

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5814830号
(P5814830)

(45) 発行日 平成27年11月17日(2015.11.17)

(24) 登録日 平成27年10月2日(2015.10.2)

(51) Int. Cl. F I
 HO 4 L 12/701 (2013.01) HO 4 L 12/701
 HO 4 L 12/743 (2013.01) HO 4 L 12/743

請求項の数 4 (全 10 頁)

<p>(21) 出願番号 特願2012-47731 (P2012-47731) (22) 出願日 平成24年3月5日(2012.3.5) (65) 公開番号 特開2013-183397 (P2013-183397A) (43) 公開日 平成25年9月12日(2013.9.12) 審査請求日 平成26年7月25日(2014.7.25)</p> <p>(出願人による申告)平成23年度、独立行政法人情報通信研究機構「新世代ネットワークを支えるネットワーク仮想化基盤技術の研究開発 課題イ サービス合成可能なネットワークプラットフォームの研究開発」、産業技術力強化法第19条の適用を受ける特許出願</p>	<p>(73) 特許権者 000208891 K D D I 株式会社 東京都新宿区西新宿二丁目3番2号 (74) 代理人 100092772 弁理士 阪本 清孝 (74) 代理人 100084870 弁理士 田中 香樹 (74) 代理人 100119688 弁理士 田邊 壽二 (72) 発明者 松本 延孝 埼玉県ふじみ野市大原二丁目1番15号 株式会社K D D I 研究所内 (72) 発明者 林 通秋 埼玉県ふじみ野市大原二丁目1番15号 株式会社K D D I 研究所内</p> <p style="text-align: right;">最終頁に続く</p>
--	---

(54) 【発明の名称】 フロー単位パケット転送のための宛先検索装置および検索方法

(57) 【特許請求の範囲】

【請求項1】

線形検索を行う、ワイルドカードを含むエントリで構成されるフロー制御情報テーブルと、

ハッシュ検索を行う、完全一致エントリで構成されるパケット転送テーブルと、

前記フロー制御情報テーブルのワイルドカードでないフィールドに対して、受信パケット情報と一致判定をGPUで並列実施することで、一致するエントリを検索する、およびパケット転送テーブルにハッシュ検索を行うことで、一致するエントリを検索する検索処理制御手段と、

を備え、

前記検索処理制御手段は、

前記フロー制御情報テーブルから複数エントリを選択し、各エントリのワイルドカードが設定されているフィールドに1を、それ以外には0を代入した一致判定行列を生成する手段と、

前記一致判定行列のカラム単位で0が設定されているフィールドの受信パケット情報との一致判定をGPUで並列実施し、一致したフィールドに対応する前記一致判定行列の行列要素に1を代入する手段と、

前記一致判定行列で全てのフィールドが1であるエントリを一致エントリとする手段と、

を含むことを特徴とするフロー単位パケット転送のための宛先検索装置。

【請求項 2】

前記検索処理制御手段は、

前記フロー制御情報テーブルに一致するエントリが存在した場合に、前記パケット転送テーブルに当該受信パケット情報に相当するエントリを追加することを特徴とする請求項 1に記載の宛先検索装置。

【請求項 3】

前記パケット転送テーブルは、削除フラグを備え、

前記検索処理制御手段は、検索されたエントリの削除フラグを「非対象」で上書きし、前記削除フラグが「対象」であるエントリを削除し、前記削除フラグが「非対象」であるエントリの削除フラグを「対象」で上書きするパケット転送テーブル管理手段を、
さらに備えることを特徴とする請求項 1又は請求項 2に記載の宛先検索装置。

【請求項 4】

線形検索を行う、ワイルドカードを含むエントリで構成されるフロー制御情報テーブルのワイルドカードでないフィールドに対して、受信パケット情報と一致判定をGPUで並列実施することで、一致するエントリを検索するステップと、

完全一致エントリで構成されるパケット転送テーブルにハッシュ検索を行うことで、一致するエントリを検索するステップと、

前記フロー制御情報テーブルから複数エントリを選択し、各エントリのワイルドカードが設定されているフィールドに1を、それ以外には0を代入した一致判定行列を生成するステップと、

前記一致判定行列のカラム単位で0が設定されているフィールドの受信パケット情報との一致判定をGPUで並列実施し、一致したフィールドに対応する前記一致判定行列の行列要素に1を代入するステップと、

