

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4295812号  
(P4295812)

(45) 発行日 平成21年7月15日(2009.7.15)

(24) 登録日 平成21年4月17日(2009.4.17)

(51) Int.Cl.		F I			
<b>HO 4 L 12/44</b>	<b>(2006.01)</b>		HO 4 L 12/44	3 0 0	
<b>HO 4 L 12/56</b>	<b>(2006.01)</b>		HO 4 L 12/56	2 6 0 Z	

請求項の数 7 (全 15 頁)

(21) 出願番号	特願2008-283025 (P2008-283025)	(73) 特許権者	000005108
(22) 出願日	平成20年11月4日(2008.11.4)		株式会社日立製作所
(62) 分割の表示	特願2003-183934 (P2003-183934) の分割		東京都千代田区丸の内一丁目6番6号
原出願日	平成15年6月27日(2003.6.27)	(74) 代理人	110000028
(65) 公開番号	特開2009-27758 (P2009-27758A)		特許業務法人明成国際特許事務所
(43) 公開日	平成21年2月5日(2009.2.5)	(72) 発明者	河野 智彦
審査請求日	平成20年11月4日(2008.11.4)		神奈川県秦野市堀山下1番地 株式会社日 立製作所 IPネットワーク事業部内
		(72) 発明者	須貝 和雄
			神奈川県秦野市堀山下1番地 株式会社日 立製作所 IPネットワーク事業部内
		(72) 発明者	矢野 大機
			神奈川県秦野市堀山下1番地 株式会社日 立製作所 IPネットワーク事業部内

最終頁に続く

(54) 【発明の名称】 ネットワーク中継装置

(57) 【特許請求の範囲】

【請求項1】

複数の物理ポートによる通信を制御可能なネットワーク中継装置であって、  
所定数の回線を収容するとともに、該各回線に対するパケットの入出力を制御するパ  
ケット制御部を複数備え、

該各パケット制御部は、

入力したパケットがユニキャストパケットであるかマルチキャストパケットであるか  
を判定するパケット種別判定部と、

ユニキャストパケットを入力した場合に、該ユニキャストパケットを転送すべき回線  
を所定の設定に基づき選択し、該回線を収容するパケット制御部に前記ユニキャストパ  
ケットを転送するユニキャスト出力回線選択部と、

マルチキャストパケットを入力した場合に、所定の設定に基づき、前記各パケット制  
御部の少なくとも一部に該マルチキャストパケットを配布するマルチキャストパケット配  
布部と、

前記マルチキャストパケットが配布された場合に、該マルチキャストパケットを出力  
する回線を、所定の設定に基づき選択するマルチキャスト出力回線選択部と、

を備えるネットワーク中継装置。

【請求項2】

請求項1に記載のネットワーク中継装置であって、

該ネットワーク中継装置は複数の物理ポートを論理的に束ねることにより1つの論理ポ

10

20

ートとして取り扱うことが可能であり、

前記ユニキャスト出力回線選択部は、前記選択した回線が前記論理ポートであった場合に、更に、該論理ポートを構成する物理ポートのうち前記ユニキャストパケットの出力に用いる物理ポートを所定の設定に基づき選択する、ネットワーク中継装置。

【請求項 3】

請求項 1 に記載のネットワーク中継装置であって、

該ネットワーク中継装置は複数の物理ポートを論理的に束ねることにより 1 つの論理ポートとして取り扱うことが可能であり、

前記マルチキャスト出力回線選択部は、前記選択した回線が前記論理ポートであった場合に、更に、該論理ポートを構成する物理ポートのうち前記マルチキャストパケットの出力に用いる物理ポートを所定の設定に基づき選択する、ネットワーク中継装置。

10

【請求項 4】

請求項 3 に記載のネットワーク中継装置であって、

前記マルチキャスト出力回線選択部は、前記選択された物理ポートが当該パケット制御部の制御対象外である場合に、前記配布されたマルチキャストパケットを破棄する、ネットワーク中継装置。

【請求項 5】

請求項 2 に記載のネットワーク中継装置であって、

前記ユニキャスト出力回線選択部は、前記ユニキャストパケットに含まれる宛先アドレス情報と該ユニキャストパケットを用いるアプリケーションを特定するための識別情報とを用いた所定のハッシュ関数によってハッシュ値を求め、該ハッシュ値に対応付けられた物理ポートを選択する、ネットワーク中継装置。

20

【請求項 6】

請求項 3 に記載のネットワーク中継装置であって、

前記マルチキャスト出力回線選択部は、前記マルチキャストパケットに含まれる宛先アドレス情報と該マルチキャストパケットを用いるアプリケーションを特定するための識別情報とを用いた所定のハッシュ関数によってハッシュ値を求め、該ハッシュ値に対応付けられた物理ポートを選択する、ネットワーク中継装置。

【請求項 7】

請求項 3 に記載のネットワーク中継装置であって、

前記マルチキャストパケットが配布されたそれぞれのマルチキャスト出力回線部は、同じ設定に基づいて前記物理ポートを選択する、ネットワーク中継装置。

30

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、ネットワークの中継を行うネットワーク中継装置に関する。

【背景技術】

【0002】

一般的に、ネットワーク機器間を配線するためには、リピータハブやレイヤ 2 スイッチといったネットワーク中継装置が利用されている。リピータハブは、複数のネットワーク機器を電氣的に単純に接続するための装置である。一方、レイヤ 2 スイッチは、入力したパケットをその宛先アドレスに応じて最適な物理ポートに転送する機能を備えている。近年のオフィスでは、ネットワーク機器の増加に伴うトラフィックの増大を抑制するため、リピータハブではなく、レイヤ 2 スイッチを用いることが多い。また、近年では、より大規模なネットワークを構築するため、レイヤ 3 スイッチと呼ばれるネットワーク中継装置も利用されている。レイヤ 3 スイッチは、レイヤ 2 スイッチのブリッジ機能に加え、ルータ機能も備えている。

40

【0003】

上述したレイヤ 2 スイッチやレイヤ 3 スイッチの中には、リンクアグリゲーションと呼ばれる機能を備えるものもある。かかる機能は、LAN スイッチ間等に接続された複数の

