(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第5973570号
(P5973570)

(45)発行日　平成28年8月23日(2016.8.23)　(24)登録日　平成28年7月22日(2016.7.22)

(51)Int.Cl.　F I

H04L 12/715	(2013.01)	H04L	12/715
H04L 12/717	(2013.01)	H04L	12/717
H04L 12/725	(2013.01)	H04L	12/725

請求項の数 16　(全 22 頁)

(21)出願番号　特願2014-524465 (P2014-524465)
(86)(22)出願日　平成24年7月26日 (2012.7.26)
(65)公表番号　特表2014-525692 (P2014-525692A)
(43)公表日　平成26年9月29日 (2014.9.29)
(86)国際出願番号　PCT/IB2012/053833
(87)国際公開番号　WO2013/021304
(87)国際公開日　平成25年2月14日 (2013.2.14)
審査請求日　平成27年6月26日 (2015.6.26)
(31)優先権主張番号　13/208,251
(32)優先日　平成23年8月11日 (2011.8.11)
(33)優先権主張国　米国(US)

(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　100079108
弁理士　稲葉　良幸
(74)代理人　100109346
弁理士　大貫　敏史
(72)発明者　ヤダバリ，キラン
アメリカ合衆国，カリフォルニア州　９５１３５，サン　ノゼ，デラクロワ　コート　４２８０

最終頁に続く

(54)【発明の名称】分割アーキテクチャ・ネットワークにおけるＯＳＰＦの実装

(57)【特許請求の範囲】
【請求項１】
分割アーキテクチャ・ネットワークの複数の領域のうちの１つのための複数のコントローラのうちの１つとして機能するネットワーク要素内に実装される方法であって、前記コントローラは、前記分割アーキテクチャ・ネットワークの前記領域に制御プレーンを提供し、前記コントローラは、分割アーキテクチャ・ネットワークの前記領域にデータ・プレーンを提供する複数のスイッチからはリモートにあり、前記コントローラは、前記分割アーキテクチャ・ネットワークの他の領域の他のコントローラ、及び前記分割アーキテクチャ・ネットワークを含むネットワークの従来型ルータに、限定領域内リンク・コスト・データを提供することによって、前記分割アーキテクチャ・ネットワークの前記複数領域にまたがる最適化ルーティングを容易にするものであって、前記限定領域内リンク・コスト・データは、すべての内部リンク・コスト・データを提供することなく、前記コントローラの領域のそれぞれの可能な最短パスのトラバースに関するコストを提供し、
前記コントローラの前記領域内の各境界スイッチを含む前記分割アーキテクチャ・ネットワーク内の前記コントローラの前記領域のトポロジを習得するステップであって、前記コントローラの前記領域内の各境界スイッチは、前記コントローラの前記領域を、前記分割アーキテクチャ・ネットワークの別の領域、又は前記ネットワーク内の前記従来型ルータのうちの１つにリンクさせる、少なくとも１つの外部ポートを有する、習得するステップと、
前記コントローラの前記領域内の各境界スイッチ・ペア間の最短パスを計算するステッ

プと、
前記コントローラのルーティング・テーブル内の各境界スイッチ・ペア間の各最短パスのコストを記憶するステップと、
前記分割アーキテクチャ・ネットワーク内の各近隣コントローラ又は前記ネットワーク内の近隣従来型ルータを、ハロー・プロトコルを使用して識別するステップであって、各近隣コントローラは、前記コントローラの前記領域の少なくとも１つの外部ポートを通じてアクセス可能な、前記分割アーキテクチャ・ネットワークの別の領域内のスイッチを制御する、識別するステップと、
リンク状態データベースを各近隣コントローラ及び近隣従来型ルータと交換するステップであって前記リンク状態データベースは、各境界スイッチ・ペア間の各最短パスの前記コストを含む、交換するステップと、
前記コントローラをツリーのルートとして備える前記ネットワークについて最短パス・ツリーを計算するステップと、
前記最短パス・ツリーに従って転送を実装するために、前記コントローラの前記領域のスイッチ内の転送テーブルを更新するステップと、
を含む、方法。
【請求項２】
各近隣コントローラを識別するステップが、
前記コントローラの前記領域の各外部ポートでハロー・パケットを送信するステップと、
前記コントローラの前記領域の少なくとも１つの外部ポートを通じて、各近隣コントローラからハロー・パケットを受信するステップと、
を更に含む、請求項１に記載の方法。
【請求項３】
リンク状態データを交換するステップが、各近隣コントローラにリンク状態アドバタイズを送信するステップを更に含み、前記リンク状態アドバタイズは前記コントローラの各外部リンクへのコストを含む、請求項１に記載の方法。
【請求項４】
リンク状態データを交換するステップが、各近隣コントローラにリンク状態アドバタイズを送信するステップを更に含み、前記リンク状態アドバタイズは、前記領域内の各境界スイッチ・ペアに関する前記最短パスの前記コストを含む、請求項１に記載の方法。
【請求項５】
更新されたリンク状態データを各近隣コントローラにアドバタイズするステップを更に含む、請求項１に記載の方法。
【請求項６】
前記更新リンク状態データをアドバタイズするステップが、更新されたリンク状態データを伴うリンク状態アドバタイズ・メッセージを各近隣コントローラに送信するステップを更に含む、請求項５に記載の方法。
【請求項７】
転送テーブルを更新するステップが、OpenFlowプロトコルを使用して前記コントローラの前記領域内の各スイッチの転送テーブルを更新するステップを更に含む、請求項１に記載の方法。
【請求項８】
前記コントローラでオープン・ショーテスト・パス・ファースト（ＯＳＰＦ）プロトコル・データ構造を開始するステップを更に含む、請求項１に記載の方法。
【請求項９】
分割アーキテクチャ・ネットワークの複数の領域のうちの１つについての複数のコントローラのうちの１つとして機能する、ネットワーク要素であって、前記コントローラは、前記分割アーキテクチャ・ネットワークの前記領域に制御プレーンを提供し、前記コントローラは、前記分割アーキテクチャ・ネットワークの前記領域にデータ・プレーンを提供

する複数のスイッチからはリモートにあり、前記コントローラは、前記分割アーキテクチャ・ネットワークの他の領域の他のコントローラ、及び前記分割アーキテクチャ・ネットワークを含むネットワークの従来型ルータに、限定領域内リンク・コスト・データを提供することによって、前記分割アーキテクチャ・ネットワークの前記複数領域にまたがる最適化ルーティングを容易にし、前記限定領域内リンク・コスト・データは、すべての内部リンク・コスト・データを提供することなく、前記コントローラの前記領域のそれぞれの可能な最短パスのトラバースに関するコストを提供し、

ネットワークを介してデータを受信するように構成された入口モジュールと、

前記ネットワークを介してデータを送信するように構成された出口モジュールと、

前記入口モジュール及び前記出口モジュールに結合されたネットワーク・プロセッサであって、コントローラ・モジュール、トポロジ習得モジュール、最短パス計算モジュール、近隣発見モジュール、及びリンク状態管理モジュールを備える、モジュール・セットを実行するように構成される、ネットワーク・プロセッサと、

前記トポロジ習得モジュールは、前記コントローラの前記領域内の各境界スイッチを含む前記分割アーキテクチャ・ネットワーク内の前記コントローラの前記領域のトポロジを決定するように構成され、各境界スイッチは、前記コントローラの前記領域を前記ネットワークの別の領域又は前記ネットワーク内の従来型ルータにリンクさせる、少なくとも１つの外部ポートを有し、

前記コントローラ・モジュールは、前記分割アーキテクチャ・ネットワーク内のコントローラの前記領域に制御プレーン機能を提供するように構成され、

前記最短パス計算モジュールは、近隣コントローラと共有することになる前記コントローラの前記領域内の各境界スイッチと前記従来型ルータとの間の最短パスを識別すること、及び、前記コントローラを前記ツリーのルートとして備える前記ネットワークについて最短パス・ツリーを計算することを、実行するように構成され、

前記近隣発見モジュールは、ハロー・プロトコルを使用して前記ネットワーク内の各近隣コントローラ及び従来型ルータを識別するように構成され、

前記リンク状態管理モジュールは、リンク状態データベースを、前記ネットワーク内の各近隣コントローラ及び従来型ルータと交換するように構成され、

前記リンク状態データベースは、前記コントローラの前記領域内の各境界スイッチ・ペアの間の各最短経路のコストを含み、

前記ネットワーク・プロセッサに通信可能に結合されたルーティング・テーブル記憶デバイスであって、前記コントローラの前記領域について、及び前記コントローラの前記領域の境界スイッチ間の、最短パス情報を含む、コントローラ・モジュールに関するルーティング・テーブルを記憶するように構成される、ルーティング・テーブル記憶デバイスと、

を備える、ネットワーク要素。

【請求項１０】

前記近隣発見モジュールが、前記領域の各外部ポートでハロー・パケットを送信すること、及び、前記領域の少なくとも１つの外部ポートを通じて、各近隣コントローラからハロー・パケットを受信することを、実行するように構成される、請求項９に記載のネットワーク要素。

【請求項１１】

前記リンク状態管理モジュールが、各近隣コントローラにリンク状態アドバタイズを送信するように更に構成され、前記リンク状態アドバタイズは前記コントローラの各外部リンクへのコストを含む、請求項９に記載のネットワーク要素。

