

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3734704号

(P3734704)

(45) 発行日 平成18年1月11日(2006.1.11)

(24) 登録日 平成17年10月28日(2005.10.28)

(51) Int. Cl.		F I		
<b>HO4L</b>	<b>12/56</b>	<b>(2006.01)</b>	HO4L	12/56 E
<b>HO4M</b>	<b>3/00</b>	<b>(2006.01)</b>	HO4M	3/00 B

請求項の数 10 (全 24 頁)

(21) 出願番号	特願2000-378432 (P2000-378432)	(73) 特許権者	500570195
(22) 出願日	平成12年12月13日(2000.12.13)		アセンド コミュニケーションズ インコーポレーテッド
(65) 公開番号	特開2001-223749 (P2001-223749A)		アメリカ合衆国 94502 カリフォルニア, アラミーダ, ハーパー ベイ パークウェイ 1701, アセンド プラザ 1
(43) 公開日	平成13年8月17日(2001.8.17)		
審査請求日	平成14年5月16日(2002.5.16)		
(31) 優先権主張番号	09/459448	(74) 代理人	100064447
(32) 優先日	平成11年12月13日(1999.12.13)		弁理士 岡部 正夫
(33) 優先権主張国	米国 (US)	(74) 代理人	100085176
			弁理士 加藤 伸晃
		(74) 代理人	100106703
			弁理士 産形 和央
		(74) 代理人	100096943
			弁理士 臼井 伸一

最終頁に続く

(54) 【発明の名称】 パケット分類エンジン

(57) 【特許請求の範囲】

【請求項1】

パケット分類装置であって、

パケット分類要求者によって処理されているパケットからの情報を含むパケット分類要求を受信するように動作する入力インタフェースロジックと、

規則メモリエントリを記憶するように動作する規則メモリであって、前記各規則メモリエントリが演算子(operator)と判定基準メモリポイントとを含む、該規則メモリと、

判定基準メモリエントリを記憶するように動作する判定基準メモリであって、各判定基準メモリエントリが判定基準を含むような該判定基準メモリと、

前記パケット分類要求者にパケット分類結果を提供するように動作する出力インタフェースロジックと、

前記受信したパケット分類要求に応答して、

(i) 前記規則メモリからの規則メモリエントリを検索し、

(ii) 前記検索した規則メモリエントリ内の前記判定基準メモリポイントによって規定される位置において、前記判定基準メモリからの判定基準メモリエントリを検索し、

(iii) 前記検索した規則メモリエントリ内の前記演算子によって規定される演算であって、前記演算が、前記パケット分類要求からの前記パケット情報と、前記検索した判定基準メモリエントリからの前記判定基準とに基づいて実行されている、該演算を実行し、

(iv) 前記演算を実行した結果を反映するパケット分類結果を生成するように動作する

10

20

制御ロジックとを備えるパケット分類装置。

【請求項 2】

前記規則メモリは、第 1 のタイプ及び第 2 のタイプの両方の規則メモリエントリを記憶するように動作し、前記各第 1 のタイプの規則メモリエントリは演算子と判定基準メモリポインタとを含み、前記各第 2 のタイプの規則メモリエントリは、規則メモリポインタを含み、

前記制御ロジックは、前記受信したパケット分類要求に応答して、

( i ) 前記検索した規則メモリエントリが前記第 1 のタイプのエントリ、前記第 2 のタイプのエントリのいずれであるかを判定し、

( i i ) 前記検索した規則メモリエントリが前記第 1 のタイプのエントリである場合には、前記判定基準メモリエントリを検索して、前記規定された演算を実行し、

( i i i ) 前記検索した規則メモリエントリが第 2 のタイプのエントリである場合には、前記第 2 のタイプのエントリに含まれる前記規則メモリポインタによって規定される位置において別の規則メモリエントリを検索し、新しく検索された規則メモリエントリに対して、上記のステップを繰り返し行い、