50

物理ポートを1つの論理ポートとして扱うことにより帯域の拡大や冗長性の確保を実現する機能である(特許文献1参照)。

【0004】

【特許文献1】特開2002-9866号公報

【発明の開示】

【発明が解決しようとする課題】

【0005】

ネットワーク中継装置は、100Mbps、1Gbps、10Gbpsと益々高速化されるネットワークの伝送速度に対応する必要がある、また、マルチキャストを行うアプリケーションの増加やリンクアグリゲーション機能の搭載など、より複雑な処理を行う必要がある。そこで、本発明は、ネットワーク中継装置において、より効率的なネットワークの中継処理を可能とし、高速化に対応可能な技術を提供することを目的とする。

10

【課題を解決するための手段】

【0006】

上記課題を解決するため、本発明の第1のネットワーク中継装置を次のように構成した。すなわち、複数の物理ポートを論理的に束ねることにより1つの論理ポートとして取り扱うことが可能なネットワーク中継装置であって、パケットを入力するパケット入力部と、前記論理ポートに出力すべきパケットについて、該パケットに含まれる宛先アドレス情報と該パケットを用いるアプリケーションを特定するための識別情報とを抽出する抽出部と、前記論理ポートを構成する物理ポートのうち、前記パケットの出力に用いる物理ポートを前記宛先アドレス情報と前記識別情報とに応じて選択する物理ポート選択部と、前記選択された物理ポートを用いて前記パケットを前記宛先アドレスに出力するパケット出力部と、を備えることを要旨とする。

20

【0007】

アプリケーションを特定するための識別情報としては、例えば、TCP/UDPポート番号を挙げることができる。本発明では、論理ポートによってパケットを出力する際、このような識別情報とパケットの宛先アドレス情報とに応じて実際にパケットを出力する物理ポートを選択する。こうすることにより、同一アドレス宛や同一アプリケーション宛のパケットを同じフローに属するパケットとして扱い、同じ物理ポートから出力することができる。同じフローに属するパケットを異なる物理ポートから出力した場合には、僅かなタイミングのずれによって、後から送信したパケットが先に送信したパケットよりも早く送信先に到達してしまうという、いわゆるパケットの追い越し問題が発生する場合がある。パケットの追い越しがあった場合には、そのパケットを受信した装置において、再度パケットの順序を復元する処理が発生してしまう。そのため、本発明によれば、このようなパケットの追い越し問題を避けることができるため、効率的にネットワークを中継することが可能となる。

30

【0008】

上記第1のネットワーク中継装置において、前記物理ポート選択部は、前記宛先アドレス情報と前記識別情報とを用いた所定のハッシュ関数によってハッシュ値を求め、該ハッシュ値に対応付けられた物理ポートを選択するものとしてもよい。こうすることにより、宛先アドレス情報や識別情報に応じて効率的に物理ポートの使い分けを行うことが可能となる。

40

【0009】

上記課題を解決するため、本発明の第2のネットワーク中継装置を次のように構成した。すなわち、複数の物理ポートを論理的に束ねることにより1つの論理ポートとして取り扱うことが可能なネットワーク中継装置であって、所定数の回線を収容するとともに、該各回線に対するパケットの入出力を制御するパケット制御部を複数備え、該各パケット制御部は、前記論理ポートに出力すべきパケットについて、該パケットが当該パケット制御部の処理対象であるか否かを判定する判定部と、前記判定部が処理対象であると判定した場合に、前記パケットを出力するパケット出力部とを備えることを要旨とする。

50

## 【0010】

論理ポートは、複数の物理ポートによって構成される回線であるため、複数のパケット制御部に跨って設定されている場合がある。そのため、本発明では、論理ポートを用いたパケットの出力時には、各パケット制御部が、それぞれ独立してパケットを出力するか否かを判断するものとした。こうすることにより、処理の負担が分散され、効率的にパケットの出力を行うことが可能となる。

## 【0011】

上記第2のネットワーク中継装置において、前記パケット出力部は、前記判定部が処理対象でないと判定した場合に、前記パケットを破棄するものとしてもよい。このような構成であれば、冗長なパケット出力に伴う混乱を回避することが可能となる。

10

## 【0012】

パケット制御部が、出力すべきパケットが自分の処理対象であるか否かを判定するためには、例えば、上記第2のネットワーク中継装置において、前記各パケット制御部は、更に、前記論理ポートを構成する物理ポートのうち前記パケットの出力に用いる物理ポートを所定の設定に基づき選択する物理ポート選択部を備え、前記判定部は、前記選択された物理ポートが、当該パケット制御部の制御対象である場合に、前記パケットが、当該パケット制御部の処理対象であると判定するといった構成によって行うことができる。

## 【0013】

かかる構成のネットワーク中継装置において、前記物理ポート選択部は、前記パケットに含まれる宛先アドレス情報と該パケットを用いるアプリケーションを特定するための識別情報とを用いた所定のハッシュ関数によってハッシュ値を求め、該ハッシュ値に対応付けられた物理ポートを選択するものとしてもよい。こうすることにより、同じフローに属するパケットを同じ物理ポートから出力することが可能となる。

20

## 【0014】

上記課題を解決するため、本発明の第3のネットワーク中継装置を次のように構成した。すなわち、複数の物理ポートによる通信を制御可能なネットワーク中継装置であって、所定数の回線を収容するとともに、該各回線に対するパケットの入出力を制御するパケット制御部を複数備え、該各パケット制御部は、入力したパケットがユニキャストパケットであるかマルチキャストパケットであるかを判定するパケット種別判定部と、ユニキャストパケットを入力した場合に、該ユニキャストパケットを転送すべき回線を所定の設定に基づき選択し、該回線を収容するパケット制御部に前記ユニキャストパケットを転送するユニキャスト出力回線選択部と、マルチキャストパケットを入力した場合に、所定の設定に基づき、前記各パケット制御部の少なくとも一部に該マルチキャストパケットを配布するマルチキャストパケット配布部と、前記マルチキャストパケットが配布された場合に、該マルチキャストパケットを出力する回線を、所定の設定に基づき選択するマルチキャスト出力回線選択部と、を備えることを要旨とする。

