。ステップ602において、ブリッジまたはリンク障害などのネットワーク・トポロジ変更が発生した場合、ステップ604において、状態情報が、ルーティング・システム・モジュール502によって、ネットワーク内のブリッジ間で交換される。ルーティング・システム・モジュール502は、トポロジ情報をブリッジ内蔵のデータベース内に保存することによって、ネットワーク構成の同期の取れた見取図を作り上げる。その後、ステップ606において、ブリッジは、先に説明されたような最短パス・アルゴリズムを使用して、同位ブリッジ間の最短パスを決定することができる。その後、ステップ608において、F I B 504は、接続性を可能にする適切なルーティング・エントリを入力される。その後、パケットは、ブリッジによって処理されることができる。ネットワーク・トポロジ変更が発生した場合、プロセスが再び開始される。

20

【0064】

図7は、受信パケットを処理するブリッジ・ノード500の一実施形態を示している。ステップ702において、パケットがブリッジのポートで受信される。ステップ704において、パケットがマルチキャスト・パケットか、それともユニキャスト・パケットかを決定するために、宛先アドレスが使用される。パケットがユニキャストであり(ステップ704においてY E S)、ルーティング・システムが収束している場合、したがって、R P F Cが使用可能であり(ステップ706においてY E S)、ステップ708において、R P F Cが実行される。R P F Cが成功した場合(ステップ708においてY E S)、すなわち、パケットが関連送信元M A Cアドレスの予想ポートに到着した場合、ステップ710において、宛先M A C用の発信ポートのF I B内における検索が行われる。R P F Cが使用可能でない場合(ステップ706においてN O)、すなわち、ループ状態が存在し、ネットワークが収束していない場合、R P F Cはバイパスされ、ステップ710において、パケットを転送するための検索が直接行われる。関連M A Cアドレス用のエントリが存在する場合(ステップ710においてY E S)、ステップ714において、パケットはその宛先に転送される。R P F Cが成功しなかった場合(ステップ708においてN O)、すなわち、パケットが送信元アドレスに基づいた予想ポートに到着しなかった場合、ステップ712において、パケットは廃棄される。同様に、パケットが適切なエントリを持たない場合(ステップ710においてN O)、ステップ712において、パケットは廃棄される。

30

40

【0065】

パケットがマルチキャスト・パケットである場合(ステップ704においてN O)、R P F Cは常に使用可能であり、ステップ708において、R P F Cが実行される。R P F Cが成功した場合(ステップ708においてY E S)、すなわち、パケットが関連M A Cアドレスの予想ポートに到着した場合、ステップ710において、宛先M A C用の発信ポートのF I B内における検索が行われる。関連M A Cアドレス用のエントリが存在する場合(ステップ710においてY E S)、ステップ714において、パケットはその宛先に転送される。R P F Cが成功しなかった場合(ステップ708においてN O)、すなわち、パケットが送信元アドレスに基づいた予想ポートに到着しなかった場合、ステップ71

50

2において、パケットは廃棄される。同様に、パケットが適切なエントリを持たない場合（ステップ710においてNO）、ステップ712において、パケットは廃棄される。先に言及されたように、（トラフィックの廃棄によって検出される）永続的なループ状態が存在するかどうかをチェックするために、ユニキャスト用のRPF Cは、定期的に使用可能にされることができ、存在しない場合、再びオフにされる。

