

(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2015-128213

(P2015-128213A)

(43) 公開日 平成27年7月9日(2015.7.9)

(51) Int.Cl.  
H04L 12/717 (2013.01)

F I  
H04L 12/717

テーマコード (参考)  
5K030

審査請求 未請求 請求項の数 10 O L (全 22 頁)

(21) 出願番号 特願2013-272650 (P2013-272650)  
(22) 出願日 平成25年12月27日 (2013.12.27)

(特許庁注：以下のものは登録商標)

1. ETHERNET

(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 100080816  
弁理士 加藤 朝道  
(72) 発明者 戸川 修  
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 GA11 HA08 HC01 HD03 LB07  
LE01 MB01 MD01

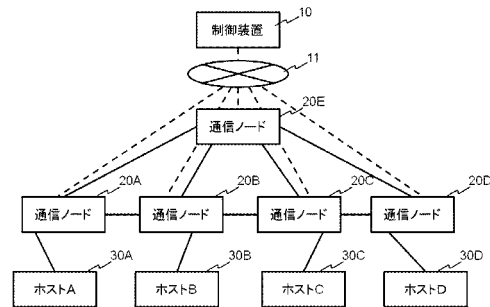
(54) 【発明の名称】 通信ノード、制御装置、通信システム、通信方法及びプログラム

(57) 【要約】

【課題】集中制御型のネットワークの通信ノードに、制御装置に障害が発生した際の新規通信の処理機能を追加する。

【解決手段】通信ノードは、制御装置から設定された第1の制御情報と、前記第1の制御情報のマッチ条件に適合しないパケットを受信した際に、前記制御装置に対する通知を指示する第2の制御情報と、前記第1の制御情報及び第2の制御情報よりも低い優先度が設定され、受信パケットを指定されたポートから出力させる第3の制御情報と、を保持する制御情報記憶部と、受信パケットに適合するマッチ条件を持つ制御情報を前記制御情報記憶部から検索して、実行するパケット処理部と、前記制御装置が通信ノードを制御できなくなったことを検出した際に、前記第3の制御情報が、前記第2の制御情報に優先して適用されるよう前記制御情報を書き換える動作モード切替部と、を備える。

【選択図】 図1



**【特許請求の範囲】****【請求項 1】**

制御装置から設定された第 1 の制御情報と、前記第 1 の制御情報のマッチ条件に適合しないパケットを受信した際に、前記制御装置に対する通知を指示する第 2 の制御情報と、前記第 1 の制御情報及び第 2 の制御情報よりも低い優先度が設定され、受信パケットを指定されたポートから出力させる第 3 の制御情報と、を保持する制御情報記憶部と、

受信パケットに適合するマッチ条件を持つ制御情報を前記制御情報記憶部から検索して、実行するパケット処理部と、

前記制御装置が通信ノードを制御できなくなったことを検出した際に、前記第 3 の制御情報が、前記第 2 の制御情報に優先して適用されるよう前記制御情報を書き換える動作モード切替部と、

を備える通信ノード。

**【請求項 2】**

前記第 3 の制御情報は、前記第 1 の制御情報のマッチ条件に適合しないパケットをフラッディングさせるための制御情報である請求項 1 の通信ノード。

**【請求項 3】**

前記第 3 の制御情報によって、所定の代替ルータに、前記第 1 の制御情報のマッチ条件に適合しないパケットを転送する請求項 2 の通信ノード。

**【請求項 4】**

さらに、ホストからのアドレス解決要求に対して応答を指示する第 4 の制御情報が設定されている請求項 1 から 3 いずれか一の通信ノード。

**【請求項 5】**

制御装置から設定された第 1 の制御情報と、前記第 1 の制御情報のマッチ条件に適合しないパケットを受信した際に、前記制御装置に対する通知を指示する第 2 の制御情報と、を保持する制御情報記憶部と、受信パケットに適合するマッチ条件を持つ制御情報を前記制御情報記憶部から検索して、実行するパケット処理部と、通信ノードにおいて制御装置が通信ノードを制御できなくなったことを検出した際に、前記第 3 の制御情報が、前記第 2 の制御情報に優先して適用されるよう前記制御情報を書き換える動作モード切替部と、を備えた通信ノードに対し、前記第 1 の制御情報及び第 2 の制御情報よりも低い優先度が設定され、受信パケットを指定されたポートから出力させる第 3 の制御情報を設定する制御装置。

**【請求項 6】**

前記第 3 の制御情報を前記通信ノードに設定することで、前記通信ノードに、前記第 1 の制御情報のマッチ条件に適合しないパケットをフラッディングさせる請求項 5 の制御装置。

**【請求項 7】**

前記第 3 の制御情報を前記通信ノードに設定することで、前記通信ノードに、前記第 1 の制御情報のマッチ条件に適合しないパケットを所定の代替ルータへと転送させる請求項 6 の制御装置。

**【請求項 8】**

さらに、前記通信ノードに、ホストからのアドレス解決要求に対して応答を指示する第 4 の制御情報を設定する請求項 5 から 7 いずれか一の制御装置。

**【請求項 9】**

制御装置から設定された第 1 の制御情報と、

前記第 1 の制御情報のマッチ条件に適合しないパケットを受信した際に、前記制御装置に対する通知を指示する第 2 の制御情報と、

前記第 1 の制御情報及び第 2 の制御情報よりも低い優先度が設定され、受信パケットを指定されたポートから出力させる第 3 の制御情報と、を保持する制御情報記憶部と、

受信パケットに適合するマッチ条件を持つ制御情報を前記制御情報記憶部から検索して、実行するパケット処理部と、

10

20

30

40

50

前記制御装置が通信ノードを制御できなくなったことを検出した際に、前記第3の制御情報が、前記第2の制御情報に優先して適用されるよう前記制御情報を書き換える動作モード切替部と、

を備える通信ノードと、

前記通信ノードに対し、前記第1の制御情報及び前記第3の制御情報を設定する制御装置と、を含む通信システム。

#### 【請求項10】

制御装置から設定された第1の制御情報と、前記第1の制御情報のマッチ条件に適合しないパケットを受信した際に、前記制御装置に対する通知を指示する第2の制御情報と、前記第1の制御情報及び第2の制御情報よりも低い優先度が設定され、受信パケットを指定されたポートから出力させる第3の制御情報と、を保持する制御情報記憶部と、

受信パケットに適合するマッチ条件を持つ制御情報を前記制御情報記憶部から検索して、実行するパケット処理部と、を備えた通信ノードが、

前記制御装置が通信ノードを制御できなくなっているか否かを確認するステップと、

前記制御装置が通信ノードを制御できなくなったことを検出した場合、前記第3の制御情報が、前記第2の制御情報に優先して適用されるよう前記制御情報を書き換えるステップと、

を含む通信方法。

#### 【発明の詳細な説明】

#### 【技術分野】

#### 【0001】

本発明は、通信ノード、制御装置、通信システム、通信方法及びプログラムに関し、特に、制御装置からの指示によって動作する通信ノード、その制御装置、通信システム、通信方法及びプログラムに関する。

#### 【背景技術】

#### 【0002】

近年、オープンフロー（OpenFlow）という技術が提案されている（非特許文献1、2参照）。オープンフローは、通信をエンドツーエンドのフローとして捉え、フロー単位で経路制御、障害回復、負荷分散、最適化を行うものである。非特許文献2に仕様化されているオープンフロースイッチは、オープンフローコントローラとの通信用のセキュアチャネルを備え、オープンフローコントローラから適宜追加または書き換え指示されるフローテーブルに従って動作する。フローテーブルには、フロー毎に、パケットヘッダと照合するマッチ条件（Match Fields）と、フロー統計情報（Counters）と、処理内容を定義したインストラクション（Instructions）と、の組が定義される（非特許文献2の「4.1 Flow Table」の項参照）。

#### 【0003】

例えば、オープンフロースイッチは、パケットを受信すると、フローテーブルから、受信パケットのヘッダ情報に適合するマッチ条件（非特許文献2の「4.3 Match Fields」参照）を持つエントリを検索する。検索の結果、受信パケットに適合するエントリが見つかった場合、オープンフロースイッチは、フロー統計情報（カウンタ）を更新するとともに、受信パケットに対して、当該エントリのインストラクションフィールドに記述された処理内容（指定ポートからのパケット送信、フラッディング、廃棄等）を実施する。一方、検索の結果、受信パケットに適合するエントリが見つからなかった場合、オープンフロースイッチは、セキュアチャネルを介して、オープンフローコントローラに対してエントリ設定の要求、即ち、受信パケットを処理するための制御情報の送信要求（Packet-Inメッセージ）を送信する。オープンフロースイッチは、処理内容が定められたフローエントリを受け取ってフローテーブルを更新する。このように、オープンフロースイッチは、フローテーブルに格納されたエントリを制御情報として用いてパケット転送を行う。

#### 【0004】

10

20

30

40

50

また、特許文献1には、上記オープンフローのような集中制御型のネットワークにおいてスイッチとコントローラ間に通信障害が発生した場合に、現用系のスイッチが現用経路上の外部スイッチとのポートのリンクダウンを実施し、冗長経路に切り替えるようにした構成が開示されている。

【0005】

また、特許文献2、3には、コントローラを複数設け、スイッチが、これら複数のコントローラの1つを経路決定者として指定する構成や、フローエントリの設定時に優先度も指定して、スイッチがフローエントリの設定可否を判断する構成が開示されている。

