(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第4680919号
(P4680919)
(45)発行日　平成23年5月11日(2011.5.11)　(24)登録日　平成23年2月10日(2011.2.10)

(51)Int.Cl.　　　F I
H04L 12/56 (2006.01)　　H04L　12/56　100A

請求項の数 16　(全 20 頁)

(21)出願番号	特願2006-535387 (P2006-535387)
(86)(22)出願日	平成16年10月15日 (2004.10.15)
(65)公表番号	特表2007-509536 (P2007-509536A)
(43)公表日	平成19年4月12日 (2007.4.12)
(86)国際出願番号	PCT/US2004/034255
(87)国際公開番号	W02005/039129
(87)国際公開日	平成17年4月28日 (2005.4.28)
審査請求日	平成19年9月27日 (2007.9.27)
(31)優先権主張番号	10/687,955
(32)優先日	平成15年10月17日 (2003.10.17)
(33)優先権主張国	米国 (US)

(73)特許権者　506128075
アイピー　インフュージョン　インコーポレイテッド
アメリカ合衆国、カルフォルニア州、サニーベール、　イースト　アーケスアベニュー、１１８８番、９４０８５
(74)代理人　100106002
弁理士　正林　真之
(72)発明者　メイ　ロバート　アルヴィン
カナダ国　ブリティッシュコロンビア　バンクーバー　ヘムロック　アパートメント　＃４　２８４９

審査官　安藤　一道

最終頁に続く

(54)【発明の名称】ネットワークノードクラスタのための冗長なルーティング機能

(57)【特許請求の範囲】
【請求項１】
相互接続ネットワークと接続し、ルーティング機能を備えたクラスタリング環境におけるアクティブなルーティングコンポーネントの障害を処理する方法であって、該アクティブなルーティングコンポーネントがネットワークデバイスのクラスタ内の第１のネットワークデバイスに存在しており、
前記方法は、
前記アクティブなルーティングコンポーネントが、前記クラスタリング環境のクラスタのアドレスで接続された前記相互接続ネットワークから受信したルーティングデータに基づき、変更されたルーティングに関連するメトリックを計算したルーティングテーブルを作成するステップと、
前記アクティブなルーティングコンポーネントが、作成した前記ルーティングテーブルを、前記クラスタ内の第２のネットワークデバイスに存在する、アクティブでないルーティングコンポーネントであるスタンドバイルーティングコンポーネントに伝達するステップと、
前記スタンドバイルーティングコンポーネントが、伝達された前記ルーティングテーブルを記憶するステップと、
前記アクティブなルーティングコンポーネントが、当該ルーティングコンポーネントの動作状態に関する状態情報を、前記スタンドバイルーティングコンポーネントに伝達するステップと、

前記スタンドバイルーティングコンポーネントが、伝達された前記状態情報を記憶するステップと、

前記障害が予定されたものである場合には、前記アクティブなルーティングコンポーネントが、前記障害が発生する前に、前記予定された障害を近隣のルーティングコンポーネントに知らせるために、前記近隣のルーティングコンポーネントに対して特別なメッセージを送信するステップと、

前記スタンドバイルーティングコンポーネントが、前記アクティブなルーティングコンポーネントの障害を検出時に、記憶した前記ルーティングテーブル及び前記状態情報を使用して前記アクティブなルーティングコンポーネントとして動作するステップと、

を有する方法。

【請求項２】

前記ルーティングデータがルート情報ベース（ＲＩＢ）を含む請求項１に記載の方法。

【請求項３】

前記ルーティングデータが転送情報ベース（ＦＩＢ）を含む請求項１に記載の方法。

【請求項４】

前記状態情報が動的な構成データを含む請求項１に記載の方法。

【請求項５】

前記状態情報が静的な構成データを含む請求項１に記載の方法。

【請求項６】

相互接続ネットワークと接続し、ルーティング機能を備えたクラスタリング環境におけるアクティブなルーティングコンポーネントの障害を処理する方法であって、該アクティブなルーティングコンポーネントがネットワークデバイスのクラスタ内の第１のネットワークデバイスに存在しており、

前記方法は、

前記アクティブなルーティングコンポーネントが、前記クラスタリング環境のクラスタのアドレスで接続された前記相互接続ネットワークから受信したルーティングデータに基づき、変更されたルーティングに関連するメトリックを計算したルーティングテーブルを作成するステップと、

前記アクティブなルーティングコンポーネントが、作成した前記ルーティングテーブルを、前記クラスタ内の第２のネットワークデバイスに存在する、アクティブでないルーティングコンポーネントであるスタンドバイルーティングコンポーネントに伝達するステップと、

前記スタンドバイルーティングコンポーネントが、伝達された前記ルーティングテーブルを記憶するステップと、

前記アクティブなルーティングコンポーネントが、当該ルーティングコンポーネントの動作状態に関する状態情報を、前記スタンドバイルーティングコンポーネントに伝達するステップと、

前記スタンドバイルーティングコンポーネントが、伝達された前記状態情報を記憶するステップと、

前記スタンドバイルーティングコンポーネントが、前記アクティブなルーティングコンポーネントの障害を検出時に、記憶した前記ルーティングテーブル及び前記状態情報を使用して前記アクティブなルーティングコンポーネントとして動作するステップと、

前記スタンドバイルーティングコンポーネントが、前記アクティブなルーティングコンポーネントの障害を検出し、再起動しているときに、前記再起動が予定されたものか又は予定外のものかを決定するステップと、

前記スタンドバイルーティングコンポーネントが、前記障害が予定外のものと決定した場合には、前記予定外の障害を近隣のルーティングコンポーネントに知らせるために、前記近隣のルーティングコンポーネントに対して特別なメッセージを送信するステップと、

を有する方法。

【請求項７】

相互接続ネットワークにルーティング機能を提供する、クラスタリング環境におけるアクティブなルーティングコンポーネントの障害を処理するルーティングコンポーネントであって、
前記ルーティングコンポーネントのうち前記アクティブなルーティングコンポーネントは、
前記クラスタリング環境のクラスタのアドレスで接続された前記相互接続ネットワークからルーティングデータを受信し、受信したルーティングデータから、変更されたルーティングに関連するメトリックを計算したルーティングテーブルを作成する動的ルーティングモジュールと、
前記動的ルーティングモジュールの動作状態に関する状態情報を記憶する構成マネージャモジュールと、
を有し、
前記障害が予定されたものである場合には、前記障害が発生する前に、前記予定された障害を近隣のルーティングコンポーネントに知らせるために、前記近隣のルーティングコンポーネントに対して特別なメッセージを送信し、
前記ルーティングコンポーネントのうち前記アクティブなルーティングコンポーネントではないスタンドバイルーティングコンポーネントは、
前記アクティブなルーティングコンポーネントから転送された前記ルーティングテーブルと前記状態情報とを記憶し、
前記アクティブなルーティングコンポーネントの障害を検出時に、記憶した前記ルーティングテーブル及び前記状態情報を使用して前記アクティブなルーティングコンポーネントとして動作する、
ルーティングコンポーネント。
【請求項８】
前記ルーティングコンポーネントは、ネットワーク対応デバイスのクラスタのために情報のルーティングを行う請求項７に記載のルーティングコンポーネント。
【請求項９】
前記動的ルーティングモジュールは、ＯＳＰＦルーティングプロトコルを実装している請求項７に記載のルーティングコンポーネント。
【請求項１０】
相互接続ネットワークにルーティング機能を提供する、クラスタリング環境におけるアクティブなルーティングコンポーネントの障害を処理する装置であって、該アクティブなルーティングコンポーネントがネットワークデバイスのクラスタ内の第１のネットワークデバイスに存在しており、
前記装置が、
前記アクティブなルーティングコンポーネントが、前記クラスタリング環境のクラスタのアドレスで接続された前記相互接続ネットワークからルーティングデータを受信し、受信したルーティングデータから、変更されたルーティングに関連するメトリックを計算したルーティングテーブルを作成する手段と、
前記アクティブなルーティングコンポーネントが、作成した前記ルーティングテーブルを、前記クラスタ内の第２のネットワークデバイスに存在するスタンドバイルーティングコンポーネントに伝達し、記憶させる手段と、
前記アクティブなルーティングコンポーネントが、当該ルーティングコンポーネントの動作状態に関する状態情報を、前記スタンドバイルーティングコンポーネントに伝達し、記憶させる手段と、
前記障害が予定されたものである場合には、前記アクティブなルーティングコンポーネントが、前記障害が発生する前に、前記予定された障害を近隣のルーティングコンポーネントに知らせるために、前記近隣のルーティングコンポーネントに対して特別なメッセージを送信する手段と、
前記スタンドバイルーティングコンポーネントが、前記アクティブなルーティングコン

ポーネントの障害を検出時に、記憶した前記ルーティングテーブル及び前記状態情報を使用して前記相互接続ネットワークに前記クラスタのアドレスでルーティングを行う手段と、を有する装置。