【0066】

これまで、単一の関心に基づくコミュニティをサポートするプロバイダ・リンク状態ブリッジ・ネットワークが説明されたが、ポートのサブセット、したがってPLSBネットワーク内のブリッジのサブセットに対する接続性のみを個々の関心に基づくコミュニティが必要とする、多数の関心に基づくコミュニティをサポートすることも可能である。必要とされるのは、関心に基づくコミュニティに関係するブリッジの組に制約されたマルチキャスト接続性および共有のユニキャスト接続性と、与えられたパケットを関心に基づくコミュニティに関連付けるためのメカニズムである。IEEE 802.1ahのI-SID（拡張サービスID）フィールドは、パケットを関心に基づくコミュニティに関連付けるメカニズムの一例である。関心に基づくコミュニティ識別子（例えばI-SID）は、ノードがI-SIDで識別される関心に基づくコミュニティの関心事を識別することができ、最終的に、各ブリッジが一意グループ・マルチキャスト・アドレスを公示された各I-SIDに関連付けるように、ルーティング・システム公示内に含まれることもできる。2つのブリッジ間の最短パス上に自らを見出したブリッジは、各ブリッジに関連付けられたユニキャストMACアドレスと、2つのブリッジに共通するすべてのI-SIDに関するマルチキャストMACアドレスとをインストールする。この結果は、与えられたエッジ・ブリッジが、すべての同位ブリッジへのユニキャスト接続性と、I-SIDで識別される各関心に基づくコミュニティに特有のマルチキャスト接続性とを有することである。これは、各同位ノードへのマルチポイント・ツー・ポイント（mp2p）ユニキャスト・ツリーのリーフである形態、および（S,G）ポイント・ツー・マルチポイント（p2mp）マルチキャスト・ツリーのルートである形態をとり、ここで、Sは送信元のアドレスであり、Gは各関心に基づくコミュニティの1組の同位ノードに対するマルチキャスト・グループ・アドレスである。ブリッジ対が共通のI-SIDを持たない場合、さらなる改良は、ユニットMACアドレスがインストールされないこととすることができる。同様に、ブリッジ対は、トランジット・ブリッジとすることができ、ノードによって終了または開始されるフローにどのようなMACアドレスも提供しないことを選択することができる。このように、マルチキャスト接続性は、特定の関心事グループに限定されるばかりでなく、この手法は、ユニキャスト接続性のための転送テーブル空間の消費も節約する。

【0067】

図8は、仮想プライベート・ネットワーク（VPN）がPLSBネットワーク上にどのようにマッピングされることができ、エッジ・ブリッジ毎のVPN毎に一意的マルチキャスト・ツリーがマッピングされることを可能にするかを示している。マルチキャストVPNシナリオでは、マルチキャスト・トラフィックは、VPNに参加するブリッジにのみ配送される。VPNグループ・マルチキャスト・アドレスが、共通なパスについてインストールされる。4つのVPNネットワークが、V1、V2、V3、V4として識別されている。多数のVPNが、ブリッジ110などのブリッジからホストされることができ、個々のVPNエンド装置とすることができる。例えばV1およびV3などの各VPNについて、一意のユニキャスト・ツリーが生成される。対応するVPNの端点を含むブリッジへの唯一の経路が識別される。例えば、V1に関するルーティング・ツリーの場合、ブリッジ116へのパスと、ブリッジ112を介するブリッジ122およびブリッジ124へのパスが必要とされる。同様に、V3に関するルーティング・ツリーの場合、ブリッジ112へのパスと、その先のブリッジ118およびブリッジ124へのパスが必要とされる。これは、V1からのVPNトラフィックが、VPN V1およびVPN V3エンド装置をホストしないブリッジに配送される可能性を排除する。各VPNは、最短パス・アルゴリズムに基づいて、VPNに特有のエッジ・ブリッジ毎のツリーを有することができる。