【請求項１２】

前記リンク状態管理モジュールが、各近隣コントローラにリンク状態アドバタイズを送信するように更に構成され、前記リンク状態アドバタイズは、前記コントローラの前記領域内の各境界スイッチ・ペアに関する前記最短パスの前記コストを含む、請求項９に記載のネットワーク要素。

【請求項１３】
前記リンク状態管理モジュールが、更新されたリンク状態データを各近隣コントローラにアドバタイズするように更に構成される、請求項９に記載のネットワーク要素。
【請求項１４】
前記リンク状態管理モジュールが、前記更新されたリンク状態データを伴うリンク状態アドバタイズ・メッセージを、各近隣コントローラに送信するように更に構成される、請求項９に記載のネットワーク要素。
【請求項１５】
前記コントローラ・モジュールが、前記0penFlowプロトコルを使用して前記コントローラの前記領域内の各スイッチの転送テーブルを更新するように更に構成される、請求項９に記載のネットワーク要素。
【請求項１６】
前記最短パス計算モジュールが、前記コントローラでオープン・ショーテスト・パス・ファースト（ＯＳＰＦ）プロトコル・データ構造を開始するように更に構成される、請求項９に記載のネットワーク要素。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明の実施形態は、分割アーキテクチャ・ネットワークにおけるパスの探知又はルーティングに関する。具体的に言えば、本発明の実施形態は、分割アーキテクチャ・ネットワークを含むネットワーク及び従来のルータにおいてデータをルーティングするために、オープン・ショーテスト・パス・ファースト・プロセスを実装するための方法及びシステムに関する。
【背景技術】
【０００２】
分割アーキテクチャ・ネットワーク設計は、制御プレーン及び転送又はデータ・プレーンとも呼ばれるネットワークの制御構成要素と転送構成要素との間に、分離を導入する。分割アーキテクチャ・ネットワークは、キャリアグレード・ネットワークのアクセス／集約ドメイン、モバイル・バックホール、クラウド・コンピューティング、及びマルチレイヤ（Ｌ３＆Ｌ２＆Ｌ１、光伝送ネットワーク（ＯＴＮ）、波長分割多重（ＷＤＭ））サポートで使用可能であり、これらはすべてネットワーク・アーキテクチャの構築ブロックの１つである。
【０００３】
転送（データ）プレーンと制御プレーンの両方を同じボックス内に統合する従来のネットワーク・アーキテクチャとは異なり、分割アーキテクチャはこれら２つの機能を切り離し、転送要素（スイッチ）とは異なる物理位置にあり得るサーバ（コントローラ）上で制御プレーンを実行する。分割アーキテクチャは、転送プラットフォームの機能及びハードウェアを簡略化し、ネットワークのインテリジェンス及び管理を、スイッチを監視するコントローラ・セットに集約する。従来のネットワーク・アーキテクチャにおける転送及び制御プレーンの密結合の結果として、通常、かなり複雑な制御プレーン及び手間のかかるネットワーク管理が生じている。これにより、新しいネットワーキング・デバイスの製作には費用が掛かり、これらデバイスの潜在的な展開のための新しいプロトコル及び技術に参入するためには高い障壁が設けられている。回線速度、ポート密度、及びパフォーマンスは急速に改善されているにもかかわらず、これらの機能を管理するためのネットワーク制御プレーン機構の進歩はかなり遅れている。
【０００４】
分割アーキテクチャ・ネットワークにおいて、コントローラはスイッチから情報を収集し、適切な転送決定を計算して、これをスイッチに配信する。コントローラ及びスイッチは、制御プレーン・プロトコルを使用して情報の通信及び交換を行う。こうしたプロトコルの例が0penFlowであり、コントローラと通信するためのオープンで標準的な方法をスイ

ッチに提供する。図１は、スイッチとコントローラとの間のOpenFlowインターフェースの概要を示す図である。OpenFlowコントローラは、転送テーブル（フロー・テーブル）を構成するためのセキュアなチャネルを用いてOpenFlowスイッチと通信する。

【０００５】

OpenFlowスイッチ内の転送テーブルには、パケット・ヘッダ内のフィールドに関する合致を定義する規則、規則によって定義された合致の検出時に実行されるアクション、及び、データ・プレーン内のデータ・パケットの処理に関する統計の収集からなる、エントリが存在する。着信データ・パケットが特定の規則に合致する場合、関連付けられたアクションがデータ・パケット上で実行される。規則は、例えば、Ethernet MACアドレス、ＩＰアドレス、ＩＰプロトコル、ＴＣＰ／ＩＰポート番号、並びに着信ポート番号等の、プロトコル・スタック内のいくつかのヘッダからのキー・フィールドを含む。同様の特徴を有するデータ・パケット・セットはフローとして管理することができる。フローは、データ・パケット内の任意数の使用可能フィールド又はその組み合わせを使用して定義することができる。望ましくないフィールドについてのワイルドカードを使用して、使用可能フィールドのサブセットに関する合致に規則を制約することも可能である。

【０００６】

分割アーキテクチャの制御プレーンとデータ・プレーンの分離により、ネットワーク制御論理を修正するタスクが容易になり、開発者が多様な新しいプロトコル及び管理アプリケーションを構築可能なプログラマチック・インターフェースが提供される。このモデルでは、データ・プレーン及び制御プレーンは独立に展開及びスケーリング可能であり、データ・プレーン要素のコストは削減される。

【発明の概要】

【発明が解決しようとする課題】

【０００７】

分割アーキテクチャ・ネットワークの複数の領域のうちの１つのための複数のコントローラのうちの１つとして機能するネットワーク要素内に実装される方法を説明する。

【課題を解決するための手段】

【０００８】

コントローラは、分割アーキテクチャ・ネットワークの領域に制御プレーンを提供し、ここでコントローラは、分割アーキテクチャ・ネットワークの領域にデータ・プレーンを提供する複数のスイッチからはリモートにある。コントローラは、分割アーキテクチャ・ネットワークの他の領域の他のコントローラ、及び分割アーキテクチャ・ネットワークを含むネットワークの従来型ルータに、限定領域内リンク・コスト・データを提供することによって、分割アーキテクチャ・ネットワークの複数領域にまたがる最適化ルーティングを容易にする。限定領域内リンク・コスト・データは、すべての内部リンク・コスト・データを提供することなく、コントローラの領域のそれぞれの可能な最短パスのトラバースに関するコストを提供する。方法は、コントローラの領域内の各境界スイッチ（border switch）を含む分割アーキテクチャ・ネットワーク内のコントローラの領域のトポロジを習得することを含むステップを含み、コントローラの領域内の各境界スイッチは、コントローラの領域を、分割アーキテクチャ・ネットワークの別の領域、又はネットワーク内の従来型ルータのうちの１つにリンクさせる、少なくとも１つの外部ポートを有する。ステップは、コントローラの領域内の各境界スイッチ・ペア間の最短パスを計算することを含む。ステップは、コントローラのルーティング・テーブル内の各境界スイッチ・ペア間の各最短パスのコストを記憶することを含む。ステップは、分割アーキテクチャ・ネットワーク内の各近隣コントローラ又はネットワーク内の近隣従来型ルータを、ハロー・プロトコル（hello protocol）を使用して識別することを含み、ここで各近隣コントローラは、コントローラの領域の少なくとも１つの外部ポートを通じてアクセス可能な、分割アーキテクチャ・ネットワークの別の領域内のスイッチを制御する。ステップは、リンク状態データベースを各近隣コントローラと交換することを含み、リンク状態データベースは、各境界スイッチ・ペア間の各最短パスのコストを含む。ステップは、コントローラをツリー

のルートとして備えるネットワークについて最短パス・ツリーを計算すること、及び、最短パス・ツリーに従って転送を実装するために、コントローラの領域のスイッチ内の転送テーブルを更新することを、含む。

【０００９】

ネットワーク要素は、分割アーキテクチャ・ネットワークの複数の領域のうちの１つについての複数のコントローラのうちの１つとして機能する。コントローラは、分割アーキテクチャ・ネットワークの領域に制御プレーンを提供し、ここでコントローラは、分割アーキテクチャ・ネットワークの領域にデータ・プレーンを提供する複数のスイッチからはリモートにある。コントローラは、分割アーキテクチャ・ネットワークの他の領域の他のコントローラ、及び分割アーキテクチャ・ネットワークを含むネットワークの従来型ルータに、限定領域内リンク・コスト・データを提供することによって、分割アーキテクチャ・ネットワークの複数領域にまたがる最適化ルーティングを容易にする。限定領域内リンク・コスト・データは、すべての内部リンク・コスト・データを提供することなく、コントローラの領域のそれぞれの可能な最短パスのトラバースに関するコストを提供する。ネットワーク要素は、ネットワークを介してデータを受信するように構成された入口（ingress）モジュールと、ネットワークを介してデータを送信するように構成された出口（egress）モジュールとを備える。ネットワーク要素は、入口モジュール及び出口モジュールに結合されたネットワーク・プロセッサも含み、ネットワーク・プロセッサは、コントローラ・モジュール、トポロジ習得モジュール、最短パス計算モジュール、近隣発見モジュール、及びリンク状態管理モジュールを備える、モジュール・セットを実行するように構成される。トポロジ習得モジュールは、コントローラの領域内の各境界スイッチを含む分割アーキテクチャ・ネットワーク内のコントローラの領域のトポロジを決定するように構成され、各境界スイッチは、コントローラの領域をネットワークの別の領域又はネットワーク内の従来型ルータにリンクさせる、少なくとも１つの外部ポートを有する。コントローラ・モジュールは、分割アーキテクチャ・ネットワーク内のコントローラの領域に制御プレーン機能を提供するように構成される。最短パス計算モジュールは、近隣コントローラと共有することになるコントローラの領域内の各境界スイッチと従来型ルータとの間の最短パスを識別すること、及び、コントローラをツリーのルートとして備えるネットワークについて最短パス・ツリーを計算することを、実行するように構成される。近隣発見モジュールは、ハロー・プロトコルを使用して分割アーキテクチャ・ネットワーク内の各近隣コントローラを識別するように構成され、リンク状態管理モジュールは、リンク状態データベースを各近隣コントローラと交換するように構成される。リンク状態データベースは、コントローラの領域内の各境界スイッチ・ペア間の、各最短経路のコストを含み、ネットワーク要素は、ネットワーク・プロセッサに通信可能に結合されたルーティング・テーブル記憶デバイスも含む。ルーティング・テーブル記憶デバイスは、コントローラの領域について、及びコントローラの領域の境界スイッチ間の、最短パス情報を含む、コントローラ・モジュールに関するルーティング・テーブルを記憶するように構成される。