【請求項１１】

前記ルーティングデータがルート情報ベース（ＲＩＢ）を含む請求項１０に記載の装置。

【請求項１２】

前記ルーティングデータが転送情報ベース（ＦＩＢ）を含む請求項１０に記載の装置。

【請求項１３】

前記状態情報が動的な構成データを含む請求項１０に記載の装置。

【請求項１４】

前記状態情報が静的な構成データを含む請求項１０に記載の装置。

【請求項１５】

相互接続ネットワークにルーティング機能を提供する、クラスタリング環境におけるアクティブなルーティングコンポーネントの障害を処理する装置であって、該アクティブなルーティングコンポーネントがネットワークデバイスのクラスタ内の第１のネットワークデバイスに存在しており、

前記装置が、

前記アクティブなルーティングコンポーネントが、前記クラスタリング環境のクラスタのアドレスで接続された前記相互接続ネットワークからルーティングデータを受信し、受信したルーティングデータから、変更されたルーティングに関連するメトリックを計算したルーティングテーブルを作成する手段と、

前記アクティブなルーティングコンポーネントが、作成した前記ルーティングテーブルを、前記クラスタ内の第２のネットワークデバイスに存在するスタンドバイルーティングコンポーネントに伝達し、記憶させる手段と、

前記アクティブなルーティングコンポーネントが、当該ルーティングコンポーネントの動作状態に関する状態情報を、前記スタンドバイルーティングコンポーネントに伝達し、記憶させる手段と、

前記スタンドバイルーティングコンポーネントが、前記アクティブなルーティングコンポーネントの障害を検出時に、記憶した前記ルーティングテーブル及び前記状態情報を使用して前記相互接続ネットワークに前記クラスタのアドレスでルーティングを行う手段と、

前記スタンドバイルーティングコンポーネントが、前記アクティブなルーティングコンポーネントの障害を検出し、再起動しているときに、前記再起動が予定されたものか又は予定外のものかを決定する手段と、

前記スタンドバイルーティングコンポーネントが、前記障害が予定外のものと決定した場合には、前記予定外の障害を近隣のルーティングコンポーネントに知らせるために、前記近隣のルーティングコンポーネントに対して特別なメッセージを送信する手段と、を有する装置。

【請求項１６】

相互接続ネットワークにルーティング機能を提供する、クラスタリング環境におけるネットワークデバイスのクラスタ内の第１のネットワークデバイスに存在する、アクティブなルーティングコンポーネントの障害を処理するための命令を機械に実行させるために、前記機械によって読み込み可能な前記命令を記憶するプログラム記憶デバイスであって、前記命令は、

前記アクティブなルーティングコンポーネントが、前記クラスタリング環境のクラスタのアドレスで接続された前記相互接続ネットワークからルーティングデータを受信し、受信したルーティングデータから、変更されたルーティングに関連するメトリックを計算したルーティングテーブルを作成するステップと、

前記アクティブなルーティングコンポーネントが、作成した前記ルーティングテーブル

を、前記クラスタ内の第２のネットワークデバイスに存在するスタンドバイルーティングコンポーネントに伝達し、記憶させるステップと、

前記アクティブなルーティングコンポーネントが、当該ルーティングコンポーネントの動作状態に関する状態情報を、前記スタンドバイルーティングコンポーネントに伝達し、記憶させるステップと、

前記障害が予定されたものである場合には、前記アクティブなルーティングコンポーネントが、前記障害が発生する前に、前記予定された障害を近隣のルーティングコンポーネントに知らせるために、前記近隣のルーティングコンポーネントに対して特別なメッセージを送信するステップと、

前記スタンドバイルーティングコンポーネントが、前記アクティブなルーティングコンポーネントの障害を検出時に、記憶した前記ルーティングテーブル及び前記状態情報を使用して前記相互接続ネットワークに前記クラスタのアドレスでルーティングを行うステップと、を有する、プログラム記憶デバイス。

【発明の詳細な説明】

【技術分野】

【０００１】

本出願は、２００３年１０月１７日に出願された同時係属中の、ロバート・メイによる米国特許出願第１０／６８７，９５５号「ネットワークノードのために冗長なルーティング機能を提供するシステムおよび方法（ＲＥＤＵＮＤＡＮＴ　ＲＯＵＴＩＮＧ　ＣＡＰＡＢＩＬＩＴＩＥＳ　ＦＯＲ　Ａ　ＮＥＴＷＯＲＫ　ＮＯＤＥ　ＣＬＵＳＴＥＲ）」の一部継続出願である。

【０００２】

本発明は、ネットワーク技術を対象としたものである。より詳細には、本発明は、論理空間内に存在する一群のコンピューティング装置に対して、バックアップのルーティングサービスを提供することを対象としたものである。

【背景技術】

【０００３】

高可用性を実現することを目的として、複数の独立したデバイスが並列に動作するように、クラスタリング環境では、多くのサービス又はアプリケーションが動作する。これにより、あるデバイスで障害が発生した場合でも、この障害の発生したデバイスの責務を引き継ぐ二次的なサービスが存在できる。図１は、クラスタの一例を示す図である。図からわかるように、２つのクラスタのメンバ（デバイス１００，１０２）が存在し、２つのネットワーク（ネットワーク１０４，１０６）がクラスタに接続されている。尚、実際には、クラスタのメンバとネットワークの数は任意でよい。

【０００４】

クラスタリング環境は多様な方法で構成することができ、これには、ネットワーク又はアプリケーションの要件に応じて、アクティブ／スタンドバイ、ロードシェアリング又はロードバランシングなどの各トポロジがある。クラスタ内のデバイスは何らかの内部メカニズムを用いて接続されており、クラスタのメンバ間で通信が可能である。しかし、そうでない場合であっても、クラスタエンティティはメンバの障害を認識する能力をもち、適切な処置を取ることができる。

【０００５】

論理的には、クラスタは外の世界（つまり、ネットワークの場合には、付設のネットワークを含む。）からみると単一のエンティティである。近隣のデバイスには、単一のエンティティ（クラスタ）しか「みえず」、それらはこの単一のエンティティと通信している。ＩＰアドレスのような特性は、クラスタを構成している個々の（障害の発生した）デバイスでなく、クラスタに帰属しているため、近隣のデバイスがこの障害に気付かないようにすることができる。

【発明の開示】

【発明が解決しようとする課題】

【０００６】
　クラスタリングを使用する多くの種類のアプリケーションおよびサービスは、そのネットワーク内のルーティングをも必要とする。この結果、クラスタ自体にルーティング機能を追加する必要があり、これにより、アプリケーション又はサービスが、ネットワーク内で適切に動作するために必要な情報（例えば、経路）をもつことになる。
【０００７】
　ルーティングコンポーネントが近隣のルーティングコンポーネントと対話する場合に、動的ルーティングが行われ、これにより各ルーティングコンポーネントが現在接続されているネットワークで、互いに情報を通知し合う。ルーティングコンポーネントは、ルーティング機能のインスタンスを生成するアプリケーション、すなわちルーティングデーモンによって実行中のルーティングプロトコルを使用して通信しなければならない。静的プロトコルとは対照的に、ルーティングテーブルに置かれる情報は、時間の経過に伴ってシステムでのルートが変わると、ルーティングデーモンによって動的に追加され、また動的に削除される。さらに、時間の経過に伴って、経路情報に他の変更が行われることもある。例えば、遅延、ルートの追加や削除、ネットワークの到達可能性（ｎｅｔｗｏｒｋ　ｒｅａｃｈａｂｉｌｉｔｙ）の問題のような、ネットワーク状態の変化に起因して、ルートプリファランス（選好）が変化することがある。
【０００８】
　ＯＳＰＦ（Ｏｐｅｎ　Ｓｈｏｒｔｅｓｔ　Ｐａｔｈ　Ｆｉｒｓｔ）は、ルーティングコンポーネントで動的ルーティングを実装するリンクステート型プロトコルである。リンクステート型プロトコルでは、各ルーティングコンポーネントが、近隣の要素のそれぞれに対してリンクステートを積極的に調べ、この情報を他の近隣の要素に送信する。このような処理は、ネットワーク内のノードの全ルーティングコンポーネントについて繰り返される。
【０００９】
　各ルーティングコンポーネントは、このリンクステート情報を取得して、完全なルーティングテーブルを構築する。この方法は、特にネットワーク内のリンクが変更された場合において、動的ルーティングシステムを迅速に導入するために使用することができる。
【００１０】
　クラスタリング環境は、近隣のデバイスと通信するためにクラスタが使用するルーティングおよび／又はシグナリングプロトコルに若干の制限を課している。第一に、プロトコルは、クラスタアドレッシング方式を使用して、付設されたネットワークと通信しなければならない。つまり、クラスタを構成している個々のデバイスに割り当てられたプライベートアドレスを、クラスタの外で共有することは許されていない。そして、第二に、近隣のデバイスは単一のエンティティ（クラスタ）しか認識していないため、クラスタ内の１つのメンバのみが、どの時点においても（クラスタアドレスを使用して）近接のデバイスとのルート交換（経路情報交換）を行うことができる。複数のデバイスが同じアドレスを使用して外部と（ｅｘｔｅｒｎａｌｌｙ）通信しようとした場合、ネットワーク上の問題が発生する。
【００１１】
　クラスタリング環境のために提唱されている解決策の１つに、プロトコル同期を使用して、アクティブなデバイスにおけるルーティングプロトコルからのデータ構造および全ての内部データを、バックアップデバイスと同期させる方法がある。これは、障害が起きている間、バックアップルーティングプロトコルがオンライン化され、何も問題が発生しなかったかのように、近隣のデバイスと通信を開始できるという発想である。この解決策の唯一の本当の利点は、あらゆる面でプライマリデバイスからバックアップデバイスにミラーリングすることによって従来の高可用性（ｈｉｇｈ　ａｖａｉｌａｂｉｌｉｔｙ：ＨＡ）が実現されることにある。このため、従来のＨＡには詳しいが、ルーティングには詳しくないユーザにとっては、この解決策が妥当であると考えるかもしれない。しかし、これには、非常に複雑で、問題があり、予測不可能な解決策であることが不都合とされ、クラ