【0068】

VPNを定義する能力と同様に、非対称接続性が構成されることができる。通常、PLSBは、(*, G)マルチキャスト接続性を生成し、Sがグループ「G」内の送信元を示す(S, G)マルチキャスト・ツリーの完全メッシュのように、*はすべての送信元を意味し、Gはマルチキャスト・グループを表す。与えられたサービス・インスタンスが(S, G)への接続性に制限されることが望ましいこともある。これは、与えられたマルチキャスト・グループの送信元、送信先、または送信元と送信先の両方のどれであることを望むかを示すアトリビュートを公示に追加することによって、マルチキャストについて容易に達成されることができる。2つの他のブリッジ間の最短パス上にあることを確定したブリッジは、送信元/送信先アトリビュートを使用して、どのマルチキャスト・グループ・アドレスがインストールされるべきかを決定する。より複雑な接続性は、異なるアトリビュート構成を用いてこの例を繰り返すことによって構成されることができる。例えば、ともに2つのVPNに参加する、ネットワーク内の装置の2つの組を想定することができる。方針の問題として、接続性は、組間(inter-set)でのみ許可され、組内(intra-set)では許可されない(実際の具体例は、本社オフィス/支社オフィス接続性である)。そのため、第1のVPNは、組Aのための送信元アトリビュートと、組Bのための送信先アトリビュートとを有する。第2のVPNは、組Aのための送信先アトリビュートと、組Bのための送信元アトリビュートとを有する。透過ブリッジング・オーバーレイが考えられる場合、そのような制限を課すことは、組Aからの未知のフラディングが組Bに、組Bからの未知のフラディングが組Aに制限されることを意味する。したがって、2つの組は、決して組内接続性を学習することはできず、組Aの装置は、組Bとのみ通信を行い、組Bの装置は、組Aとのみ通信を行う。

10

20

【0069】

図9に示されるように、ネットワークは、(本発明の譲受人に譲渡された出願US20050220096で説明されるような)PBT、および802.1ah PBBなどのその他の技術を利用して拡大されることができる。PBB904は、PLSB WANDメイン902およびPLSBメトロ・ドメイン906などのPLSBエリアを互いに結びつけるため、またはPBT908を使用して接続性を拡張するために、エリア境界ルータ(ABR: Area Border Router)の緩やかな同等物として使用されることができる。エリア境界におけるクライアント情報の検査は、同位エリアが、隣接エリアのルーティング・システムにおいて単一のMACアドレスとして単純にモデル化されること、または同位エリアには単一のB-MACとして見えるPBTファンイン・ドメイン(fan-in domain)と相互動作することを可能にし、これらの技術のどちらかが、要約を提供する。ドメイン間のループ・フリー性を保証するため、ネットワークは、ドメインの厳格な階層であり、ルーティング・ドメインのメッシュは、サポートされることができない。

30

【0070】

ネットワークの遠端において、送信元学習が、ポートIDのための送信元B-MACダブリング(source B-MAC doubling)を用いて動作する仕方と同じようにして、与えられたC-MACのための送信元B-MACが知られる。この手順は、リンク状態ブリッジングを用いて動作するようにわずかに変更される。C-MACとB-MACの対応を学習する手順は、変更されない。C-MACに対するB-MACが学習されていない場合、関心に基づくコミュニティ(典型的にはクライアントVPN)に適したブリッジのマルチキャスト・アドレスが使用され、これが、C-MACブロードキャストのPLSB空間において必要とされるエミュレーションを提供する。

40

【0071】

PLSBは、スパニング・ツリー・プロトコルの不利な側面の大部分が除去された、MAC-in-MACブリッジ・ネットワークを提供する。これは、メッシュ接続性のはるかに良い利用をもたらし、各装置がリンク状態データベースを有するので、はるかに速い収束をもたらす。ユニキャスト接続性は、ネットワークの再収束の最中に中断されない。

50

加えて、PLSBは、(異なるVID範囲を使用して)PBTと並列して動作する能力、または(ハブ実装において)PBTと連結されて動作する能力を提供し、イーサネット・アトリビュートは完全に保存されて、クライアント・レイヤに完璧なエミュレーションを提供する。

【0072】

上で説明された本発明の実施形態は、単に説明的であることが意図されている。したがって、本発明の範囲は、添付の特許請求の範囲によってのみ限定されることが意図されている。