【先行技術文献】

【特許文献】

10

【0006】

【特許文献1】国際公開第2012/165446号

【特許文献2】特開2011-160363号公報

【特許文献3】特開2011-166384号公報

【非特許文献】

【0007】

【非特許文献1】Nick McKeownほか7名、“OpenFlow: Enabling Innovation in Campus Networks”、[online]、[平成25(2013)年8月7日検索]、インターネット URL: <http://www.openflow.org/documents/openflow-wp-latest.pdf>

20

【非特許文献2】“OpenFlow Switch Specification” Version 1.1.0 Implemented (Wire Protocol 0x02)、[online]、[平成25(2013)年8月7日検索]、インターネット URL: <http://www.openflow.org/documents/openflow-spec-v1.1.0.pdf>

【発明の概要】

【発明が解決しようとする課題】

【0008】

以下の分析は、本発明によって与えられたものである。特許文献1にも指摘されているように、上記したオープンフローに代表される集中制御型のネットワークでは、制御装置と、スイッチとの間のセキュアチャネルに障害が起きるとスイッチの制御ができなくなってしまうという問題点がある。とりわけ、制御装置に障害が起きると、上記Packet-Inメッセージの処理ができなくなってしまうため、新規通信も不可能となる。

30

【0009】

この点、特許文献1では、複数のスイッチのうち、外部通信装置と接続された現用系のスイッチが、制御装置(コントローラ)との通信障害を検出した際に、外部通信装置との接続ポートのリンクダウンを実施することで、待機系のスイッチに切り替えるものであり(経路の冗長化)、制御装置の障害全般に対応できるものとはなっていない。

【0010】

特許文献2、3は、複数の制御装置を用意するものであり(機器の冗長化)、コストの増大を伴ってしまうのはもちろんとして、制御装置単体での障害対策とはなっていない。

40

【0011】

本発明は、集中制御型のネットワークにおいて、制御装置に障害が発生した際の新規通信の処理機能を追加した通信ノード、制御装置、通信システム、通信方法及びプログラム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】

【0012】

第1の視点によれば、制御装置から設定された第1の制御情報と、前記第1の制御情報のマッチ条件に適合しないパケットを受信した際に、前記制御装置に対する通知を指示する第2の制御情報と、前記第1の制御情報及び第2の制御情報よりも低い優先度が設定され、受信パケットを指定されたポートから出力させる第3の制御情報と、を保持する制御

50

情報記憶部と、受信パケットに適合するマッチ条件を持つ制御情報を前記制御情報記憶部から検索して、実行するパケット処理部と、前記制御装置が通信ノードを制御できなくなったことを検出した際に、前記第3の制御情報が、前記第2の制御情報に優先して適用されるよう前記制御情報を書き換える動作モード切替部と、を備える通信ノードが提供される。

【0013】

第2の視点によれば、制御装置から設定された第1の制御情報と、前記第1の制御情報のマッチ条件に適合しないパケットを受信した際に、前記制御装置に対する通知を指示する第2の制御情報と、を保持する制御情報記憶部と、受信パケットに適合するマッチ条件を持つ制御情報を前記制御情報記憶部から検索して、実行するパケット処理部と、通信ノードにおいて制御装置が通信ノードを制御できなくなったことを検出した際に、前記第3の制御情報が、前記第2の制御情報に優先して適用されるよう前記制御情報を書き換える動作モード切替部と、を備えた通信ノードに対し、前記第1の制御情報及び第2の制御情報よりも低い優先度が設定され、受信パケットを指定されたポートから出力させる第3の制御情報を設定する制御装置が提供される。

10

【0014】

第3の視点によれば、上記した通信ノードと、制御装置との組み合わせにより、耐障害性を向上させた通信システムが提供される。

【0015】

第4の視点によれば、制御装置から設定された第1の制御情報と、前記第1の制御情報のマッチ条件に適合しないパケットを受信した際に、前記制御装置に対する通知を指示する第2の制御情報と、前記第1の制御情報及び第2の制御情報よりも低い優先度が設定され、受信パケットを指定されたポートから出力させる第3の制御情報と、を保持する制御情報記憶部と、受信パケットに適合するマッチ条件を持つ制御情報を前記制御情報記憶部から検索して、実行するパケット処理部と、を備えた通信ノードが、前記制御装置が通信ノードを制御できなくなっているか否かを確認するステップと、前記制御装置が通信ノードを制御できなくなったことを検出した場合、前記第3の制御情報が、前記第2の制御情報に優先して適用されるよう前記制御情報を書き換えるステップと、を含む通信方法が提供される。本方法は、前記第1～第3の制御情報に従って受信パケットを処理する通信ノードという、特定の機械に結びつけられている。

20

30

【0016】

第5の視点によれば、制御装置から設定された第1の制御情報と、前記第1の制御情報のマッチ条件に適合しないパケットを受信した際に、前記制御装置に対する通知を指示する第2の制御情報と、前記第1の制御情報及び第2の制御情報よりも低い優先度が設定され、受信パケットを指定されたポートから出力させる第3の制御情報と、を保持する制御情報記憶部と、受信パケットに適合するマッチ条件を持つ制御情報を前記制御情報記憶部から検索して、実行するパケット処理部と、を備えた通信ノードに搭載されたコンピュータに、前記制御装置が通信ノードを制御できなくなっているか否かを確認する処理と、前記制御装置が通信ノードを制御できなくなったことを検出した場合、前記第3の制御情報が、前記第2の制御情報に優先して適用されるよう前記制御情報を書き換える処理と、を実行させるプログラムが提供される。なお、このプログラムは、コンピュータが読み取り可能な（非トランジエントな）記憶媒体に記録することができる。即ち、本発明は、コンピュータプログラム製品として具現することも可能である。

40

【発明の効果】

【0017】

本発明によれば、集中制御型のネットワークの通信ノードに、制御装置に障害が発生した際の新規通信の処理機能を追加することが可能となる。

【図面の簡単な説明】

【0018】

【図1】第1の実施形態の通信システムの構成を示す図である。

50

- 【図 2】第 1 の実施形態の通信ノードの詳細構成を示すブロック図である。
- 【図 3】第 1 の実施形態の通信ノードに設定される制御情報の一例を示す図である。
- 【図 4】第 1 の実施形態の通信ノードに設定される制御情報の別の一例を示す図である。
- 【図 5】第 3 の制御情報に対応するフラッディング用経路を示す図である。
- 【図 6】図 3 の状態から第 2 の制御情報を削除した状態を示す図である。
- 【図 7】制御装置に障害が発生しているときのホスト A からの新規通信パケットの転送経路を示す図である。
- 【図 8】第 2 の実施形態の通信システムの構成を示す図である。
- 【図 9】第 2 の実施形態の通信ノードに設定される制御情報の一例を示す図である。
- 【図 10】第 2 の実施形態の通信ノードに設定される制御情報の別の一例を示す図である。
- 【図 11】制御装置内の仮想ルータと、代替ルータへのサブネットとアドレスの設定例を示す図である。
- 【図 12】図 8 に、図 11 の情報を付記した図である。
- 【図 13】第 2 の実施形態における第 2 の制御情報に対応するフラッディング用経路を示す図である。
- 【図 14】図 9 の状態から第 2 の制御情報を削除した状態を示す図である。
- 【図 15】図 10 の状態から第 2 の制御情報を削除した状態を示す図である。
- 【図 16】第 2 の実施形態の制御装置に障害が発生したときの新規通信の処理を説明するための図である。
- 【図 17】ホスト A からの IP アドレス R 1 の MAC アドレスを問い合わせる ARP パケットの例を示す図である。
- 【図 18】IP アドレス R 1 の MAC アドレスを問い合わせる ARP パケットに対する ARP 応答パケットの例を示す図である。
- 【図 19】ホスト A からのホスト B 宛ての IP パケットの例を示す図である。
- 【図 20】代替ルータによる書き換えられたパケットヘッダを示す図である。
- 【図 21】第 2 の実施形態の制御装置に障害が発生したときの既存通信の処理を説明するための図である。
- 【図 22】ホスト C からの IP アドレス R 1 の MAC アドレスを問い合わせる ARP パケットの例を示す図である。
- 【図 23】IP アドレス R 1 の MAC アドレスを問い合わせる ARP パケットに対する ARP 応答パケットの例を示す図である。
- 【図 24】第 2 の実施形態の制御装置が、通信ノードに設定する第 4 の制御情報の例を示す図である。
- 【図 25】ホスト C からのホスト D 宛ての IP パケットの例を示す図である。
- 【図 26】代替ルータによる書き換えられたパケットヘッダを示す図である。
- 【図 27】第 2 の実施形態の変形実施形態を説明するための図である。
- 【図 28】第 2 の実施形態の通信システムの変形構成を示す図である。
- 【図 29】図 4 のテーブルの変形例を示す図である。
- 【発明を実施するための形態】
- 【0019】
- はじめに本発明の一実施形態の概要について図面を参照して説明する。なお、この概要に付記した図面参照符号は、理解を助けるための一例として各要素に便宜上付記したものであり、本発明を図示の態様に限定することを意図するものではない。
- 【0020】
- 本発明は、その一実施形態において、制御情報を保持する制御情報記憶部（図 2 の 201）と、受信パケットに適合するマッチ条件を持つ制御情報を前記制御情報記憶部から検索して、実行するパケット処理部（図 2 の 202）と、前記制御装置との通信障害を検出した際に、前記制御情報記憶部から、前記第 2 の制御情報を削除する動作モード切替部（図 2 の 203）と、を備えた通信ノード（図 2 の 20A）にて実現できる。