【００１０】

本発明は、添付の図面の図中で、限定的ではなく単なる例示として示されており、図面内の同じ参照番号は同様の要素を示す。本開示における「実施形態」又は「一実施形態」という異なる言い方は、必ずしも同じ実施形態を指しておらず、こうした言い方は少なくとも１つを意味することに留意されたい。更に、特定の機能、構造、又は特徴が実施形態に関連して説明される場合、明示的に説明されるか否かにかかわらず、こうした機能、構造、又は特徴を他の実施形態に関連して実行することは、当業者の知識の範囲内であるものと考えられる。

【図面の簡単な説明】

【００１１】

【図１】単純なOpenFlowネットワークに関する例示アーキテクチャの一実施形態を示す図である。

【図２】オープン・ショーテスト・パス・ファースト（ＯＳＰＦ）パケット・ヘッダを示

す図である。
【図３】ＯＳＰＦハロー・パケットを示す図である。
【図４】リンク状態アドバタイズ（advertisement）（ＬＳＡ）ヘッダ・フォーマットを示す図である。
【図５】ルータＬＳＡメッセージを示す図である。
【図６】ＯＳＰＦ外部リンク・コスト情報が交換される、例示ネットワークを示す図である。
【図７】ＯＳＰＦ内部リンク・コスト情報が交換される、例示ネットワークを示す図である。
【図８】複数コントローラを備えた例示の複数領域OpenFlowネットワークの一実施形態を示す図である。
【図９】ＯＳＰＦプロセスを実装するネットワーク要素の一実施形態を示す図である。
【図１０】分割アーキテクチャ領域を備えたネットワーク内のＯＳＰＦルーティングに関するプロセスの一実施形態を示すフローチャートである。
【図１１】ＯＳＰＦリンク状態アドバタイズ・ヘッダ・フォーマットを示す図である。
【図１２】ルータＬＳＡメッセージを示す図である。
【図１３】分割アーキテクチャＯＳＰＦハロー・メッセージを示す図である。
【図１４】ネットワーク内で分割アーキテクチャ・コントローラから従来型ルータへハロー・メッセージを送信するプロセスの、一実施形態を示す図である。
【発明を実施するための形態】
【００１２】
以下の説明では、多数の特定の細部が示される。しかしながら、本発明の実施形態はこれら特定の細部なしで実施可能であることを理解されよう。その他の場合、良く知られた回路、構造、及び技法は、この説明に関する理解を不明瞭にしないために詳細には示されていない。しかしながら、当業者であれば、こうした特定の細部なしに本発明が実施できることを理解されよう。当業者であれば、過度の実験なしに、含められた説明を用いて適切な機能を実装することができよう。
【００１３】
フローチャートの動作は、図７～図９及び図１４の例示の実施形態を参照しながら説明される。しかしながら、図１０のフローチャートの動作は、図７～図９及び図１４を参照しながら考察される実施形態以外の本発明の実施形態によって実行可能であり、図７～図９及び図１４を参照しながら考察される実施形態は、図１０のフローチャートを参照しながら考察される動作とは異なる動作を実行可能であることを理解されたい。
【００１４】
図内に示される技法は、１つ又は複数の電子デバイス（例えばエンド・ステーション、ネットワーク要素、サーバ、又は同様の電子デバイス）上に記憶され、実行される、コード及びデータを使用して実装可能である。こうした電子デバイスは、非一過性の機械読み取り可能又はコンピュータ読み取り可能記憶媒体（例えば、磁気ディスク、光ディスク、ランダム・アクセス・メモリ、読み取り専用メモリ、フラッシュ・メモリ・デバイス、及び相変化メモリ）等の、非一過性の機械読み取り可能又はコンピュータ読み取り可能媒体を使用して、コード及びデータを記憶及び（内部的に、及び／又は、ネットワークを介して他の電子デバイスと）通信する。加えてこうした電子デバイスは、通常、１つ又は複数の記憶デバイス、ユーザ入力／出力デバイス（例えば、キーボード、タッチ・スクリーン、及び／又はディスプレイ）、及びネットワーク接続等の、１つ又は複数の他の構成要素に結合された、１つ又は複数のプロセッサのセットを含む。プロセッサのセットと他の構成要素との結合は、通常、１本又は複数のバス及びブリッジ（バス・コントローラとも呼ばれる）を介する。記憶デバイスは、１つ又は複数の非一過性の機械読み取り可能又はコンピュータ読み取り可能記憶媒体、及び、非一過性の機械読み取り可能又はコンピュータ読み取り可能通信媒体を表す。したがって、所与の電子デバイスの記憶デバイスは、通常、その電子デバイスの１つ又は複数のプロセッサのセット上で実行するための、コード及

び／又はデータを記憶する。もちろん、本発明の実施形態の１つ又は複数の部分は、ソフトウェア、ファームウェア、及び／又はハードウェアの異なる組み合わせを使用して実装可能である。

【００１５】

本明細書で使用される場合、ネットワーク要素（例えば、ルータ、スイッチ、ブリッジ、又は同様のネットワーキング・デバイス）は、ネットワーク上の他の機器（例えば、他のネットワーク要素、エンド・ステーション、又は同様のネットワーキング・デバイス）を通信可能に相互接続するハードウェア及びソフトウェアを含む、１つのネットワーキング機器である。いくつかのネットワーク要素は、複数のネットワーキング機能（例えば、ルーティング、ブリッジング、スイッチング、レイヤ２集約、セッション境界制御、マルチキャスティング、及び／又は加入者管理）に対するサポートを提供する、及び／又は、複数のアプリケーション・サービス（例えばデータ収集）に対するサポートを提供する、「複数サービス・ネットワーク要素」である。

【００１６】

分割アーキテクチャ領域

単一のアクセス／集約ネットワークは、複数の従来型ルータと協働する複数の個別の分割アーキテクチャ領域で構成され得る。本明細書で使用される場合、分割アーキテクチャ領域とは、ドメインに類似した、別々のルーティングを伴う分割アーキテクチャ・ネットワークのセクションである。これは、ネットワークの堅固さ又は制御プレーンのスケーラビリティに関する、広範な地理的領域にわたる管理を簡略化するために実行することができる。各分割アーキテクチャ領域は、別々のコントローラによって管理可能である。特定のアプリケーションに応じて、これら別個の分割アーキテクチャ領域のコントローラは、分割アーキテクチャ・ネットワークの適切な管理のために、何らかの情報を共有及び交換する必要がある。

【００１７】

オープン・ショーテスト・パス・ファースト・ルーティング

従来型ネットワーク及び分割アーキテクチャ・ネットワークは、どちらも、スイッチとネットワークによってサービスが提供される他のデバイスとの間の経路を計算しなければならない。オープン・ショーテスト・パス・ファースト（ＯＳＰＦ）は、内部ゲートウェイ・ルーティング・プロトコルである。ＯＳＰＦ（ＲＦＣ２３２８に定義）は、ルータがその近隣のリンク状態情報をルーティング・ドメイン内のすべてのノードにブロードキャストする、リンク状態プロトコルである。この情報を使用して、あらゆるルータがドメイン内にネットワーク全体のトポロジ・マップを構築する。各ルータは、全ネットワーク・トポロジを反映するリンク状態データベースを維持する。このトポロジ・マップ及びリンク・コスト・メトリクスに基づき、ルータは、ダイクストラ法アルゴリズムを使用してすべての他のルータへの最短パスを決定する。次にこの情報を使用し、インターネット・プロトコル（ＩＰ）パケットの転送に使用されるルーティング・テーブルが作成される。

