(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2015-138987
(P2015-138987A)
(43)公開日　平成27年7月30日(2015.7.30)

(51)Int.Cl.			F I		テーマコード（参考）
H04L	12/717	(2013.01)	H04L	12/717	5K030
H04L	12/775	(2013.01)	H04L	12/775	
H04L	12/703	(2013.01)	H04L	12/703	

審査請求 未請求　請求項の数 8　OL　（全 18 頁）

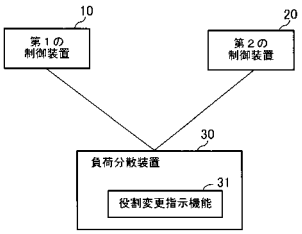
(21)出願番号　特願2014-7773(P2014-7773)
(22)出願日　平成26年1月20日(2014.1.20)

(71)出願人　000004237
日本電気株式会社
東京都港区芝五丁目７番１号
(74)代理人　100103090
弁理士　岩壁　冬樹
(74)代理人　100124501
弁理士　塩川　誠人
(72)発明者　小川　英輝
東京都港区芝五丁目７番１号　日本電気株式会社内

Fターム(参考)　5K030 GA12　HD03　KA01　LB07　LE03
LE10　MA12　MB01　MD02

(54)【発明の名称】通信システムおよび通信システムにおけるサービス復旧方法

(57)【要約】
【課題】ビザンチン型故障の発生を検出した場合に、早期に故障状態を解消しパケット転送を再開可能とする通信システムおよび通信システムにおけるサービス復旧方法を提供する。
【解決手段】フロー制御機能と仮想ネットワークの設定情報とをもとに複数のパケット転送装置を制御する第１の制御装置１０と、動作実績があるフロー制御機能と、運用実績がある仮想ネットワークの設定情報とを有し、第１の制御装置１０に故障が発生した場合に切り替え先となる第２の制御装置２０と、複数のパケット転送装置と各制御装置との間に配置された負荷分散装置３０とを備え、負荷分散装置３０は、第１の制御装置１０においてビザンチン型故障が発生した場合に、複数のパケット転送装置の制御を第２の制御装置２０に実行させる役割変更指示機能３１を有する。
【選択図】図８



【特許請求の範囲】
【請求項１】
フロー制御機能と仮想ネットワークの設定情報とをもとに複数のパケット転送装置を制御する第１の制御装置と、
動作実績があるフロー制御機能と、運用実績がある仮想ネットワークの設定情報とを有し、前記第１の制御装置に故障が発生した場合に切り替え先となる第２の制御装置と、
前記複数のパケット転送装置と各制御装置との間に配置された負荷分散装置とを備え、
前記負荷分散装置は、前記第１の制御装置においてビザンチン型故障が発生した場合に、前記複数のパケット転送装置の制御を前記第２の制御装置に実行させる役割変更指示機能を有する
ことを特徴とする通信システム。
【請求項２】
負荷分散装置の役割変更指示機能は、第１の制御装置のＣＰＵ使用率が所定の閾値を超えたときに、当該第１の制御装置においてビザンチン型故障が発生したと判断する
請求項１に記載の通信システム。
【請求項３】
通常運用時、第１の制御装置は、複数のパケット転送装置から送信される全てのメッセージを受信するマスタとして動作し、第２の制御装置は、前記複数のパケット転送装置からトポロジ検出に必要なメッセージのみを受信するスレーブとして動作し、
負荷分散装置の役割変更指示機能は、第１の制御装置においてビザンチン型故障が発生した場合に、前記第１の制御装置をスレーブとして動作させ、第２の制御装置をマスタとして動作させる
請求項１または請求項２に記載の通信システム。
【請求項４】
各制御装置は、負荷分散装置からの役割変更指示に基づいて、自装置をマスタとして動作させるか、スレーブとして動作させるかを管理する役割管理機能を有する
請求項３に記載の通信システム。
【請求項５】
第２の制御装置の役割管理機能は、スレーブ動作中に、トポロジ検出に必要なメッセージを受信すると、当該メッセージをもとにマスタとして動作を開始する際に必要なフローを作成する
請求項４に記載の通信システム。
【請求項６】
各制御装置は、パケット転送装置から受信したメッセージがトポロジ検出に必要なメッセージであると判断した場合に、他の制御装置に当該メッセージを通知する
請求項１に記載の通信システム。
【請求項７】
第１の制御装置が、複数の制御装置によりクラスタ構成された
請求項１から請求項６のうちのいずれか１項に記載の通信システム。
【請求項８】
フロー制御機能と仮想ネットワークの設定情報とをもとに複数のパケット転送装置を制御する第１の制御装置と、動作実績があるフロー制御機能と、運用実績がある仮想ネットワークの設定情報とを有し、前記第１の制御装置に故障が発生した場合に切り替え先となる第２の制御装置と、前記複数のパケット転送装置と各制御装置との間に配置された負荷分散装置とを備えた通信システムにおいて、
前記負荷分散装置が、前記第１の制御装置においてビザンチン型故障が発生した場合に、前記複数のパケット転送装置の制御を前記第２の制御装置に実行させる
ことを特徴とするサービス復旧方法。
【発明の詳細な説明】
【技術分野】

【０００１】
本発明は、複数のパケット転送装置を制御する制御装置を含む通信システムおよび通信システムにおけるサービス復旧方法に関する。
【背景技術】
【０００２】
オープンフローなどのＳＤＮ（Ｓｏｆｔｗａｒｅ－Ｄｅｆｉｎｅｄ Ｎｅｔｗｏｒｋｉｎｇ）分野では、ネットワーク（ＮＷ）におけるパケット転送を単一のコントローラにより制御する。このような分野においては、これまでもコントローラの可用性を高めるための技術が提案されてきた。例えば、特許文献１には、オープンフローネットワークにおいて、複数配置されたコントローラの負荷を均等化する技術が記載されている。
【０００３】
オープンフロー（ＯｐｅｎＦｌｏｗ）では、データプレーンとコントロールプレーンが分離され、ＯＦＣ（ＯｐｅｎＦｌｏｗ Ｃｏｎｔｒｏｌｌｅｒ）がネットワークを集中的に管理する。このような構成において可用性をより高めるためには、ＯＦＣに故障が発生した場合に、速やかに業務を復旧するための手段を備えている必要がある。しかし、コンピュータシステムで発生する故障、特にビザンチン型故障からの復旧に関する技術についての提案は現時点では少ない。
【０００４】
一般的に、コンピュータシステムで発生する故障について、「沈黙型故障」と「ビザンチン型故障」の二つに分類することができる。
【０００５】
沈黙型故障は、他ノードからの要求に対し応答を返却できない状態に陥った場合を指す。一方、ビザンチン型故障は、入力に対する出力が正しかったり誤っていたりして、入力に対する出力が信用できない状態に陥った場合を指す。
【０００６】
ＯＦＣにおける沈黙型故障とは、ＯＦＳ（ＯｐｅｎＦｌｏｗ Ｓｗｉｔｃｈ）や管理クライアントなどからの要求に対し、応答を返却できない状態であると考えられる。ＯＦＣにおける沈黙型故障は、致命的な障害が発生した場合や、過剰な負荷が掛かった場合などに発生する。この沈黙型故障からの復旧手段については、既に考案・実用化されている。例えば、クラスタソフトウェア（クラスタＳＷ）によりＯＦＣを冗長化しておき、沈黙型故障発生時にはスタンバイノードにフェイルオーバーして処理を継続する方法である。
【０００７】
一方で、ＯＦＣにおけるビザンチン型故障とは、なんらかの要因によりＯＦＳに意図しないフローを設定し、意図しないパケット転送処理が行われるようになった状態であると考えられる。
【０００８】
ビザンチン型故障対策の詳細に移る前に、先ず背景となる技術として、（Ｉ）シングルＯＦＣ構成における通常運用処理、（ＩＩ）シングルＯＦＣ構成における再起動処理、（ＩＩＩ）クラスタＯＦＣ構成におけるフェイルオーバー処理、について説明する。
【０００９】
（Ｉ）シングルＯＦＣ構成における通常運用処理
【００１０】
シングルＯＦＣ構成、すなわちＯＦＣが冗長化されていない場合のシステム構成について説明する。図９は、オープンフローシステムにおけるシングルＯＦＣ構成の一例を示すブロック図である。ＯＦＣ１００は、複数のＯＦＳ１０１～１０３を管理する。ＯＦＳ１０１～１０３は、それぞれ固有のフローテーブル１１３～１１５を保持する。各フローテーブルには、有限個のフローエントリが格納される。フローエントリには、パケット転送ルールが記述される。ＯＦＣ１００は、コンフィグ１１０、トポロジ１１１、および、フローテーブル１１２を保持する。
【００１１】