前記一致判定行列で全てのフィールドが1であるエントリを一致エントリとするステップと、

を有することを特徴とするフロー単位パケット転送のための宛先検索方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、フロースイッチの宛先検索において、ワイルドカードを含むフローテーブル上の宛先検索を行うための、宛先検索装置および検索方法に関するものである。

【背景技術】

【0002】

近年、通信制御をフロー単位で行うことにより、より粒度の細かいサービス品質制御を可能とするフローベースネットワークングが検討されている。フローとは、IPとポート番号で識別される、アプリケーション毎の一連の通信のまとまりであり、フローベースネットワークングではユーザやアプリケーションに応じたQoS制御や経路制御が可能となる。

【0003】

フローベースネットワークングの中継ノードであるフロースイッチでは、フローテーブルに登録されたエントリを参照して、パケットを適切な転送先へ送出する。フローエントリは、エントリの優先度、受信IFやヘッダの各情報に相当するフィールド（一致条件）、転送先情報によって構成される。

【0004】

フローに対する柔軟な制御を可能とするため、GPU等により高速化したソフトウェアベースのフロースイッチに関する検討が進められている。

【0005】

従来の宛先検索に関する手法としては、ソフトウェアルータにおける経路表検索、フロールータにおける完全一致検索、線形検索といった方法が取られてきた。

【0006】

10

20

30

40

50

非特許文献1では、Patricia-Treeという二分木検索の応用手法が述べられている。本手法はIP経路検索における最長プレフィックス一致を実現する検索方法として、LinuxやFreeBSD等OSの宛先検索に利用されている。

【0007】

非特許文献2では、商用のフロールータにおけるフロー検索方法が述べられている。全てのフローを完全一致エントリとすることで、ハッシュによるフローの高速検索を可能としている。

【0008】

上記課題を解決するため本発明のフロー単位パケット転送のための宛先検索装置は、線形検索を行う、ワイルドカードを含むエントリで構成されるフロー制御情報テーブルと、ハッシュ検索を行う、完全一致エントリで構成されるパケット転送テーブルと、前記フロー制御情報テーブルのワイルドカードでないフィールドに対して、受信パケット情報と一致判定をGPUで並列実施することで、一致するエントリを検索する、およびパケット転送テーブルにハッシュ検索を行うことで、一致するエントリを検索する検索処理制御手段と、を備え、

前記検索処理制御手段は、前記フロー制御情報テーブルから複数エントリを選択し、各エントリのワイルドカードが設定されているフィールドに1を、それ以外には0を代入した一致判定行列を生成する手段と、前記一致判定行列のカラム単位で0が設定されているフィールドの受信パケット情報との一致判定をGPUで並列実施し、一致したフィールドに対応する前記一致判定行列の行列要素に1を代入する手段と、前記一致判定行列で全てのフィールドが1であるエントリを一致エントリとする手段と、を含む。

【先行技術文献】

【非特許文献】

【0009】

【非特許文献1】Donald R. Morrisonら著 "PATRICIA: Practical Algorithm To Retrieve Information Coded in Alpha numeric" Journal of the ACM, Volume 15 Issue 4 1968年10月

【非特許文献2】L. G. Roberts著 "The Next Generation of IP - Flow Routing" SSGR R 2003S 2003年7月

【非特許文献3】Sangjin Hanら著 "PacketShader: a GPU-accelerated Software Router" Proceedings of ACM SIGCOMM 2010 2010年8月

【発明の概要】

【発明が解決しようとする課題】

【0010】

しかしながら、非特許文献1に代表される従来のIP経路検索は、最長プレフィックス一致検索を前提としており、複数フィールドを一致条件に含むフローテーブルの検索には適用できない。非特許文献2に示される完全一致検索を適用するには、全てのフローに対してフローエントリを定義する必要がある。運用を考えると大量のフローエントリの設定はスケール性に乏しい。非特許文献3ではワイルドカード検索も考慮されているが、線形検索のためエントリ数が増えた場合に性能劣化の影響が大きくなる。

【0011】

したがって、本発明では、ソフトウェアベースのフロースイッチにおいて、ワイルドカードを含むフローテーブルの検索処理を高速化するフロー単位パケット転送のための宛先検索装置および検索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】

【0012】

上記課題を解決するため本発明のフロー単位パケット転送のための宛先検索装置は、線形検索を行う、ワイルドカードを含むエントリで構成されるフロー制御情報テーブルと、ハッシュ検索を行う、完全一致エントリで構成されるパケット転送テーブルと、前記フロー制御情報テーブルのワイルドカードでないフィールドに対して、受信パケット情報と一