【００１８】

ＯＳＰＦは、ルーティング・ドメインを、管理しやすいように領域境界ルータ（ＡＢＲ）によって分離されている複数の「領域」に分割することができる。各領域は、ＯＳＰＦネットワークのコア又はバックボーン用に予約された識別子「０」を備える、３２ビット識別子、通常は領域内のメイン・ルータのＩＰアドレスによって、識別される。リンク状態情報のブロードキャストは領域に限定され、領域を越えて送信されることはない。ＯＳＰＦプロセスにおいて、各ルータ（スイッチ）は、それが属する各領域に対するＯＳＰＦプロトコル・プロセスの別々のコピーを実行する。ルータは、異なる領域に属する複数のインターフェースを有する場合、プロセスの複数のコピー、各インターフェースについて１つを実行する。起動時、ルータのＯＳＰＦプロセスは、すべてのルーティング・プロトコル・データ構造を開始し、そのルータのアクティブ・インターフェースに関する情報を下位レイヤから取得する。その後ルータは、ＯＳＰＦのハロー・プロトコル・パケットを使用して、その近隣を検出する。ルータはその近隣にハロー・パケットを送信し、その近

隣からハロー・パケットを受信する。ブロードキャスト及びポイント・ツー・ポイント・ネットワークでは、ハロー・パケットはマルチキャスト・アドレス２２４．０．０．５に送信される。非ブロードキャスト・ネットワークでは、ユーザ構成は近隣を開発するために必要である。ブロードキャスト及び非ブロードキャスト多重アクセス・ネットワーク（ＮＢＭＡ）では、ネットワークのセグメントに「指定ルータ」及び「バックアップ指定ルータ」を選択するために、ハロー・プロトコルが使用される。

【００１９】

近隣を検出すると、ルータはその新しく検出された近隣を用いて「隣接（adjacencies）」を確立するように試行することになる。隣接は、領域内のルーティング情報の分布を決定する。ルーティング更新は、隣接上でのみ送受信される。隣接を確立すると、ルータはその「リンク状態データベース」を、隣接の他端上の対応するルータと同期させることになる。ブロードキャスト及びＮＢＭＡネットワークの場合、指定ルータはどのルータが隣接になるかを決定する。

【００２０】

ルータは、リンク状態アドバタイズ（ＬＳＡ）を使用して、「リンク状態」とも呼ばれるその状態を定期的にアドバタイズする。リンク状態は、ルータの状態が変化した時にもアドバタイズされる。ＬＳＡは、アドバタイズ・ルータの隣接も含む。ルータは、領域全体にそのＬＳＡをフラッディングする（flood）。フラッディング・アルゴリズムは、領域内のすべてのルータが同じ正確なリンク状態データベース確実に有するように、情報の信頼性を保証する。リンク状態データベースは、領域に属する各ルータによって発せられるＬＳＡの集合からなる。

【００２１】

各ルータは、このデータベースを使用して、それ自体をルートとして備える最短パス・ツリーを計算する。次に最短パス・ツリーを使用して、ルータのルーティング・テーブルが作成される。ＯＳＰＦメッセージは、ＴＣＰ又はＵＤＰ等の移送レイヤ・プロトコルを使用せずに、直接伝送され、プロトコル番号８９と共にＩＰパケットに封入される。加えてＯＳＰＦは、それ自体の誤り検出及び訂正を使用する。

【００２２】

ＯＳＰＦは、以下の５つの異なるパケット・タイプを定義する。ハロー・パケット：ＯＳＰＦのハロー・プロトコル・パケットは、近隣関係を発見及び維持するために使用される。データベース記述パケット：これらのパケットは、隣接を形成するために使用される。リンク状態要求パケット：これらのパケットは、隣接ルータ間でリンク状態データベースをダウンロードするために使用される。リンク状態更新パケット：これらのパケットは、ＯＳＰＦの信頼できる更新機構に使用される。単一のリンク状態更新パケットは、いくつかのルータのＬＳＡを含む。リンク状態Ａｃｋパケット：これらのパケットは、ＯＳＰＦの信用できる更新機構に関するリンク状態更新パケットと共に使用される。ＯＳＰＦプロトコル・パケット（ハロー・パケットを除く）は、隣接を介してのみ送信される。したがってすべてのＯＳＰＦプロトコル・パケットは、ソース・アドレスとして１つのルータのＩＰアドレス、及び、宛先アドレスとして他のルータのＩＰアドレスを用いて、単一のインターネット・プロトコル（ＩＰ）ホップを移動する。

【００２３】

ＲＦＣ　２３２８に定義されたＯＳＰＦは、以下の５つの異なるタイプのＬＳＡを指定する。ルータＬＳＡ：これらのＬＳＡは、領域内のすべてのルータによって送信される。各ＬＳＡは、領域に対するルータのインターフェースの状態を含む。ルータＬＳＡは単一の領域全体のみでフラッディングされる。ネットワークＬＳＡ：これらのＬＳＡは、ブロードキャスト又はＮＢＭＡネットワークの指定ルータによって発信される。このＬＳＡは、ネットワークに接続されたルータのリストを含む。ルータＬＳＡと同様に、このＬＳＡも単一の領域全対のみでフラッディングされる。ネットワーク要約ＬＳＡ：このＬＳＡは、領域境界ルータ（ＡＢＲ）によって発信され、各ＬＳＡは、領域外部であるが依然として自律システム（ＡＳ）内部にある宛先ネットワークへの経路を記述する。境界要約ＬＳ

Ａ：ネットワーク要約ＬＳＡと同様に、このＬＳＡは境界領域ルータ（ＡＢＲ）によって発信され、ＡＳ境界ルータへの経路を記述する。ＡＳ外部ＬＳＡ：このＬＳＡは、ＡＳ境界ルータによって発信され、ＡＳ全体でフラッディングされる。これらのＬＳＡのそれぞれが、他のＡＳ内の宛先への経路を記述する。ＯＳＰＦパケット・ヘッダ・フォーマットは、図２に示される通りである。ハロー・パケット・フォーマットは図３に示されている。ＬＳＡヘッダ・フォーマットは図４に示される通りである。ルータＬＳＡパケット・フォーマットは図５に示される通りである。

【００２４】

分割アーキテクチャ・ネットワークにおけるＯＳＰＦ

分割アーキテクチャ・ネットワークにおいて、ルーティング・メッセージは、すべての他の制御メッセージと同様に、コントローラ間で交換される。分割アーキテクチャにおいてＯＳＰＦを実装するための単純な方法の１つは、内部コストと、外部コストに関する交換情報のみと、すなわち、異なるネットワーク領域を接続するリンクのコストを、無視することである。このように実行することによって、従来のＯＳＰＦメッセージはコントローラ間で交換可能であり、各分割アーキテクチャ領域は単一のノードとみなすことができる。

【００２５】

図６に示されたネットワークでは、例えばコントローラＢが、１０単位のコストを伴う領域Ｃへのリンクを有することを告知する。しかしながら、この情報に基づいて探知された経路は、準最適である。この実装では、内部コストが他のコントローラに提供されないため、コントローラは実際のエンド・ツー・エンド・コストに基づいて決定することはできない。その結果、１つの領域内の内部コストが内部パスによってかなり異なる場合、準最適な経路が生じる。

【００２６】

代替実施形態において、上記の準最適経路の問題は、コントローラ間で交換される情報に内部コストを含めることによって解決できる。各コントローラは、内部リンク・コストを計算し、その境界スイッチのうちの任意の２つの間の最短パス（すなわち最も安価なパス）を探知することができる。例えば、再度図６を参照すると、コントローラＣは、それ自体の領域とコントローラＦの領域との間のリンク・コストを告知した場合、境界スイッチＳ１とＳ２の間の最短パス・コストを追加し、それをコントローラＢに送信されるコスト情報に追加する。同様に、コントローラＣはスイッチＳ２とＳ３の間の最短パス・コストを、領域ＣとＦの間のコストに追加して、この情報をコントローラＤに送信する。

【００２７】

この代替実施形態は問題も発生させる。この実施形態は、同じリンクについて複数の（及び恐らくは異なる）リンク・コスト・メッセージを送信する必要がある。例えば、領域Ｆに到達するためにコントローラＣによって告知されるコストは、この情報がコントローラＢに送信される場合とコントローラＤに送信される場合とでは、それぞれが異なる内部パス・コストに基づいているため、異なる（例えば、Ｓ１とＳ２の間の内部最短パス・コストがＳ２とＳ３の間のコストと異なる場合、コントローラＣはそれ自体の領域とコントローラＦの領域との間の２つの異なるリンク・コストを告知することになる）。最終的に、これら２つの矛盾するメッセージが他のコントローラによって受信されることになり、これらのメッセージは、２つの別々のメッセージとして解釈されるのではなく、第２のメッセージは第１の受信されたメッセージの更新として解釈されることになる。

【００２８】

例えば、複数のコストを同じリンクに起因させることができるようにＯＳＰＦプロトコルを修正することによって、この複数コスト問題を解決しようとする他の実施形態は、分割アーキテクチャ内のコントローラによって探知される最短パスが、最適でない可能性がある。例えば、図７に示されたシナリオでは、各リンクの番号がいずれかの方向のリンクのコストを示し、準最適経路が決定されることになる。領域Ａから領域Ｆへの最短パスを計算する際には、以下のステップが実行される。（１）ＣはＢ及びＤに、Ｆに到達するた

めのコスト２（１＋）を告知し、（２）ＢはＡに、Ｃに到達するためのコスト１１（１０＋１）を告知し、（３）ＤはＢに、Ｃに到達するためのコスト２１（１＋２０）を告知し（領域Ｄ内の境界スイッチＳ４とＳ５の間には１つの内部パスのみが存在することに留意されたい）、（４）ＢはＡに、Ｃに到達するためのコスト２（１＋１）を告知する。