ＯＦＣ１００は、配下のＯＦＳの数だけフローテーブルを保持する。また、ＯＦＣ１００が保持するフローテーブル１１２とＯＦＳ１０１～１０３が保持するフローテーブル１１３～１１５の内容は、常時同期される。また、ＯＦＣ１００のフローテーブル１１２は、ＯＦＣの再起動等の処理において揮発する。

【００１２】

コンフィグ１１０は、コンフィグを示す情報（コンフィグ情報）であって、仮想ネットワークの設定が記述されている。オープンフローネットワーク管理者は、適宜、ＯＦＣを操作することで、仮想ネットワークの設定を変更することができる。つまり、コンフィグは、ＯＦＣの運用中に変更される情報である。また、通常、コンフィグはセーブすることができる。従って、コンフィグは、ＯＦＣの再起動等の処理において揮発しない。

【００１３】

トポロジ１１１は、トポロジを示す情報（トポロジ情報）であって、ＯＦＣがトポロジ検出機能を用いて作成する情報である。トポロジ１１１は、スイッチ増設や冗長化など、ネットワーク構成を変更した場合に更新される。また、トポロジ１１１は、ＯＦＣの再起動等の処理において揮発する。

【００１４】

ここで、ＯＦＣ１００の基本的な動作を説明する。図１０は、ＯＦＣ１００の構成の一例を示すブロック図である。

【００１５】

ＯＦＣ１００は、通信手段２０１と、ＯＦ（ＯｐｅｎＦｌｏｗ）プロトコル処理手段２０２と、トポロジ検出手段２０３と、コンフィグ管理手段２０５と、経路管理手段２０７とを含む。

【００１６】

ＯＦプロトコル処理手段２０２は、通信手段２０１を用いて、複数のＯＦＳとの間でオープンフロープロトコルに則った通信を行う。

【００１７】

トポロジ検出手段２０３は、ＯＦプロトコル処理手段２０２を用いて、ネットワークのトポロジを示す情報（図１０に示すトポロジ２０４）を作成する。トポロジ検出手段２０３は、ＬＬＤＰ（Ｌｉｎｋ　Ｌａｙｅｒ　Ｄｉｓｃｏｖｅｒｙ　Ｐｒｏｔｃｏｌ）パケットを配下の全てのＯＦＳ宛に送信する。ＯＦＳは、受け取ったＬＬＤＰパケットを全ポートから送出する。ＯＦＳのポートにケーブルが接続され、当該ケーブルが他のＯＦＳに接続されている場合、接続先ＯＦＳはＬＬＤＰパケットのＰＡＣＫＥＴ＿ＩＮメッセージをＯＦＣ１００に対して通知する。ＯＦＣ１００は、このＰＡＣＫＥＴ＿ＩＮメッセージを受け取り、ＯＦＳポート間の接続情報を得る。ＯＦＣ１００は、全ての接続情報をまとめることで、ネットワークのトポロジを取得できる。トポロジ検出手段２０３は、定期的にＬＬＤＰを送出することで、ＯＦＳの追加やケーブルの抜き差しによる、トポロジの変化を検出できる。そのたびに、トポロジ検出手段２０３は、トポロジを更新する。

【００１８】

コンフィグ管理手段２０５は、ユーザによる仮想ネットワークの設定変更操作に応じて、コンフィグを示す情報（図１０に示すコンフィグ２０６）を更新する。

【００１９】

経路管理手段２０７は、経路計算手段２０９と、経路設定手段２１０と、オーディット手段２１１とを含む。

【００２０】

オープンフローによりフロー制御を行う場合、始めに少なくともブロードキャストやマルチキャストされたパケットを適切に転送するためのフローをＯＦＳに設定しておく必要がある。このフロー（以下、ＢＣ／ＭＣ（ＢｒｏａｄＣａｓｔ／ＭａｌｔｉＣａｓｔ）フローと呼称する）は、トポロジやコンフィグの変更に合わせて、更新される。経路計算手段２０９は、システム開始時や、トポロジまたはコンフィグの更新時にＢＣ／ＭＣフローを計算し、フローテーブル（図１０に示すフローテーブル２０８）に設定する。

【００２１】
フローテーブルにフローエントリが追加された場合、経路設定手段２１０は、ＯＦＳに対し、フローエントリの追加を通知する。
【００２２】
ＯＦＳにパケットが到着したとき、当該パケットのヘッダに記された宛先アドレスや送信元アドレスなどの情報（ヘッダ情報）が適合条件にマッチするフローエントリが存在する場合、ＯＦＳは、そのフローエントリに記述された通りにパケットを処理する。当該処理をハード転送と呼ぶ。フローエントリには、パケット転送ルールとして、上記適合条件や、該当パケットをあるポートから送出する、などの処理内容が記述されている。
【００２３】
一方で、ＯＦＳにパケットが到着したとき、当該パケットのヘッダ情報にマッチするフローエントリが存在しない場合、ＯＦＳはＯＦＣ１００に対して、当該パケットの処理方法を問い合わせる。ＯＦＳからの問い合わせメッセージを受け取った経路管理手段２０７は、トポロジ、コンフィグ、および、当該パケットのヘッダ情報を、経路計算手段２０９に入力し、当該パケットの処理方法を定義するフローエントリを導出する。経路管理手段２０７がこのフローエントリをフローテーブル２０８に追加したのち、経路設定手段２１０は，ＯＦＳのフローテーブルに当該フローエントリを追加する。ＯＦＳは、追加されたフローエントリに従って、転送や廃棄等の処理を行う。当該処理をソフト転送と呼ぶ。ソフト転送は、フローエントリー導出処理を含む分、ハード転送に比べて性能が悪い。
【００２４】
また、ＯＦＳのフローテーブルについて、一定時間以上ヒットしなかったフローエントリを削除する機能がある。この機能では、ＯＦＳが、既定時間ヒットしないフローエントリを識別し削除する。その後、ＯＦＳは、フローエントリを削除した旨をＯＦＣ１００に通知する。当該通知を受けた経路管理手段２０７は、フローテーブルから当該フローエントリを削除する。
【００２５】
このようにして、ＯＦＣのフローテーブルとＯＦＳのフローテーブルは同期を保ったまま更新される。しかし、ＯＦＣとＯＦＳの間の接続が切れた場合などは、テーブルの同期が失われる。従って、ＯＦＣとＯＦＳとが再度接続した際には、オーディット手段２１１によって、再度テーブルを同期させる処理が実行される。この処理をオーディット（Ａｕｄｉｔ）処理と呼ぶ。オーディット手段２１１は、ＯＦＣ側で持つフローテーブルを主とし、従であるＯＦＳ側フローテーブルに生じた差分を修正する。
【００２６】
以上のように、オープンフローシステムがフロー制御を行うためには、ＯＦＣがコンフィグ、トポロジ、フローテーブルを保持し、且つ、ＯＦＳがＯＦＣと同期されたフローテーブルを保持している必要がある。
【００２７】
（ＩＩ）シングルＯＦＣ構成における再起動処理
【００２８】
（Ｉ）のシングルＯＦＣ構成において、ＯＦＣを再起動する場合の処理について説明する。図１１は、シングルＯＦＣ構成において、沈黙型故障が発生した際のＯＦＣ再起動によるサービス停止時間、つまりＯＦＣが再起動してからパケット転送を再開するまでの時間を示す説明図である。
【００２９】
ＯＦＣのＯＳを再起動すると、ＯＦＣとＯＦＳの間の通信が切断される。切断を検出したＯＦＳは、ＯＦＣへの接続をリトライし続ける。ＯＦＣが起動すると、ＯＦＣが保持するコンフィグ、トポロジ、フローテーブルは、何れも空の状態になる。従って、ＯＦＣでは、まず、セーブされているコンフィグのリロード（コンフィグリロード処理３００）、トポロジの作成（トポロジ情報作成処理３０１）を行う必要がある。これらの処理が完了すると、ＯＦＣは、ＢＣ／ＭＣフローエントリの導出（ＢＣ／ＭＣフローエントリ導出処