30

## 【0015】

本発明によれば、入力したパケットがユニキャストパケットであれば、入力側のパケット制御部でそのパケットの出力を行う回線を選択し、該回線が接続される出力側のパケット制御部にパケットを転送する。マルチキャストパケットであれば、入力側のパケット制御部から、マルチキャストパケット配布部経由で、一つあるいは複数の出力側のパケット制御部にパケットを転送する。各々の出力側のパケット制御部では、出力を行う一つあるいは複数の回線を選択する。従って、マルチキャストパケットは、マルチキャストパケット配布部と、出力側のパケット制御部の両方で、複数に展開されるため、装置内のトラフィックの増加を抑え、高性能なマルチキャスト中継を実現することができる。

40

## 【0016】

上記第3のネットワーク中継装置において、該ネットワーク中継装置は複数の物理ポートを論理的に束ねることにより1つの論理ポートとして取り扱うことが可能であり、前記ユニキャスト出力回線選択部は、前記選択した回線が前記論理ポートであった場合に、更に、該論理ポートを構成する物理ポートのうち前記ユニキャストパケットの出力に用いる

50

物理ポートを所定の設定に基づき選択するものとしてもよい。こうすることにより、論理ポートを用いてユニキャストパケットを出力することができる。

【0017】

また、上記第3のネットワーク中継装置において、該ネットワーク中継装置は複数の物理ポートを論理的に束ねることにより1つの論理ポートとして取り扱うことが可能であり、前記マルチキャスト出力回線選択部は、前記選択した回線が前記論理ポートであった場合に、更に、該論理ポートを構成する物理ポートのうち前記マルチキャストパケットの出力に用いる物理ポートを所定の設定に基づき選択するものとしてもよい。こうすることにより、論理ポートを用いてマルチキャストパケットを出力することができる。

【0018】

かかる構成のネットワーク中継装置において、前記マルチキャスト出力回線選択部は、前記選択した物理ポートが当該パケット制御部の制御対象外である場合に、前記配布されたマルチキャストパケットを破棄するものとしてもよい。このような構成であれば、複数の物理回線を束ねて一つの回線のごとく扱い、一つに束ねられた複数の物理回線のうちのいずれかの物理回線に障害が発生したときに出力先を、束ねられた複数の物理回線内の他の物理回線に切り替える際に、束ねられた複数の物理回線のいずれかを収容する全てのパケット制御部にパケットを送っておき、出力回線選択部で、束ねられた複数の物理回線のうちの一本だけからパケットを出力し、他の物理回線へは出力する前にパケットを廃棄するように制御することにより、高速に物理回線を切り替えることができる。

【0019】

本発明において上述した種々の態様は、適宜、組み合わせたり、一部を省略して適用することができる。また、本発明は、上述した種々のネットワーク中継装置としての構成に限られず、例えば、ネットワークの中継方法やネットワークを中継するためのコンピュータプログラム、かかるプログラムを記録したコンピュータ読み取り可能な記録媒体等としても構成することができる。いずれの構成においても、上述した各態様を適宜、適用可能である。コンピュータが読み取り可能な記録媒体としては、例えば、フレキシブルディスクやCD-ROM、DVD-ROM、光磁気ディスク、メモリカード、ハードディスク等の種々の媒体を利用することができる。

【発明を実施するための最良の形態】

【0020】

以下、本発明の実施の形態について実施例に基づき次の順序で説明する。

A．ネットワーク中継装置の概略構成：

B．各テーブルの構造：

(B1) 経路テーブル：

(B2) リンクアグリゲーションテーブル：

C．ネットワーク中継装置の動作概要：

(C1) ユニキャスト中継時：

(C2) マルチキャスト中継時：

D．パケット入力処理：

E．パケット出力処理：

【0021】

A．ネットワーク中継装置の概略構成：

図1は、実施例としてのネットワーク中継装置100の概略構成を示す説明図である。本実施例のネットワーク中継装置100は、いわゆるLANスイッチであり、ブリッジ機能やルータ機能を備えるとともに、リンクアグリゲーション機能も備えている。リンクアグリゲーションとは、複数の物理ポートを束ねて1つの論理ポートとして取り扱うことが可能な機能であり、帯域幅の拡大や冗長性の確保を図ることができる。図1では、5本の物理ポートを束ねて構成した論理ポートLA1と、2本の物理ポートを束ねて構成した論理ポートLA2とを示した。各論理ポートは、リンクアグリゲーションID(図中の「LA-ID」と呼ばれる識別番号によって判別することができる。

10

20

30

40

50

## 【 0 0 2 2 】

ネットワーク中継装置 1 0 0 は、大まかに分けて、制御ユニット 1 1 0 と、クロスバスイッチ 1 2 0 と、パケット制御部 1 3 0 (以下、「 P P U 1 3 0 」と記載。)と、ネットワークインターフェース部 1 4 0 (以下、「 N I F 1 4 0 」と記載。)により構成されている。本実施例では、クロスバスイッチ 1 2 0 に 3 つの P P U 1 3 0 が接続され、各 P P U 1 3 0 には、それぞれ、 1 つまたは 2 つの N I F 1 4 0 が接続されているものとした。

## 【 0 0 2 3 】

制御ユニット 1 1 0 は、ネットワーク中継装置 1 0 0 の全体制御を行うユニットであり、内部には、 C P U やメモリなどを備えている。制御ユニット 1 1 0 は、図示していない P C I バスによって各 P P U 1 3 0 と接続されており、この P C I バス経由で各 P P U 1 3 0 の制御を行う。また、制御ユニット 1 1 0 は、ネットワークによって接続された管理用コンピュータに対して、 W e b ベースの設定用ユーザインタフェースを提供する。管理者は、このユーザインタフェースを利用することにより、経路テーブルなどの設定を行うことができる。その他、制御ユニット 1 1 0 は、各 P P U 1 3 0 に対する経路テーブルの配布やルーティングプロトコルの処理などを行う。

10

## 【 0 0 2 4 】

P P U 1 3 0 は、パケットの入出力を制御するためのモジュールである。各 P P U 1 3 0 には、図中に # 0、 # 1、 # 2 と示したように、それぞれに識別番号が付されている。各 P P U 1 3 0 は、 C P U 1 3 1、 R O M 1 3 2、 R A M 1 3 3、 検索エンジン 1 3 4、 テーブル記憶部 1 3 5 を備えている。 R O M 1 3 3 には、 P P U 1 3 0 を制御するためのプログラムが記録されており、 C P U 1 3 1 は、 R A M 1 3 3 をワークエリアとして用いることによりかかるプログラムを実行する。