【００２９】

コントローラＡが領域Ｆへのその最短パスを計算する場合、上記情報に基づいて、コストが最も小さい１４であることから、他のパスよりもパスＡ－Ｂ－Ｃ－Ｆを選択する。しかしながら、図７に示されるすべてのコスト情報に基づく、ＡとＦの間の最適なルーティングは、Ａ－Ｂ－Ｄ－Ｅ－Ｄ－Ｃ－Ｆであり、コストは１０のみである。コントローラＡは、コントローラ間で交換される情報がこうした選択にとって不十分であることから、このパスは選択され得ない。したがって、最適なルーティング・ソリューションの場合、コントローラ間で交換されるＯＳＰＦメッセージは、以下でより詳細に考察されるように、内部リンク・コストに関する更なる情報を伝搬する必要がある。

【００３０】

抽出された（abstracted）内部領域パス・コストを備えるＯＳＰＦ

本発明の実施形態は、従来技術の欠点を回避するための方法及びシステムを提供する。従来技術及び前述の分割アーキテクチャ・ネットワークにおけるＯＳＰＦの単純な実施は、最短パスが必ずしも正確に決定されない、及び／又は、非効率的又は拡張性がない経路を識別するために過剰な情報が提供される、準最適ルーティング・ソリューションを提供する。

【００３１】

本発明の実施形態は、従来技術のこれらの欠点を克服する。本発明の実施形態は、境界スイッチの各ペア間の内部領域パス・コストを、コスト値に関連付けられた直接リンクとして抽出する。このソリューションは、最適なパスを提供し、効率的に実行可能であり、拡張性があり、非分割アーキテクチャにおける従来のルータとの後方互換性がある。

【００３２】

一実施形態において、ＯＳＰＦルーティング・プロトコルは、ネットワーク内の任意の２つの転送要素間に最適な（最短の）パスを確立するために、分割アーキテクチャにおけるコントローラ間で実装される。実施形態は、内部領域パス・コスト及び内部領域リンク・コストの両方の必要な情報を、複数領域分割アーキテクチャ・ネットワーク全体に公開する。前述のように、実施形態は、最適性、効率性、拡張性、及び後方互換性を提供する。

【００３３】

ルーティング・プロトコルは、各分割アーキテクチャ・コントローラが、任意の他の宛先に到達するために最適なパスを個別に計算できるようにする、十分かつ正確な情報を提供する。これは、従来のドメイン内ルーティング・プロトコル（ＯＳＰＦ、ＩＳ－ＩＳ）が従来型ネットワーク内で提供する、最適性特性である。しかしながら、最適パスの定義は、分割アーキテクチャ・ネットワークで使用される場合はわずかに異なる。従来型ネットワークでは、これは最小のルータ間コストを伴うパスである。分割アーキテクチャ・ネットワークにおいて、最適パスは、最小ルータ間コストとルータ内コスト（又はＳＡ領域内コスト）とを伴うパスである。

【００３４】

分割アーキテクチャＯＳＰＦ（ＳＡ－ＯＳＰＦ）ルーティング・プロトコルの実施形態は、各ルータが最短パス決定を独立に計算できるようにするものである。ＯＳＰＦは高速で収束する。言い換えれば、分割アーキテクチャＯＳＰＦプロトコルは、従来型ネットワークＯＳＰＦと比べて、追加の収束オーバヘッドを招かない。分割アーキテクチャＯＳＰＦプロトコルは、何百ものスイッチを備える大規模ネットワークに拡張する。拡張性は、交換されるメッセージの数のオーバヘッドとコントローラに関する記憶要件の両方によって量子化することができる。増分の展開は、ネットワーク・プロトコルに関する任意の新しい提案の適合にとって重要である。分割アーキテクチャＯＳＰＦプロトコルは、従来型

ルータとの後方互換性がある。この特性により、実際のネットワーク環境におけるその潜在的な使用率を増加させる。

【００３５】

分割アーキテクチャＯＳＰＦの実施形態は、コスト値に関連付けられた直接リンクとして、境界スイッチの任意のペア間の内部領域パス・コストを抽出する。この機能を容易にする分割アーキテクチャＯＳＰＦのいくつかの態様が存在する。内部領域コストは、後方互換性を保証するために従来のＯＳＰＦメッセージの形で組み込まれる。領域内コストは、領域間コストとは別に、すべての領域におけるすべてのコントローラに伝搬される。これにより、最短パス計算の最適性を保証する。これは、各コントローラが、領域間及び領域内の両方の、全ネットワークの完全なピクチャを有することを保証する。こうした十分な領域がある場合、従来のＯＳＰＦと同様に、最短パスの計算はダイクストラ法アルゴリズムを使用して容易に実行可能である。

【００３６】

分割アーキテクチャＯＳＰＦは、どの内部情報が外部コントローラに伝搬されることになるかが慎重に管理されるという点で、拡張可能である。１つの生来の手法は、分割アーキテクチャ・スイッチの任意のペア間の内部コストを送信することである。しかしながら、これは拡張性の問題を生じさせる。更にこの情報のほとんどが、外部コントローラでの決定に有用ではない。分割アーキテクチャＯＳＰＦは、境界分割アーキテクチャ・スイッチの任意のペアの合計集約コストのみを伝搬する。

【００３７】

図８は、分割アーキテクチャ・ネットワークの例示の一実施形態を示す図である。例示の分割アーキテクチャ・ネットワークは、別々の分割アーキテクチャ領域（ＳＡ）８０１Ａ～Ｃに分割される。各領域８０１Ａ～Ｃはスイッチ・セットを含む。同じ領域内のすべてのスイッチは、単一の論理コントローラ８０３Ａ～Ｃによって制御される。一実施形態において、ＳＡは、冗長性の目的で、１次コントローラ及びバックアップ・コントローラとして実装可能である。

【００３８】

各ＳＡにおけるスイッチは、分割アーキテクチャ・ネットワークのデータ・プレーンを実装可能な、任意のタイプのルータ、スイッチ、又は同様のネットワーキング・デバイスとすることができる。スイッチは、境界分割アーキテクチャ・スイッチ及び内部分割アーキテクチャ・スイッチを含むことができる。境界分割アーキテクチャ・スイッチは、異なるＳＡ領域内の別のスイッチに接続するインターフェースで、分割アーキテクチャ機能をサポートする。境界分割アーキテクチャ・スイッチは、通常、単一のＳＡ領域のコントローラによって制御される。他の実施形態において、境界分割アーキテクチャ・スイッチは、複数のＳＡ内に存在することが可能であり、各それぞれのＳＡコントローラによって制御されるインターフェースを有する。内部分割アーキテクチャ・スイッチは、分割アーキテクチャ・プロトコルをサポートする。これは、その領域内のコントローラによって制御される。そのすべての近隣は同じＳＡ領域内にある。

【００３９】

スイッチは、リンク・セットを介して互いに通信している。これらのリンクは、有線又は無線の通信媒体、及びそれらの任意の組み合わせを含む、任意のタイプの通信媒体とすることができる。リンクは、内部リンク又は外部リンクのいずれかとして分類可能である。内部リンクは、ＳＡ内の２つのスイッチ間のリンクであり、これらのスイッチは、同じＳＡ領域に属する境界スイッチ又は内部ＳＡ領域スイッチのいずれかとすることができる。外部リンクは、異なるＳＡ領域に属する２つのＳＡスイッチ間のリンクである。この場合、どちらのＳＡスイッチも境界ＳＡスイッチである。

【００４０】

リンク状態アドバタイズ（ＬＳＡ）８０５は、分割アーキテクチャＯＳＰＦを実装するＬＳＡの例示セットである。各ＬＳＡは、境界スイッチのペア及びこれら２つの境界スイッチの間のＳＡ領域内をトラバースする関連付けられたコストの抽出である。これらのＬ

ＳＡは各コントローラによって生成され、隣接するコントローラに伝送される。例示のＬＳＡセット８０５は、ＳＡ領域８０１Ｂに関するコントローラ８０３ＢのＬＳＡのセットである。

【００４１】

図９は、コントローラを実装するネットワーク要素の一実施形態を示す図である。一実施形態において、コントローラ９０１はルータ、スイッチ、又は同様のネットワーキング・デバイスである。コントローラ９０１は、入口モジュール９０３、出口モジュール９０５、ネットワーク・プロセッサ９０７、及び記憶デバイス９１１を含むことができる。入口モジュール９０３は、物理及びリンク・レベルで着信データ・トラフィックを処理し、このデータを更なる処理のためにネットワーク・プロセッサに提供する。同様に、出口モジュール９０５は、物理及びリンク・レベルで発信データ・トラフィックを処理し、接続されたネットワークを介してこれを他のデバイスに伝送する。これら２つのモジュール機能がまとまって、ネットワークを介した他のデバイスとの通信を実行可能にする。

【００４２】

ネットワーク・プロセッサ９０７は、ネットワークのデータ・プレーンを支配するネットワークの制御プレーンに関する機能のそれぞれを含むネットワーク要素の機能を実行する、処理デバイス又は処理デバイスのセットである。ネットワーク・プロセッサ９０７は、近隣発見モジュール９１３、ＯＳＰＦモジュール９１５、トポロジ習得モジュール９１７、リンク状態管理モジュール９１９、及びOpenFlowコントローラ９２１等のコントローラ・モジュールを含む、モジュール・セットを実行することができる。