理３０２）が可能になる。ＯＦＣは、この処理が完了したのち、ＯＦＳからの接続要求を受け付ける。ＯＦＣとＯＦＳの通信が切断されている間に、ＯＦＣとＯＦＳのそれぞれのフローテーブルに差異が発生するため、ＯＦＣは、オーディット（オーディット処理３０３）を行う。このオーディット処理により、再起動前に設定されていたフローエントリは削除され、ＢＣ／ＭＣフローのみが登録された状態となる。

【００３０】

以上のように、シングルＯＦＣ構成において、ＯＦＣを再起動する場合、上記の処理３００～３０３が完了した後にオープンフローによるフロー制御が可能となる。なお、サービス再開後は、ソフト転送となる。

【００３１】

（ＩＩＩ）クラスタＯＦＣ構成におけるフェイルオーバー処理

【００３２】

ＯＦＣでは、沈黙型故障に対応するため、既にクラスタ方式が実用化されている。ここで、クラスタＯＦＣ構成、すなわちＯＦＣがクラスタ構成（冗長化）されたシステム構成について説明する。図１２は、オープンフローシステムにおけるクラスタＯＦＣ構成の一例を示すブロック図である。図１２を参照し、クラスタＯＦＣ構成でのフェイルオーバー処理について説明する。

【００３３】

ＯＦＣ（ＡＣＴ）４００、ＯＦＣ（ＳＢＹ）４０１の２台のＯＦＣノードは、クラスタミドルウェアによりクラスタシステムを構成する。なお、“ＡＣＴ（ＡＣＴＩＶＥ）”は、現用系であることを表す。“ＳＢＹ（ＳＴＡＮＤＢＹ）”は、待機系であることを表す。ＯＦＣ（ＡＣＴ）４００およびＯＦＣ（ＳＢＹ）４０１には、それぞれ固有のアドレスが割り当てられる。また、クラスタシステムには、固有の仮想アドレスが割り当てられる。

【００３４】

始めに、クラスタミドルウェアは、ＯＦＣ（ＡＣＴ）４００のアドレスに仮想アドレスを対応付ける。ＯＦＳ４０３～４０５は仮想アドレスに接続することで、結果的にＯＦＣ（ＡＣＴ）４００と接続される。こうして、ＯＦＣ（ＡＣＴ）４００とＯＦＳ４０３～４０５によりオープンフローによるフロー制御が行われる。ＯＦＣ（ＡＣＴ）４００は、コンフィグ４１０、トポロジ４１１、フローテーブル４１２を保持する。これらの情報は、上記「（Ｉ）シングルＯＦＣ構成における通常運用処理」と同様に更新される。また、ＯＦＣ（ＳＢＹ）４０１も、コンフィグ４１３、トポロジ４１４、フローテーブル４１５を保持する。ＯＦＣ（ＳＢＹ）４０１が保持するこれらの情報は、クラスタミドルウェアが備えるミラーディスク機能やメモリ同期機能により、ＯＦＣ（ＡＣＴ）４００が保持する、対応する情報と常に同期される。

【００３５】

ＯＦＣ（ＡＣＴ）４００とＯＦＣ（ＳＢＹ）４０１は、ＬＡＮケーブルにより接続されていて、定期的に通信（ハートビート通信）することで、対向ノードが正常に稼働していることを確認する。

【００３６】

沈黙型故障が発生した場合、ＯＦＣ（ＡＣＴ）４００は、ハートビート通信を継続できず、タイムアウトなどにより通信が切断する。ハードビート通信の切断を検出したクラスタミドルウェアは、仮想アドレスとＯＦＳとの間で確立していた通信を切断する。ＯＦＳは、通信切断を検出すると、再度ＯＦＣと接続するべく、仮想アドレスへの通信確立を繰り返しリトライする。

【００３７】

一方で、クラスタミドルウェアは、ＯＦＣ（ＡＣＴ）４００側でサービスを停止、ＯＦＣ（ＳＢＹ）４０１側でサービスを開始させ、仮想アドレスをＯＦＣ（ＳＢＹ）４０１のアドレスに対応付ける。これにより、ＯＦＳは、ＯＦＣ（ＳＢＹ）４０１と通信を確立できるようになる。ＯＦＣ（ＳＢＹ）４０１のコンフィグ４１３、トポロジ４１４、および

、フローテーブル４１５は、ＯＦＣ（ＡＣＴ）４００側のものと同期されているため最新の情報になっている。従って、ＯＦＣ（ＳＢＹ）４０１において、（Ｉ）のケースで必要だった、コンフィグリロードやトポロジ検出、ＢＣ／ＭＣフローエントリの導出、といった処理は不要である。つまり、ＯＦＣ（ＳＢＹ）４０１では、ＯＦＳとＯＦＣの間の通信が切断されていた間に生じた差分を解消するためのオーディット処理のみが必要となり、オーディット処理完了後にはパケット転送可能となる。なお、サービス再開時からハードウェア転送になる。

【００３８】

図１３は、クラスタＯＦＣ構成において、沈黙型故障が発生した際のフェイルオーバー処理におけるサービス停止時間、つまりノード切り替えを開始してからパケット転送を再開するまでの時間を示す説明図である。図１３に示すように、クラスタＯＦＣ構成では、クラスタミドルウェアによるノード切り替え処理５００と、オーディット処理５０１だけで、サービス再開できる。

【先行技術文献】

【特許文献】

【００３９】

【特許文献１】再表ＷＯ２０１３／１１４４９０号

【発明の概要】

【発明が解決しようとする課題】

【００４０】

上記（Ｉ）～（ＩＩＩ）に記載した、基本的なＯＦＣの動作を踏まえた上で、ＯＦＣにおけるビザンチン型故障について考える。

【００４１】

ＯＦＣにおけるビザンチン型故障は、例えば、システム管理者が意図した通りにパケット転送されることもあれば、意図しない形でパケット転送されることもあるという、フロー制御が信用できない状態に陥った状態である。つまり、意図しないフローエントリが導出され、フローテーブルに登録されてしまった状態である。このような状態は、以下のような場合に発生し得る。

【００４２】

（Ａ）不正なコンフィグが設定されてしまった場合。例えば、コンフィグ設定作業にミスがあった場合や、内部犯行者により意図的にセキュリティホールを含むような論理ネットワークが設定された場合。

【００４３】

（Ｂ）ＯＦＣ機能（例えば、経路導出機能やトポロジ検出機能）が不正である場合。例えば、ＯＦＣに潜在的なバグがある場合や、不正使用者によりＯＦＣ機能が改ざんされた場合。