スタメンバおよび内部クラスタネットワークに大きな影響を及ぼす場合がある。ルーティング及びシグナリングプロトコルは、このような方法で実行するように設計されていなかったため、この設計の実現可能性は疑わしい。しかし、この解説策においてより重要なのは、近隣のルーティングデバイスがアクティブなルーティングデバイスの障害を検出し、それから新しい情報によって自分自身のルーティングテーブルを作り直すことであり、これはシームレスな、つまりスムーズに繋がる移行とは言いがたい。即ち、大規模なネットワークでは、近隣のデバイスの数は膨大であり、それらのデバイスのルーティングテーブルのサイズは非常に大きく、このためフェールオーバ（障害迂回）中に、著しい負担がネットワークにかかってしまうことになる。

【００１２】

また、提唱されている別の解決策として、等コストロードバランシングをサポートし得るクラスタに対して、高性能のルータを導入する方法がある。この新しいクラスタルータ（ｃｌｕｓｔｅｒ　ｒｏｕｔｅｒ：ＣＲ）は、クラスタアドレスに代わって外部ネットワークデバイスとの全ルーティングの通信を行う役割を果たす。各クラスタメンバは、ＣＲとのルート交換を容易にするために、標準のＯＳＰＦを実行している。ＣＲは、全クラスタメンバに亘って等コストロードバランシングを行う。しかし、この解決策はかなりのコストを要し、非常に複雑である。さらに、ＣＲは、ネットワークの機能にリスクを負わす、１つの障害点となってしまう。

【００１３】

そこで、クラスタリング環境のルーティング機能を効率的かつ効果的な方法で提供する解決策が求められている。

【課題を解決するための手段】

【００１４】

本発明を簡潔に説明すると、クラスタリング環境においてルーティング機能を提供するために、ルーティング及びシグナリングプロトコルのグレースフルかつヒットレスの再起動機能が、クラスタメンバ間の同期と共に使用される。通常動作中には、アクティブなクラスタメンバがルーティングプロトコルを動作させ、クラスタのアドレスを使用して近隣のコンポーネントと通信する。アクティブなメンバが送信先までのルートを学習すると、ルーティングデータが、内部通信メカニズムを介して、スタンドバイ（待機）クラスタメンバに伝えられる。そして、ルーティングコンポーネントの構成情報も、スタンドバイクラスタメンバに伝えられる。アクティブなルーティングコンポーネントに障害が発生すると、前記クラスタの外に存在する近隣のルーティングコンポーネントがこの障害に基づいてネットワークトポロジを再計算しないように、スタンドバイルーティングコンポーネントが起動される。本願明細書においてヒットレスの再起動は、このようなスタンドバイルーティングコンポーネントの起動を指し、ネットワークルーティングプロトコルは、ヒットレス再起動のためにグレースフル再起動信号の送信を含むことができる。これにより、近隣のコンポーネントに影響を及ぼさずに、かつシステムに過度の負荷をかけることなく、フェールオーバ（障害迂回）が実現される。

【発明を実施するための最良の形態】

【００１５】

この明細書に組み込まれ、その一部を構成している添付図面は、本発明の１つ以上の実施形態を示しており、その詳細な説明と共に、本発明の原理および実装形態を説明する役目を果たしている。

【００１６】

本発明の実施形態について、コンピュータ、サーバおよびソフトウェアのシステムに即してここに記載する。尚、当業者には理解されるように、本発明にかかる以下の詳細な説明が、発明の実例に過ぎず、当該説明はいかなる限定をも意図するものではない。また、本発明の他の実施形態は、本開示の利益を受けるこのような当業者にとって容易に明らかとなる。以降、添付図面に図示されるように、本発明の各種実装の細部に対して符号を付している。そして、同一又は同様の部分を示すために、図面ならびに下記の詳細な説明を

通して同じ参照符号を用いることにする。

【００１７】

尚、明確化のために、本明細書に記載する実装形態について通常の特徴全てを、図示しかつ記載してはいない。当然、実際の実装の開発においては、アプリケーションに関わる制約およびビジネスに関わる制約に適合させるなど、開発者の具体的な目的を達成するために、実装に固有の判断が数多く必要とされ、これは実装および開発者ごとに変わるということが理解される。さらに、この種の開発作業は複雑かつ時間がかかるものであるが、本開示の利益を受ける当業者にとっては日常的な技術的作業であるということが理解されよう。

【００１８】

本発明によれば、コンポーネント、処理ステップおよび／又はデータ構造は、様々なタイプのオペレーティングシステム、コンピューティングプラットホーム、コンピュータプログラムおよび／又は汎用機を使用して実装することができる。さらに、当業者には理解されるように、ハードワイヤー式の、配線により接続されるデバイスや、現場でプログラム可能なゲートアレイ（ＦＰＧＡ）、特定用途向け集積回路（ＡＳＩＣ）のような、汎用性の低いデバイスであっても、本明細書に開示した発明的概念の範囲および趣旨から逸脱することなく使用できる。

【００１９】

本発明では、クラスタリング環境においてルーティング機能を提供するために、ルーティング及びシグナリングプロトコルにおいて一般的なグレースフルかつヒットレスの再起動機能を、クラスタメンバ間の同期と共に使用する。通常動作中に、アクティブなクラスタメンバがルーティングプロトコルを動作させ、クラスタのアドレスを使用して近隣のデバイスと通信する。アクティブなメンバが送信先までのルートを学習すると、このルートが、内部通信メカニズムを介して、スタンドバイクラスタメンバに伝達される。ルートの伝達は、個々のルーティングプロトコルの外部とされる、一元化されたプロセス又はタスクとして実行される。次に、この外部のプロセスは、一般に、ルート情報ベース（Ｒｏｕｔｅ　Ｉｎｆｏｒｍａｔｉｏｎ　Ｂａｓｅ：ＲＩＢ）および／又は転送情報ベース（Ｆｏｒｗａｒｄｉｎｇ　Ｉｎｆｏｒｍａｔｉｏｎ　Ｂａｓｅ：ＦＩＢ）の形式の全ての経路情報の他、ルーティングプロトコルのグレースフルかつヒットレスの再起動に必要とされる他の任意の情報を、クラスタメンバに伝達する役目をもつ。

【００２０】

この解決策によって、全ての構成コマンドおよびデータが全てのクラスタメンバに伝達され、全メンバは同じ方法で外部デバイスと通信できることが保証される。アクティブなメンバに障害が生じている間、グレースフルかつヒットレスの再起動機能を使用して、ルーティングプロトコルがスタンドバイデバイスで開始される。これらの機能によって、ルータは、ネットワークトポロジに影響を与えずに再起動することができ、近隣のデバイスはクラスタへのパケットの転送を続けることができる。スタンドバイメンバは、既に見つけている全ルートを学習しているため、プロトコルの再起動中もパケットの転送を継続することができる。

【００２１】

一般には、グレースフルかつヒットレスの再起動機能をサポートしているか、あるいはプロトコルの固有の機能によりこのような拡張を必要としないルーティング及びシグナリングプロトコルであれば、どのようなものでも使用することができる。しかし、本発明の一実施形態では、ＯＳＰＦプロトコルが使用される。尚、本明細書では、ここで提供される解決策を、「クラスタルーティング拡張（Ｃｌｕｓｔｅｒ　Ｒｏｕｔｉｎｇ　Ｅｘｔｅｎｓｉｏｎ：ＣＲＸ）」と呼ぶことがある。

【００２２】

正常動作中において、ＯＳＰＦは、近隣のデバイスと通信を行って、ネットワークトポロジを学習する。そして、ＣＲＸは、クラスタのアクティブなメンバからスタンドバイメンバに対して、動的なデータを同期させる。この動作により、カーネル転送テーブルがク