## 【 0 0 2 1 】

より具体的には、制御情報記憶部（図 2 の 2 0 1 ）には、制御装置から設定された第 1 の制御情報と、前記第 1 の制御情報のマッチ条件に適合しないパケットを受信した際に、前記制御装置に対する通知を指示する第 2 の制御情報と、前記第 1 の制御情報及び第 2 の制御情報よりも低い優先度が設定され、受信パケットを指定されたポートから出力させる第 3 の制御情報と、が設定される（図 3 参照）。そして、制御装置が通信ノードを制御できなくなっていること（例えば、通信障害）を検出した際に、動作モード切替部（図 2 の 2 0 3 ）は、前記第 3 の制御情報が、前記第 2 の制御情報に優先して適用されるよう前記制御情報を書き換える（例えば、第 2 の制御情報の削除、第 2 の制御情報のマッチ条件の変更（一時無効化）、第 2 の制御情報の優先度の引き下げ、前記第 3 の制御情報の優先度の引き上げ、前記第 2、第 3 の制御情報の優先度の入れ替え等）。これにより、第 1 の制御情報による既存通信の実行のみならず、第 3 の制御情報による新規通信の処理（パケットを指定されたポートから出力）が開始される。

10

## 【 0 0 2 2 】

以上のようにして、通信ノードに、制御装置に障害が発生した際の新規通信の処理機能が追加される。

## 【 0 0 2 3 】

## [ 第 1 の実施形態 ]

続いて、レイヤ 2（L 2）転送の論理ネットワークにおいて、冗長化によらず、制御装置に障害が発生した場合においても、新規通信を処理できるようにした第 1 の実施形態について説明する。

20

## 【 0 0 2 4 】

図 1 は、本発明の第 1 の実施形態の通信システムの構成を示す図である。図 1 を参照すると、所定のネットワークトポロジで接続された通信ノード 2 0 A ~ 2 0 E と、これら通信ノード 2 0 A ~ 2 0 E を制御チャネル 1 1 を介して制御する制御装置 1 0 と、を含む構成が示されている。また、図 1 の例では、通信ノード 2 0 A ~ 2 0 D にそれぞれホスト 3 0 A ~ 3 0 D が接続可能となっている。以下の実施形態では、これらのホスト間の通信を実現する例を挙げて説明する。

## 【 0 0 2 5 】

図 2 は、第 1 の実施形態の通信ノード 2 0 A の詳細構成を示すブロック図である。図 2 を参照すると、通信ノード 2 0 A は、制御情報記憶部 2 0 1 と、パケット処理部 2 0 2 と、制御メッセージ送受信部（動作モード切替部）2 0 3 とを備えている。なお、通信ノード 2 0 B ~ 2 0 E は通信ノード 2 0 A と同様の構成であるので省略する。また以下、特に個々の通信ノードを区別しないときは、「通信ノード 2 0」とも記す。

30

## 【 0 0 2 6 】

制御情報記憶部 2 0 1 は、受信パケットを処理するための制御情報を記憶する。図 3、図 4 は、通信ノードに設定される制御情報の一例を示す図である。制御情報は、受信パケットのヘッダ等と照合して処理対象のパケットを特定するマッチ条件と、このマッチ条件に適合するパケットに適用する処理を示すインストラクションとを対応づけて構成される。また、図 3、図 4 の例では、各制御情報のエントリの優先度を示す優先度フィールドを設けている。例えば、あるパケットに適合するマッチ条件を持つ制御情報が 2 つ以上ある場合、優先度の高い制御情報が選択される。なお、図 3、図 4 の例では、明示的に優先度フィールドを設けているが、制御情報を格納するテーブル上の位置が優先度を表すものとしてもよい。例えば、優先度の高い順に制御情報エントリを格納することで、テーブルの上位から受信パケットに適合する制御情報を探索していき、該当する制御情報が見つかったところで、探索を終了することができる。なお、制御情報記憶部 2 0 1 は、非特許文献 2 のフローエントリを格納するフローテーブルによって構成することもできる。

40

## 【 0 0 2 7 】

図 3、図 4 の例では、第 1 の制御情報 ~ 第 3 の制御情報が配置されている。第 1 の制御情報は、制御装置によって設定された L 2 転送用（ユニキャスト用）の制御情報である。

50

図3は、図1の通信ノード20Aに設定される制御情報を示している。この場合、通信ノード20Aは、MAC(Media Access Control)アドレス等によって特定されるホストAからホストB宛てのパケットを受信すると、通信ノード20Bに転送する動作を行う。同様に、通信ノード20Bが制御情報を参照して、ホストAからホストB宛てのパケットをホストBに転送することで、ホストAからのパケットがホストBに届けられることになる。図4は、図1の通信ノード20Eに設定される制御情報を示している。この場合、通信ノード20Eは、通信ノード20Cを介して、ホストCからホストD宛てのパケットを受信すると、通信ノード20Dに転送する動作を行う。同様に、通信ノード20Dが制御情報を参照して、ホストCからホストD宛てのパケットをホストDに転送することで、ホストCからのパケットがホストDに届けられることになる。

10

**【0028】**

図3、図4の第2の制御情報は、上記した第1の制御情報のマッチ条件に適合しない、すなわち、新規に発生した通信のパケットを捕捉して、制御装置10に通知するための制御情報である。図3、図4の例では、マッチ条件をANY(任意/ワイルドカード)とし、インストラクションとして、制御装置へのパケットの転送(非特許文献2のPacket-Inメッセージに相当)を指示するエントリが設定されている。第2の制御情報のマッチ条件はANYなので、通常、新規に発生した通信のパケットは、第2の制御情報にマッチすることになる。従って、第2の制御情報が削除されない限り、より低い優先度が設定された第3の制御情報が使用されることはない。なお、第2の制御情報のマッチ条件は、ANY(任意)とせず、特定のパケットを破棄(Drop)するようにしてもよい。

20

**【0029】**

図3、図4の第3の制御情報は、第2の制御情報が削除される(制御装置に障害が発生することによって、パケット処理部202から参照されるようになる新規通信の処理用の制御情報である。本実施形態では、図5に示すようなフラッディング転送を行う経路に従って、新規通信パケットを転送させる制御情報を設定する。このようにすることで制御装置に障害が発生した際の新規通信を処理可能となる。図3の例では、第3の制御情報として、第1の制御情報にマッチしなかったパケット(新規通信パケット)のうち、ホストA側から受信したパケットを通信ノード20Eに転送させる制御情報と、通信ノード20E側から受信したパケットをホストAに転送させる制御情報とが設定されている。同様に、図4の例では、第3の制御情報として、第1の制御情報にマッチしなかったパケット(新規通信パケット)を、それぞれ受信ポート以外のポートに接続された通信ノードに転送させる制御情報が設定されている。同様に、通信ノード20B、20C、20Dにも、それぞれ第1の制御情報にマッチしなかったパケット(新規通信パケット)を、フラッディング転送を行う第3の制御情報を設定することにより、制御装置に障害が発生していても、ホストAからの新規通信パケットを宛先とするホストに転送することが可能となる。なお、図3においては省略されているが、図5に示すように、各通信ノードにおいてホストが接続されていないポートにもパケットを転送するような第3の制御情報を設定しておくことで、制御装置に障害が発生した後にネットワークに接続したホスト間の通信にも対応することが可能となる。

30

**【0030】**

パケット処理部202は、上記した第1～第3の制御情報が格納された制御情報記憶部201を参照して受信パケットを処理する手段である。なお、制御情報記憶部201及びパケット処理部202自体は、非特許文献2に規定されているオープンフロースイッチのフローテーブルを用いたパケット処理機能を用いることができる。

40

**【0031】**

制御メッセージ送受信部(動作モード切替部)203は、制御装置10と制御用メッセージを授受して、例えば、制御情報記憶部201に制御情報を格納する動作を行う。また、本実施形態の制御メッセージ送受信部(動作モード切替部)203は、動作モード切替部としても機能し、制御装置10に障害が発生したことを検出すると、制御情報記憶部201から第2の制御情報を削除する動作を行う。なお、制御装置10の障害発生の検出方

50



法は、制御メッセージの受信状態や、所定の監視パケットの送出等種々の方法を用いることができる。

#### 【0032】

前記制御メッセージ送受信部（動作モード切替部）203による、制御情報記憶部201から第2の制御情報の削除の方式としては、様々なものが考えられる。例えば、各制御情報のエントリにタイムアウト値（例えば数秒）を設定しておき、所定時間内に制御装置10により更新（リセット）されない場合にタイムアウト成立により削除する方法が考えられる。あるいは、通信ノード20が、制御チャネル11を介して制御装置と通信可能であることを定期的に確認して、制御装置が通信不可状態に陥ったと判断した際に削除するなどの方法もある。

10

#### 【0033】

なお、図2に示した転送ノード20の各部（処理手段）は、転送ノード20に搭載されたコンピュータに、そのハードウェアを用いて、上記した各処理を実行させるコンピュータプログラムにより実現することもできる。

#### 【0034】

続いて、本実施形態の動作について図面を参照して詳細に説明する。以下、制御装置10によって行われる障害発生前の準備段階と、障害発生時の通信ノードの動作に分けて順番に説明する。