【００４４】

（Ｉ）～（ＩＩＩ）で説明したような構成では、ビザンチン型故障から復旧することが難しい。（Ａ）の場合、一度ＯＦＣを再起動し、修正したコンフィグを再度読み込ませてサービスを再開することにより、ＯＦＳに設定された不正なフローが一掃でき、故障状態を解消できる。しかし、少なくとも図１１に示す程度のサービス停止時間が発生する。（Ｂ）の場合、バグや改ざんを取り除いたＯＦＣを新たに用意し、コンフィグをロードするなど、必要となる処理を行うことにより、サービスを再開できる。しかし、少なくともバグや改ざんを除去する時間と図１１に示す時間とを合計したサービス停止時間が発生する。何れの場合も、図１１に示す程度のサービス停止時間が発生する。

【００４５】

そこで、本発明は、ビザンチン型故障の発生を検出した場合に、早期に故障状態を解消しパケット転送を再開可能とする通信システムおよび通信システムにおけるサービス復旧方法を提供することを目的とする。

【課題を解決するための手段】

【００４６】
本発明による通信システムは、フロー制御機能と仮想ネットワークの設定情報とをもとに複数のパケット転送装置を制御する第１の制御装置と、動作実績があるフロー制御機能と、運用実績がある仮想ネットワークの設定情報とを有し、第１の制御装置に故障が発生した場合に切り替え先となる第２の制御装置と、複数のパケット転送装置と各制御装置との間に配置された負荷分散装置とを備え、負荷分散装置は、第１の制御装置においてビザンチン型故障が発生した場合に、複数のパケット転送装置の制御を第２の制御装置に実行させる役割変更指示機能を有することを特徴とする。
【００４７】
本発明によるサービス復旧方法は、フロー制御機能と仮想ネットワークの設定情報とをもとに複数のパケット転送装置を制御する第１の制御装置と、動作実績があるフロー制御機能と、運用実績がある仮想ネットワークの設定情報とを有し、第１の制御装置に故障が発生した場合に切り替え先となる第２の制御装置と、複数のパケット転送装置と各制御装置との間に配置された負荷分散装置とを備えた通信システムにおいて、負荷分散装置が、第１の制御装置においてビザンチン型故障が発生した場合に、複数のパケット転送装置の制御を第２の制御装置に実行させることを特徴とする。
【発明の効果】
【００４８】
本発明によれば、通信システムにおいて、ビザンチン型故障の発生を検出した場合に、早期に故障状態を解消しパケット転送を再開することができる。
【図面の簡単な説明】
【００４９】
【図１】本発明による通信システムの第１の実施形態の構成を示すブロック図である。
【図２】アドレス変換表の一例を示す説明図である。
【図３】第１の実施形態におけるＯＦＣの構成を示すブロック図である。
【図４】ロードバランサにおけるＯＦＣの切り替え動作を示すフローチャートである。
【図５】ステップＳ５において更新されたアドレス変換表の一例を示す説明図である。
【図６】第１の実施形態の通信システムにおいてビザンチン型故障が発生した際の復旧に伴うサービス停止時間を示す説明図である。
【図７】本発明による通信システムの第２の実施形態の構成を示す説明図である。
【図８】本発明による通信システムの最小構成を示すブロック図である。
【図９】オープンフローシステムにおけるシングルＯＦＣ構成の一例を示すブロック図である。
【図１０】ＯＦＣの構成の一例を示すブロック図である。
【図１１】シングルＯＦＣ構成において、沈黙型故障が発生した際のＯＦＣ再起動によるサービス停止時間を示す説明図である。
【図１２】オープンフローシステムにおけるクラスタＯＦＣ構成の一例を示すブロック図である。
【図１３】クラスタＯＦＣ構成において、沈黙型故障が発生した際のフェイルオーバー処理におけるサービス停止時間を示す説明図である。
【発明を実施するための形態】
【００５０】
実施形態１．
以下、本発明の第１の実施形態を図面を参照して説明する。
【００５１】
本実施形態では、オープンフロープロトコルＶｅｒ１．２で実装されたマルチコントローラ機能を用いて、現用系でオープンフロー制御を行いつつ、待機系でも常時トポロジ情報とＢＣ／ＭＣフローエントリーの導出を行う通信システムを例として説明する。
【００５２】
本実施形態では、クラスタ構成されたＯＦＣ（以下、第１のＯＦＣと呼称する）とは別

に、ビザンチン型故障が発生した場合に切り替え先となるＯＦＣ（以下、第２のＯＦＣと呼称する）を配置する。また、ＯＦＳと、第１のＯＦＣ／第２のＯＦＣとの間にロードバランサ（ＬＢ）を配置し、ロードバランサの操作によりＯＦＳの接続先ＯＦＣを切り替えられるようにする。

【００５３】

第２のＯＦＣとして、経路導出やトポロジ検出機能にバグや改ざんがないＯＦＣ、つまり信用がおけるＯＦＣ機能を有するＯＦＣを用意する。例えば、以下のような考え方で第２のＯＦＣを用意する。

【００５４】

・第１のＯＦＣを構成するＯＦＣノードにバグが見つかった場合、そのバグを改修したＯＦＣを用意し、第２のＯＦＣとして使用する。或いは、既に十分な動作実績があり、品質が安定している、古いバージョンのＯＦＣを第２のＯＦＣとして使用する、など。

・第１のＯＦＣより厳しいセキュリティポリシーを適用し、機能改ざんなどの不正操作に対する対策が施されたＯＦＣを第２のＯＦＣとして使用する。例えば、第２のＯＦＣにはより限定されたユーザのみアクセスできるよう設定する、第２のＯＦＣをより保護されたネットワークに配置する、など。

【００５５】

また、第２のＯＦＣには、信用がおけるコンフィグ情報をロードしておく。例えば、以下のような考え方で信用がおけるコンフィグを定義する。

【００５６】

・十分運用実績があるコンフィグ
・更新できないように設定したコンフィグ
・シンプルな構成とし、容易に改ざんを見抜けるようなコンフィグ

【００５７】

以上のように、本実施形態では、第２のＯＦＣが、常に最も信用がおけるＯＦＣ機能と、最も信用がおけるコンフィグ情報とを有するように管理する。第１のＯＦＣと第２のＯＦＣは非対称になる。

【００５８】

なお、トポロジ情報は、ＯＦＣで管理できるものではなく、ＬＬＤＰパケットをＯＦＳに送信しその応答を得て生成する情報である。本実施形態では、このトポロジ情報を、第１のＯＦＣ、第２のＯＦＣで常に最新の情報を取得できるように、通信システムを構成する。それにより、第２のＯＦＣでは、信用がおけるＯＦＣ機能およびコンフィグ情報と、最新のトポロジ情報とから、常にＢＣ／ＭＣフローエントリーを導出できる状態とすることができる。その結果、通信システムは、切り替え時のサービス停止時間を短縮できる。

【００５９】

図１は、本発明による通信システムの第１の実施形態の構成を示すブロック図である。なお、図１において点線で示す構成要素以外の要素については、図１２に示す要素と同様であるため、詳細な説明を省略する。

【００６０】

図１に示すように、通信システムは、第１のＯＦＣ６００と、第２のＯＦＣ６０３と、ロードバランサ６０４とを備える。なお、図１には、ロードバランサ６０４に３台のＯＦＳ（ＯＦＳ６０８～６１０）が接続されているが、ＯＦＳは、ロードバランサ６０４にいくつ接続されていてもよい。

【００６１】

第１のＯＦＣ６００は、クラスタミドルウェアにより冗長化構成されたＯＦＣノードである。第１のＯＦＣ６００は、ＯＦＣ（ＡＣＴ）６０１とＯＦＣ（ＳＢＹ）６０２とを含む。なお、第１のＯＦＣ６００は、２台以上のＯＦＣで冗長化構成されていてもよい。