ラスタメンバ全体に亘って同期化されることが保証される。フェールオーバ時には、スタンドバイデバイスが、パケットを転送するのに必要な全ルートを有している。これを通して、ＣＲＸは、アクティブなＮＳＭからスタンドバイＮＳＭに対して、全ＲＩＢルート（およびＦＩＢルート）を同期させる。次に、ＣＲＸは、ネットワークサービスモジュール（Ｎｅｔｗｏｒｋ　Ｓｅｒｖｉｃｅ　Ｍｏｄｕｌｅ：ＮＳＭ）を介して、アクティブなプロトコルのグレースフルかつヒットレスの再起動に必要な全データを同期させる。ＣＲＸは、アクティブなメンバからスタンドバイメンバに対して、静的な構成の変更及び動的な構成の変更を同期させ、これにより、スタンドバイデバイスは、障害の発生したデバイスと同一の構成情報を有することが保証される。

【００２３】

予想外の障害は、ネットワークに影響を与える最も深刻なタイプの障害であり、企業がクラスタリングを採用する主な理由となっている。フェールオーバの間、近隣のルータは、クラスタにパケットを転送し続けなければならず、新たにアクティブとなったメンバは、障害の発生したノードによってそれ以前には伝達可能であった全ての送信先に対して、これらのパケットを転送できなければならない。これらの要求については、ＣＲＸのＲＩＢ同期機能を使用して、全ての経路情報を同期させると共に、トポロジの再計算を回避する、プロトコルのグレースフルかつヒットレスの再起動機能によって満たすことができる。障害が予定されていたか又は予想外であったかに関わらず、フェールオーバ時には、スタンドバイＯＳＰＦが立ち上がり、その近隣の要素とともにグレースフルな再起動期間を経る。近隣のルータは、再起動中のルータが完全に隣接しているかのように、それら自身のリンクステートアドバタイズメント（Ｌｉｎｋ－Ｓｔａｔｅ　Ａｄｖｅｒｔｉｓｅｍｅｎｔ）で再起動中のルータに通報し続け、これにより、ＯＳＰＦが完全に再起動されるまでの間、ネットワークの途絶が生じることはない。スタンドバイ要素はプライマリルータから同期化されたＲＩＢとＦＩＢを持っているため、全てのパケットが、途絶えることなくクラスタを介して流れ続ける。

【００２４】

事前に計画されていた不稼働時間（停止時間）において、クラスタ内のアクティブなルータは、この移行の発生する時間を前もって知っているため、近隣のルータをこの事象（イベント）に備えて準備させることができる。このときのＯＳＰＦルータの基本動作は、シャットダウン前に、特別なメッセージを送信して、この不稼動時間を近隣の要素に通知することである。近隣の要素は、ＯＳＰＦルータがシャットダウンして、その後復旧することを知っているため、ネットワークトポロジの再計算を実行して、再起動中のルータを除外する必要がなくなる。

【００２５】

また、計画外の障害が起きている間、アクティブなルータが、近隣の要素に再起動の事象を通知することができないことは明らかである。しかし、プロトコルの再起動しているときに、アクティブなルータは、グレースフルスタートアップであること、およびそれが予定された再起動であったか否かについて、その両方を決定するのに十分な情報をＮＳＭから取得することができる。障害が計画外とされる場合には、再起動中のルータが、近隣の要素に対して再起動の事象を通知し、これにより、近隣の要素はネットワークトポロジの再計算を実行することがなくなる。

【００２６】

ＣＲＸについては、３つの論理的なコンポーネントを有するものと考えることができる。第一に、アクティブなメンバとスタンドバイメンバとの間でＲＩＢを同期させるために動的同期コンポーネントを使用することができ、これにより両メンバ間で、暗黙裡にＦＩＢも同期化される。障害の間、新たにアクティブとなったメンバは、タイムアウト期間中、全てのＲＩＢ／ＦＩＢルートを保持しうる。このタイムアウトについては、設定可能なタイマーであっても、あるいはルーティングプロトコルの再起動機能に組み込まれているものであってもよい。また、このコンポーネントによって、プロトコルのグレースフルかつヒットレスの再起動動作にかかる訂正動作に必要な、任意のデータを同期させるように

してもよい。
【００２７】
　第二には、全ての動的構成情報および静的構成情報を、アクティブなメンバからスタンドバイメンバへと同期させるために、構成同期コンポーネントを使用することができる。アクティブなメンバにて実行されるコマンドは、スタンドバイメンバでも並行して実行される。両メンバにおけるアクティブなコンポーネントに対して、これは直接的である。しかし、フェールオーバ中にスタンドバイメンバで開始されるプロトコルコンポーネントについては、デバイスが、このようなプロトコルのそれぞれについて、アクティブなメンバに関する最新の構成情報を維持しなければならない。フェールオーバ時に、開始されるプロトコルは、障害の発生したデバイスにおいてそれ以前アクティブであったプロトコルと同じ構成とされる。
【００２８】
　第三に、動作及び制御コンポーネントは、ＣＲＸ設計の動作のためのタイミングおよびシーケンスを指定することができ、これには、どのクラスタメンバでどのプロトコルがアクティブになっているか、その開始順序などが含まれる。このコンポーネントは、グレースフルかつヒットレスの再起動の要件など、プロトコルがどのように動作すべきかをも指定する。また、このコンポーネントは、本発明と、それが動作している個々のクラスタメンバとの間の統合点（Ｉｎｔｅｇｒａｔｉｏｎ　Ｐｏｉｎｔ）を形成しうる。
【００２９】
　図２は、本発明の実施形態による、アクティブなルーティングコンポーネントの障害を処理する方法を示すフローチャート図である。本方法における各動作は、ソフトウェア、ハードウェア又はこれらの任意の組み合わせにおいて実行することができる。アクティブなルーティングコンポーネントは、ネットワークデバイスのクラスタにおいて、第１のネットワークデバイスに存在しうる。２００において、アクティブなルーティングコンポーネントからのルーティングデータは、クラスタにある第２のネットワークデバイスに存在するスタンドバイルーティングコンポーネントと同期する。このデータには、ＲＩＢおよび／又はＦＩＢが含まれる。そして、２０２において、グレースフルかつヒットレスの再起動に必要とされる、アクティブなルーティングコンポーネントからのデータが、スタンドバイルーティングコンポーネントと同期する。さらに２０４において、アクティブなルーティングコンポーネントからの構成データが、スタンドバイルーティングコンポーネントと同期する。これには、動的な構成データと静的な構成データが含まれる。そして、２０６において、アクティブなルーティングコンポーネントで障害が発生すると、クラスタの外に存在する近隣のルーティングコンポーネントがこの障害に基づいてネットワークトポロジを再計算しないように、スタンドバイルーティングコンポーネントが起動される。このとき、グレースフルかつヒットレスの再起動が実行される。また、障害を知らせる特別なメッセージを近隣のルーティングコンポーネントに送信してもよい。尚、障害が予想されるものである場合、この特別なメッセージの送信は、障害が発生する前に行われてもよい点に留意する必要がある。
【００３０】
　図３は、本発明の一実施形態に従う場合において、ルーティングコンポーネントを用いてサービスを受ける一群のデバイスを有するネットワークの模式図である。このネットワークの各要素は、ソフトウェア、ハードウェア又はこれらの任意の組み合わせにおいて実装することができる。クラスタ３００は、幾つかのネットワーク対応デバイス３０２ａ～ｄを有している。電子回路網を通してデバイスにアクセス可能な他のデバイスは、相互接続ネットワーク３０４を介してクラスタ３００におけるデバイス３０２ａ～ｄのサービス、又はこれらに存在するデータにアクセスすることができる。代表的な例では、ＴＣＰ／ＩＰのようなネットワークプロトコルを使用して、サービスやデータへの要求が作成され、クラスタ３００に送信される。ネットワークプロトコルは、数層のレベルで発生することも勿論あり、このレベルには、アプリケーション層、プレゼンテーション層、セッション層、トランスポート層、ネットワーク層、データリンク層および物理層がある。さらに

、多くのネットワークプロトコルおよび／又はアドレッシング方式がこれらのレベルで利用可能である。そして、当業者には理解されるように、特定のネットワークプロトコル又はアドレッシングプロトコルについての上記言及に限定されることなく、本発明を実施する上で、これらのネットワークプロトコルのうちのいずれをも使用することができる。

【００３１】

各ネットワーク対応デバイス３０２ａ～ｄについては、単一のアドレスによってアクセスされる。したがって、ルーティングコンポーネント３０６は、クラスタ３００の「入口」に近いネットワークデバイス３０２ａに置かれており、これにより、クラスタ３００内の適切なデバイスに対してメッセージおよび／又はデータの適切なルーティングが行われる。また、このルーティングコンポーネントは、デバイス３０２ａ～ｄのいずれかから、相互接続ネットワーク３０４に結合されている別のデバイスに送られるメッセージおよび／又はデータについての適切なルーティングを行うことができる。