致判定をGPUで並列実施することで、一致するエントリを検索する、およびパケット転送テーブルにハッシュ検索を行うことで、一致するエントリを検索する検索処理制御手段とを備える。

【0013】

また、前記検索処理制御手段は、最初に前記パケット転送テーブルを検索し、該当するエントリが存在しない場合に前記フロー制御情報テーブルを検索することも好ましい。

【0015】

また、前記検索処理制御手段は、前記フロー制御情報テーブルに一致するエントリが存在した場合に、前記パケット転送テーブルに当該受信パケット情報に相当するエントリを追加することも好ましい。

【0016】

また、前記パケット転送テーブルは、削除フラグを備え、前記検索処理制御手段は、検索されたエントリの削除フラグを「非対象」で上書きし、前記削除フラグが「対象」であるエントリを削除し、前記削除フラグが「非対象」であるエントリの削除フラグを「対象」で上書きするパケット転送テーブル管理手段をさらに備えることも好ましい。

【0017】

上記課題を解決するため本発明のフロー単位パケット転送のための宛先検索方法は、線形検索を行う、ワイルドカードを含むエントリで構成されるフロー制御情報テーブルのワイルドカードでないフィールドに対して、受信パケット情報と一致判定をGPUで並列実施することで、一致するエントリを検索するステップと、完全一致エントリで構成されるパケット転送テーブルにハッシュ検索を行うことで、一致するエントリを検索するステップと、前記フロー制御情報テーブルから複数エントリを選択し、各エントリのワイルドカードが設定されているフィールドに1を、それ以外には0を代入した一致判定行列を生成するステップと、前記一致判定行列のカラム単位で0が設定されているフィールドの受信パケット情報との一致判定をGPUで並列実施し、一致したフィールドに対応する前記一致判定行列の行列要素に1を代入するステップと、前記一致判定行列で全てのフィールドが1であるエントリを一致エントリとするステップとを有する。

【発明の効果】

【0018】

本発明のワイルドカードの適用により、フローテーブル上の管理エントリ数が削減されるという特徴によれば、アプリケーション単位や宛先ネットワーク単位等でフローエントリを集約して設定できるため、運用上のスケール性向上という効果が得られる。

【0019】

また、ワイルドカードを含むフローテーブルをGPUにより高速検索するという特徴によれば、処理スループットの向上により、コア網等、より大容量なネットワークへ適用可能になるという効果が得られる。

【図面の簡単な説明】

【0020】

【図1】本発明のフロー宛先検索装置のシステム構成図を示す。

【図2】宛先検索手順を示すフローチャートを示す。

【図3】フロー制御情報テーブルの検索フローチャートを示す。

【図4】パケット転送テーブルの検索フローチャートを示す。

【図5】本発明の実施例における、フロー制御情報テーブルと一致判定行列の構成及び検索時の更新内容を示す。

【図6】本発明の実施例における、パケット転送テーブルの構成及び検索時の更新内容を示す。

【図7】本発明の実施例における、パケット転送テーブルの不要エントリの削除時の更新内容を示す。

【発明を実施するための形態】

【0021】

10

20

30

40

50

本発明を実施するための最良の実施形態について、以下では図面を用いて詳細に説明する。図 1 に、本発明のフロー宛先検索装置のシステム構成図を示す。

【 0 0 2 2 】

本発明のフロー宛先検索装置 1 は、パケット送受信部 2 からパケット情報を取得し、2 つのフローテーブル（フロー制御情報テーブル 1 3、パケット転送テーブル 1 5）を参照し、一致するエントリを検索して、パケットの宛先情報を返す。また、制御ポリシーに従いフロー制御情報テーブル 1 3 を管理する。

【 0 0 2 3 】

フロー宛先検索装置 1 は、検索処理制御部 1 1、フロー制御情報テーブル管理部 1 2、フロー制御情報テーブル 1 3、パケット転送テーブル管理部 1 4、パケット転送テーブル 1 5 で構成される。フロー宛先検索装置 1 はパケット送受信部 2 と接続し、受信パケットの取得及び転送先の通知を行う。また、フロー宛先検索装置 1 は制御ポリシー管理部 3 と接続し、フロー制御情報の更新を受け付ける。

10

【 0 0 2 4 】

検索処理制御部 1 1 は、受信パケットの情報に基づき適切な転送先情報を取得するため、パケット転送テーブル 1 5 及びフロー制御情報テーブル 1 3 を検索し、得られた転送先情報をパケット送受信部 2 に通知する。また、パケット転送テーブル 1 5 に必要なエントリが無かった場合、エントリ追加処理を行う。