【図面の簡単な説明】

【0073】

【図1】スパンニング・ツリー・プロトコル(STP)を使用するメッシュ・ネットワークの概略図である。

【図2a】装置Aから示されたPLSBを実施するメッシュ・ネットワークの概略図である。

【図2b】装置Bから示されたPLSBを実施するメッシュ・ネットワークの概略図である。

【図3】パケット・ループ・シナリオの概略図である。

【図4】収束中のポート・ブロッキングの概略図である。

【図5】PLSBブリッジの概略図である。

【図6】PLSBブリッジを構成するための方法のフローチャートである。

【図7】PLSBブリッジを動作させるための方法のフローチャートである。

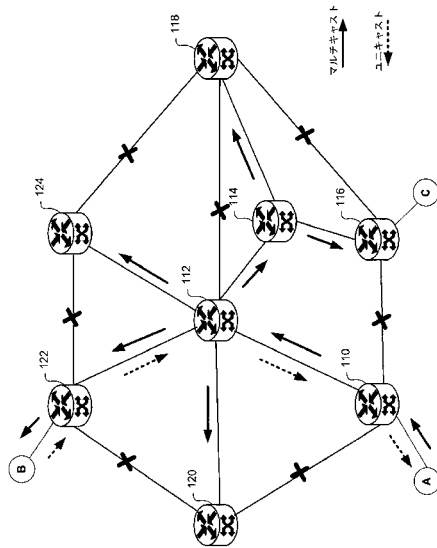
【図8】PLSBを使用するVPNオーバーレイの概略図である。

【図9】ハイブリッドPBBおよびPBTネットワークと組み合わされたPLSBのネットワーク図である。

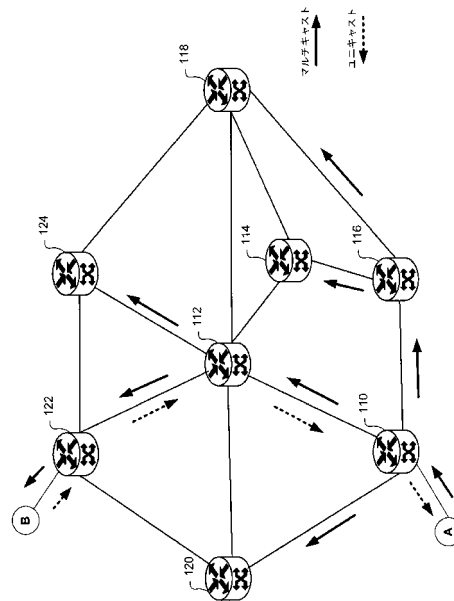
10

20

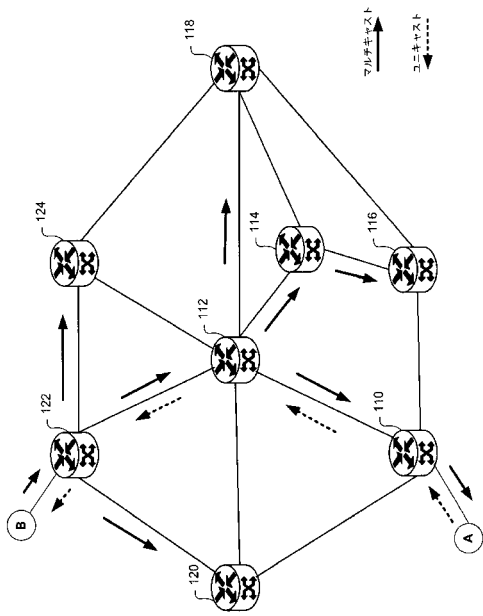
【図1】



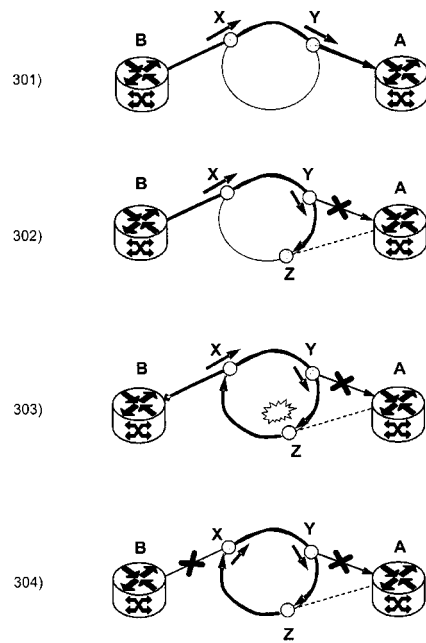
【図2a】



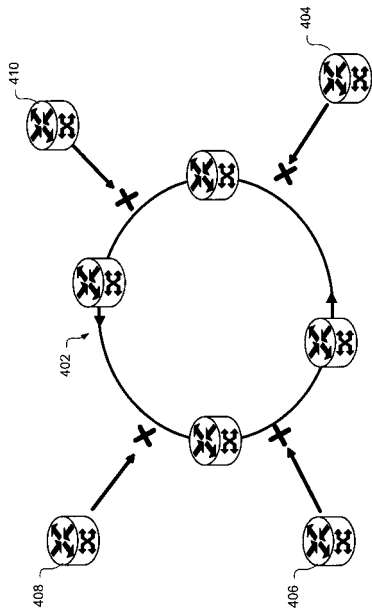
【 図 2 b 】



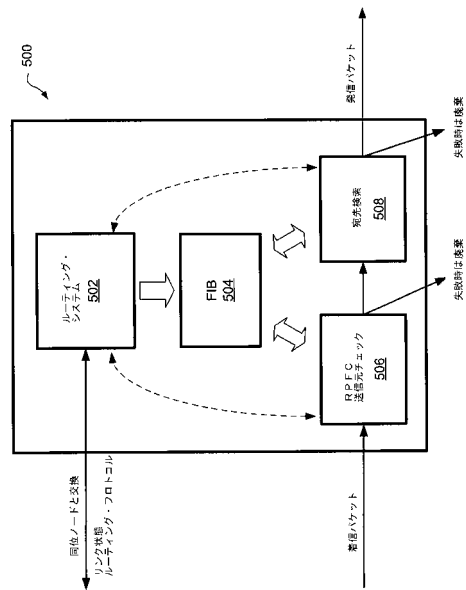
【 図 3 】



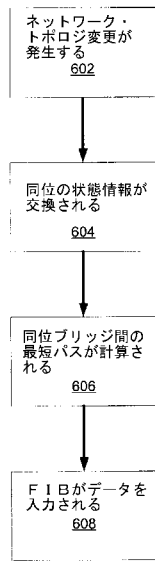
【 図 4 】



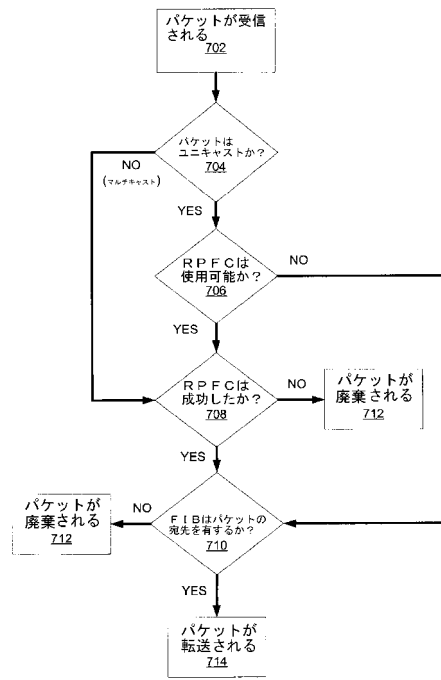
【 図 5 】



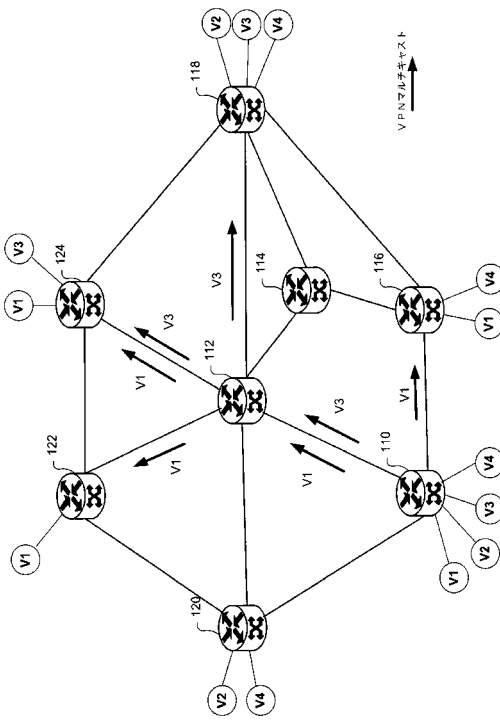
【 図 6 】



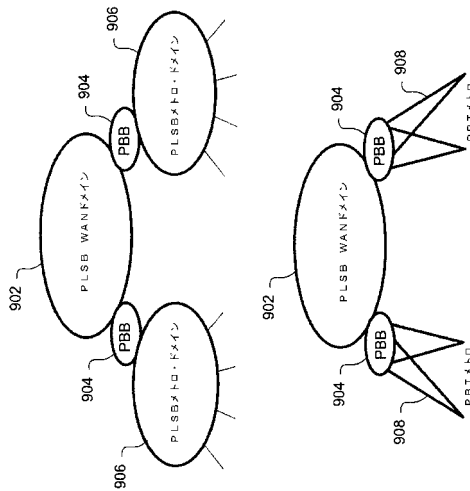
【 図 7 】



【 図 8 】



【 図 9 】



【 国際調査報告 】

INTERNATIONAL SEARCH REPORT		International application No. PCT/CA2006/001614												
<p>A. CLASSIFICATION OF SUBJECT MATTER IPC: <i>H04L 12/413</i> (2006.01) , <i>H04L 12/56</i> (2006.01) , <i>H04L 29/06</i> (2006.01) According to International Patent Classification (IPC) or to both national classification and IPC</p>														
<p>B. FIELDS SEARCHED</p> <p>Minimum documentation searched (classification system followed by classification symbols) IPC: <i>H04L</i> (2006.01)</p> <p>Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched</p> <p>Electronic database(s) consulted during the international search (name of database(s) and, where practicable, search terms used) Canadian Patent Database, Delphion, Derwent, IEEEXplore, Google. Keywords: Ethernet, link state, bridging/switching, MAC address, Reverse Path Forwarding Check, Forwarding Information Database, unicast, multicast, frames/packets, same ingress port, egress port, network topology, mesh network, WAN/MAN, Metro Ethernet.</p>														