#### 【0035】

##### [ 障害発生前の準備 ]

20

特に障害が発生しておらず制御装置10からの通信ノード20の制御が正常に行えているとき、以下の動作が行われる。

#### 【0036】

##### (1) 新規通信パケット処理用の制御情報（第2、第3の制御情報）の設定

制御装置10は、ネットワーク立ち上げ時、あるいは、新規に通信ノードと制御チャネルを介して接続されると、初期化手順の中で、通信ノード20に、新規通信パケット処理用の制御情報（第2、第3の制御情報）を設定する。

#### 【0037】

第2の制御情報は、図3、図4に示したとおり、任意のパケットを制御装置10に転送させる制御情報である。第3の制御情報は、例えば、以下の手順で作成される。

30

#### 【0038】

まず、制御装置10は、図5に示すようなフラッディング転送用の論理ネットワークを構築する。図5に示すようなフラッディング転送用の論理ネットワークは、各通信ノード20の各ポートのうちホストが接続されている、もしくは、その可能性があるものを頂点に含むスパニング木を求めることで作成できる。このスパニング木に属するポートにパケットを送信すると、パケットはスパニング木に属する他の全てのポートに転送されることになる。

#### 【0039】

次に、制御装置10は、各通信ノード20に、前記フラッディング転送用の論理ネットワークに沿ったパケット転送を行わせる第3の制御情報を生成し、それぞれの通信ノード20に設定する。

40

#### 【0040】

なお、フラッディング転送用の論理ネットワークは、通信ノード20によって構成されるネットワークトポロジの変化を検出した際に、その都度更新されることが望まれる。即ち、制御装置10は、ネットワークトポロジに応じて、フラッディング転送用の論理ネットワークに対応する適切なスパニング木を維持することになる。スパニング木の作成とトポロジの検出は一般的な技術なので詳細は省略する。

#### 【0041】

##### (2) 新規通信の発生と既存通信

通信ノード20に接続したホストが新たに通信を行おうとすると、通信ノード20には

50

、当該通信を処理するための制御情報が設定されていないので、当該ホストから送信されたパケットは、第2の制御情報にマッチして、制御装置10に送られる。

【0042】

前述の第2の制御情報により新規通信パケットを受信すると、制御装置10は、当該パケットを宛先に転送する経路を計算し、経路上の通信ノード20に、後続するパケットを処理するための制御情報(第1の制御情報)を設定する。この手順は、オープンフローコントロールに代表される集中制御型ネットワークの制御装置の動作と同様である。

【0043】

以後は、前記ホストからの通信は既存通信となり、制御装置10が介在することなく、通信ノード20によりパケット転送が行われる。

10

【0044】

[障害発生時の通信ノードの動作]

ここで、制御装置10又は制御チャネルの障害によって、通信ノード20が制御装置10からの制御を受けられなくなったとする。

【0045】

通信ノードの制御メッセージ送受信部(動作モード切替部)203は、タイムアウトによって第2の制御情報を削除する。図6は、図3の状態から第2の制御情報を削除した状態を示す図である。このように第2の制御情報が削除されることにより、通信ノード20Aに接続するホストAが新たに通信を行おうとすると、第2の制御情報は削除済みなので、制御装置10への転送は行われずに、第3の制御情報が適用される。この結果、ホストAから送信されたパケットは、通信ノード20Eに転送される。通信ノード20Eにおいても第2の制御情報は削除済みなので、制御装置10への転送は行われずに、第3の制御情報が適用される。通信ノード20Aから受信したパケットは、通信ノード20B、20C、20Dに転送され、最終的には、図7に示すようにスパニング木全体に転送され、宛先のホストBにて受信される。以後、制御装置10の障害が解消するまで、新規通信パケットは、フラッディングされる。

20

【0046】

[障害からの回復動作]

その後、制御装置10又は制御チャネルの障害が解消して、通信ノード20が再び制御装置10の制御を受けられるようになると、前述の[障害発生前の準備]のとおり、第2の制御情報が再設定される。この結果、障害中にフラッディング転送が行われていた通信についても、回復後の最初のパケットが第2の制御情報にマッチして、制御装置10に転送されることになる。制御装置10は、当該パケットを処理するための第1の制御情報を作成して経路上の通信ノード20に設定する。以降は、制御装置10が介在することなく、通信ノード20のみでパケット転送が行われようになり、本来の転送動作に回復する。

30

【0047】

以上のように、本実施形態によれば、冗長化構成を採ることなく、制御装置10や制御チャネルの障害時でも新規通信のパケットを転送することが可能となる。

【0048】

上記した第1の本実施形態の有効性に関し通信量についても注目すべきである。障害発生の直後、新規通信が全体に占める割合は少ない。通信量の多くは既存通信であり、それらは通信ノードのみでパケット転送が行われる。第3の制御情報による転送を利用するとパケットを無関係な通信ノードやホストに転送するので一見不効率であるが、新規通信の割合が少なければシステム全体に与える影響は許容できる。

40

【0049】

[第2の実施形態]

続いて、レイヤ3(L3)転送の論理ネットワークにおいて、冗長化によらず、制御装置に障害が発生した場合においても、新規通信を処理するとともに、既存通信を維持できるようにした第2の実施形態について説明する。

【0050】

50

図 8 は、第 2 の実施形態の通信システムの構成を示す図である。第 1 実施形態との構成上の相違は、通信ノード 20E に接続する代替ルータ 21 が追加され、制御装置 10A 内に仮想ルータ 12 が追加されている点である。

【0051】

仮想ルータ 12 は、制御装置 10A 内で仮想的に動作するルータである。代替ルータ 21 は、制御装置 10A の障害時に、仮想ルータ 12 に代わって動作するルータである。それぞれ一般的なルータと同様であるが、一部異なる動作を行う。その詳細は、後の動作の説明において説明する。

【0052】

図 9、図 10 は、第 2 の実施形態の通信ノードに設定される制御情報の一例を示す図である。第 1 の実施形態の制御情報と異なる点は、第 1 の制御情報よりも低優先度かつ第 2 の制御情報よりも高優先度の第 4 の制御情報が追加されている点である。

10

【0053】

第 1 の制御情報は、制御装置によって設定された L2 転送又は L3 転送用（ユニキャスト用）の制御情報である。図 9 は、図 8 の通信ノード 20A に設定される制御情報を示している。この場合、通信ノード 20A は、MAC アドレスや IP（Internet Protocol）アドレス等によって特定されるホスト A からホスト B 宛てのパケットを受信すると、通信ノード 20B に転送する動作を行う。同様に、通信ノード 20B が制御情報を参照して、ホスト A からホスト B 宛てのパケットをホスト B に転送することで、ホスト A からのパケットがホスト B に届けられることになる。図 10 は、図 1 の通信ノード 20E に設定される制御情報を示している。通信ノード 20E は、通信ノード 20C を介して、ホスト C からホスト D 宛てのパケットを受信すると、通信ノード 20D に転送する動作を行う。同様に、通信ノード 20D が制御情報を参照して、ホスト C からホスト D 宛てのパケットをホスト D に転送することで、ホスト C からのパケットがホスト D に届けられることになる。

20

【0054】

図 9、図 10 の第 4 の制御情報は、ホストから送信された ARP（Address Resolution Protocol）要求に対して、制御装置 10A に代わって通信ノードに ARP 応答を送信させるための制御情報である。個々の制御情報は、第 2 の制御情報にて制御装置 10A に転送された ARP パケットと、制御装置 10A が応答する ARP 応答パケットとの対応関係を用いて作成される。

30

【0055】

図 9、図 10 の第 2 の制御情報は、第 1 の実施形態と同様である。図 9、図 10 の第 3 の制御情報は、第 1 の実施形態と同様のフラディング転送用の制御情報である（ただし、代替ルータ 21 もトポロジに含まれるため、転送先に設定される。）。

【0056】

以下の説明では、図 11 に示すように、制御装置 10A 内の仮想ルータ 12 と、代替ルータ 21 へのサブネットとアドレスが設定されているものとして説明する。図 12 は、図 8 に、図 11 の情報を付記した図であるので、以下、図 12 をベースに説明する。

【0057】

40

続いて、本実施形態の動作について図面を参照して詳細に説明する。以下、制御装置 10 によって行われる障害発生前の新規通信に対する準備段階、障害発生時の通信ノードの新規通信に対する動作、障害発生前の既存通信に対する準備段階、障害発生時の通信ノードの既存通信に対する動作、に分けて順番に説明する。

【0058】

[ 障害発生前の準備（新規通信） ]

特に障害が発生しておらず制御装置 10 からの通信ノード 20 の制御が正常に行えているとき、以下の動作が行われる。

【0059】

図 12 に示すように仮想ルータ 12 と代替ルータ 21 の IP アドレスは共通（R1）な

50

ので、制御装置 10 A の障害によって仮想ルータ 12 から代替ルータ 21 に切り替わる際でも、各ホストのデフォルトゲートウェイは変更不要である。

【0060】

(1) 新規通信パケット処理用の制御情報(第2、第3の制御情報)の設定

制御装置 10 A は、ネットワーク立ち上げ時、あるいは、新規に通信ノードと制御チャネルを介して接続されると、初期化手順の中で、通信ノード 20 に、新規通信パケット処理用の制御情報(第2、第3の制御情報)を設定する。