【 0 0 2 5 】

フロー制御情報テーブル 1 3 は、フロー制御情報の定義をエントリとして保持している。各エントリは任意のフィールドにワイルドカードを指定可能であり、全てのエントリには一致優先度を示すエントリ優先度が設定される。本テーブルに対する検索は線形検索に基づき行われる。

20

【 0 0 2 6 】

パケット転送テーブル 1 5 は、現在流れているフローの完全一致エントリ情報を保持している。パケットの一致情報から得られるハッシュ値をキーとする完全一致エントリと削除可能なフロー情報であるかどうかを示す削除フラグで構成されており、ハッシュによる検索が行われる。

【 0 0 2 7 】

フロー制御情報テーブル管理部 1 2 は、制御ポリシー管理部 3 からの要求に基づき、フロー制御情報テーブル 1 3 の更新を行う。

30

【 0 0 2 8 】

パケット転送テーブル管理部 1 4 は、パケット転送テーブル 1 5 上で既に使われなくなったフロー情報がないかを監視し、不要エントリの削除を行う。

【 0 0 2 9 】

フロー制御情報テーブル 1 3 及びパケット転送テーブル 1 5 は、データアクセスの高速性のため、GPU 上のメモリに展開しておいてもよい。

【 0 0 3 0 】

制御ポリシー管理部 3 は、フロー制御情報管理テーブル 1 3 へのエントリ追加、削除を要求する。要求形態は、例えば運用者画面からの手動設定、ルーティングデーモン等からの経路情報更新、およびシグナリングデーモンからのフロー情報設定である。

40

【 0 0 3 1 】

図 2 は、宛先検索手順を示すフローチャート図である。以下、本フローチャートに基づき、フロー宛先検索装置 1 の動作を説明する。

【 0 0 3 2 】

ステップ 1：パケット情報の受信、パケット送受信部 2 は、パケット情報（受信 I F、M A C、V L A N、I P、ポート番号）を受信し、検索処理制御部 1 1 に送信する。

ステップ 2：ハッシュ値計算、検索処理制御部 1 1 は、パケット送受信部 2 から受け取った受信パケット情報に基づき、パケット転送テーブル検索用のハッシュ値計算を行う。

ステップ 3：パケット転送テーブル検索、検索処理制御部 1 1 は、算出されたハッシュ

50

値に基づきパケット転送テーブル15を検索する。

ステップ4：一致エン트리有無の判定、パケット転送テーブル15に該当するエントリアあった場合には、ステップ5に進み、該当するエントリア無かった場合には、ステップ6に進む。

ステップ5：該当エントリアの転送先情報を通知、検索処理制御部11は、該当するエントリアの転送先情報をパケット送受信部2に通知する。

ステップ6：フロー制御情報テーブルの検索、検索処理制御部11は、フロー制御情報テーブル13上で受信パケットに相当するエントリアの検索を行う。

ステップ7：一致エン트리有無の判定、フロー制御情報テーブル13に該当するエントリアあった場合には、ステップ8に進み、該当するエントリア無かった場合には、ステップ10に進む。

10

ステップ8：パケット転送テーブルへの新規エン트리追加、検索処理制御部11は、受信パケット情報と該当した転送先情報に基づきパケット転送テーブル15に新規エントリアを追加する。

ステップ9：該当エントリアの転送先情報を通知、検索処理制御部11は、該当するエントリアの転送先情報をパケット送受信部2に通知する。

ステップ10：不一致応答、転送先がないものとして、パケットは破棄される。

【0033】

図3は、フロー制御情報テーブル13の検索フローチャートを示す。以下、本フローチャートに基づいて、フロー制御情報テーブル13の検索を説明する。

20

【0034】

ステップ31：一致判定行列生成、最初に優先度の高い方から複数エントリアを選択し、それらエン트리群に対する一致判定行列を生成する。一致判定行列は、各フィールドの一致状況を保持するための行列である。各エントリアのワイルドカードが設定されているフィールドに相当する行列要素には1を、それ以外には0を代入する。

ステップ32：一致判定必要フィールド選出、一致判定行列上の各列において、0が設定されている部分は、エントリアとの一致判定が必要となるフィールドであるため、選出する。

ステップ33：フィールド一致判定処理、同一インストラクションで実行可能なカラム単位の受信パケット情報との一致判定を、GPU上で並列実施する。エン트리群の全てのエントリアに対しフィールドの一致判定が終了したら、一致したフィールドに相当する一致判定行列の行列要素に1を代入する。