【００４３】

加えて、ネットワーク・プロセッサ９０７は、記憶デバイス９１１内に記憶されたデータにアクセスすることができる。記憶デバイス９１１に記憶されたデータは、ルーティング・テーブル９２３及びリンク状態データベース９２５を含むことができる。他の実施形態において、記憶デバイス９１１は、任意数の別々のローカル又は分散型の記憶デバイス、及び、これらのデバイス全体にわたって記憶されたデータの任意の配置構成を含むことができる。ネットワーク・プロセスによって実行される他のモジュールは、記憶デバイス９１１からロードすること、又は記憶デバイス９１１上に記憶することも可能である。

【００４４】

近隣発見モジュール９１３は、近隣コントローラのそれぞれに関する情報を取得して、コントローラによって管理されるＳＡ領域のスイッチ間での適切な通信及びそれらスイッチの構成を実行可能にするために、ハロー・プロトコル又は同様のプロトコルを使用して、ネットワーク内の他のデバイスと通信するためのプロトコルを管理することができる。任意のハロー・プロトコル又はプロセスを使用して、ＳＡ領域に関して隣接するコントローラ及びスイッチを識別することができる。

【００４５】

トポロジ習得モジュール９１７は、内部でコントローラが動作するネットワークのトポロジを決定するために、近隣発見モジュール９１３によって収集された情報を使用する。このトポロジ情報は、ネットワークを介して最適なルートを計算するために、ＯＳＰＦモジュール９１５によって使用される。トポロジ情報は、ネットワーク内のリンク・コストを追跡及び決定するために、リンク状態管理モジュール９１９によっても使用される。

【００４６】

ＯＳＰＦモジュール９１７は、ネットワーク内のソース又は発信デバイスから宛先デバイスの間の最適な経路を計算する。ＯＳＰＦモジュール９１７は、ルーティング・テーブル・セット９２３内にルーティング情報を記憶することができる。ＯＳＰＦモジュール９１７は、トポロジ習得モジュール９１７によって生成されたトポロジ情報を使用して、ネットワーク要素間の経路を計算することができる。いくつかの実施形態において、ＯＳＰＦモジュール９１７は、他の各デバイスへの経路を確立するために、ルートにコントローラを備えたネットワークに関するスパニング・ツリーを計算することもできる。

【００４７】

リンク状態管理モジュール９１９は、リンク状態データベース９２５内のネットワーク又はネットワークの領域に関するリンク状態情報を管理する。リンク状態管理モジュール９１９は、コントローラと他のコントローラの間並びにコントローラのＳＡ領域内の接続に関する情報を提供する隣接コントローラに広められることになる、リンク状態アドバタイズを生成することもできる。この情報は、ネットワーク内の他のコントローラに送信される各境界スイッチ・ペアに関するリンク状態アドバタイズ・セットとして、パッケージングすることができる。

【００４８】

コントローラ９２１は、ＳＡ領域の制御プレーンを管理するための、任意のタイプの分割アーキテクチャ・コントローラとすることができる。コントローラは、分割アーキテクチャ・ネットワークを管理するための、0penFlowプロトコル又は同様のプロトコルを実装することができる。コントローラ・モジュール９２１は、データ・プレーン上でのパケットの転送を構成するために、ＳＡ領域内のスイッチと通信することができる。コントローラ９２１は、近隣情報、リンク状態アドバタイズを交換するため、及び同様の情報をピアに提供するために、他のコントローラとも通信する。

【００４９】

図１０は、分割アーキテクチャ・ネットワーク内のＯＳＰＦを実装及びサポートするためのコントローラの動作の一実施形態を示すフローチャートである。プロセスは、わかりやすくするために、コントローラによって実行されていると考えて説明されるが、プロセスは、特定のコントローラ・モジュール（例えば0penFlowコントローラ）と共にコントローラの構成要素（例えばＯＳＰＦモジュール）によって実行されることが可能である。プロセスは、コントローラが活動化又はリセットされた時点で開始することができる。他の実施形態において、プロセスは連続的又は定期的である。コントローラは、ＯＳＰＦルーティング・プロトコル関連データ構造等のデータを記憶及び操作するために使用されることになる、データ構造セットを開始することができる（ブロック１００１）。

【００５０】

コントローラは、境界スイッチであるＳＡ領域内のスイッチを識別することを含む、その割り当てられたＳＡ領域のトポロジを習得する（ブロック１００３）。トポロジは、リンク状態アドバタイズ又はリンク状態データベースの他のコントローラ及び従来型ルータとの交換を通じて、並びに同様の機構を通じて、習得することができる。異なるＳＡ領域内のコントローラがＳＡ－ＯＳＰＦプロトコルと通信可能である。交換されるメッセージは、ハロー・メッセージ、データベース記述、リンク状態要求、リンク状態更新、及びリンク状態肯定応答を含む、従来のＯＳＰＦメッセージと同様である。特に図８では、リンク状態更新メッセージ８０５が示されている。単一のリンク状態更新パケットは、いくつかのリンク状態アドバタイズ（ＬＳＡ）を含むことができる。ＳＡ－ＯＳＰＦの基本ルーティング・アルゴリズムの別々のコピーは、各領域内で実行する。複数の領域へのインターフェースを有するルータは、アルゴリズムの複数のコピーを実行する。このデータは、ＳＡ領域及びＳＡ領域が位置する境界ネットワークのトポロジカル・マップにコンパイルされる。

【００５１】

コントローラ（例えばコントローラのＯＳＰＦモジュール）は、コントローラのＳＡ領域内の境界スイッチの各ペア間の最短パスを計算する（ブロック１００５）。これは、スイッチ間の他の領域内経路を計算することと共に実行可能である。境界スイッチ・ペア・コストは、他のコントローラ及び従来型ルータと共有され、それらのコントローラ及び従来型ルータがコントローラのＳＡ領域を横断する最適なパスを決定できるようにするものである。最短パスは、ＳＡ領域の習得トポロジを使用して計算される。一実施形態において、ダイクストラ法アルゴリズムは、任意の境界スイッチ・ペア間の最短パスを計算するために使用される。各ペアについて計算された最短パスのコストは、その後、トポロジカル・データと共に、又はＯＳＰＦルーティング・データと共に、記憶される（ブロック１００７）。

【００５２】
次にコントローラは、近隣コントローラを識別する（ブロック１００９）。一実施形態において、コントローラは、近隣を獲得するためにＳＡ－ＯＳＰＦのハロー・プロトコルを使用することができる。コントローラは、その近隣の領域又は近隣の従来型ルータ内の他のコントローラにハロー・パケットを送信し、次に他のコントローラ及び従来型ルータからハロー・パケットを受信する（ブロック１１１１）。ブロードキャスト及びポイント・ツー・ポイント・ネットワークでは、コントローラは、そのハロー・パケットをマルチキャスト・アドレスに送信することによって、その近隣領域内のコントローラ又は従来型ルータを動的に検出する。非ブロードキャスト・ネットワークでは、近隣コントローラを発見するために、いくつかの構成情報を提供することができる。

【００５３】
次にコントローラは、その新しく獲得された近隣コントローラのいくつかと共に、隣接の形成を試行することになる。リンク状態データベースは、隣接コントローラとＬＳＡメッセージを交換すること（ブロック１１１５）によって、隣接コントローラのペア間で同期される（ブロック１１１３）。各ＬＳＡメッセージは、その外部リンク全体のコストを含む。コントローラは、近隣に接続しているリンクのコストをこの特定の近隣に送信するのみならず、リンクのコストをすべての近隣にも送信する。外部リンクのコストの送信に加えて、コントローラは内部リンクのコストも送信することになる。ＬＳＡメッセージは、２つの境界スイッチのＳＡスイッチＩＤを、２つの関連インターフェースＩＤと共に使用して構築することができる。コントローラは、リンク状態とも呼ばれるそのＳＡ領域の状態を、定期的にアドバタイズすることができる。リンク状態は、領域の（コントローラの）状態が変化した場合にもアドバタイズされる。領域の隣接は、そのＬＳＡの内容に反映される。この隣接とリンク状態との間の関係により、プロトコルが適時に障害を検出できるようになる。

【００５４】
ＬＳＡは、複数のＳＡ領域にまたがってフラッディングされる。フラッディング・プロセスは信頼できるものであり、ネットワーク全体のすべてのコントローラが確実に同じリンク状態データベースを有することを保証する。このデータベースは、領域間トポロジ及び領域内トポロジの両方を含む各領域によって発信されたＬＳＡの集合からなる。

【００５５】
各コントローラは、このリンク状態データベースから、それ自体をルートとして備えるネットワーク全体について最短パス・ツリーを計算する（ブロック１１１７）。次にこの最短パス・ツリーは、ＯＳＰＦプロトコルに関するルーティング・テーブルを生み出す。コントローラはこのルーティング情報に基づき、計算された最短パスを使用して、ＳＡ領域を介してデータ・パケットをルーティングするように、そのＳＡ領域内のスイッチをプログラミングする（ブロック１１１９）。スイッチの更新は、セキュア・チャネルを介する制御プレーン・プロトコル（例えばOpenFlow）を使用するか、又は同様のプロセスを使用して、達成可能である。