20

## 【 0 0 2 5 】

検索エンジン 1 3 4 は、テーブル記憶部 1 3 5 に記録された各種テーブルを用いて経路検索を行う。本実施例において、経路検索とは、入力したパケットの宛先アドレス等に応じてそのパケットを出力するインタフェースを検索することをいう。検索エンジン 1 3 4 は、ソフトウェア的に実現することも可能であるが、処理を高速に行うために、経路検索用に専用に設計された A S I C によって構成することが好ましい。

## 【 0 0 2 6 】

テーブル記憶部 1 3 5 には、経路テーブルやリンクアグリゲーションテーブル等が記録されている。これらのテーブルの詳細については後述する。テーブル記憶部 1 3 5 は、例えば、 C A M (Content-Address Memory) とよばれる特殊なメモリや D R A M 等によって構成することができる。

30

## 【 0 0 2 7 】

N I F 1 4 0 は、 1 0 B A S E - 5 や 1 0 B A S E - T、 1 0 0 B A S E - T X 等のネットワーク回線をネットワーク中継装置 1 0 0 に対して物理的に接続するためのモジュールである。 N I F 1 4 0 は、パケット転送エンジン 1 4 1 (以下、「 P F P 」と称する。)と、ネットワーク回線を接続するための物理ポート 1 4 2 とを備えている。 P F P 1 4 1 は、 P P U 1 3 0 からの指示に基づいて物理ポート 1 4 2 に対するパケットの転送を制御する。各 P F P 1 4 1 および物理ポート 1 4 2 には、図示するようにそれぞれ識別番号が付されている。

40

## 【 0 0 2 8 】

クロスバスイッチ 1 2 0 は、 P P U 1 3 0 間でパケットの転送を行う際に、その伝送路を高速にスイッチングするための素子である。 P P U 1 3 0 は、クロスバスイッチ 1 2 0 を介してパケットの転送を行う際に、そのパケットに対して、ビットマップ、 M C フラグ、出力先情報といった付加情報を添付する。伝送路のスイッチングは、こうして付加されたビットマップを解析することによって行う。

## 【 0 0 2 9 】

図 2 は、付加情報とパケットの概略構造を示す説明図である。図示するビットマップの 3 つのビットは左側から順に、それぞれ P P U # 0、 P P U # 1、 P P U # 2 にパケット

50

を転送するか否かを表している。そのため、例えば、ビットマップとして「001」と設定されれば、そのパケットはクロススイッチ120によってPPU#2に転送される。また、「110」と設定されれば、PPU#0とPPU#1の両者に転送される。「MCフラグ」は、転送するパケットの種別を表すためのフラグである。「0」であればユニキャストパケット、「1」であればマルチキャストパケットであることを表す。出力先情報とは、そのパケットを出力するPPU130、PFP141、物理ポート142の各識別番号を指定するための情報である。この出力先情報は、ユニキャストパケットを転送する場合に用いられる。

#### 【0030】

パケットには、宛先MACアドレス、送信元MACアドレス、宛先IPアドレス、送信元IPアドレス、宛先TCPポート番号、送信元TCPポート番号、データ等の情報が含まれる。本実施例の説明において「宛先アドレス」とは、これらのうち宛先MACアドレスのことをいうものとする。しかし、宛先IPアドレスを適用しても特に差し支えはない。また、本実施例における「パケット」という用語は、いわゆる「イーサネット(登録商標)フレーム」と同義であるものとする。

#### 【0031】

B. 各テーブルの構造：

(B1) 経路テーブル：

図3は、テーブル記憶部135に記憶された経路テーブルの概略構造を示す説明図である。経路テーブルは、入力したパケットの宛先アドレスをキーとしてその出力先インタフェースを検索するためのテーブルである。

#### 【0032】

図示した例では、宛先アドレスAに対しては、出力先インタフェースとして、PPU#0、PFP#0、物理ポート#1が対応付けられ、宛先アドレスBに対しては、リンクアグリゲーションID#1の論理ポートが対応付けられているものとした。ここで、各PPU130に割り当てられるPPU番号は、2進数で「0000」から「1110」までの値とし、「1111」は、その出力先インタフェースが論理ポートである旨を示す識別子として取り扱うものとした。

#### 【0033】

図示する宛先アドレスCは、マルチキャストアドレスであるものとした。マルチキャストとは、予め登録した複数の宛先に同一内容のパケットを複製して配布することをいう。そのため、マルチキャストアドレスに対しては図示するように複数の出力先インタフェース対応付けられている場合がある。図では、以下の3つの出力先インタフェースが1つのマルチキャストアドレスCに登録されている例を示した。

(1) PPU#1, PFP#0, 物理ポート#0；

(2) PPU#1, PFP#1, 物理ポート#0；

(3) リンクアグリゲーションID#2の論理ポート；

#### 【0034】

マルチキャストアドレスに対しては、クロススイッチ120のスイッチングに用いるビットマップも経路テーブル内に予め登録されるものとした。こうすることにより、PPU130は、出力先インタフェースとなる各PPU130に迅速にマルチキャストパケットを転送することができる。

#### 【0035】

以上で説明した経路テーブルは、基本的に、上述した管理用コンピュータを用いて管理者が設定する。しかし、ここでは詳しい説明は省略するが、ユニキャストアドレスについてはMACアドレス学習処理によって、マルチキャストアドレスについてはIGMPスヌーピングによって自動的に設定することも可能である。MACアドレス学習処理を行う場合には、図4に示すようなポートテーブルが用いられる。ポートテーブルには、物理ポート番号と、その物理ポートが属するPPU番号やPFP番号もしくはリンクアグリゲーションIDが対応付けて記録されている。従って、各PPU130は、入力したパケットの

10

20

30

40

50

送信元アドレスと、ポートテーブルを参照して得られた P P U 番号やリンクアグリゲーション ID とを対応付けて経路テーブルに記録することにより、自動的に経路テーブルを設定することができる。

#### 【 0 0 3 6 】

( B 2 ) リンクアグリゲーションテーブル :