【0061】

第2の制御情報は、図9、図10に示したとおり、任意のパケットを制御装置 10 A に転送させる制御情報である。第3の制御情報は、例えば、以下の手順で作成される。

【0062】

まず、制御装置 10 A は、図13に示すような代替ルータ 21 を含むフラッディング転送用の論理ネットワークを構築する。図13に示すようなフラッディング転送用の論理ネットワークは、代替ルータ 21 及び各通信ノード 20 の各ポートのうちホストが接続されている、もしくは、その可能性があるものを頂点に含むスパニング木を求めることで作成できる。このスパニング木に属するポートにパケットを送信すると、パケットはスパニング木に属する他の全てのポートに転送されることになる。

【0063】

次に、制御装置 10 A は、第1の実施形態と同様、各通信ノード 20 に、前記フラッディング転送用の論理ネットワークに沿ったパケット転送を行わせる第3の制御情報を生成し、それぞれの通信ノード 20 に設定する。

【0064】

なお、本実施形態においても、フラッディング転送用の論理ネットワークは、通信ノード 20 によって構成されるネットワークトポロジの変化を検出した際に、その都度更新されることが望まれる。即ち、制御装置 10 は、ネットワークトポロジに応じて、フラッディング転送用の論理ネットワークに対応する適切なスパニング木を維持することになる。スパニング木の作成とトポロジの検出は一般的な技術なので詳細は省略する。

【0065】

なお、制御装置 10 A が正常に動作している場合の新規通信に対する処理は第1の実施形態と同様である。

【0066】

[ 障害発生時の通信ノードの新規通信に対する動作 ]

ここで、制御装置 10 A 又は制御チャネルの障害によって、通信ノード 20 が制御装置 10 からの制御を受けられなくなったとする。

【0067】

通信ノードの制御メッセージ送受信部(動作モード切替部) 203 は、タイムアウトによって第2の制御情報を削除する。図14、図15は、図9、図10の状態から第2の制御情報を削除した状態を示す図である。このように第2の制御情報が削除された状態で、通信ノード 20 A に接続するホスト A が新たに通信を行おうとした場合を考える。

【0068】

ここでは、図16に示すように、ホスト A の MAC アドレスが「MAC アドレス A」、IP アドレスが「IP アドレス A」、サブネットが「サブネット1」であり、ホスト B の MAC アドレスが「MAC アドレス B」、IP アドレスが「IP アドレス B」、サブネットが「サブネット2」であるものとする。このように、ホスト A とホスト B は異なるサブネットに属している。

【0069】

このとき、前述のとおり、第2の制御情報は削除済みなので、制御装置 10 A への転送は行われずに、第3の制御情報が適用される。この結果、ホスト A から送信されたパケットは、図16に示すようにスパニング木全体に転送され、例えば、宛先のホスト B にて受信される(ホスト B からホスト A 宛てのパケットも同様)。以後、制御装置 10 A の障害

10

20

30

40

50

が解消するまで、新規通信パケットは、フラッディングされる。

【 0 0 7 0 】

なお、ホスト A が A R P 要求パケットを送信した場合において、ホスト A が接続された通信ノード 2 0 A に第 4 の制御情報が設定されていない場合、次の手順で A R P 応答が返される。まず、例えばホスト A が、図 1 7 に示すように、ホスト A のデフォルトゲートウェイ ( I P アドレス R 1 ) の M A C アドレスを問い合わせたものとする。

【 0 0 7 1 】

前記 A R P 要求パケットが代替ルータ 2 1 に到達すると、代替ルータ 2 1 は、図 1 8 に示すような A R P 応答パケットを送信する。この A R P 応答パケットには、図 1 8 のとおり、 I P アドレス R 1 に、 M A C アドレス R 1 が対応することを通知するものとなっている。

10

【 0 0 7 2 】

前記 A R P 応答パケットは、 A R P 要求パケットと同様に、通信ノード 2 0 の第 3 の制御情報により、図 1 6 に示すようにスパニング木全体に転送される。前記 A R P 応答パケットが、ホスト A に到達すると、ホスト A は、 I P アドレス R 1 と M A C アドレス R 1 の対応関係を学習する。なお、ホスト A が接続された通信ノード 2 0 A に、第 4 の制御情報が設定済みである場合、通信ノード 2 0 A が第 4 の制御情報に従って A R P 応答パケットを返すことになる。この場合もホスト A は、 I P アドレス R 1 と M A C アドレス R 1 の対応関係を学習することができる。

【 0 0 7 3 】

その後、ホスト A は、図 1 9 に示すように、 I P アドレス B 宛ての I P パケットを送信する際に、宛先 M A C アドレスとして前記学習した M A C アドレス R 1 を宛先に指定する。

20

【 0 0 7 4 】

上記のようにフラッディング転送される結果、ホスト A が送信した I P アドレス B 宛ての I P パケットはホスト B に到達することになる。しかし、宛先 M A C アドレスがホスト B の M A C アドレス ( M A C アドレス B ) と異なるため、ホスト B はホスト A から送信された I P パケットを無視する。

【 0 0 7 5 】

一方、ホスト A が送信した I P アドレス B 宛ての I P パケットは代替ルータ 2 1 にも到達することになる。代替ルータ 2 1 は、宛先 M A C アドレスが代替ルータ 2 1 の M A C アドレスと一致するので、このパケットを受信して転送処理を行う。具体的には、代替ルータ 2 1 は、前記ホスト A からホスト B 宛ての I P パケットを受信すると、図 2 0 に示すように、 I P パケットの宛先 M A C アドレスと送信元 M A C アドレスをそれぞれ変更して、通信ノード 2 0 E に送信する。

30

【 0 0 7 6 】

前記宛先 M A C アドレスと送信元 M A C アドレスが書き換えられた I P アドレス B 宛ての I P パケットは、再度ホスト B に到達する。ホスト B は、宛先 M A C アドレスがホスト B の M A C アドレス ( M A C アドレス B ) であるため、パケットを適切に処理する。

【 0 0 7 7 】

[ 障害からの回復動作 ]

その後、制御装置 1 0 A 又は制御チャネルの障害が解消して、通信ノード 2 0 が再び制御装置 1 0 の制御を受けるようになると、前述の [ 障害発生前の準備 ] のとおり、第 2 の制御情報が再設定される。この結果、障害中にフラッディング転送が行われていた通信についても、回復後の最初のパケットが第 2 の制御情報にマッチして、制御装置 1 0 A に転送されることになる。制御装置 1 0 は、当該パケットを処理するための第 1 の制御情報を作成して経路上の通信ノード 2 0 に設定する。以降は、制御装置 1 0 A が介在することなく、通信ノード 2 0 のみでパケット転送が行われようになり、本来の転送動作に回復する。

40

【 0 0 7 8 】

50

## [ 障害発生前の準備 ( 既存通信 ) ]

続いて、本実施形態が、L3の既存通信に対しても有効であることを説明する。特に障害が発生しておらず制御装置10Aからの通信ノード20の制御が正常に行えているとき、以下の動作が行われる。ここでは、図21に示すようにホストCとホストDがそれぞれ通信ノード20C、20Dを介して通信している例を挙げて説明する。また、以下の説明では、ホストCのMACアドレスが「MACアドレスC」、IPアドレスが「IPアドレスC」、サブネットが「サブネット1」であり、ホストDのMACアドレスが「MACアドレスD」、IPアドレスが「IPアドレスD」、サブネットが「サブネット2」であるものとする。このように、ホストCとホストDは異なるサブネットに属している。

## 【0079】

当初の時点では、ホストCとホストD間の通信は行なわれておらず、既存通信にはなっていないものとする。ここで、送信元ホスト(例えば、ホストC)が、図22に示すようデフォルトゲートウェイのMACアドレスを問い合わせるARP要求パケットを送信すると、当該ARP要求パケットを受信した通信ノード(例えば、通信ノード20C)は、制御装置10Aに対して、ARP要求パケットを転送する(Packet-In)。

## 【0080】

制御装置10AはARP要求パケットを解析して、図23に示すようなARP応答パケットを作成する。そして、制御装置10Aは、通信ノード20Cに対して、ARP応答パケットをホストCに送信するように指示する(Packet-Outメッセージを利用)。ホストCは、通信ノード20Cから、前記ARP応答パケットを受信する。

## 【0081】

次に、制御装置10Aは、通信ノード20に、ARP応答用の制御情報(第4の制御情報)を設定する。本動作は、本実施形態の仮想ルータ12が、一般的な仮想ルータと異なる点である。

## 【0082】

図24は、上記第4の制御情報の詳細内容を示す図である。図24の例では、第4の制御情報のマッチ条件として、EthernetTypeフィールドにARP要求パケットであることを示すEthernetType=0x0806、オペコード=2が設定される。また、図24の例では、宛先MACアドレスがブロードキャストアドレス、宛先IPアドレスがIPアドレスR1、送信元MACアドレス及び送信元IPアドレスがホストCのアドレスであることが設定されている。そして、インストラクションフィールドには、受信したARPパケットを図23に示すARP応答パケットに書き換えてから送信する動作が記述されている。なお、第4の制御情報のタイムアウト値は、他の制御情報より長くするか、特に設定しなくともよい。

## 【0083】

以上のようにして、ARP応答パケットが返されると、送信元ホスト(ホストC)は、図25に示すような宛先ホスト(ホストD)宛てのIPパケットを送信する。なお、ホストCとホストDは異なるサブネットに属するので、図25のIPパケットの宛先MACアドレスフィールドには、前記学習したIPアドレスR1に対応するMACアドレスR1が設定される。