【００３２】

単一のアドレスを使用してクラスタ３００内のデバイス３００ａ～ｄのうち、任意のデバイスを識別する例では、相互接続ネットワーク３０４から入ってくるメッセージ又はデータパケットが、ルーティングコンポーネント３０６で受信される。ルーティングコンポーネント３０６は、メッセージ又はデータがある場合、その宛先がデバイス３０４ａ～ｄのいずれであるかを確定する。次に、ルーティングコンポーネント１６は、クラスタ３００内の適切なデバイスに向けてメッセージを中継するか、あるいは該メッセージを他の場所に渡す。

【００３３】

送信メッセージについては、クラスタ３００内のネットワークデバイス３０２ａ～ｄが、ルーティングコンポーネント３０４に宛ててメッセージおよび／又はデータの送信を指図する。ルーティングコンポーネント３０４は、送信メッセージおよび／又はデータを受信すると、そのメッセージに含まれる情報に基づいて、メッセージを送信するための適切なルートを決定する。

【００３４】

クラスタ３００内のデバイス３０２ａ～ｄがリソース（資源）の共有プール内に存在するか、あるいはクラスタ３００内のデバイス３０２ａ～ｄが、負荷分散がなされたリソースの組を代表するものであってもよい点に留意すべきである。いずれの場合も、ルーティングコンポーネント３０６は同じ原理で動作する。

【００３５】

また、クラスタ３００内のデバイス３０２ａ～ｄについては、どのようなネットワーク対応デバイスであってもよい点にも留意すべきである。このようなデバイスのタイプには、サーバ、ルータ、汎用コンピューティング装置、キオスクタイプ（ｋｉｏｓｋ－ｔｙｐｅ）のように駅、空港等でインターネットを利用可能なデバイス、スマートカード（登録商標）対応デバイス、無線対応デバイスなどがある。また、他の多くのネットワークデバイスが可能であり、それらが発明の形態に関連して使用できることは当業者の理解するところである。そして、図３には４台のデバイスのみを図示しているが、ネットワーク対応デバイスの台数は幾つでも構わないと解釈されるべきである。この場合も、当業者には、任意の台数のネットワークデバイスを、現時点で特許請求の範囲に記載されている発明的形態に関連して使用できることが理解される。

【００３６】

ルーティングコンポーネントの接続に関する新しい情報や、相互接続ネットワークにおける様々なポイント内、あるいは該ポイントへの及び該ポイントからのメッセージのルーティングに関連する、他のメトリック（数的指標）が利用可能である場合に、ルーティングコンポーネント３０６がそのルーティング動作を調整することで、これらの変更が許容される。該ルーティングコンポーネント３０６は、動的ルーティングモジュール３０８を有する。この動的ルーティングモジュール３０８は、相互接続ネットワーク３０４に結合されているデバイスのための適切なルーティングデータを受信する。そして、動的ルーテ

ィングモジュールは、このデータに関わる任意のエントリに関連した、あらゆるメトリックを再計算する。これらの動作によって動的ルーティングテーブルが得られるが、その際、新しいデータが受信されて、新しいメトリックが計算され、当該テーブル内のルーティングエントリが適切に更新される。

【００３７】

さらに、動的ルーティングモジュール３０８は、ルーティングコンポーネント３０６と連絡をとる新しいルーティングコンポーネント（例えば、近隣のルーティングコンポーネント）に対して応答しうる。この場合、動的ルーティングモジュール３０８は、その近隣のルーティングコンポーネントに関連するメトリックを決定し、将来使用するために、当該ルーティングコンポーネントとその関連の経路についての情報をルーティングテーブルに記述する。

【００３８】

通常、動的ルーティングモジュール３０８は設定可能であり得る。例えば、動的ルーティングモジュール３０８がＯＳＰＦパッケージのインスタンス（ｉｎｓｔａｎｔｉａｔｉｏｎ）である場合、その動作特性をコマンドラインインタフェースから定義することができる。本発明の一実施形態においては、構成コマンドが動的ルーティングモジュール３０８に送信されて、動的ルーティングモジュール３０８のネットワーク性能についてのパラメータの設定に使用される。さらに、構成コマンドを使用して動的ルーティングモジュール３０８で数多くの機能が実現され、動的ルーティングモジュール３０８の挙動が正確かつ詳細に指定される。例えば、これらのコマンドは、ＯＳＰＦエリア又はスタブエリアを作成し又は削除するために使用され、また、エリア境界でルートのサマリを作成するために使用される。また、このようなコマンドは、ＯＳＰＦエリアのパスワード保護を追加又は削除するために使用され、ＯＳＰＦインタフェースを使用可能にし又は使用不可能にするために使用される。さらには、インタフェースにメトリックを割り当てるためや、デッドインターバル（すなわち、近隣のルーティングコンポーネントが動作していないことをスイッチが宣言する前に、そのスイッチが近隣のルーティングコンポーネントからハロー（ｈｅｌｌｏ）パケットを受信するまで待機する時間）を割り当てるために上記コマンドが使用される。また、ハロータイムインターバル（すなわち、スイッチが別のハローパケットを発行するまで待機する時間）を割り当てるためや、新しく指定されたルーティングコンポーネントを決定する際に、動的ルーティングモジュール３０８がインタフェースに使用する優先レベルを指定するために上記コマンドを使用し、あるいはデータベースエントリアナウンスメント（すなわちリンクステートアナウンスメント（ｌｉｎｋ　ｓｔａｔｅ　ａｎｎｏｕｎｃｅｍｅｎｔ：ＬＳＡ））間の時間を設定するために上記コマンドを使用することができる。尚、動的ルーティングモジュール３０８を動作させているルーティングコンポーネント３０６の性能を調整するために、他の構成コマンドおよび設定を用いてもよいことは当業者の理解するところである。さらに、当業者はこの他の構成上の設定が本発明の範囲内で使用できることを理解する。

【００３９】

ルーティングコンポーネント３０６内で動作している動的ルーティングモジュール３０８と協働するのが、構成マネージャモジュール３１０である。この構成マネージャモジュール３１０は、動的ルーティングモジュール３０８の動作状態に関する状態情報を記憶する。さらに、構成マネージャモジュール３１０は、動的ルーティングモジュール３０８の構成に対する変更をも記憶しうる。このため、動的ルーティングモジュール３０８に対して構成要求が行われると、構成マネージャモジュール３１０は、この構成要求を適用した後、この要求又は動的ルーティングモジュール３０８の動作特性の表現（ｒｅｐｒｅｓｅｎｔａｔｉｏｎ）を記憶する。

【００４０】

本発明の一実施形態において、ルーティングコンポーネント３０６内で動作している動的ルーティングモジュール３０８に適用された構成設定は、構成マネージャモジュール３１０に中継される。この場合、動的ルーティングモジュール３０８のための構成情報が、

構成マネージャモジュール３１０の動作により記憶される。ある実装形態では、構成要求が構成マネージャモジュール３１０に中継され、構成マネージャモジュール３１０がこの構成要求を記憶する。この実施形態では、構成マネージャモジュール３１０が、この構成要求を動的ルーティングモジュール３０８に中継する。

【００４１】

また、本発明の別の実施形態では、構成要求を「分岐」させることができる。つまり、この代替の実施形態では、構成要求が、動的ルーティングモジュール３０８と構成マネージャモジュール３１６の両方に送信される。更に別の実施形態では、動的ルーティングモジュール３０８のメッセージング部が、構成マネージャモジュール３１０に構成要求を中継する。

【００４２】

動的ルーティングモジュール３０８が構成要求を受信すると、動的ルーティングモジュール３０８はこの構成要求を処理して、該構成要求を適用する。この構成要求の処理時に、動的ルーティングモジュール３０８は、要求された方法でその挙動又は動作特性を変更する。

【００４３】

構成要求が何らかの理由で失敗した場合に、動的ルーティングモジュール３０８はこの障害状態を構成マネージャモジュール３１０に中継する。この場合、失敗した構成要求が記憶されることはないが、その理由は、コマンドが動作中の動的ルーティングモジュール３０８において失敗したためである。又は、別の実施形態において、構成マネージャモジュール３１０が構成要求を動的ルーティングモジュール３０８に中継してもよい。構成要求が正しく適用されたことが示された場合、構成マネージャモジュール３１０は、次にこの構成要求を、動的ルーティングモジュール３０８の動作状態の表現に適用する。

【００４４】

さらに構成要求が行われると、構成マネージャモジュール３１０によって保持されている、動的ルーティングモジュール３０８の動作状態の表現が変更される。例えば、起動動作の後の僅かな時間に、動的ルーティングモジュール３０８への新しいデッドタイムが要求されるとする。この場合、構成マネージャモジュール３１０は、新しいデッドタイムに対する要求を記録する。その後に、別のデッドタイムが要求されたとすると、構成マネージャモジュール３１０によって保持されている動作状態の表現には、この新しいデッドタイムが、動的ルーティングモジュール３０８の動作特性として反映される。この種の、構成のトラッキング（追跡）は、動的ルーティングモジュール３０８の様々な動作特性について行われる。