図 5 は、テーブル記憶部 1 3 5 に記録されたリンクアグリゲーションテーブルの概略構造を示す説明図である。リンクアグリゲーションテーブルには、リンクアグリゲーション ID 毎に、その論理ポートを構成する物理ポートの組とその数 ( パス数 ) が定義されている。図 5 には、図 1 で例示した論理ポート L A 1 のリンクアグリゲーションテーブルを示した。図中、「 E n 」とあるのは、「 E n a b l e 」の意であり、そのエントリが有効であるか否かを示すためのフラグである。「 1 」であれば有効であり、「 0 」であれば無効である旨を表す。このテーブルによれば、以下に挙げる 5 つの物理ポートによって論理ポート L A 2 が構成されていることとなる。

#### 【 0 0 3 7 】

- ( 1 ) P P U # 1、P F P # 1、物理ポート # 1 ;
- ( 2 ) P P U # 1、P F P # 1、物理ポート # 2 ;
- ( 3 ) P P U # 2、P F P # 0、物理ポート # 0 ;
- ( 4 ) P P U # 2、P F P # 0、物理ポート # 1 ;
- ( 5 ) P P U # 2、P F P # 0、物理ポート # 2 ;

#### 【 0 0 3 8 】

ネットワーク中継装置 1 0 0 は、以上で説明した経路テーブルやリンクアグリゲーションテーブルを用いて経路検索を行うことにより、入力したパケットを最適な出力先インタフェースから出力することができる。

#### 【 0 0 3 9 】

C . ネットワーク中継装置の動作概要 :

( C 1 ) ユニキャスト中継時 :

図 6 は、ユニキャストパケット入力時のネットワーク中継装置 1 0 0 の基本的な中継動作の概要を示す説明図である。まず、ユニキャストパケットを入力した入力側 P P U # 0 は、そのパケットの宛先アドレスに対応付けられた出力先インタフェースを経路テーブルから検索する。そして、入力側 P P U # 0 は、出力先インタフェースを付加情報によって指定して出力側の P P U # 2 にパケットを転送する。パケットの転送を受けた出力側 P P U # 2 は、こうして指定された出力先インタフェースからパケットを出力する。

#### 【 0 0 4 0 】

図 7 は、上述した入力側 P P U # 0 の経路検索において、出力先インタフェースとしてリンクアグリゲーション ID が検索された場合の動作を示す説明図である。この場合には、入力側 P P U # 0 は、更に、そのリンクアグリゲーション ID に対応したリンクアグリゲーションテーブルを参照して、登録された出力先インタフェースの中から所定の方法により 1 つの出力先インタフェースを選択する。かかる選択方法については後述する。そして、P P U # 0 は、選択された出力先インタフェースを付加情報によって指定して出力側 P P U # 2 にパケット転送する。パケットの転送を受けた P P U # 2 は、こうして指定された出力先インタフェースからパケットを出力する。

#### 【 0 0 4 1 】

以上で説明したように本実施例のネットワーク中継装置 1 0 0 は、ユニキャストパケットを入力した場合には、そのパケットを入力した入力側 P P U 1 3 0 で、リンクアグリゲーションに伴う処理も含めた詳細な経路検索を行う。そのため、出力側 P P U 1 3 0 は、論理ポートの使用の有無にかかわらず、入力側 P P U 1 3 0 によって指定された物理ポートからパケットを出力すればよい。従って、出力側 P P U 1 3 0 の負荷を軽減することができる。

#### 【 0 0 4 2 】

( C 2 ) マルチキャスト中継時 :



図8は、マルチキャストパケット入力時のネットワーク中継装置100の基本的な中継動作の概要を示す説明図である。まず、マルチキャストパケットを入力した入力側PPU#0は、経路テーブルからビットマップを検索する。そして、そのビットマップを付加してパケットをクロスバススイッチ120に転送する。クロスバススイッチは、パケットに付加されたビットマップを解析し、そのビットマップに対応する各PPU130に、パケットを複製して配布する。こうして、パケットの配布を受けた出力側PPU#1、#2は、そのパケットの宛先アドレスをキーとして経路テーブルから出力先インタフェースを検索し、検索された物理ポートからパケットを出力する。ただし、マルチキャストパケットについては経路テーブルに複数の出力先インタフェースが対応付けられているため、各PPU130は、検索された出力先インタフェースのうち、自分が管理する出力インタフェース

10

**【0043】**

図9は、出力側PPU#1、#2の経路検索において出力先インタフェースとしてリンクアグリゲーションIDが検索された場合の動作を示す説明図である。この場合には、出力側PPU#1、#2は、更に、そのリンクアグリゲーションIDに対応したリンクアグリゲーションテーブルを参照して、登録された出力先物理ポートの中から所定の方法により一つの出力先物理ポートを選択する。出力側PPU#1、#2は、こうして選択された出力先物理ポートからパケットを出力する。ただし、選択された出力先物理ポートが、自分の管理下でない場合にはパケットを破棄する(図のPPU#2参照)。

**【0044】**

20

このようにリンクアグリゲーションの出力先物理ポート選択処理を行う場所を、ユニキャストの場合は入力側PPU#0であるのに対し、マルチキャストの場合は出力側PPU#1、#2に変えている理由を以下で説明する。

**【0045】**

ユニキャストの場合には、パケットを一つ受信したときに、装置から一つだけパケットを送出すれば良いため、リンクアグリゲーションの出力先物理ポート選択処理を1回だけ行えばよく、入力側PPU#0で処理を行っても処理性能ネックにならない。

**【0046】**

これに対してマルチキャストの場合には、パケットを一つ受信したときに、一般に複数の物理ポート、および、複数のリンクアグリゲーションポートからパケットを送出する必要があり、さらに、複数のリンクアグリゲーションポート毎に、それぞれ一つの出力先物理ポートを選択する処理を行う必要があるため、入力側PPU#0で、全出力リンクアグリゲーションポートに対応する回数だけ、出力先物理ポートを選択する処理を行うと入力側PPU#0のパケット処理性能が、出力リンクアグリゲーションポート数に比例して低下してしまう、という問題があり、これを解決するためである。

30

**【0047】**

出力側PPU#1、#2で自分が管理する出力先物理ポートについてのみ、リンクアグリゲーションポートの出力先物理ポートを選択する処理を行うようにすることにより、マルチキャストで、複数のリンクアグリゲーションポートにパケットを送出する場合でも、出力先物理ポートを選択する処理を複数のPPU#1、#2で分散して行うことができ、装置のマルチキャスト中継性能を向上させることができる。