## 【0084】

通信ノード20Cには、このIPパケットに対応する第1の制御情報が設定されていないため、通信ノード20Cは、第2の制御情報により、受信したIPパケットを制御装置10Aに転送する(Packet-In)。

## 【0085】

制御装置10Aは、受信したIPパケットを解析して、送信元ホスト(ホストC)から宛先ホスト(ホストD)までの経路を計算し、さらに、経路上の通信ノード20に、このIPパケットを転送するための第1の制御情報を設定する。

## 【0086】

例えば、ホストCが直接接続している通信ノード20Cには、以下のような制御情報が

10

20

30

40

50

設定される。

・マッチ条件

Ethernet Type : 0x0800 (IPv4)

宛先MACアドレス : MACアドレスR1

宛先IPアドレス : IPアドレスD

・インストラクション

動作1:送信元MACアドレス及び宛先MACアドレスを変更

送信元MACアドレス : MACアドレスR2

宛先MACアドレス : MACアドレスD

動作2:経路上の次の通信ノードに又は宛先ホストと接続しているポートから送信

【0087】

例えば、経路上のその他通信ノード(例えば、通信ノード20D)には、以下のような制御情報が設定される。

・マッチ条件

Ethernet Type : 0x0800 (IPv4)

宛先MACアドレス : MACアドレスD

宛先IPアドレス : IPアドレスD

・インストラクション

動作1:経路上の次の通信ノードに又は宛先ホストと接続しているポートから送信

【0088】

さらに、制御装置10Aは、ホストDと接続している通信ノード20Dに対し、前記通信ノード20Cから受信したホストCからホストD宛てのIPパケットの出力を指示する。これは、前述の第1の制御情報の設定のみでは、Packet-Inで受け取ったIPパケットの送信が行われなために行われるものである。従って、送信するパケットの内容は、前述の通信ノード20Cに設定する第1の制御情報によって書き換えられる内容と同一となる。

【0089】

以後は、ホストCからホストD宛てのIPパケットは既存通信となり、制御装置10Aを介することなく、転送される。また、宛先ホスト(ホストD)も同様に、ARP要求パケットとIPパケットとを送信することで、双方向の通信が実現される。

【0090】

[障害発生時の通信ノードの既存通信に対する動作]

ここで、制御装置10A又は制御チャネルの障害によって、通信ノード20が制御装置10からの制御を受けられなくなったとする。

【0091】

通信ノードの制御メッセージ送受信部(動作モード切替部)203は、タイムアウトによって第2の制御情報を削除する。図14、図15は、図9、図10の状態から第2の制御情報を削除した状態を示す図である。このように第2の制御情報が削除された状態で、通信ノード20Cに接続するホストCが新たに通信を行おうとした場合を考える。

【0092】

ホストCから、図22に示すようなARP要求パケットが送信されると、通信ノード20Cは、前述の第4の制御情報に従い、ホストCに図23に示すARP応答パケットを送信する。なお、ホストDからのARP応答も同様に処理される。

【0093】

次に、ホストCがホストDに対し、図25に示すIPパケットを送信すると、通信ノード20Cに設定されている第1の制御情報により、図26に示すように送信元MACアドレス及び宛先MACアドレスをそれぞれMACアドレスR2、MACアドレスDに書き換えた上で、通信ノード20Dに転送する動作が行われる。

【0094】

さらに、経路上の通信ノード20Dが、第1の制御情報により、宛先ホスト(ホストD

10

20

30

40

50

)にパケットを送信する。最終的に図26に示すIPパケットは、宛先ホスト(ホストD)に到達することになる。逆方向も同様に、宛先ホスト(ホストD)がARP要求パケットとIPパケットとを送信することで、双方向の通信が実現される。

【0095】

以上のように、本実施形態によれば、制御装置10Aの障害時でもレイヤ3の新規通信のパケットを転送できる。また、本実施形態によれば、レイヤ3の既存通信の維持も可能となる。

【0096】

また、本実施形態の副次的な効果としては、障害が未発生のおきも制御装置10AへのARP問合せ負荷を軽減できる点が挙げられる。

10

【0097】

上記した第2の実施形態も、第3の制御情報による転送を行って、パケットを無関係な通信ノードやホストに転送するので一見不効率であるが、新規通信の割合が少なければシステム全体に与える影響は許容できる。また、代替ルータ21に要求されるパケット転送性能は、通信システム全体の性能に比べて小さくなり、代替ルータ21は相対的に安価なもので済む。障害発生から長時間が経過すると新規通信の割合が増していくことになる。第3の制御情報による転送がシステム全体に影響を与える、あるいは、代替ルータ21のパケット転送性能の上限に到達するといった状況がいずれ訪れる。しかしながら、その前に障害を復旧してシステム本来の性能を取り戻せばよい。そのようにして障害発生から復旧までの猶予期間を作ることも本発明の効果の一つである。

20

【0098】

以上、本発明の各実施形態を説明したが、本発明は、上記した実施形態に限定されるものではなく、本発明の基本的技術的思想を逸脱しない範囲で、更なる変形・置換・調整を加えることができる。例えば、各図面に示したネットワーク構成や要素の構成は、本発明の理解を助けるための一例であり、これらの図面に示した構成に限定されるものではない。

【0099】

また例えば、上記した実施形態では、ホスト毎に第4の制御情報を個別に作成、設定するものとして説明したが、制御情報のインストラクションフィールドの書き方次第で、制御情報のエントリ数を削減することもできる(第3の実施形態)。例えば、入力パケットの値を参照して書き換えを行うようにすることが考えられる。この場合、マッチ条件及びインストラクションフィールドに、要求送信元ホストのMACアドレス及びIPアドレスを指定せず、その代わり通信ノードが入力パケットから書き換え後の値を用いることになる(図27参照)。このようにすることで、各通信ノード20に設定する第4の制御情報の数をサブネットの個数分に抑えることが可能となる。

30

【0100】

また、上記した実施形態では、各通信ノード20は、非特許文献1、2のオープンフロースイッチと同様の動作を行うものとして説明したが、通信ノード20が一般的なレイヤ2、レイヤ3スイッチの機能を併せ持ってもよい(非特許文献2の「ハイブリッドスイッチ」参照)。例えば、上記した第2の実施形態の代替ルータ21に代えて、そのようなルータ機能を持つ通信ノードを用いることもできる(図28参照)。

40

【0101】

また、上記した実施形態では、仮想ルータ12と代替ルータ21は同一のIPアドレスを持つものとして説明したが仮想ルータ12と代替ルータ21のIPアドレスが異なっても構わない。その場合は、第3の制御情報によるフラッディング転送におけるホストと代替ルータ間のパケット転送を行う際にIPアドレスの書き換えを行えばよい。

【0102】

また、上記した実施形態では、仮想ルータと代替ルータのMACアドレスが同一であるものとして説明したが、両者のMACアドレスが異なってもよい。この場合、障害発生時、第2の制御情報が削除されて、第3の制御情報によるフラッディング転送に切り替

50



わった際に、GARP (Gratuitous ARP) パケットを発信して各ホストのARPキャッシュを更新させる。このGARPパケットは、代替ルータのIPアドレスとMACアドレス情報を持つ。これによって、各ホストは、ルータ経由の通信をする際に、宛先MACアドレスを代替ルータのMACアドレスにしたパケットを発信することができるようになる。代替ルータは、これらパケットを受信した場合、自身のMACアドレスであることを確認して、上記した実施形態と同様に転送を行うことになる。

【0103】

また、上記した実施形態では、各通信ノード20に設定する制御情報に優先度フィールドを設けるものとして説明したが、テーブル等に優先度順に制御情報を格納することで、優先度を表すものとしてもよい。

10

【0104】

また、上記した実施形態では、第2の制御情報を削除することで、第3の制御情報が優先的に適用されるようにしたが、その他の方法も採用可能である。例えば、第2の制御情報のマッチ条件の特定のフィールドに値を追加することで、実質上、第2の制御情報のマッチするパケットが存在しないようにすることもできる(一時的な無効化)。また、第2の制御情報の優先度の引き下げ(配置順序の入れ替えを含む)、前記第3の制御情報の優先度の引き上げ(配置順序の入れ替えを含む)、前記第2、第3の制御情報の優先度の入れ替え(配置位置の入れ替えを含む)等も適宜採用可能である。

【0105】

また、上記した各実施形態では、単一のテーブルに第1~第3(第4)の制御情報を設定するものとして説明したが、非特許文献2に準拠した通信ノードを用いる場合、複数のテーブルに第1~第3(第4)の制御情報を設定することができる。図29は、第4のテーブルを2つのテーブルで構成した例である。このように複数のテーブルを用いることで、テーブルの管理や更新が容易化されるほか、通信ノードにおける制御情報の検索時間も短縮される。

20

【0106】

最後に、本発明の好ましい形態を要約する。

[第1の形態]

(上記第1の視点による通信ノード参照)

[第2の形態]

第1の形態の通信ノードにおいて、

前記第3の制御情報は、前記第1の制御情報のマッチ条件に適合しないパケットをフラグディングさせるための制御情報である通信ノード。

30

[第3の形態]

第2の形態の通信ノードにおいて、

前記第3の制御情報によって、所定の代替ルータに、前記第1の制御情報のマッチ条件に適合しないパケットを転送する通信ノード。

[第4の形態]