<p>C. DOCUMENTS CONSIDERED TO BE RELEVANT</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 10%;">Category</th> <th style="width: 70%;">Citation of document, with indication, where appropriate, of the relevant</th> <th style="width: 20%;">Relevant to claim No.</th> </tr> </thead> <tbody> <tr> <td style="text-align: center;">A</td> <td>Garcia et al.: "LSOM: A Link State protocol Over MAC addresses for metropolitan backbones using Optical Ethernet switches", Proceedings of the second IEEE International Symposium on Network Computing and Applications (NCA 2003) 16-18 April 2003, pages: 315-321. *entire document*</td> <td style="text-align: center;">1-33</td> </tr> <tr> <td style="text-align: center;">A</td> <td>Bolton et al.: "Analyses of the reverse path forwarding routing algorithm", Proceedings of the 2004 International Conference on Dependable Systems and Networks, (DSN 2004), 28 June - 1 July 2004, pages: 485-494. *entire document*</td> <td style="text-align: center;">1-33</td> </tr> <tr> <td style="text-align: center;">A</td> <td>Ibanez et al.: "Alternative multiple spanning tree protocol (AMSTP) for optical Ethernet backbones", Proceedings of the 29th Annual IEEE International Conference on Local Computer Networks, (LCN 2004), 16-18 Nov. 2004, pages: 744-751. *entire document*</td> <td style="text-align: center;">1-33</td> </tr> </tbody> </table>			Category	Citation of document, with indication, where appropriate, of the relevant	Relevant to claim No.	