【００４５】

動的ルーティングモジュール３０８の動作特性を記憶することについては、これを数多くの方法で行うことができる。様々な動作特性や、構成要求の特性に対する適用の状況に関するフィールドを含むファイルが保持される。又は、構成要求の記録を保持してもよく、この場合、以前に実行された構成パラメータを変更する構成要求が上書きされる。あるいは、構成要求がデータベースエントリの形で記憶されてもよい。特に、これらの方法を用いて、動的ルーティングモジュール３０８の動作状態の表現を記憶できることは、当業者の理解するところである。

【００４６】

動作中、動的ルーティングモジュール３０８は正常に動作しうる。ルーティングコンポーネント３０６は、近隣のルーティングコンポーネントの一団（ａｓｓｏｒｔｍｅｎｔ）と連絡をとる。この対話を通じて、相互接続ネットワークで利用可能な他のポイントに対するルーティングテーブルが、入手可能となる。付設されるネットワークのルーティングトポロジが変更されるか、又はネットワーク内のメトリックが変更されると、動的ルーティングモジュール３０８はその内部ルーティングデータを変更して、当該変更を反映させる。

【００４７】

スタンドバイルーティングコンポーネント３１２については、別のネットワークデバイス３０２ｂ上に存在しうる。その動作の過程において、ルーティング情報に対するあらゆる変更も、スタンドバイルーティングコンポーネント３１２に伝えられる。そして、スタンドバイルーティングコンポーネント３１２は、ルーティングコンポーネント３０６の動作に関連する、あらゆるルーティング情報を、相互接続ネットワーク３０４へのトラフィック及び該ネットワークからのトラフィックと共に、記憶し、かつ更新する。このようにして、スタンドバイルーティングコンポーネント３１２は、相互接続ネットワーク３０４に関連する、他のネットワークデバイスに対するシステムの動作において、更新されたルーティング情報を保持する。

【００４８】

スタンドバイルーティングコンポーネント３１２に関連するルーティング情報の変更については、これを様々な方法で行うことができる。本発明の一実施形態では、動的ルーティングモジュール３０８に到着するメッセージが分岐して、そのうちの一方の経路がスタンドバイルーティングコンポーネント３１２に向かうようになっている。また、本発明の別の実施形態においては、動的ルーティングモジュール３０８が、それ自身のルーティング情報の記憶部に情報を保存する間、又はその前若しくはその後に、該情報の転送を開始する。いずれの方法でも、相互接続ネットワーク３０４を介して他のデバイスと対話するのに使用される、最新又はほぼ最新のルーティング情報の、動作可能なコピーが、スタンドバイルーティングコンポーネント３１２に記憶される。このため、スタンドバイルーティングコンポーネント３１２に格納されている情報、又はスタンドバイルーティングコンポーネント３１２からアクセス可能な情報は、相互接続ネットワーク３０４を介して他のデバイスと通信するために、ルーティングコンポーネント３０６によって用いられるルーティング情報の現時点での状態を表すもの（状態の最新の表現）である。

【００４９】

最後に、ルーティングコンポーネント３０６又は動的ルーティングモジュール３０８が機能を停止した場合を想定する。通常であれば、現在機能していないルーティングコンポーネント３０６を有するデバイス３０２ａに接続している、他のデバイス３０２ｂ～ｄは、相互接続ネットワーク３０４との間の情報の送信又は受信を停止することになる。さらに、近隣のルーティングコンポーネントであって、相互接続ネットワークを介してルーティングコンポーネント３０６に接続し、かつこれと連絡をとってルーティングコンポーネント３０６との間で情報を送信し又は受信しているルーティングコンポーネントは、ルーティングコンポーネント３０６が機能しなくなったことを検出することとなる。そして通常であれば、これらの近隣のルーティングコンポーネントは、それ自身のルーティングテーブルを再作成することになる。

【００５０】

一般に、ルーティングテーブルを再作成するには、特定のコンポーネントが認識している全ルーティングコンポーネントに対するコンタクトを伴う。このことはまた、ネットワークに結合されている他のルーティングコンポーネントに対してネットワークを介して信号を送信することや、このような他のルーティングコンポーネントからのメッセージを受信することを伴う。関連する情報を受信すると、それ自身のテーブルを再作成しようとしているルーティングコンポーネントに戻される情報に基づいて、ルーティングテーブルが再作成される。このように、ルーティングコンポーネントの障害状態は、ネットワークルーティングコンポーネントに相当な労力を強いることになるが、これはネットワークルーティングコンポーネント間で情報が同期するように保証しなければならないからである。

【００５１】

さらに、ルーティングコンポーネント３０６は、オンラインに復帰すると、通常、相互接続ネットワーク３０４を介してルーティングコンポーネント３０６と結合されている他のルーティングコンポーネントとの同期処理を受ける必要がある。この場合に、近隣のルーティングコンポーネントとルーティングコンポーネント３０６は、ルーティングコンポーネント３０６の不稼動時間（ダウンタイム）と、これに続く稼動時間（アップタイム）

に起因する、ネットワークトポロジの変更を処理するために、相当量の資源の消費を強いられる可能性がある。

【００５２】

本発明の一実施形態において、スタンドバイルーティングコンポーネント３１２は、障害状態を検出して、動的ルーティングモジュール３１４を開始させる。動的ルーティングモジュール３１４のインスタンス生成時に、スタンドバイルーティングコンポーネント３１２は、動作サイクルの早い時点でルーティングコンポーネント３０６によって提供されたルーティング情報を使用し、クラスタ３００内のデバイス３０２ａ～ｄのためのバックアップルーティングコンポーネントとして機能を果たすことができる。

【００５３】

また、本発明の一実施形態では、スタンドバイルーティングコンポーネント３１２が記憶媒体を有し、そこに動的ルーティングモジュール３１４の実行可能なコピーが置かれる。ルーティングコンポーネント３０６がクラスタ３００内のデバイス３０２ａ～ｄのためにルーティング機能を果たすことができない状況に遭遇したと判断された場合に、スタンドバイルーティングコンポーネント３１２は、動的ルーティングモジュール３１４をメモリに読み込んで、動的ルーティングモジュール３１４を実行する。

【００５４】

動的ルーティングモジュール３１４が動作を開始すると、スタンドバイルーティングコンポーネント３１２は、構成情報を動的ルーティングモジュール３１４が利用できるようにするか、又は該情報を動的ルーティングモジュール３１４に適用するか、あるいはこの両方を行う必要があるかどうかを決定する。本発明の一実施形態において、構成マネージャモジュール３１６は、スタンドバイルーティングコンポーネント３１２上いつでも使用することができる。構成マネージャモジュール３１６は、元の動的ルーティングモジュール３０８が機能を停止した時点で、又はその近傍の時点で、ルーティングコンポーネント３０６内の動的ルーティングモジュール３０８が動作していたのと同じ構成でもって、動的ルーティングモジュール３１４を動作させるために必要な情報を、利用できるようになっている。

【００５５】

本発明の一実施形態では、構成マネージャモジュール３１０の記憶している情報が、構成マネージャモジュール３１６に中継される。あるいは、構成マネージャモジュール３１０の記憶している情報を、構成マネージャモジュール３１６が利用できるものとされる。

【００５６】

次に、スタンドバイルーティングコンポーネント３１２は、構成マネージャモジュール３１０と構成マネージャモジュール３１６との対話を通して、スタンドバイルーティングコンポーネント３１２にて現在動作している動的ルーティングモジュール３１４のインスタンスに、構成情報を適用する。このため、スタンドバイルーティングコンポーネント３１２で動作している動的ルーティングモジュール３１４は、ルーティングコンポーネント３０６で動作している動的ルーティングモジュール３０８がその動作を停止する前の、動的ルーティングモジュール３０８と少なくともほぼ同じ構成となる。

【００５７】

動的ルーティングモジュール３１４がスタンドバイルーティングコンポーネント３１２で動作しているため、新しい構成パラメータが、スタンドバイルーティングコンポーネント３１２の動作上で実装される。したがって、スタンドバイルーティングコンポーネント３１２に関連して動作している構成マネージャモジュール３１６は、新しい構成状態の変更を保存するために、上記と同じステップを実行する。これにより、ルーティングコンポーネント３０６が動作を再開すると、動的ルーティングモジュール３０８がルーティングコンポーネント３０６で再開される際に、動的ルーティングモジュール３１４の制御状態における、新しい構成の変更のすべてが、動的ルーティングモジュール３０８に適用される。