( i v ) 全ての検索した第 1 のタイプのエントリによって規定される各演算を実行した結果を反映するパケット分類結果を生成するように動作する、請求項 1 に記載のパケット分類装置。

【請求項 3】

前記制御ロジックはさらに、終了の指示があるまで、さらに別の規則メモリエントリに対してステップ ( i ) ~ ( i i i ) を繰り返し行うように動作する、請求項 1 に記載のパケット分類装置。

【請求項 4】

前記終了の指示は、検索した規則メモリエントリ内のアサートされた D O N E ビットである、請求項 3 に記載のパケット分類装置。

【請求項 5】

前記制御ロジックは、前記検索したパケット分類要求に含まれる規則メモリアドレスに基づいて、前記規則メモリエントリを検索するように動作する、請求項 1 に記載のパケット分類装置。

【請求項 6】

前記制御ロジックによって検索された前記規則メモリエントリは、第 1 の規則メモリエントリであり、前記制御ロジックは、前記受信したパケット分類要求に含まれる第 1 の規則メモリアドレスに基づいて、前記第 1 の規則メモリエントリを選択するように動作し、前記制御ロジックはさらに、前記受信したパケット分類要求に同様に含まれる第 2 の規則メモリアドレスに基づいて、第 2 の規則メモリエントリを選択し、前記第 2 の規則メモリエントリに対してステップ ( i ) ~ ( i i i ) を繰り返し行うように動作する、請求項 1 に記載のパケット分類装置。

【請求項 7】

前記規則メモリエントリは、前記規則メモリエントリが、第 2 の規則メモリエントリとの複合規則を形成する第 1 の規則メモリエントリであるか否かを示す C A R R Y 指示子を含み、前記制御ロジックは、前記パケット分類結果が、前記第 1 及び前記第 2 の規則メモリエントリの両方に対する前記演算の結果を反映するように、前記第 2 の規則メモリエントリに対してステップ ( i ) ~ ( i i i ) を繰り返し行い、ステップ ( i v ) において前記パケット分類結果を生成するように動作する、請求項 1 に記載のパケット分類装置。

【請求項 8】

前記判定基準メモリエントリ内の前記判定基準、及び前記要求内の前記情報は、ネットワークアドレスである、請求項 1 に記載のパケット分類装置。

【請求項 9】

ハードウェアサーチエンジンにサーチ判定基準を保持するために用いられる 1 組の判定基準メモリ内の空間を管理する方法であって、

10

20

30

40

50

前記判定基準メモリの種々の構成の順序を保持する過程であって、前記構成が前記各構成を用いて判定基準メモリエントリを記憶するために必要とされる判定基準メモリ記憶空間の総量に準じて順序付けされるような該保持する過程と、  
前記種々の構成に従って判定基準メモリエントリを記憶するための前記判定基準メモリ内の空間の利用可能性を指示する１組のリストを保持する過程と、  
所与のサーチを実行するために、前記構成のいずれが、前記サーチに必要とされる判定基準メモリエントリを記憶する判定基準メモリ記憶空間の最低量しか必要としない最小構成であるかを判定する過程と、  
前記リストによって指示されるように、最小の利用可能な構成をサーチする過程であって、前記最小構成で開始し、最初の前記利用可能な最小構成が見つかるまで、判定基準メモリ空間の消費が増加する順序に進行するようなサーチする過程と、  
前記サーチのために必要とされる前記判定基準メモリエントリを記憶するために、前記最小の利用可能な構成を割り当て、前記割り当てられた構成が利用可能でなくなることを示すために、利用可能性のリストを更新する過程とを有する、方法。

10

**【請求項 10】**

前記最小構成以外の構成が割り当てられる場合には、前記割り当てられた構成のうち不要な部分が、別の判定基準メモリエントリに割り当てることができる異なる構成として用いることができるか否かを判定する過程と、前記異なる構成の前記利用可能性を示すために前記利用可能性のリストを更新する過程とをさらに有する、請求項 9 に記載の方法。