30

ステップ34：一致エン트리抽出、ここで、全フィールドが一致しているエントリア(全てが1になったエントリア)があれば、そのなかで最も優先度の高いエントリアを当該パケットに対する一致エントリアとする。一致エントリアが存在しなければ、フロー制御情報テーブル13の次のエン트리群に対する一致判定を繰り返していく。なお、エントリアが全フィールド一致しているかの判定は、フィールドのビット論理積を全エントリアに対して行うことから、GPUによる並列処理を行うことが望ましい。

【0035】

図4に、パケット転送テーブル15の検索フローチャートを示す。以下、本フローチャートに基づいて、パケット転送テーブル15の検索を説明する。

40

【0036】

ステップ41：ハッシュ値による検索、検索処理制御部11は受信パケット情報に対するハッシュ値をキーとして、パケット転送テーブル15を検索する。

ステップ42：一致条件判定、該当するハッシュ値に紐づくエントリアがあった場合、一致条件がすべて等しいか判定する。等しい場合はステップ43に進み、等しくない場合は、ステップ44に進む。

ステップ43：削除フラグ上書、等しければ一致エントリアとし、テーブルの削除フラグを0(削除非対象)で上書きする。

ステップ44：再検索、等しくなければ、適用しているハッシュのアルゴリズムに基づ

50

き、一致エントリの検索を続ける。例えば、オープンハッシュ法の場合、リハッシュによりハッシュ値を再計算して検索する。チェーンハッシュ法の場合、当該ハッシュ値に紐づくエントリで構成されるリストを順にたどって検索する。

【 0 0 3 7 】

なお、パケット転送テーブル 15 は、新規フローが到着するたびにフロー制御情報テーブル 13 の情報に基づきエントリが追加されるため、不要エントリを削除しないとテーブルの容量上限に達してしまい、新規エントリの追加ができなくなる。よって、下記の手順で不要エントリを削除する。パケット転送テーブル管理部 14 は、宛先検索とは別に、各エントリの削除フラグを順次チェックし、削除フラグが 1 (削除対象) の場合は不要エントリとして情報を削除する。削除フラグが 0 (削除非対象) の場合はエントリを削除せず、削除フラグを 1 に上書きして次のエントリのチェックへ進む。そのため、次回チェック時まで参照されなかったエントリは、不要エントリとして削除される。

10

【 0 0 3 8 】

以下に、本発明の実施例を示す。図 5 は、本発明の実施例における、フロー制御情報テーブル検索を示す。図 5 (a) は、フロー制御情報テーブルの例であり、図 5 (b)、図 5 (c) は、一致判定行列の例を示す。

ステップ 31、図 5 (b) は、図 5 (a) のフロー制御情報テーブルから生成された一致判定行列のエントリ群を示す。エントリ優先度 10 から 40 までが示される。図 5 (a) のワイルドカードフィールドには、1 が代入され、それ以外は 0 が代入されている。なお、本例で、エントリ群は 4 行しかないが、実際には、もっと多くの行が選択され、多くの多重度で GPU の並列計算が行われる。

20

ステップ 32、図 5 (b) で、四角で囲まれているフィールドが選出されたエントリである。

ステップ 33、図 5 (c) は、GPU の一致判定演算で、一致したフィールドに 1 が代入された一致判定行列を示す。図 5 (c) で、四角で囲まれているフィールドが、1 が代入されたエントリである。

ステップ 34、図 5 (c) の例では、網掛けのエントリが、全フィールドが一致しているエントリである。もっと優先度が高い、優先度 20 のエントリが一致エントリとして、抽出される。

【 0 0 3 9 】

図 6 は、本発明の実施例における、パケット転送テーブル検索を示す。

30

ステップ 41、受信パケット情報 (I F = e t h 1 , D - M A C = D : D : D , S - M A C = E : E : E , . . .) から、ハッシュ値 H (受信パケット情報) = 0 x 0 0 0 1 を求め、検索する。

ステップ 42、ハッシュ値 0 x 0 0 0 1 は、図 6 で点線が引かれているエントリである。受信パケット情報と一致条件がすべて等しいか判定する。ここでは、全て等しいため、ステップ 43 に進む。