【００５８】

本発明の一実施形態において、ネットワークルーティングプロトコルは「ヒットレス再起動」機能を実装することができる。この概念では、ルーティングコンポーネント３０６がグレースフル再起動信号を送信する。この信号は、１つ以上の近隣の要素に中継される。そして近隣の要素は、ルーティングコンポーネント３０６があたかも連続して動作を続けていたかのように、ルーティングコンポーネント３０６との関係をアドバタイズ（通知）し続ける。つまり、近隣の要素は、ルーティングコンポーネント３０６との隣接関係（ａｄｊａｎｃｙ）の現在の同期状態の如何に関わらず、ネットワークセグメントを介してルーティングコンポーネント３０６に近接していることをリストし続ける。

【００５９】

この機能を更に活用するために、スタンドバイルーティングコンポーネント３１２は、ネットワークにある適切なルーティングコンポーネントに、「ヒットレス再起動」のメッセージを送信する。この場合に、このような他のルーティングコンポーネントは、ネットワークトポロジで使用される各種のルーティングデータベースを再作成するための情報を再計算して、これを再ブロードキャストする必要はない。同様に、スタンドバイルーティングコンポーネント３１２で動作している動的ルーティングモジュール３１４は、相互接続ネットワーク３０４でアクセス可能な残りのルーティングコンポーネントに影響を及ぼす、不必要なオーバーヘッドを生じさせずに、クラスタ３００内のデバイス３０２ａ～ｄへのトラフィックの指示及び該デバイスからのトラフィックの指示を送受する動作を開始しうる。さらに、代替ルーティングコンポーネントとしてのスタンドバイルーティングコンポーネント３１２の動作に対する影響が最小限に抑えられる。尚、前述の機能に関して記載したメカニズムは、動的ルーティングモジュール３０８の計画的な停止に関しても、また動的ルーティングモジュール３０８の計画外の障害に関しても使用できるという点に留意されたい。

【００６０】

計画的な再起動の場合、ネットワークマネージャ３１８は、ルーティングコンポーネント３０６のサービスの提供を停止させ、スタンドバイルーティングコンポーネント３１２に、これに対応してクラスタ３００内のデバイス３０２ａ～ｄに対するサービスの提供を開始させる。

【００６１】

特定の時点において、ネットワークマネージャ３１８は、ルーティングコンポーネント３０６にコマンドを発行して、動的ルーティングモジュール３０８の動作を停止させることができる。このコマンドは、該モジュールを直ちに停止させる形式のものであってもよいし、未来の所定の時点で効力を生じるものであってもよい。つまり、この所定の時点については、未来の特定の時点（すなわち２月１７日午後６時０２分）であっても、またオフセット（すなわちコマンドから２０分後）であってもよい。あるいは事象の発生時（すなわち、ネットワークデバイス内のアプリケーションでエラーが発生したとき）であってもよいし、このような事象が発生した時点からのオフセット（すなわち、ネットワークトラフィックが所定量を下回ってから３０秒後）であってもよい。これと並行して、ネットワークマネージャ３１８は、動的ルーティングモジュール３０８の動作停止に対応する時点で、代替の動的ルーティングモジュール３１４を起動するため、対応するコマンドをスタンドバイルーティングコンポーネント３１２に発行することができる。

【００６２】

また、上記対応する時点において、相互接続ネットワーク３０４にある任意の隣接ルーティングコンポーネントに対して、ヒットレス再起動が開始されてもよい。ヒットレス再起動は、動的ルーティングモジュール３０８、動的ルーティングモジュール３１４、又はネットワークマネージャ２８によって送信され又は開始される。

【００６３】

動的ルーティングモジュール３０８が動作を停止すると、スタンドバイルーティングコンポーネント３１２は、ほぼシームレスな移行で起動することができる。この場合に、ルーティングコンポーネント３０６は、特定の事象の発生時に、又は特定の時点において、

サービスを提供されるか、あるいはアップグレードされる。
【００６４】
　さらに、相互接続ネットワーク３０４に存在する他のルーティングコンポーネントからみた場合の、ルーティングコンポーネント３０６からスタンドバイルーティングコンポーネント３１２への動作移行については、クラスタ３００との間で流れるネットワークトラフィックの、ほぼシームレスな動作によって達成される。このことは、近隣のルーティングコンポーネントが、何らかの事象の発生を認識しているが、その事象が、これらのコンポーネントが現在有している情報に対するネットワークトポロジには必ずしも影響を与える訳でないと認識しているという性質によるものである。また、クラスタ３００内のデバイス３０２ａ～ｄのネットワークサービスの移行がなされるが、その際、スタンドバイルーティングコンポーネント３１２のネットワーク情報に関連する情報の再作成又は再送信のような、資源の消費を要しない。
【００６５】
　また、ネットワークマネージャ３１８から見た場合に、これを使用して、動的ルーティングモジュール３０８と動的ルーティングモジュール３１４との間を、よりシームレスに移行させることができる。ネットワークマネージャ３１８は、情報についての、「間際の（ｌａｓｔ－ｓｅｃｏｎｄ）」移行を達成するために使用されるが、この情報とは、スタンドバイルーティングコンポーネント３１２が使用できる情報であって、動的ルーティングモジュール３０８からの情報である。この場合に、動的ルーティングモジュール３０８と動的ルーティングモジュール３１４の間の移行は、可能な限り最新の状態に保たれる。
【００６６】
　図４は、本発明の一実施形態に従う場合に、ルーティングコンポーネントを用いてサービスの提供を受ける一群のデバイスを有するネットワークの論理的平面を示す図である。このネットワークの各要素は、ソフトウェア、ハードウェア又はこれらの任意の組み合わせにおいて実装することができる。大抵の場合、ルーティングコンポーネント４００の機能を、制御プレーンモジュール４０２と転送プレーンモジュール４０４とに分けることができる。即ち、制御プレーンモジュール４０２がノードに関連する制御および管理の機能を有するのに対して、転送プレーンモジュール４０４はノードを通過しているパケットについてのパケット毎の処理を行うものである。制御プレーンアプリケーションの例として、ＯＳＰＦのようなルーティングプロトコルと、ＳＮＭＰのような管理プロトコルが挙げられる。
【００６７】
　より詳細に説明すると、ルーティングコンポーネント４００は、クラスタ４０６に関連するルーティング機能を実現するものとして示されている。スタンドバイルーティングコンポーネント４０８は、ルーティングコンポーネント４００と結びつけられている。このスタンドバイルーティングコンポーネント４０８は、制御プレーンモジュール４１０、４１２の他、転送プレーンモジュール４１４をも有する。そして、スタンドバイルーティングコンポーネント４０８は、アクティブな制御プレーンモジュール４１２と、アクティブでない制御プレーンモジュール４１０の両方を有しうる。
【００６８】
　動作中には、ルーティングコンポーネント４００が転送プレーンモジュール４０４のネットワーク特性に関する情報を、スタンドバイルーティングコンポーネント４０８に転送する。これを受けて、スタンドバイルーティングコンポーネント４０８はこのネットワーク情報を転送プレーンモジュール４１４に伝達する。したがって、ルーティングコンポーネント４００に関連する転送プレーンに関する情報の（たとえそれが完全に更新されていないとしても）、緊密に更新されたコピーが、ルーティングコンポーネント４００とは独立して保持される。
【００６９】
　さらに、制御プレーンモジュール４０２に関する情報が、スタンドバイルーティングコンポーネント４０８に中継される。スタンドバイルーティングコンポーネント４０８（あ

るいは別のリモートデバイスの可能性もある）が、ルーティングコンポーネント４００内の制御プレーンモジュール４０２の動作特性に関連する制御プレーン情報を保持する。オーバーヘッドを減らすために、スタンドバイルーティングコンポーネント４０８内の制御プレーンモジュール４１２は、アクティブでない状態４１０に保持されており、ハードディスクに置かれるなど、実行のためメモリにロードされていなくてもよい。

【００７０】

当業者には分かるように、アクティブな制御プレーンモジュール４１２又はアクティブでない制御プレーンモジュール４１０のいずれかについては、幾つかの実行可能モジュールから構成されてもよい。また、スタンドバイルーティングコンポーネント４０８において、フェールオーバの事象が生じた時に、制御プレーンモジュールが１つも実行されていないこともあれば、その一部又は全てが実行されていることもある。したがって、フェールオーバの事象時には、それが計画されたものであるか計画外であるかに関わらず、完全に機能している制御プレーンを有する残りのモジュールが、スタンドバイルーティングコンポーネントにおいて実行される。

【００７１】

さらに、各種場面（ｆｏｒｕｍ）において、動的に設定可能かつ動作可能な別のルーティングシステムが使用されてもよいことは、当業者の理解するところである。本開示はＯＳＰＦに限定されず、この種の動的に設定可能な他のルーティングシステムも同様に含むという点を当業者は理解する。この趣旨において、パケットをクラスタに転送し続けるようにネットワークに知らせるＯＳＰＦのヒットレス再起動は、他のタイプのプロトコルでは必要ない（あるいは利用可能でなくてもよい）ことがある。さらに、ヒットレス再起動の二次的な機能（すなわち、１つのコンポーネントの障害により、ネットワークの内部データの表現が再作成されないようにする機能）が、現行又は将来登場するルーティングモジュールにおいて、あるいは他の動的ルーティングモジュールの拡張において実装される可能性がある。また、他の同等の機能的特徴は、同等の動作を実行するため、このような他のルーティングプロトコルにおいて利用できるようになる可能性があるか、又はこのようなプロトコルの将来的な機能強化においてこのような特徴が含まれる可能性がある。