40

**【0048】**

ここで、一つのリンクアグリゲーションポートに属する複数の物理ポートが、複数のPPU#1、#2に跨って存在する場合には、リンクアグリゲーションポートの出力先物理ポートを選択する処理を、複数のPPU間で同一のアルゴリズムに従って行う必要がある。

**【0049】**

これは、複数のPPU#1、#2間で、もし異なるアルゴリズムに従って行ってしまうと、どのPPU#1、#2でも、自分の管理下の物理ポートが選択されず、該リンクアグリゲーションポートに属する物理ポートからパケットが1個も送出不される、という問題

50

や、逆に、複数の P P U # 1、# 2 で、それぞれ自分の管理下の物理ポートが選択され、一つのリンクアグリゲーションポートに属する複数の物理ポートから、パケットが合計で複数回送われてしまう、という問題があるためである。

【 0 0 5 0 】

以上で説明したように、本実施例のネットワーク中継装置 1 0 0 は、マルチキャストパケットを入力した場合には、そのパケットの出力側となる P P U 1 3 0 で詳細な経路検索およびリンクアグリゲーションに伴う処理を行う。従って、入力側 P P U 1 3 0 の負担を軽減し、負荷の分散を図ることが可能となる。

【 0 0 5 1 】

D . パケット入力処理 :

以下では、上述した各中継動作を実現するための詳細な処理について説明する。図 1 0 は、各 P P U 1 3 0 が実行するパケット入力処理のフローチャートである。まず、P P U 1 3 0 は、N I F 1 4 0 を介してパケットを入力すると (ステップ S 1 0 )、そのパケットの宛先アドレスをキーとして経路テーブルを検索する (ステップ S 1 1 )。検索の結果、ビットマップが検索された場合には (ステップ S 1 2 : 1 )、入力したパケットがマルチキャストパケットであると判定し、付加情報の M C フラグをオンにして (ステップ S 1 3 )、ステップ S 1 7 に処理を移行する。P P U 番号、P F P 番号、物理ポート番号が検索された場合には (ステップ S 1 2 : 2 )、ステップ S 1 6 に処理を移す。

【 0 0 5 2 】

上記ステップ S 1 1 の検索において、リンクアグリゲーション I D が検索された場合には (ステップ S 1 2 : 3 )、次のような処理を行う。まず、入力したパケットの宛先アドレスと宛先 T C P ポート番号を用いて、以下の手順によりハッシュ値を計算する (ステップ S 1 4 )。

【 0 0 5 3 】

( 1 ) 宛先アドレスと宛先 T C P ポート番号とを 8 ビットずつに区切り、それぞれを加算する。ただし、桁上がりは無視する。

( 2 ) 加算値をビット逆順し、ハッシュ値とする。ハッシュ値の取り得る範囲は 0 ~ 2 5 5 である。

【 0 0 5 4 】

更に、このハッシュ値を用いて次の計算を行う。ただし、計算結果の小数点以下は切り捨てるものとする。

【 0 0 5 5 】

エントリ番号 = ( ( ハッシュ値 ) × ( 論理ポートを構成する物理ポートの数 ) ) / 2 5 6 ;

【 0 0 5 6 】

そして、リンクアグリゲーションテーブルから、このエントリ番号に対応する P P U 番号、P F P 番号、物理ポート番号を検索し (ステップ S 1 5 )、検索結果を出力先情報として付加情報に設定する (ステップ S 1 6 )。上記ステップ S 1 1 において P P U 番号、P F P 番号、物理ポート番号が検索された場合には、かかる検索結果を付加情報に設定する。

【 0 0 5 7 】

次に、P P U 1 3 0 は、上記ステップ S 1 1 または S 1 5 によって検索された P P U 番号に応じてビットマップを付加情報に設定する (ステップ S 1 7 )。例えば、P P U # 0 が検索された場合には、ビットマップは「 1 0 0 」となり、P P U # 1 が検索された場合には「 0 1 0 」となる。上記ステップ S 1 1 においてビットマップが検索された場合には、検索されたビットマップを付加情報に設定する。

【 0 0 5 8 】

最後に、P P U 1 3 0 は、付加情報とともにパケットをクロスバスイッチ 1 2 0 に転送する (ステップ S 1 8 )。クロスバスイッチ 1 2 0 は、パケットの転送を受けると、付加情報内のビットマップにより指定された P P U 1 3 0 に対してパケットと付加情報を転送

10

20

30

40

50

する。ビットマップによって複数の P P U 1 3 0 が指定されていた場合には、付加情報とパケットを複製してそれぞれの P P U 1 3 0 に配布する。

【 0 0 5 9 】

E . パケット出力処理 :

図 1 1 は、各 P P U 1 3 0 が実行するパケット出力処理のフローチャートである。まず、P P U 1 3 0 は、クロススイッチ 1 2 0 からパケットを入力すると (ステップ S 3 0 )、そのパケットの付加情報に設定された M C フラグの状態をチェックする (ステップ S 3 1 )。M C フラグがオフの場合 (ステップ S 3 1 : N o )、すなわち、入力したパケットがユニキャストパケットである場合には、その付加情報には上記パケット入力処理によって既に出力先情報が設定されているため、P P U 1 3 0 は、付加情報を解析して (ステップ S 3 2 )、出力先インタフェースとして指定された P F P 1 4 1、物理ポート 1 4 2 からパケットを出力する (ステップ S 3 3 )。

【 0 0 6 0 】

M C フラグがオンの場合には (ステップ S 3 1 : Y e s )、入力したパケットはマルチキャストパケットであるため、入力側 P P U 1 3 0 によって出力先インタフェースは指定されていない。そこで、以下の処理により出力側 P P U 1 3 0 で経路検索を行う。まず、パケットの宛先アドレスをキーとして経路テーブルから出力先インタフェースを検索する (ステップ S 3 4 )。マルチキャストパケットについては、図 3 で示したように、複数の出力先インタフェースが検索される場合がある。次に、検索された出力先インタフェースの中に、リンクアグリゲーション I D が存在するか否かを判定する (ステップ S 3 5 )。リンクアグリゲーション I D が存在する場合には (ステップ S 3 5 : Y e s )、入力したパケットの宛先アドレス、宛先 T C P ポート番号を用いて上述した方法によりハッシュ値を計算して (ステップ S 3 6 )、更に、エントリ番号を計算する。そして、このエントリ番号に対応した P P U 番号、P F P 番号、物理ポート番号をリンクアグリゲーションテーブルから検索する (ステップ S 3 7 )。リンクアグリゲーション I D が含まれない場合には (ステップ S 3 5 : N o )、これらの処理 (ステップ S 3 6 および S 3 7 ) は行わない。