【００５６】
ネットワークに変化が生じた場合、コントローラによって更新リンク状態アドバタイズを送信することができる（ブロック１１２１）。ＬＳＡメッセージは、隣接するコントローラがそれらのルーティングを必要に応じて更新できるように、コントローラのＳＡ領域内の変化を通知するために隣接のコントローラに送信される。

【００５７】
分割アーキテクチャＯＳＰＦは、従来型ルータがネットワーク内の他のルータには単一のルータのように見えるという点で、従来のＯＳＰＦとは異なる。しかしながらＳＡ領域は、ネットワーク内の他の領域／ルータにはＮ個のルータのように見えることになり、ここでＮは、領域の境界分割アーキテクチャ・スイッチの数である。相違は、従来のＯＳＰＦは領域あたり単一のルータ上で単一のプロセスを実行し、単一のルータのように見える一方で、ＳＡ－ＯＳＰＦはＳＡ領域の単一のコントローラ上で単一のプロセスを実行し、

他のコントローラにはＮ個のルータのように見えるという点である。

【００５８】

一実施形態において、単一のＳＡ－ＯＳＰＦプロセスは、従来型ルータにおいてＯＳＰＦプロセスによって送信されるあらゆるハロー・パケットについて、Ｎ個の境界ＳＡスイッチのそれぞれを表すために、Ｎ個の異なるハロー・パケットを送信する。これらのハロー・パケットは、他のルータ／ＳＡ領域に接続されたインターフェース上で送信される。同様に、単一のＳＡ－ＯＳＰＦプロセスは、従来型ルータにおいてＯＳＰＦプロセスによって送信されるあらゆるルータＬＳＡパケットについて、Ｎ個の境界ＳＡスイッチのそれぞれを表すために、Ｎ個の異なるルータＬＳＡパケットを送信する。これらのＬＳＡパケットは、他のルータ／ＳＡ領域に接続されたインターフェース上で送信される。

【００５９】

端的に言えば、ＳＡ領域に関するコントローラは、Ｎ個のＯＳＰＦコントローラ・プロセスを有するかのようにＯＳＰＦを管理し、ここでＮは、ＳＡ領域内の境界分割アーキテクチャ・スイッチの数である。これにより、接続された従来型ルータ上で実行中のＯＳＰＦとの後方互換性を保証する。したがってコントローラは、Ｎ個のハロー・パケット、Ｎ個のＬＳＡメッセージ等を送信する。従来型ネットワークにおいて同じ数のルータの場合、ＳＡ－ＯＳＰＦは、分割アーキテクチャ・ネットワーク内で従来のＯＳＰＦに比べて少ない数のプロトコル・パケットを使用する。同じＳＡ領域に属する境界ＳＡスイッチ間でＯＳＰＦプロトコル・パケットを送信する必要がないことから、節約が生じる。

【００６０】

ＬＳＡメッセージ

図１１は、ＯＳＰＦリンク状態アドバタイズ・ヘッダ・フォーマットを示す図である。ＳＡ－ＯＳＰＦに後方互換性があることを保証するために、標準ＯＳＰＦメッセージ・フォーマットと同じフォーマット（すなわち、同じフィールド・シーケンス及びフィールド長さ）が使用される。しかしながら、いくつかのフィールドに対する値の割り当ては異なる可能性がある。

【００６１】

図１２は、ＳＡ－ＯＳＰＦ　ＬＳＡメッセージ、具体的に言えばルータＬＳＡタイプ・メッセージを示す図である。ルータＬＳＡは領域内の各スイッチによって生成される。これは、領域内の領域インターフェースの状態を記述する。このメッセージはすべてのコントローラによって発信される。このＬＳＡは、複数ドメインＳＡネットワークへのルータのインターフェースの集合状態を記述する。これは、すべてのコントローラ全体にフラッディングされる。

【００６２】

一実施形態において、ルータＬＳＡメッセージは、自己スイッチＩＤ、自己インターフェースＩＤ、近隣インターフェースＩＤ、及び近隣スイッチＩＤの、メッセージの４つのフィールドにおいて、従来型ルータＬＳＡメッセージと異なる（図１２に図示）。スイッチＩＤは、分割アーキテクチャ内の各スイッチに割り当てられる固有の３２ビット識別子である。このＩＤは、複数領域ＳＡドメイン内のすべてのスイッチの中で固有でなければならない。一実施形態において、スイッチの最高位ＩＰアドレスがスイッチＩＤに使用される。自己スイッチＩＤはアドバタイズ・スイッチの識別子であり、近隣スイッチＩＤは近隣スイッチの識別子である。インターフェースＩＤは、スイッチの各インターフェースに割り当てられる３２ビット識別子である。この識別子は単一のスイッチにおいてのみ固有であればよい。自己インターフェースＩＤはアドバタイズ・スイッチのインターフェースＩＤであり、近隣スイッチＩＤは（近隣スイッチの）近隣インターフェースのインターフェースＩＤである。

【００６３】

ハロー・メッセージの処理

ＯＳＰＦハロー・プロトコルは、ルータが近隣ルータを伴う隣接を確立及び維持できるようにする機構である。隣接ルータはハロー・メッセージを交換する。ブロードキャスト

及びポイント・ツー・ポイント・ネットワークにおいて、ルータは、ハロー・メッセージをマルチキャスト・アドレスに送信することによって、その近隣ルータを動的に検出することができる。ルータはハロー・パケットを受信すると、近隣内に隣接を形成する。リンク状態データベースは、隣接ルータのペア並びにルーティング更新間で同期される。隣接が確立された後、依然としてルータは、活動中であることを示すためにハロー・メッセージを定期的に交換する必要がある。要約すると、ハロー・メッセージの目的は、近隣関係を発見及び維持することである。

【００６４】

図１３は、ＳＡ－ＯＳＰＦにおけるハロー・メッセージを示す図である。ＳＡ－ＯＳＰＦプロトコルにおいて、ハロー・メッセージは未変更とすることができる。しかしながら、分割アーキテクチャ・コントローラは、近隣ルータとそれに接続する分割アーキテクチャ・スイッチとの間で隣接を維持するために、追加のハロー・メッセージを構築する必要がある。

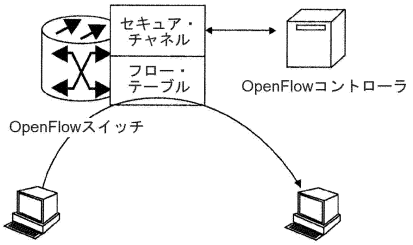
【００６５】

図１４は、分割アーキテクチャ・コントローラと従来型ルータとの間で交換されるハロー・メッセージを示す図である。ＳＡ－ＯＳＰＦにおいて、コントローラは、境界スイッチの代わりに第１にＯＳＰＦハロー・メッセージを構築する。これは、スイッチＩＤのフィールドで境界スイッチの最高位ＩＰアドレスを使用する。パケットは第１にスイッチＳ１に投入され、その後、出口リンク上の隣接ルータに送信される。

【００６６】

このように、分割アーキテクチャ・ネットワークにおいてＯＳＰＦを実装するための方法、システム、及び装置を説明した。上記の説明は例示的であり、制約的ではないことが意図されることを理解されよう。当業者であれば、上記の説明を読み、これを理解することで、多くの他の実施形態が明らかとなろう。したがって本発明の範囲は、添付の特許請求の範囲、並びにこうした特許請求の範囲が権利を認められる等価物の全範囲を参照しながら、決定されるものとする。