【００６２】

ＯＦＣ（ＡＣＴ）６０１が保持する、コンフィグ６１１、トポロジ６１２、フローテーブル６１３は、それぞれ、ＯＦＣ（ＳＢＹ）６０２が保持する、コンフィグ６１４、トポ

ロジ６１５、フローテーブル６１６と同期されている。
【００６３】
本実施形態では、第１のＯＦＣ６００が沈黙型故障に対処することに加えて、第２のＯＦＣ６０３が、ビザンチン型故障に対処する。第２のＯＦＣ６０３は、前述したように、信用できるＯＦＣ機能とコンフィグ情報とを有する。なお、第２のＯＦＣは１台に限定されず、複数台配置されていてもよい。
【００６４】
ＯＦＳ６０８～６１０がＯＦＣに接続する際、各ＯＦＳは、複数のＯＦＣと通信可能な方法で接続する。各ＯＦＳは、例えば、オープンフロープロトコルＶｅｒ１．２規格で定義されているマルチコントローラ機能などを用いて、複数のＯＦＣと通信する。
【００６５】
マルチコントローラ機能では、それぞれのＯＦＣには役割が与えられる。「マスタ（Ｍａｓｔｅｒ）」の役割が与えられたＯＦＣは、ＯＦＳからのメッセージを受け取り、そのメッセージに応答し、ＯＦＳが持つフローエントリを更新する権限を有する。一方、「スレーブ（Ｓｌａｖｅ）」の役割が与えられたＯＦＣは、ＯＦＳからのメッセージを受け取ることができるが、メッセージに対する応答を返さない。つまり、スレーブのＯＦＣは、ＯＦＳのフローテーブルを更新することはできない。
【００６６】
ロードバランサ６０４は、後述する役割変更指示機能６０５を有する。なお、役割変更指示機能６０５は、例えば、プログラムに従って動作するコンピュータのＣＰＵによって実現される。当該プログラムは、例えば、ロードバランサ６０４の記憶装置（図示せず）に記憶される。ロードバランサ６０４のＣＰＵは、そのプログラムを読み込み、そのプログラムに従って、役割変更指示機能６０５として動作する。
【００６７】
ロードバランサ６０４は、マスタのＯＦＣとスレーブのＯＦＣに、それぞれ仮想アドレスを与える。また、ロードバランサ６０４は、ＯＦＣの実際のアドレス（以下、実アドレスという）との対応付けを、アドレス変換表６０６で管理する。ロードバランサ６０４のＮＷアドレス変換機能６０７は、アドレス変換表６０６を参照して、パケットヘッダの送信先・宛先情報を書き換える。図２は、アドレス変換表の一例を示す説明図である。
【００６８】
初期状態では、アドレス変換表６０６は、図２に示すようになっている。ＯＦＳは、ロードバランサ６０４の仮想アドレス１にアクセスすることで、結果的に第１のＯＦＣ６００にアクセスできる。また、ロードバランサ６０４の仮想アドレス２にアクセスすることで、結果的に第２のＯＦＣ６０３にアクセスできる。
【００６９】
各ＯＦＳは、接続するＯＦＣのアドレス情報として、マスタとスレーブの二つのアドレス情報をもつ。各ＯＦＳは、マスタのＯＦＣのアドレスとしてロードバランサ６０４の仮想アドレス１を設定し、スレーブのＯＦＣとして仮想アドレス２を設定する。こうすることで、第１のＯＦＣ６００とＯＦＳの間でオープンフロープロトコルによる通信が成立し、フロー制御が可能になる。一方、第２のＯＦＣ６０３では、ＯＦＳから発行されたオープンフローメッセージを受信可能になる。
【００７０】
本実施形態におけるＯＦＣ（第１のＯＦＣ６００のＯＦＣ（ＡＣＴ）６０１、ＯＦＣ（ＳＢＹ）６０２、第２のＯＦＣ６０３）は、「役割管理機能」と、「拡張ＯＦプロトコル処理手段」とを有する。
【００７１】
図３は、第１の実施形態におけるＯＦＣの構成を示すブロック図である。図２を参照して、ＯＦＣの機能について説明する。なお、図３において点線で示す構成要素以外の要素については、図１０に示す要素と同様であるため、詳細な説明を省略する。
【００７２】

図３に示すＯＦＣ７００は、図１０に示すＯＦＣ１００に含まれる構成要素に加え、拡張ＯＦプロトコル処理手段７０１と、役割管理機能７０２とを有する。

【００７３】

役割管理機能７０２は、当該ＯＦＣ（ＯＦＣ７００）に、マスタとしての動作を行わせるか、スレーブとしての動作を行わせるかを管理する機能である。

【００７４】

ＯＦＣ７００がマスタとして動作するように設定された場合、役割管理機能７０２は、拡張ＯＦプロトコル処理手段７０１に対し、全てのオープンフローメッセージを受け付け、必要となる処理を行い、応答を返却するよう指示する。

【００７５】

一方、ＯＦＣ７００がスレーブとして動作するように設定された場合、役割管理機能７０２は、拡張ＯＦプロトコル処理手段７０１に対し、受信したオープンフローメッセージの内、ＬＬＤＰのＰＡＣＫＥＴ＿ＩＮメッセージ以外を破棄するよう指示する。こうすることで、スレーブ側、つまり「スレーブ」の役割が与えられたＯＦＣ７００では、トポロジ検出に必要なメッセージのみを受信できるようになる。ＬＬＤＰのＰＡＣＫＥＴ＿ＩＮメッセージを受け取ったスレーブ側では、トポロジが更新されると、それに伴ってＢＣ／ＭＣフローエントリーが導出され、フローテーブルに設定される。

【００７６】

なお、拡張ＯＦプロトコル処理手段７０１および役割管理機能７０２は、例えば、プログラムに従って動作するコンピュータのＣＰＵによって実現される。当該プログラムは、例えば、ＯＦＣ７００の記憶装置（図示せず）に記憶される。ＯＦＣ７００のＣＰＵは、そのプログラムを読み込み、そのプログラムに従って、拡張ＯＦプロトコル処理手段７０１および役割管理機能７０２として動作する。また、拡張ＯＦプロトコル処理手段７０１および役割管理機能７０２が別々のハードウェアで実現されていてもよい。

【００７７】

次に、本実施形態の動作を説明する。

【００７８】

まず、ビザンチン型故障が発生した場合に、ロードバランサ６０４が、コントローラを第１のＯＦＣから第２のＯＦＣに切り替える動作を説明する。

【００７９】

なお、ビザンチン型故障の発生を検出する手段は様々あり得るが、例として以下のような場合に、ビザンチン型故障の発生を検出することができる。

【００８０】

・ＯＦＣ（ＡＣＴ）のＣＰＵ使用率が異常に上がった場合。この場合、なんらかの要因で、多数の意図しないフロー設定が行われた可能性がある。
・ＯＦＳのフローテーブルに設定されたフローエントリの数が異常に増加した場合。この場合、多数の意図しないフローが設定された可能性がある。

【００８１】

ビザンチン型故障が発生した場合、ロードバランサ６０４は、以下のようにしてＯＦＣを切り替える。図４は、ロードバランサ６０４におけるＯＦＣの切り替え動作を示すフローチャートである。

【００８２】

まず、システム管理者は、ロードバランサ６０４の役割変更指示機能６０５を使用して、ロードバランサ６０４に対して、各コントローラの役割を変更するように指示する。つまり、役割変更指示機能６０５は、システム管理者から操作部（図示せず）等を介して、各コントローラの役割の変更指示を入力する（ステップＳ１）。