【 0 0 6 1 】

次に、P P U 1 3 0 は、上述の処理により検索された出力先インタフェース毎に、そのインタフェースが自分の制御対象のインタフェースであるか否かをチェックする (ステップ S 3 8 )。かかるチェックは、検索された出力先インタフェースの番号が、自分の P P U 番号および自分の管理する P F P 番号であるか否かをチェックすることにより行う。出力先インタフェースが、自分の制御対象でない場合には (ステップ S 3 8 : N o )、そのインタフェース宛のパケットは破棄する (ステップ S 3 9 )。自分の制御対象である場合には (ステップ S 3 8 : Y e s )、そのインタフェースからパケットを出力する (ステップ S 4 0 )。

【 0 0 6 2 】

以上で説明したように、パケットの出力を行う P P U 1 3 0 は、ユニキャストパケットの転送を受けた場合には、予め入力側の P P U 1 3 0 によって指定されたインタフェースからパケットの出力を行い、マルチキャストパケットの転送を受けた場合には、リンクアグリゲーションに伴う処理も含めた詳細な経路検索を行う。こうすることにより、パケットの種別に応じて効率的に処理を分散させることができる。

【 0 0 6 3 】

なお、上述したパケット入力処理および出力処理では、ハッシュ値の計算において、宛先アドレスと宛先 T C P ポート番号を利用するものとした。しかし、ハッシュ値の計算は、他の情報、例えば、宛先 M A C アドレスや送信元 M A C アドレス、宛先 I P アドレス、送信元 I P アドレス、宛先 T C P ポート番号、送信元 T C P ポート番号、M P L S プロトコルによる通信においてはラベル情報などを併せて用いるものとしてもよい。

【 0 0 6 4 】

[ 発明の効果 ]

10

20

30

40

50

本発明によれば、パケットを論理ポートから出力する場合には、その宛先アドレスやTCPポート番号に応じてハッシュ値を計算し、実際に出力する物理ポートを選択する。従って、同じフローに属するパケットは同じ物理ポートから出力されることとなり、パケットの追い越しを避けることができる。

【0065】

また、本発明によれば、入力したパケットがユニキャストパケットであるかマルチキャストパケットであるかによって経路検索を行う処理を各パケット制御部に分散させることができる。

【0066】

また、本発明によれば、各パケット制御部は、マルチキャストパケット出力時に、他のパケット制御部宛に対するパケットの破棄を行う。従って、冗長なパケット出力による混乱を回避することができる。

10

【0067】

以上のことから、本発明のネットワーク中継装置は、効率的にネットワークの中継を行うことができ、処理の高速化およびスループットの向上を図ることが可能となる。

【図面の簡単な説明】

【0068】

【図1】ネットワーク中継装置100の概略構成を示す説明図である。

【図2】付加情報とパケットの概略構造を示す説明図である。

【図3】経路テーブルの概略構造を示す説明図である。

20

【図4】ポートテーブルの概略構造を示す説明図である。

【図5】リンクアグリゲーションテーブルの概略構造を示す説明図である。

【図6】ユニキャストパケット入力時の基本的な中継動作の概要を示す説明図である。

【図7】ユニキャストパケット入力時に出力先インタフェースとしてリンクアグリゲーションIDが検索された場合の動作を示す説明図である。

【図8】マルチキャストパケット入力時の基本的な中継動作の概要を示す説明図である。

【図9】マルチキャストパケット入力時に出力先インタフェースとしてリンクアグリゲーションIDが検索された場合の動作を示す説明図である。

【図10】パケット入力処理のフローチャートである。

【図11】パケット出力処理のフローチャートである。

30

【符号の説明】

【0069】

100...ネットワーク中継装置

110...制御ユニット

120...クロスバスイッチ

130...パケット制御部

131...CPU

132...ROM

133...RAM

134...検索エンジン

40

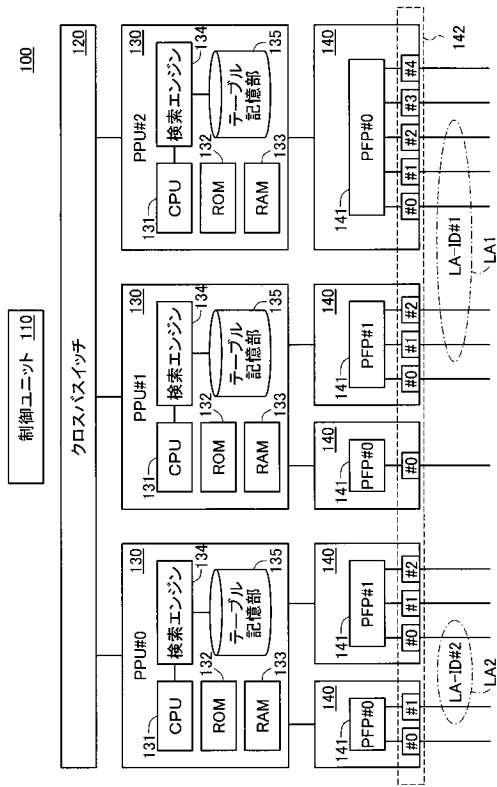
135...テーブル記憶部

140...ネットワークインターフェース部

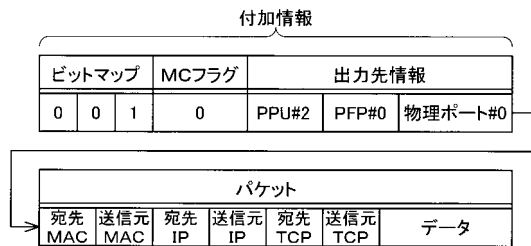
141...パケット転送エンジン

142...物理ポート

【図1】



【図2】



【図3】

経路テーブル

出力先インターフェース			宛先アドレス
PPU#0	PFP#0	物理ポート#1	A
1111	LA-ID#1		B
PPU#1	PFP#0	物理ポート#0	ビットマップ (マルチキャスト)
PPU#1	PFP#1	物理ポート#0	
1111	LA-ID#2		