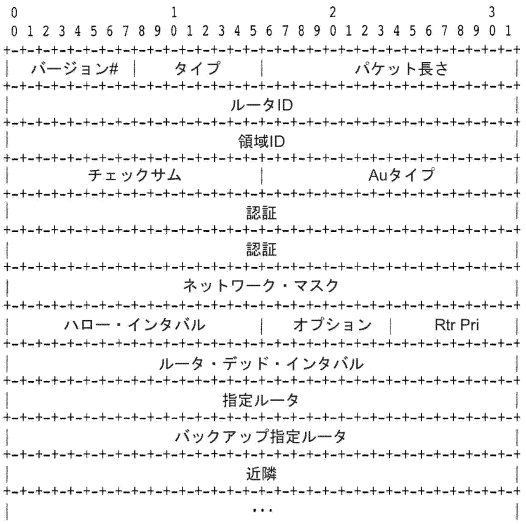
【図１】



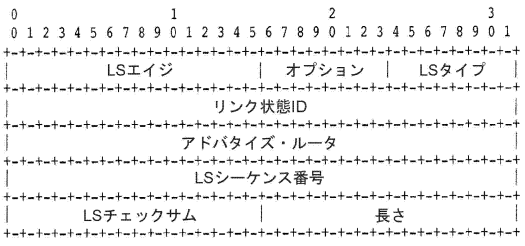
（従来技術）

【図２】

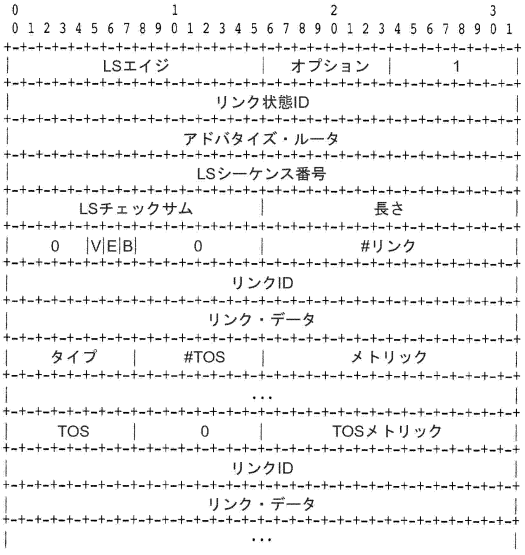
【図３】



【図４】



【図５】



【図６】

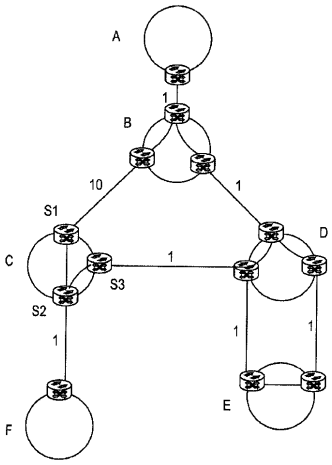


FIG. 6

【図７】

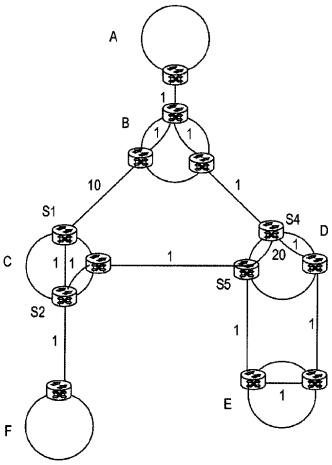
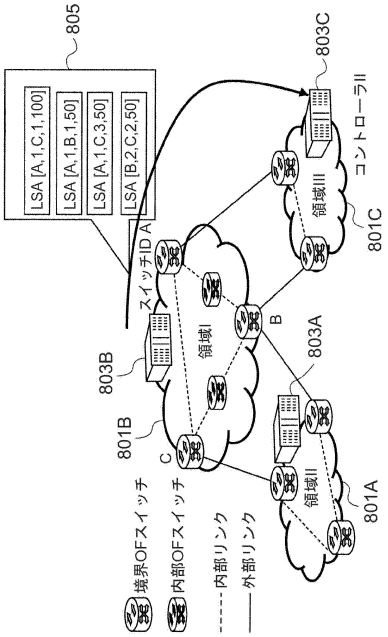
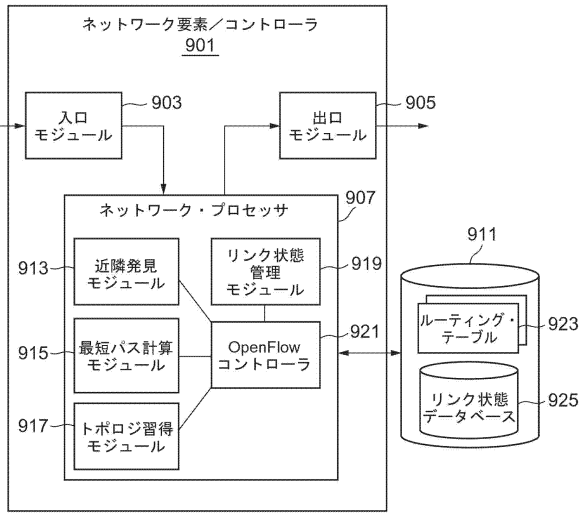


FIG. 7

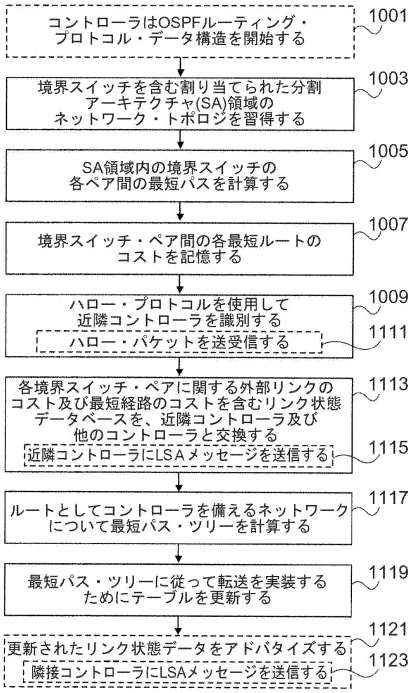
【図８】



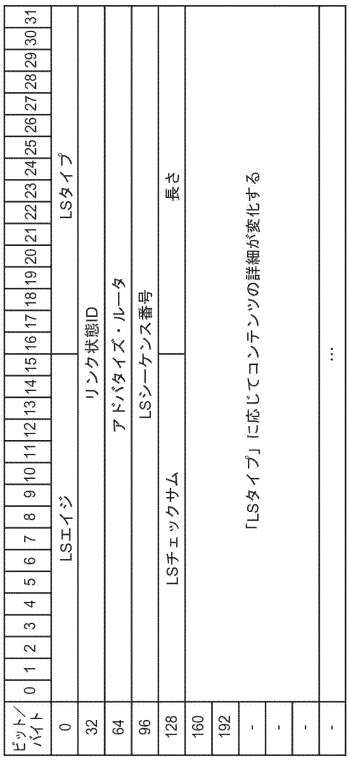
【図９】



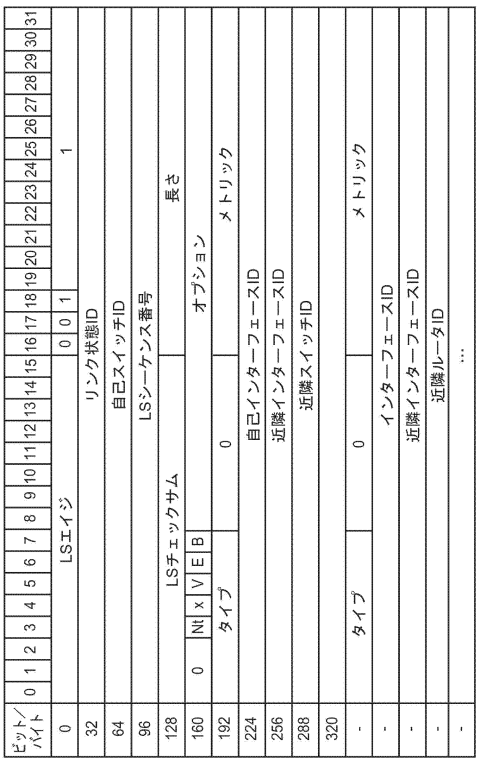
【図１０】



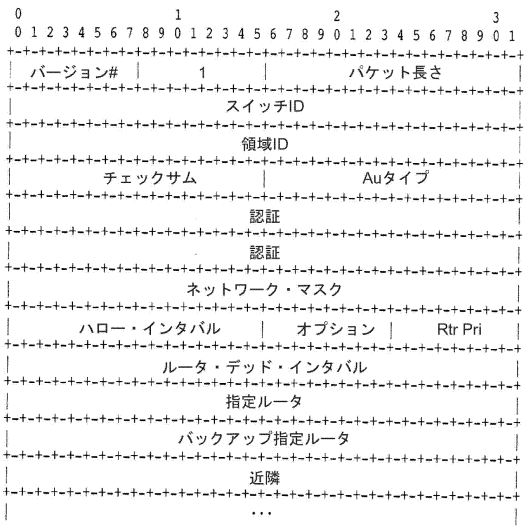
【図１１】



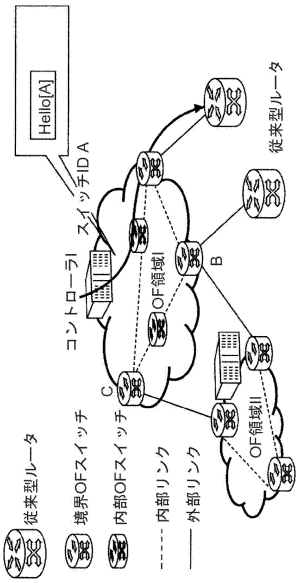
【図１２】



【図１３】



【図１４】


Hello[A]
スイッチID A
コントローラI
OF領域I
OF領域II
A
B
C
従来型ルータ
境界OFスイッチ
内部OFスイッチ
内部リンク
外部リンク
従来型ルータ

フロントページの続き

(72)発明者　ベヘシュチーサバレ，ネダ
アメリカ合衆国，カリフォルニア州　９５１３４，サン　ノゼ，パルミラ　ドライブ　３５００，ユニット　１０２６
(72)発明者　チャン，イン
アメリカ合衆国，カリフォルニア州　９４５５５，フリーモント，ミッデイ　コモン　５４７４

審査官　安藤　一道

(56)参考文献　特表２０１０－５４１４２６（ＪＰ，Ａ）
米国特許出願公開第２００９／０１３８５７７（ＵＳ，Ａ１）
欧州特許出願公開第０２３５５４２３（ＥＰ，Ａ１）
特開２０１０－１９９８８２（ＪＰ，Ａ）
特表２０１１－５２４７２８（ＪＰ，Ａ）
国際公開第２０１０／０９６５５２（ＷＯ，Ａ１）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｈ０４Ｌ　１２／７１５
Ｈ０４Ｌ　１２／７１７
Ｈ０４Ｌ　１２／７２５