ステップ 43、削除フラグを 0 で上書きする。

【 0 0 4 0 】

図 7 に、本発明の実施例における、パケット転送テーブルの不要エントリ削除を示す。ハッシュ値 0 x 0 0 0 1 と 0 x 0 0 1 0 のエントリは、削除フラグが 1 であるため、巡回チェックで削除される。ハッシュ値 0 x 0 0 0 0 と 0 x 0 0 1 1 のエントリは、削除フラグが 0 であるため、1 に上書きされる。

40

【 0 0 4 1 】

本発明の宛先検索装置を用いた高速フロースイッチを通信キャリアのコア網に適用することにより、従来のインターネットでは不可能であったエンドツーエンドでのフロー単位制御が可能になる。それにより、アプリケーション毎に適した通信品質の提供が可能となり、サービス品質の向上を実現することができる。また、ネットワーク資源利用の効率化が可能となり、線路設備コストの抑制につながる。

【 0 0 4 2 】

50

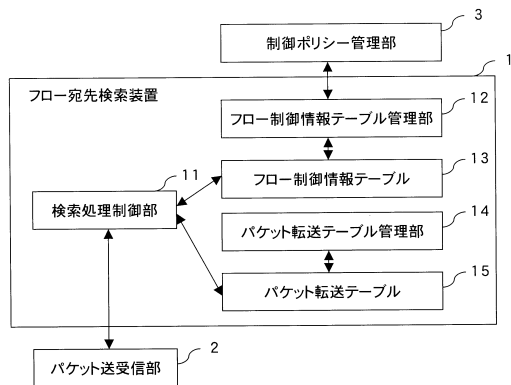
また、以上述べた実施形態は全て本発明を例示的に示すものであって限定的に示すものではなく、本発明は他の種々の変形態様および変更態様で実施することができる。従って本発明の範囲は特許請求の範囲およびその均等範囲によってのみ規定されるものである。

【符号の説明】

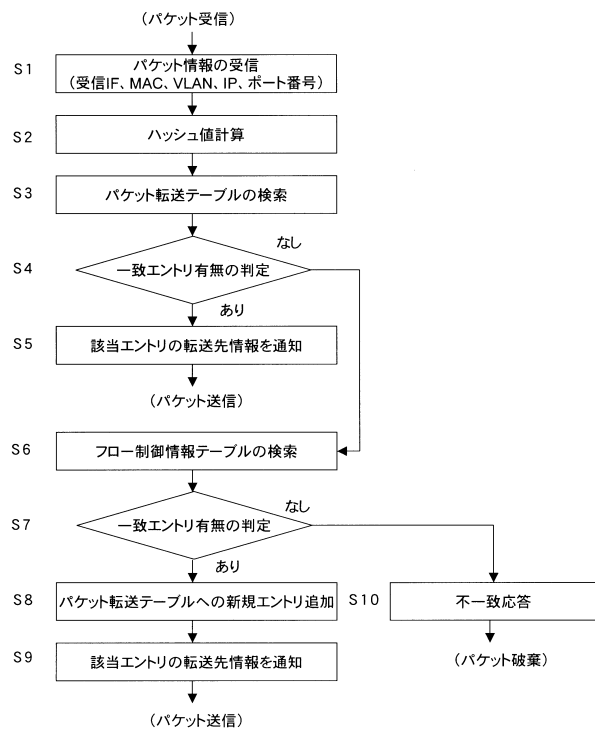
【0043】

- 1 フロー宛先検索装置
- 2 パケット送受信部
- 3 制御ポリシー管理部
- 11 検索処理制御部
- 12 フロー制御情報テーブル管理部
- 13 フロー制御情報テーブル
- 14 パケット転送テーブル管理部
- 15 パケット転送テーブル

【図1】



【図2】



フロントページの続き

審査官 浦口 幸宏

- (56)参考文献 特表2007-534203(JP,A)
特開2005-086668(JP,A)
特開2009-017439(JP,A)
米国特許出願公開第2011/0271082(US,A1)
Sangjin Han et al., PacketShader: A GPU-Accelerated Software Router, Proceedings of the ACM SIGCOMM 2010 conference, ACM New York, 2010年 8月30日, p.195~206, URL, <http://dl.acm.org/citation.cfm?doi=1851182.1851207>
岡川博一, ネットワーク技術者のための実践講座 ルータ&スイッチ ハンズオンセミナー, NETWORK WORLD, 日本, 株式会社アイ・ディ・ジー・ジャパン, 2002年12月 1日, 第7巻, 第12号, p.110~115

(58)調査した分野(Int.Cl., DB名)

H04L 12/00 - 12/26
12/50 - 12/955