【００８３】

なお、ロードバランサ６０４が、ビザンチン型故障の発生を検出して、ステップＳ２以降の処理を開始するようにしてもよい。例えば、ロードバランサ６０４が、ＯＦＣのＣＰＵ使用率を監視可能な場合には、当該ＣＰＵ使用率が所定の閾値を超えたときに、ビザン

チン型故障が発生したと判断することができる。また例えば、ロードバランサ６０４が、ＯＦＳのフローエントリの数を取得可能な場合には、当該フローエントリの数が所定の閾値を超えたときに、ビザンチン型故障が発生したと判断することができる。

【００８４】

ロードバランサ６０４は、ロードバランサ６０４とＯＦＳの間で確立していた通信を切断する（ステップＳ２）。このとき、ロードバランサ６０４は、ＯＦＳからの接続要求の受け入れを停止し、ステップＳ６が完了するまで、ＯＦＳからの接続要求に応じない。

【００８５】

役割変更指示機能６０５は、第１のＯＦＣ６００の役割管理機能に対し、スレーブとして動作するよう通知する（ステップＳ３）。これ以降、第１のＯＦＣ６００は、ＬＬＤＰのＰＡＣＫＥＴ＿ＩＮメッセージのみを受信するようになる。

【００８６】

役割変更指示機能６０５は、第２のＯＦＣ６０３の役割管理機能に対し、マスタとして動作するよう通知する（ステップＳ４）。これ以降、第２のＯＦＣ６０３は、すべてのオープンフローメッセージを受信するようになる。

【００８７】

役割変更指示機能６０５は、アドレス変換表６０６を更新し、マスタの実アドレスとして、第２のＯＦＣ６０３の物理アドレスを設定する。また、スレーブの実アドレスとして、第１のＯＦＣ６００の仮想アドレスを設定する（ステップＳ５）。図５は、ステップＳ５において更新されたアドレス変換表６０６の一例を示す説明図である。

【００８８】

ロードバランサ６０４は、ＯＦＳからの接続要求の受け入れを再開する（ステップＳ６）。

【００８９】

ＯＦＳと第２のＯＦＣ６０３との間で接続が確立され、オープンフロープロトコルによる通信が成立する（ステップＳ７）。ＯＦＳと第１のＯＦＣ６００との間でも接続が確立され、第１のＯＦＣ６００はＬＬＤＰのＰＡＣＫＥＴ＿ＩＮメッセージのみを受信するようになる。

【００９０】

第２のＯＦＣ６０３は、オーディット処理を実行する（ステップＳ８）。これにより、第２のＯＦＣ６０３のフローテーブル６１９が、ＯＦＳのフローテーブル６２０～６２２に同期される。つまり、ＯＦＳにはＢＣ／ＭＣフローのみが設定された状態となり、ＯＦＣの切り替え前に設定されていた意図しないフローは一掃される。

【００９１】

上記ステップＳ１～Ｓ８の処理により、図１に示す通信システムは、信用できるＯＦＣとコンフィグ情報により、フロー制御を再開できるようになる。なお、サービス再開直後は、ソフト転送となる。

【００９２】

以上に説明したように、本実施形態では、マスタのＯＦＣにおいてビザンチン型故障が発生した場合に、最も信用がおけるＯＦＣ機能と最も信用がおけるコンフィグ情報とを有するスレーブのＯＦＣをマスタとして動作させる。それにより、上記（Ａ）に示すコンフィグ情報の不正、上記（Ｂ）に示すＯＦＣ機能の不正、という二つの原因に由来するビザンチン型故障からの復旧が可能となる。

【００９３】

また、本実施形態では、スレーブのＯＦＣが、トポロジ検出に必要なメッセージを受信し、ＢＣ／ＭＣフローエントリーを導出する。それにより、ＯＦＣの役割を「スレーブ」から「マスタ」に切り替える際、当該ＯＦＣにおいてＢＣ／ＭＣフローエントリ導出処理等が不要となり、サービス停止時間を短縮できる。ビザンチン型故障からの復旧に伴うサービス停止時間につき、特に対策しない場合は少なくとも図１１に示す程度のサービス停止時間が発生するが、本実施形態によれば、図６に示す程度のサービス停止時間となる。

図６は、第１の実施形態の通信システムにおいてビザンチン型故障が発生した際の復旧に伴うサービス停止時間を示す説明図である。

【００９４】

また、第２のＯＦＣを、第１のＯＦＣとは離れた場所、例えば、第１のＯＦＣと異なるラック、異なるフロア、異なるＤＣに配置することで、本発明をディザスタリカバリ対策としても応用できる。

【００９５】

実施形態２．

以下、本発明の第２の実施形態を図面を参照して説明する。

【００９６】

本実施形態では、マルチコントローラ機能を使用しないＯＦＳを制御するＯＦＣを含む通信システムを例にする。マルチコントローラ機能を使用しないＯＦＳは、単一のコントローラにしかメッセージを送信できない。つまり、第１のＯＦＣのみでトポロジ情報を検出できる。図７は、本発明による通信システムの第２の実施形態の構成を示す説明図である。

【００９７】

図７に示す通信システムの構成は、第１の実施形態と同様である。

【００９８】

ただし、ロードバランサ１１０７は、役割変更指示機能を有さない。また、第１のＯＦＣ１１００のＯＦＣ（ＡＣＴ）１１０１、ＯＦＣ（ＳＢＹ）１１０２、第２のＯＦＣ１１０３はそれぞれ、役割管理機能および拡張ＯＦプロトコル処理手段の代わりに、トポロジ変更イベント送受信機能１１０４、１１０５、１１０６を有する。

【００９９】

ロードバランサ１１０７は、ＮＷアドレス変換機能１１０８により、単一の仮想アドレスを管理する。つまり、ロードバランサ１１０７は、アドレス変換表１１０９で、当該単一の仮想アドレスとＯＦＣの実アドレスとの対応付けを管理する。図７に示す例では、当該単一の仮想アドレスとして、「ＬＢの仮想ＩＰアドレス」が設定されている。また、ＯＦＣの実アドレスとして、「第１のＯＦＣの仮想アドレス」が設定されている。ＯＦＳ１１０８～１１１０は、ロードバランサ１１０７の当該単一の仮想アドレスに接続する。

【０１００】

ロードバランサ１１０７は、アドレス変換表１１０９に従い、第１のＯＦＣ１１００または第２のＯＦＣ１１０３にパケットを中継する。

【０１０１】

ＯＦＳ１１１０～１１１２と接続されたＯＦＣは、オープンフロー制御を行う。また、各ＯＦＣは、受信したオープンフローメッセージがＬＬＤＰのＰＡＣＫＥＴ＿ＩＮメッセージだった場合は、トポロジ変更イベント送受信機能を用いて、対向のＯＦＣノードにメッセージを通知する。これにより、オープンフローメッセージを直接受信していないＯＦＣノードでもトポロジ情報を生成することが可能になり、結果、常時ＢＣ／ＭＣフローエントリーを導出することが可能となる。

【０１０２】

このように、ロードバランサが保持するアドレス変換表の実アドレスに、切り替え先ＯＦＣのアドレスを設定することで、ノード切り替えを実行できる。つまり、本実施形態によれば、ＯＦＣのＯＦプロトコル処理手段等にバグや改ざんなどがなく、対向ノードに伝達するメッセージに誤りが含まれる可能性がなければ、図７に示すような簡易な構成により、第１の実施形態と同様の効果を得ることができる。

【０１０３】

なお、各ＯＦＣのトポロジ変更イベント送受信機能（トポロジ変更イベント送受信機能１１０４、１１０５、１１０６）は、例えば、プログラムに従って動作するコンピュータのＣＰＵによって実現される。当該プログラムは、例えば、各ＯＦＣの記憶装置（図示せず）に記憶される。各ＯＦＣのＣＰＵは、そのプログラムを読み込み、そのプログラムに