A	Garcia et al.: "LSOM: A Link State protocol Over MAC addresses for metropolitan backbones using Optical Ethernet switches", Proceedings of the second IEEE International Symposium on Network Computing and Applications (NCA 2003) 16-18 April 2003, pages: 315-321. *entire document*	1-33	A	Bolton et al.: "Analyses of the reverse path forwarding routing algorithm", Proceedings of the 2004 International Conference on Dependable Systems and Networks, (DSN 2004), 28 June - 1 July 2004, pages: 485-494. *entire document*	1-33	A	Ibanez et al.: "Alternative multiple spanning tree protocol (AMSTP) for optical Ethernet backbones", Proceedings of the 29th Annual IEEE International Conference on Local Computer Networks, (LCN 2004), 16-18 Nov. 2004, pages: 744-751. *entire document*	1-33
Category	Citation of document, with indication, where appropriate, of the relevant	Relevant to claim No.												
A	Garcia et al.: "LSOM: A Link State protocol Over MAC addresses for metropolitan backbones using Optical Ethernet switches", Proceedings of the second IEEE International Symposium on Network Computing and Applications (NCA 2003) 16-18 April 2003, pages: 315-321. *entire document*	1-33												
A	Bolton et al.: "Analyses of the reverse path forwarding routing algorithm", Proceedings of the 2004 International Conference on Dependable Systems and Networks, (DSN 2004), 28 June - 1 July 2004, pages: 485-494. *entire document*	1-33												
A	Ibanez et al.: "Alternative multiple spanning tree protocol (AMSTP) for optical Ethernet backbones", Proceedings of the 29th Annual IEEE International Conference on Local Computer Networks, (LCN 2004), 16-18 Nov. 2004, pages: 744-751. *entire document*	1-33												
<p><input checked="" type="checkbox"/> Further documents are listed in the continuation of Box <input checked="" type="checkbox"/> See patent family annex.</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <tbody> <tr> <td style="width: 50%;"> * 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 but 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 </td> <td style="width: 50%;"> "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 </td> </tr> </tbody> </table>			* 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 but 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										