【００７２】

本発明の実施形態および用途を図示かつ記載したが、本明細書に記載した発明的概念から逸脱することなく、前述した以外の数多くの変形例が可能であることは、本開示の利益を受ける当業者にとって明らかである。このため、本発明は、添付した特許請求の範囲の趣旨以外によって限定されることはない。

【図面の簡単な説明】

【００７３】

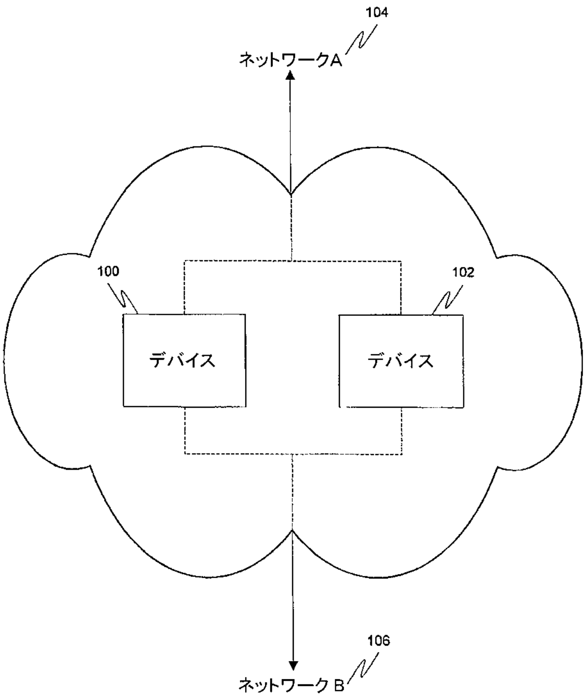
【図１】クラスタの一例を示す図である。

【図２】本発明の実施形態に従って、アクティブなルーティングコンポーネントに障害が生じたときの処理方法を示すフローチャート図である。

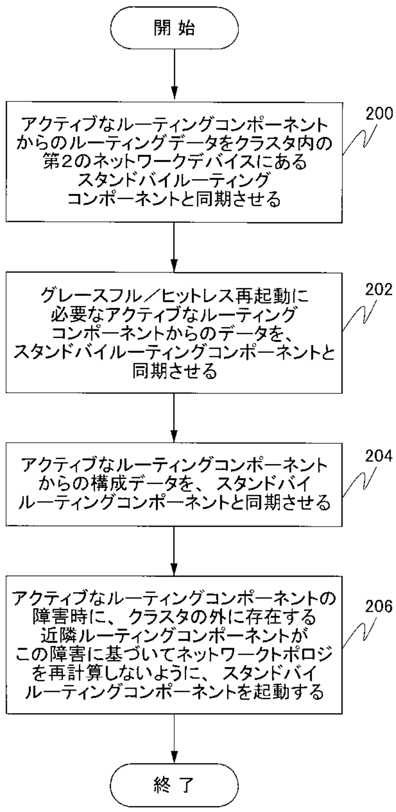
【図３】本発明の一実施形態による、ルーティングコンポーネントを用いてサービスの提供を受ける一群のデバイスを有するネットワークの模式図である。

【図４】本発明の一実施形態による、ルーティングコンポーネントを用いてサービスの提供を受ける一群のデバイスを有するネットワークの論理構成面を示す図である。

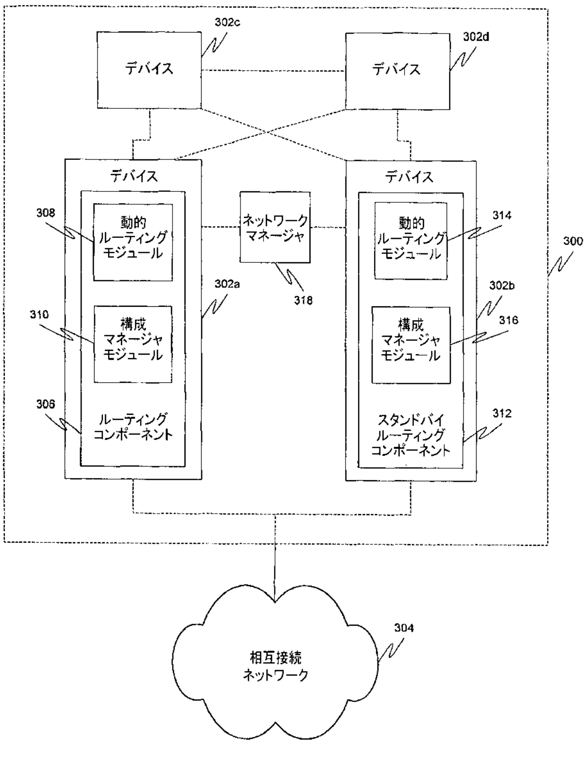
【図１】


104
ネットワークA
100
102
デバイス
デバイス
106
ネットワーク B

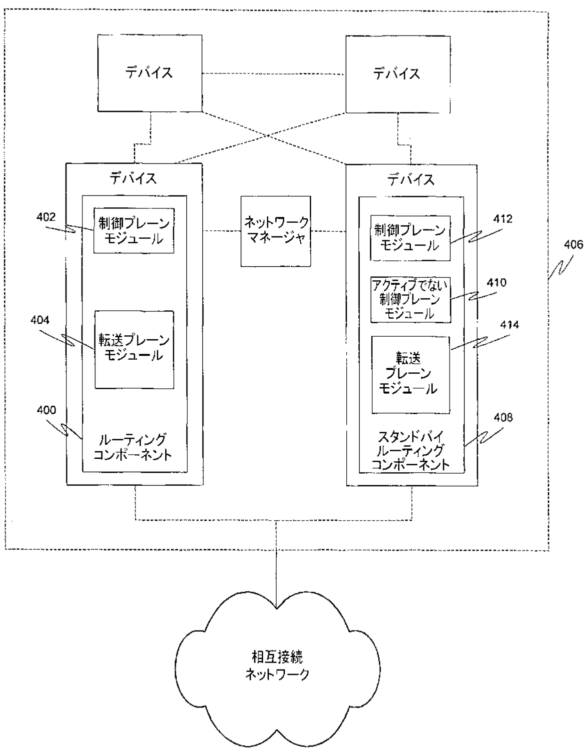
【図２】


開 始
アクティブなルーティングコンポーネントからのルーティングデータをクラスタ内の第2のネットワークデバイスにあるスタンドバイルーティングコンポーネントと同期させる
200
グレースフル／ヒットレス再起動に必要なアクティブなルーティングコンポーネントからのデータを、スタンドバイルーティングコンポーネントと同期させる
202
アクティブなルーティングコンポーネントからの構成データを、スタンドバイルーティングコンポーネントと同期させる
204
アクティブなルーティングコンポーネントの障害時に、クラスタの外に存在する近隣ルーティングコンポーネントがこの障害に基づいてネットワークトポロジを再計算しないように、スタンドバイルーティングコンポーネントを起動する
206
終 了

【図３】


302c
デバイス
302d
デバイス
デバイス
デバイス
308
動的ルーティングモジュール
ネットワークマネージャ
動的ルーティングモジュール
314
300
318
302a
302b
310
構成マネージャモジュール
構成マネージャモジュール
316
306
ルーティングコンポーネント
スタンドバイルーティングコンポーネント
312
相互接続ネットワーク
304

【図４】


デバイス
デバイス
デバイス
デバイス
402
制御プレーンモジュール
ネットワークマネージャ
制御プレーンモジュール
412
406
アクティブでない制御プレーンモジュール
410
404
転送プレーンモジュール
414
転送プレーンモジュール
400
ルーティングコンポーネント
408
スタンドバイルーティングコンポーネント
相互接続ネットワーク

フロントページの続き

(56)参考文献　米国特許出願公開第２００３／００５６１３８（ＵＳ，Ａ１）
特表２００２－１３５３２８（ＪＰ，Ａ）
米国特許出願公開第２００３／０１４０１６７（ＵＳ，Ａ１）
米国特許出願公開第２００２／０１９１５４７（ＵＳ，Ａ１）
SANGLI SRIHARI R., Graceful Restart Mechanism for BGP, DRAFT-IETF-IDR-RESTART-06.TXT, IETF, NETWORK WORKING GROUP, ２００３年　7月３１日, ＵＲＬ, http://www.ietf.org/proceedings/03mar/I-D/draft-ietf-idr-restart-06.txt

(58)調査した分野(Int.Cl.，ＤＢ名)
H04L　12/56