従って、トポロジ変更イベント送受信機能（トポロジ変更イベント送受信機能１１０４、１１０５、１１０６）として動作する。

【０１０４】

以上、各実施形態においてオープンフローに適用される通信システムを例にしたが、本発明はオープンフロー以外にも適用可能である。例えば、各実施形態におけるＯＦＣは、オープンフローにおけるコントローラ以外の制御装置であってもよい。また例えば、各実施形態におけるＯＦＳは、オープンフローにおけるスイッチ以外のパケット転送装置であってもよい。つまり、通信ネットワーク上の各パケット転送装置を制御装置が集中管理する構成の通信システムであれば、本発明を適用することができる。

【０１０５】

次に、本発明の概要を説明する。図８は、本発明による通信システムの最小構成を示すブロック図である。本発明による通信システムは、フロー制御機能（例えば、オープンフローにおけるＯＦＣ機能）と仮想ネットワークの設定情報（例えば、オープンフローにおけるコンフィグ情報）とをもとに複数のパケット転送装置を制御する第１の制御装置１０（図１に示す第１のＯＦＣ６００に相当）と、動作実績があるフロー制御機能と、運用実績がある仮想ネットワークの設定情報とを有し、第１の制御装置１０に故障が発生した場合に切り替え先となる第２の制御装置２０（図１に示す第２のＯＦＣ６０３に相当）と、複数のパケット転送装置と各制御装置との間に配置された負荷分散装置３０（図１に示すロードバランサ６０４に相当）とを備え、負荷分散装置３０は、第１の制御装置１０においてビザンチン型故障が発生した場合に、複数のパケット転送装置の制御を第２の制御装置２０に実行させる役割変更指示機能３１（図１に示すロードバランサ６０４における役割変更指示機能６０５に相当）を有する。

【０１０６】

そのような構成によれば、第１の制御装置においてビザンチン型故障が発生した場合に、最も信用がおけるフロー制御機能と最も信用がおけるネットワークの設定情報とを有する第２の制御装置を動作させることができる。それにより、例えば、オープンフローにおいて、コンフィグ情報の不正、ＯＦＣ機能の不正、という二つの原因に由来するビザンチン型故障からの復旧が可能となる。従って、通信システムにおいて、ビザンチン型故障の発生を検出した場合に、早期に故障状態を解消しパケット転送を再開することができる。

【０１０７】

また、負荷分散装置３０の役割変更指示機能３１は、第１の制御装置１０のＣＰＵ使用率が所定の閾値を超えたときに、当該第１の制御装置においてビザンチン型故障が発生したと判断してもよい。そのような構成によれば、ビザンチン型故障をより確実に検出することができ、より早期に故障状態を解消しパケット転送を再開することができる。

【０１０８】

また、通常運用時、第１の制御装置１０は、複数のパケット転送装置から送信される全てのメッセージを受信するマスタとして動作し、第２の制御装置２０は、複数のパケット転送装置からトポロジ検出に必要なメッセージのみを受信するスレーブとして動作し、負荷分散装置３０の役割変更指示機能３１は、第１の制御装置１０においてビザンチン型故障が発生した場合に、第１の制御装置１０をスレーブとして動作させ、第２の制御装置２０をマスタとして動作させてもよい。そのような構成によれば、第２の制御装置は、マスタとしてフロー制御を開始する際に必要なブロードキャストやマルチキャストされたパケットを適切に転送するための情報（例えば、オープンフローにおけるＢＣ／ＭＣフロー）を、スレーブ動作中に計算し保持しておくことができる。

【０１０９】

また、各制御装置は、負荷分散装置３０からの役割変更指示に基づいて、自装置をマスタとして動作させるか、スレーブとして動作させるかを管理する役割管理機能（図３に示す役割管理機能７０２に相当）を有していてもよい。そのような構成によれば、負荷分散装置は、役割変更指示を出力するだけで各制御装置の役割を変更することが可能となる。

【０１１０】

また、第２の制御装置２０の役割管理機能は、スレーブ動作中に、トポロジ検出に必要なメッセージ（例えば、オープンフローにおけるＬＬＤＰのＰＡＣＫＥＴ＿ＩＮメッセージ）を受信すると、当該メッセージをもとにマスタとして動作を開始する際に必要なフロー（例えば、オープンフローにおけるＢＣ／ＭＣフロー）を作成してもよい。そのような構成によれば、第２の制御装置２０は、マスタとしてフロー制御を開始する際に、マスタとして動作を開始する際に必要なフローを作成する必要がないので、サービス停止時間をより短縮することができる。

【０１１１】

また、各制御装置は、パケット転送装置から受信したメッセージがトポロジ検出に必要なメッセージであると判断した場合に、他の制御装置に当該メッセージを通知するトポロジ変更イベント送受信機能（図７に示すトポロジ変更イベント送受信機能１１０４～１１０６に相当）を有してもよい。そのような構成によれば、より簡易な構成で、ビザンチン型故障の発生を検出した場合の、故障状態の早期解消およびパケット転送の再開を行うことができる。

【０１１２】

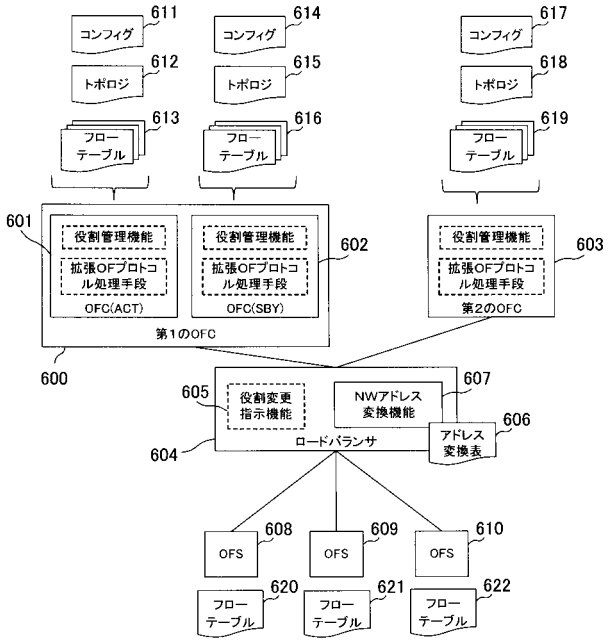
また、第１の制御装置１０が、複数の制御装置（例えば、図１に示すＯＦＣ（ＡＣＴ）６０１、ＯＦＣ（ＳＢＹ）６０２）によりクラスタ構成されていてもよい。そのような構成によれば、通信システムを、ビザンチン型故障だけでなく、沈黙型故障にも対応させることができる。

【符号の説明】

【０１１３】

１０　第１の制御装置
２０　第２の制御装置
３０　負荷分散装置
３１、６０５　役割変更指示機能
１００、７００　ＯＦＣ
１０１～１０３、４０３～４０５、６０８～６１０、１１１０～１１１２　ＯＦＳ
１１０、２０６、４１０、４１３、６１１、６１４、６１７　コンフィグ
１１１、２０４、４１１、４１４、６１２、６１５、６１８　トポロジ
１１２、１１３～１１５、２０８、４１２、４１５、４１７～４１９、６１３、６１６、６１９、６２０～６２２　フローテーブル
２０１　通信手段
２０２　ＯＦプロトコル処理手段
２０３　トポロジ検出手段
２０５　コンフィグ管理手段
２０７　経路管理手段
２０９　経路計算手段
２１０　経路設定手段
２１１　オーディット手段
４００、６０１、１１０１　ＯＦＣ（ＡＣＴ）
４０１、６０２、１１０２　ＯＦＣ（ＳＢＹ）
６００、１１００　第１のＯＦＣ
６０３、１１０３　第２のＯＦＣ
６０４、１１０７　ロードバランサ
６０７、１１０８　ＮＷアドレス変換機能
６０６、１１０９　アドレス変換表
７０１　拡張ＯＦプロトコル処理手段
７０２　役割管理機能
１１０４～１１０６　トポロジ変更イベント送受信機能