* 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 but 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													
<p>Date of the actual completion of the international search 19 December 2006 (19-12-2006)</p>		<p>Date of mailing of the international search report 18 January 2007 (18-01-2007)</p>												
<p>Name and mailing address of the ISA/CA Canadian Intellectual Property Office Place du Portage I, C114 - 1st Floor, Box PCT 50 Victoria Street Gatineau, Quebec K1A 0C9 Facsimile No.: 001-819-953-2476</p>		<p>Authorized officer Georges Matar 819-994-6366</p>												

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CA2006/001614

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant	Relevant to claim No.
A	Brockners et al.: "Metro Ethernet - Deploying the extended Campus using Ethernet Technology", Cisco Syst. Inc., San Jose, CA, USA, This paper appears in: Proceedings of the 28th Annual IEEE International Conference on Local Computer Networks (LCN 2003), 20-24 Oct. 2003, pages: 594-604. *entire document*	1-33
A	WO 02/078250 A2 (Pegrum et al.) 3 October 2002 (03-10-2002) *entire document*	1-33

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CA2006/001614

Patent Document Cited in Search Report	Publication Date	Patent Family Member(s)	Publication Date
WO 02/078250 A2	03-10-2002	CA2441470 A1	03-10-2002
		EP1374500 A2	02-01-2004
		US2002141429 A1	03-10-2002
		US2003198182 A1	23-10-2003

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

(特許庁注：以下のものは登録商標)

1. イーサネット

(72)発明者 ブラッグ, ナイジェル

イギリス国シー・ビー・1 5・エヌ・エックス、ケンブリッジシャー州ウエストン・コルヴィル
、ホームワード・チャペル・ロード 26

Fターム(参考) 5K033 AA05 CB01 CB08 CC02 DA02 DB19 EA05