第1~第3いずれか一の形態の通信ノードにおいて、

さらに、ホストからのアドレス解決要求に対して応答を指示する第4の制御情報が設定されている通信ノード。

40

[第5の形態]

(上記第2の視点による制御装置参照)

[第6の形態]

第5の形態の制御装置において、

前記第3の制御情報を前記通信ノードに設定することで、前記通信ノードに、前記第1の制御情報のマッチ条件に適合しないパケットをフラグディングさせる制御装置。

[第7の形態]

第6の形態の制御装置において、

前記第3の制御情報を前記通信ノードに設定することで、前記通信ノードに、前記第1

50

の制御情報のマッチ条件に適合しないパケットを所定の代替ルータへと転送させる制御装置。

[ 第 8 の形態 ]

第 5 ~ 第 7 いずれか一の形態の制御装置において、

さらに、前記通信ノードに、ホストからのアドレス解決要求に対して応答を指示する第 4 の制御情報を設定する制御装置。

[ 第 9 の形態 ]

( 上記第 3 の視点による通信システム参照 )

[ 第 10 の形態 ]

( 上記第 4 の視点による通信方法参照 )

10

[ 第 11 の形態 ]

( 上記第 5 の視点によるプログラム参照 )

なお、上記第 9 ~ 第 11 の形態は、第 1 の形態と同様に、第 2 ~ 第 4 の形態に展開することが可能である。

【 0107 】

なお、上記の特許文献および非特許文献の各開示を、本書に引用をもって繰り込むものとする。本発明の全開示（請求の範囲を含む）の枠内において、さらにその基本的技術思想に基づいて、実施形態ないし実施例の変更・調整が可能である。また、本発明の請求の範囲の枠内において種々の開示要素（各請求項の各要素、各実施形態ないし実施例の各要素、各図面の各要素等を含む）の多様な組み合わせ、ないし選択が可能である。すなわち、本発明は、請求の範囲を含む全開示、技術的思想にしたがって当業者であればなし得るであろう各種変形、修正を含むことは勿論である。特に、本書に記載した数値範囲については、当該範囲内に含まれる任意の数値ないし小範囲が、別段の記載のない場合でも具体的に記載されているものと解釈されるべきである。

20

【 符号の説明 】

【 0108 】

10、10A 制御装置

11 制御チャネル

12 仮想ルータ

20A ~ 20E 通信ノード

30

21 代替ルータ

21a ルータ機能

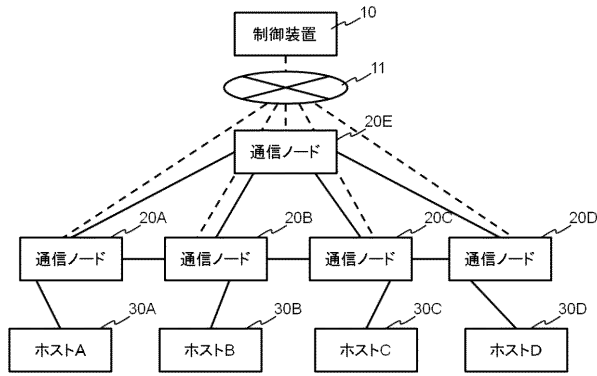
30A ~ 30D ホスト

201 制御情報記憶部

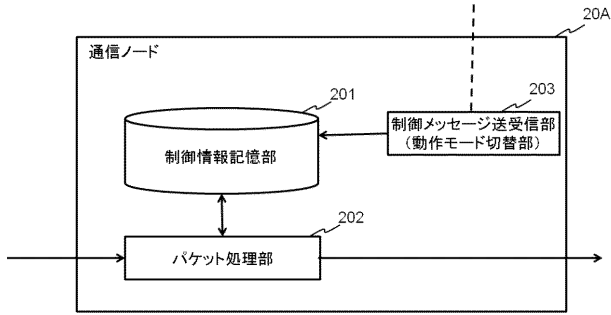
202 パケット処理部

203 制御メッセージ送受信部

【 図 1 】



【 図 2 】



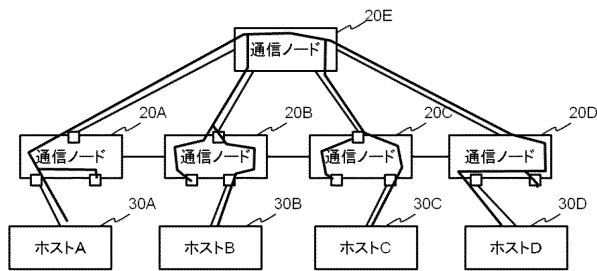
【 図 3 】

優先度	マッチ条件	インストラクション	
高	ホストAからホストB宛てのバケット	通信ノード20Bに転送	第1の制御情報
:	ANY	制御装置に転送 (Packet-In)	
低	ホストAと接続したポートで受信	通信ノード20Eに転送	第3の制御情報
:	通信ノード20Eと接続したポートで受信	ホストAに転送	

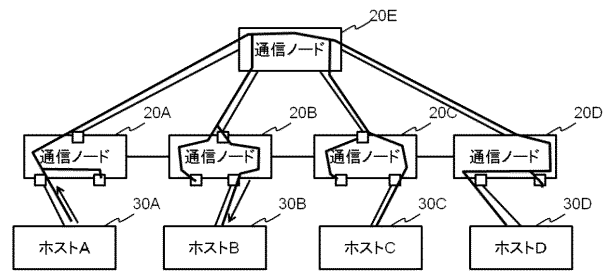
【 図 4 】

優先度	マッチ条件	インストラクション		
高	ホストCからホストD宛てのバケット	通信ノード20Dに転送	第1の制御情報	
:	ANY	制御装置に転送 (Packet-In)		第2の制御情報
低	通信ノード20Aと接続したポートで受信	通信ノード20B, 20C, 20Dに転送	第3の制御情報	
:	通信ノード20Bと接続したポートで受信	通信ノード20A, 20C, 20Dに転送		
:	通信ノード20Cと接続したポートで受信	通信ノード20A, 20B, 20Dに転送		
:	通信ノード20Dと接続したポートで受信	通信ノード20A, 20B, 20Cに転送		

【 図 5 】



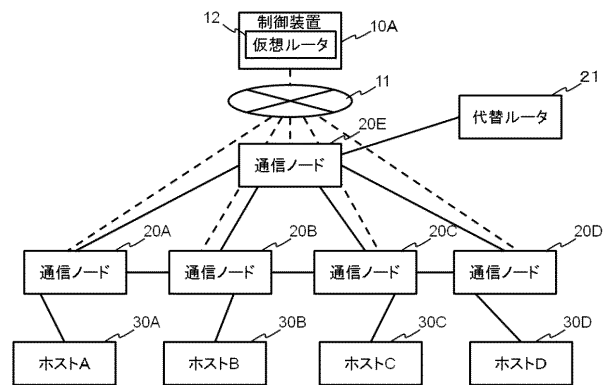
【 図 7 】



【 図 6 】

優先度	マッチ条件	インストラクション	
高	ホストAからホストB宛てのバケット	通信ノード20Bに転送	第1の制御情報
:			
低	ホストAと接続したポートで受信	通信ノード20Eに転送	第3の制御情報
:	通信ノード20Eと接続したポートで受信	ホストAに転送	

【 図 8 】



【 図 9 】

優先度	マッチ条件	インストラクション	
高	ホストAからホストB宛てのバケット	通信ノード20Bに転送	第1の制御情報
:	ARP	ARP応答を送信	
:	ANY	制御装置に転送(Packet-In)	第2の制御情報
低	ホストAと接続したポートで受信	通信ノード20Eに転送	第3の制御情報
:	通信ノード20Eと接続したポートで受信	ホストAに転送	

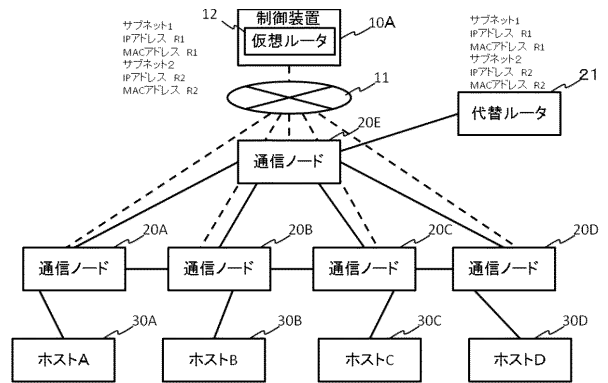
【 図 1 1 】

	サブネット1		サブネット2	
	IPアドレス	MACアドレス	IPアドレス	MACアドレス
代替ルータ	IPアドレス R1	MACアドレスR1	IPアドレス R2	MACアドレスR2
仮想ルータ	IPアドレス R1	MACアドレスR1	IPアドレス R2	MACアドレスR2