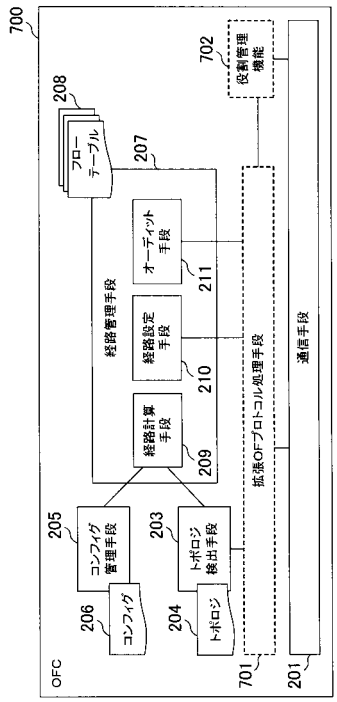
【図１】



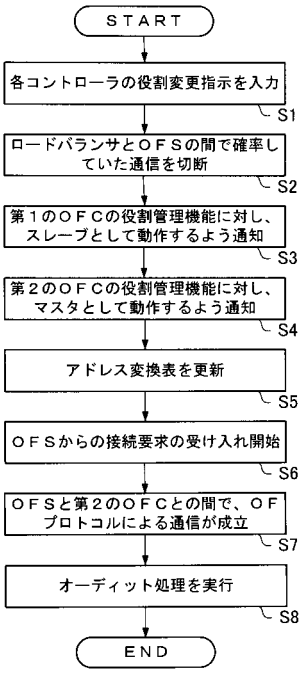
【図２】

OFCの役割	仮想アドレス	実アドレス
Master	LBの仮想アドレス1	第1のOFCの仮想アドレス
Slave	LBの仮想アドレス2	第2のOFCの物理アドレス

【図３】



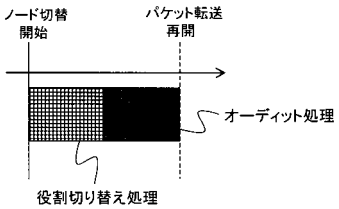
【図４】



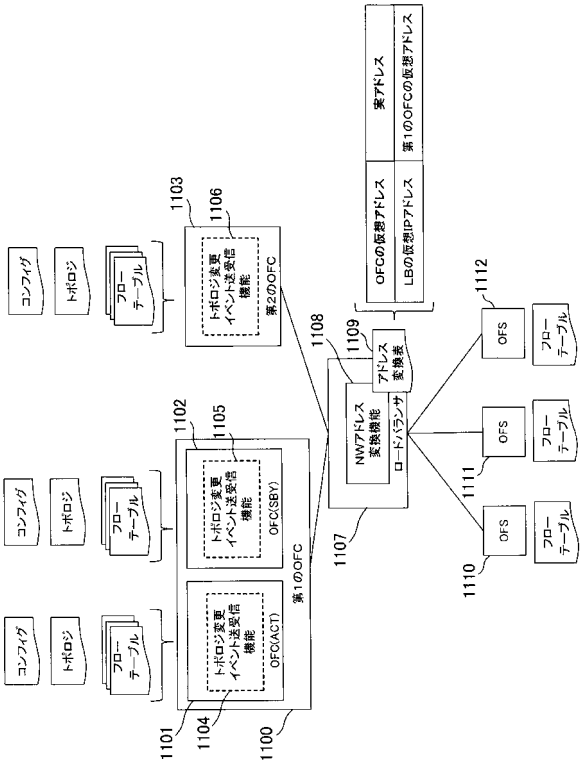
【図５】

OFCの役割	仮想アドレス	実アドレス
Master	LBの仮想アドレス1	第2のOFCの物理アドレス
Slave	LBの仮想アドレス2	第1のOFCの仮想アドレス

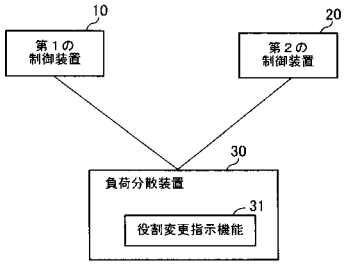
【図６】



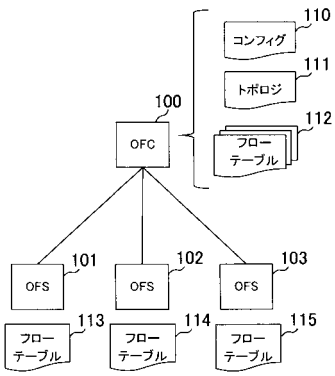
【図７】



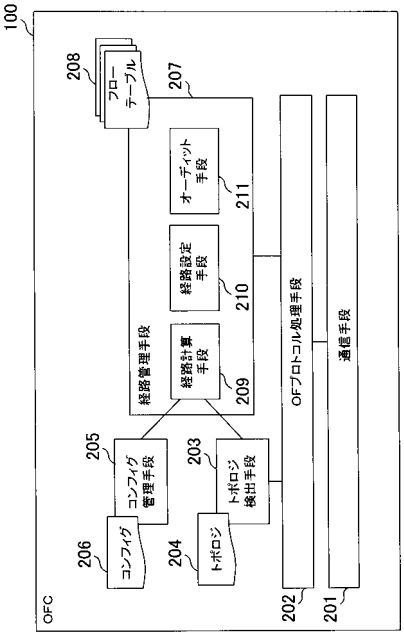
【図８】



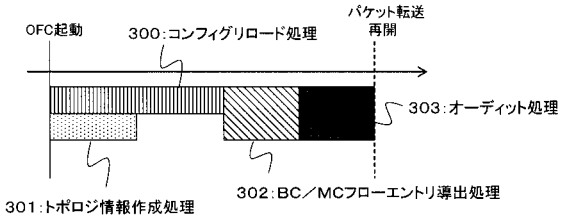
【図９】



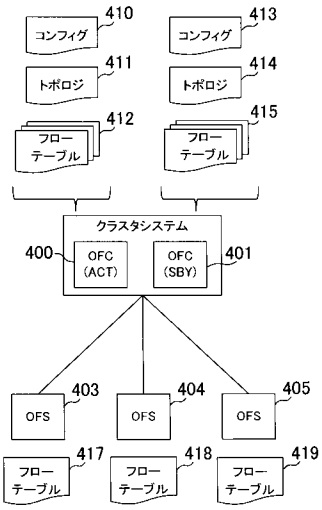
【図１０】



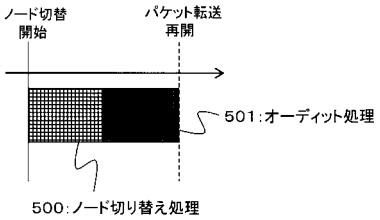
【図１１】


OFC起動
300:コンフィグリロード処理
パケット転送
再開
303:オーディット処理
301:トポロジ情報作成処理
302:BC／MCフローエントリ導出処理

【図１２】


410
コンフィグ
411
トポロジ
412
フロー
テーブル
413
コンフィグ
414
トポロジ
415
フロー
テーブル
クラスタシステム
400
OFC
(ACT)
OFC
(SBY)
401
403
OFS
404
OFS
405
OFS
417
フロー
テーブル
418
フロー
テーブル
419
フロー
テーブル

【図１３】


ノード切替
開始
パケット転送
再開
501:オーディット処理
500:ノード切り替え処理