【図4】

ポートテーブル

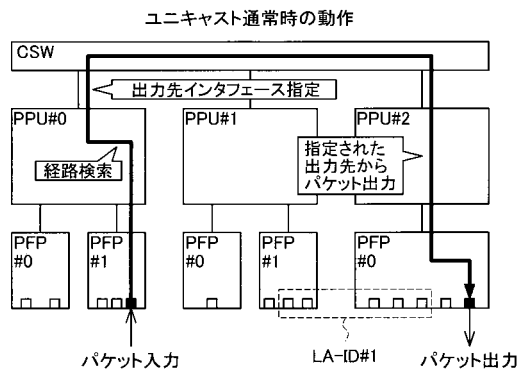
物理ポート#0	PPU#0	PFP#0	物理ポート#0
物理ポート#1	1111		LA-ID#2

【図5】

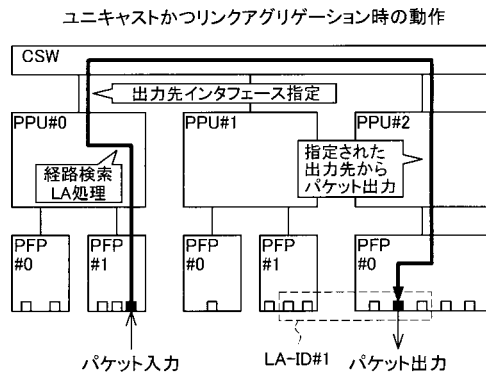
リンクアグリゲーションテーブル

LA-ID#0 パス数:5本				
No.	En	PPU#	PFP#	物理ポート#
0	1	1	1	1
1	1	1	1	2
2	1	2	0	0
3	1	2	0	1
4	1	2	0	2
...	...	...	...	...
15	0	0	0	0

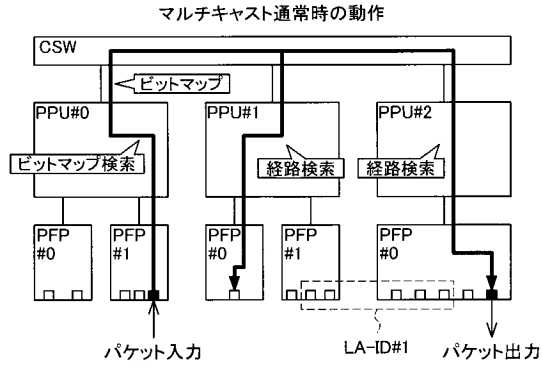
【図6】



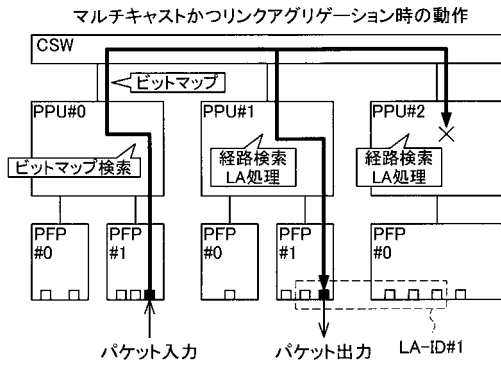
【図7】



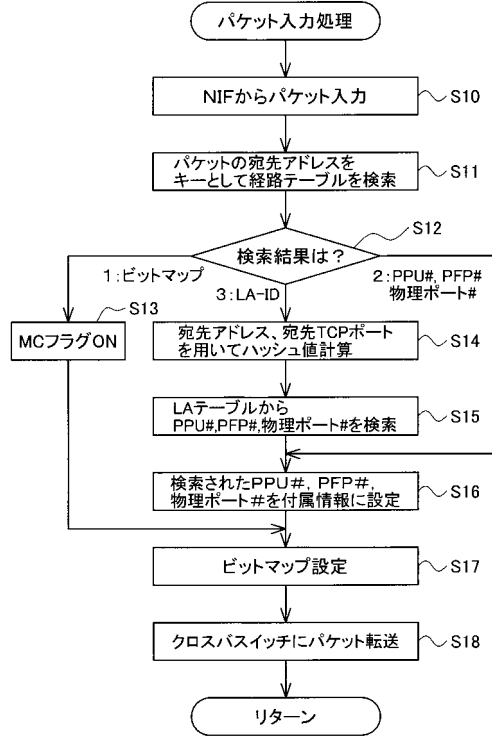
【 図 8 】



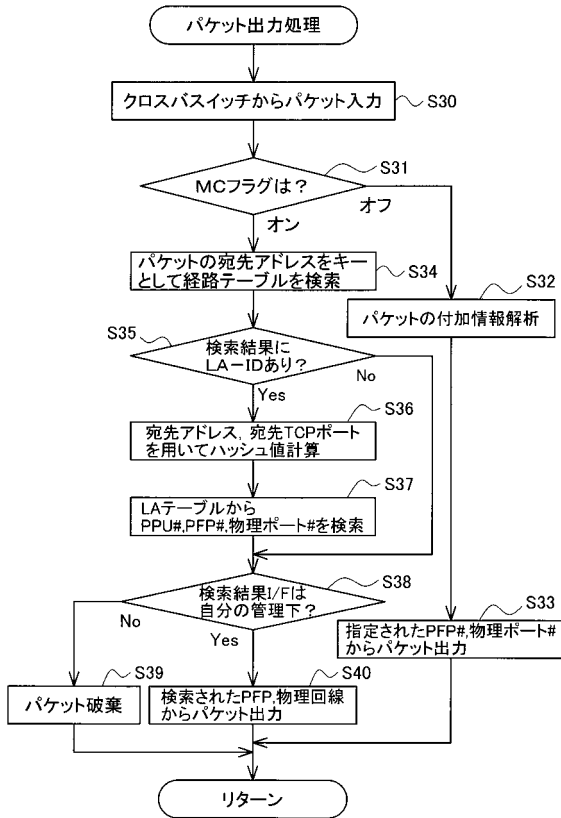
【 図 9 】



【 図 10 】



【 図 11 】



フロントページの続き

審査官 中木 努

(56)参考文献 特開2000-115243(JP,A)  
特開2003-18214(JP,A)  
特開2000-244574(JP,A)

(58)調査した分野(Int.Cl., DB名)  
H04L 12/28-46  
H04L 12/56