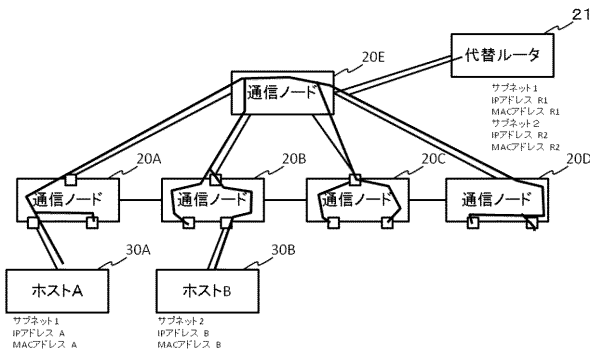
【 図 1 0 】

優先度	マッチ条件	インストラクション	
高	ホストCからホストD宛てのバケット	通信ノード20Dに転送	第1の制御情報
:	ARP	ARP応答を送信	
:	ANY	制御装置に転送(Packet-In)	第2の制御情報
低	通信ノード20Aと接続したポートで受信	通信ノード20B,20C,20D,代替ルータに転送	第3の制御情報
:	通信ノード20Bと接続したポートで受信	通信ノード20A,20C,20D,代替ルータに転送	
:	通信ノード20Cと接続したポートで受信	通信ノード20A,20B,20D,代替ルータに転送	
:	通信ノード20Dと接続したポートで受信	通信ノード20A,20B,20C,代替ルータに転送	
:	代替ルータと接続したポートで受信	通信ノード20A,20B,20C,20Dに転送	

【 図 1 2 】



【 図 1 3 】



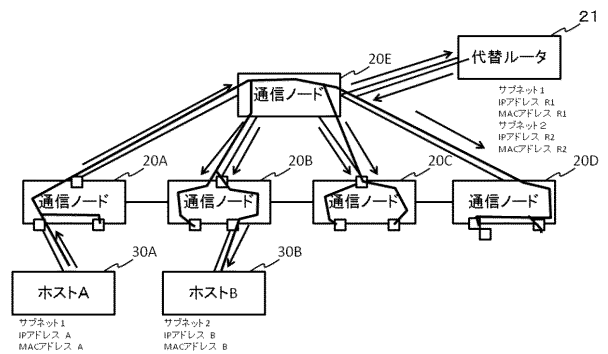
【 図 1 5 】

優先度	マッチ条件	インストラクション	
高	ホストCからホストD宛てのバケット	通信ノード20Dに転送	第1の制御情報
:	ARP	ARP応答を送信	
:			第4の制御情報
低	通信ノード20Aと接続したポートで受信	通信ノード20B,20C,20D,代替ルータに転送	第3の制御情報
:	通信ノード20Bと接続したポートで受信	通信ノード20A,20C,20D,代替ルータに転送	
:	通信ノード20Cと接続したポートで受信	通信ノード20A,20B,20D,代替ルータに転送	
:	通信ノード20Dと接続したポートで受信	通信ノード20A,20B,20C,代替ルータに転送	
:	代替ルータと接続したポートで受信	通信ノード20A,20B,20C,20Dに転送	

【 図 1 4 】

優先度	マッチ条件	インストラクション	
高	ホストAからホストB宛てのバケット	通信ノード20Bに転送	第1の制御情報
:	ARP	ARP応答を送信	
:			第4の制御情報
低	ホストAと接続したポートで受信	通信ノード20Eに転送	第3の制御情報
:	通信ノード20Eと接続したポートで受信	ホストAに転送	

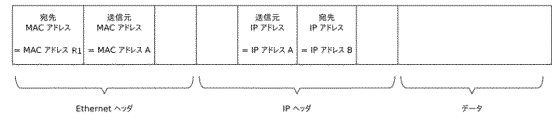
【 図 1 6 】



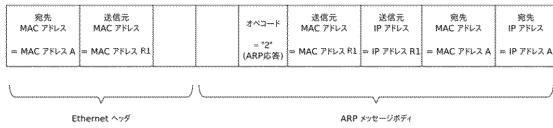
【 図 17 】



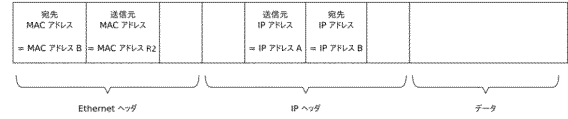
【 図 19 】



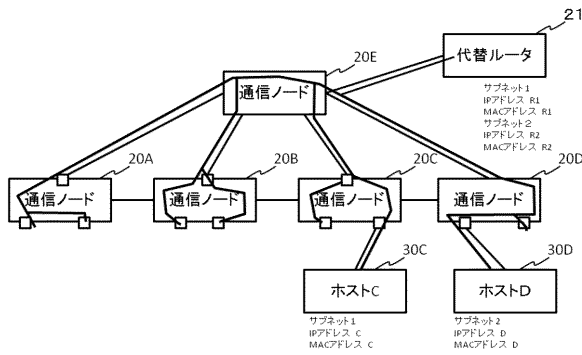
【 図 18 】



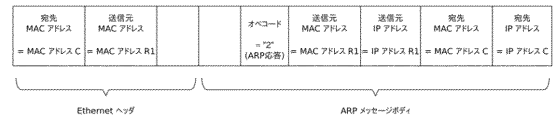
【 図 20 】



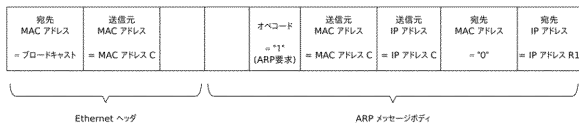
【 図 21 】



【 図 23 】



【 図 22 】



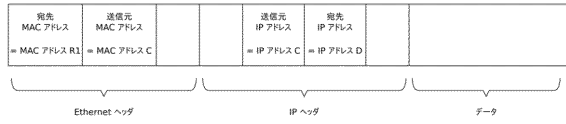
【 図 24 】

優先度	マッチ条件	インスタラクション
高	第1の制御情報	
	ARP	ARP応答を送信
:	第20の制御情報	
低	第30の制御情報	

マッチ条件	動作	タイムアウト
・Ethernet ヘッダの条件 Type = ARP 宛先 MAC アドレス = MAC アドレス C 送信元 MAC アドレス = MAC アドレス C 送信元 IP アドレス = MAC アドレス R1 ・ARP メッセージボディの条件 OPCODE = 1 (ARP 要求) 送信元 MAC アドレス = MAC アドレス C 送信元 IP アドレス = IP アドレス C 宛先 IP アドレス = IP アドレス R1	・Ethernet ヘッダの書換 宛先 MAC アドレス = MAC アドレス C 送信元 MAC アドレス = MAC アドレス R1 ・ARP メッセージボディの書換 OPCODE = 2 (ARP 応答) 送信元 MAC アドレス = MAC アドレス R1 送信元 IP アドレス = IP アドレス R1 宛先 MAC アドレス = MAC アドレス C 宛先 IP アドレス = IP アドレス C ・入力ポートに送信	長時間 or 無し

【 図 2 5 】



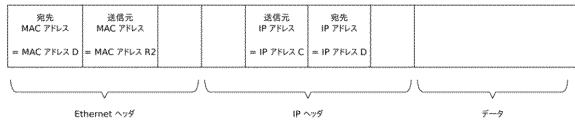
【 図 2 7 】

優先度	マッチ条件	インスタレーション
高	第1の制御情報	
:	ARP	ARP応答を送信
:	第2の制御情報	
低	第3の制御情報	

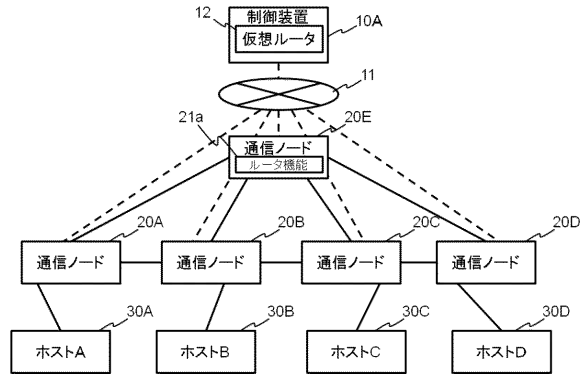
  

マッチ条件	動作	タイムアウト
・Ethernet ヘッダの条件 Type = ARP 宛先 MAC アドレス = プロトコル 送信元 MAC アドレス = 任意 *1 ・ARP メッセージボディの条件 OPCODE = 1 (ARP 要求) 送信元 MAC アドレス = 任意 *2 送信元 IP アドレス = 任意 *3 宛先 IP アドレス = ホスト R1 の IP アドレス	・Ethernet ヘッダの書換 宛先 MAC アドレス = *1 を参照 送信元 MAC アドレス = MAC アドレス R1 ・ARP メッセージボディの書換 OPCODE = 2 (ARP 応答) 送信元 MAC アドレス = *1 送信元 IP アドレス = IP アドレス R1 宛先 MAC アドレス = *1 または *2 を参照 宛先 IP アドレス = *3 を参照 ・入力ポートに送信	長時間 or 無し

【 図 2 6 】



【 図 2 8 】



【 図 2 9 】

テーブル#0

優先度	マッチ条件	インスタレーション
高	ホストCからホストD宛ての packets	通信ノード20Dに転送
:	ANY	制御装置に転送 (Packet In)
低	ANY	テーブル #N を参照

第1の制御情報  
第2の制御情報  
第3の制御情報

テーブル#N

優先度	マッチ条件	インスタレーション
低	通信ノード20Aと接続したポートで受信	通信ノード20B, 20C, 20Dに転送
低	通信ノード20Bと接続したポートで受信	通信ノード20A, 20C, 20Dに転送
低	通信ノード20Cと接続したポートで受信	通信ノード20A, 20B, 20Dに転送
低	通信ノード20Dと接続したポートで受信	通信ノード20A, 20B, 20Cに転送

第3の制御情報 (詳細)