20

**【発明の詳細な説明】****【0001】****【発明の属する技術分野】**

本発明は、データ通信ネットワークの分野に関する。

**【0002】****【従来の技術】**

データ通信ネットワークでは、交換器のようなネットワーク装置を用いて、パケットをネットワーク内にルーティングする。各交換器は典型的には多数の回線インタフェースを有し、各インタフェースは、異なるネットワークセグメントに接続される。所与の回線インタフェースにおいてパケットが受信されるとき、転送ロジックは、そのパケットがどの回線インタフェースから送信されるべきかを判定し、そのパケットは適当な出力回線インタフェースに移送されて、ネットワーク内の行先に向けて送出される。

30

**【0003】**

交換器のようなネットワーク装置においては、パケットフィルタリングを実行することが知られている。パケットフィルタリングを用いて、トラフィックモニタ及びセキュリティ目的のような、種々のネットワーク管理を行うことができる。フィルタリング判定基準はネットワーク管理者によって確立され、フィルタリング動作を実行する交換器或いは他の装置に提供される。交換器によって受信されたパケットは、その特性が任意の確立されたフィルタについての判定基準に適合するか否かを判定するために検査される。１つ或いは複数のフィルタのための判定基準を満足するパケットの場合、そのフィルタに関連する所定の動作が実行される。例えば、ある状況下では、所与のネットワークノードから発信するパケットは、ネットワーク内に転送されずに破棄されることが望ましい場合もある。パケットソースアドレスが、そのパケットが破棄されるべきであるノードのアドレスである特定の値に正確に一致するという判定基準を有するフィルタを定義することができる。そのフィルタに関連する動作は、パケットを破棄することである。ソースアドレスがこの判定基準を満足するパケットを受信するとき、そのパケットは、通常通りに転送されずに破棄される。

40

**【0004】**

パケットをフィルタリングするために用いる場合がある、多数の異なる種類の判定基準が存在する。これらの判定基準は、厳密な一致と、範囲検査、すなわちパケット内の値がある範囲内の値に入るか否かの検査とを含む。ソースアドレス、行先アドレス、ポート識別

50

子、サービスタイプ等のような多数のパケットパラメータを判定基準として用いることができる。実用的にするためには、パケットフィルタリングプロセスによって、そのフィルタが、これらの判定基準と他の判定基準からなる異なる組み合わせを用いて、柔軟に定義されなければならない。

**【 0 0 0 5 】**

パケットフィルタリングに固有の複雑さのため、この処理は従来から、パケットフィルタリングをサポートする交換器或いは他のネットワーク装置内のソフトウェアにおいて、広範に或いは排他的に実行されてきた。しかしながら、ソフトウェア系のフィルタリングは、高いパケット転送性能が要求されるときに欠点を露呈する。例えば、従来のシステムは回線速度で信頼性のある (robust) パケットフィルタリングを実行できなかったため、ネットワーク管理者は、ネットワーク応答性とネットワークセキュリティとの間でトレードオフを余儀なくされており、問題となっていた。

10

**【 0 0 0 6 】****【 発明が解決しようとする課題 】**

本発明によれば、パケットフィルタリングを行うために高速のパケット分類を提供する、ネットワーク装置内のパケット処理ロジックが開示される。その分類装置のアーキテクチャは、複雑なフィルタ判定基準の定義において実質的な柔軟性を提供する。十分に高い速度でロバストフィルタリングを実行し、パケット転送性能が低下するのを避けることができる。

**【 0 0 0 7 】****【 課題を解決するための手段 】**

パケット分類装置は、規則メモリと判定基準メモリとを備える。1つのタイプの規則メモリエントリは、判定基準メモリエントリに対する演算子及びポインタを含む。演算子は、E Q U A L (厳密な一致) 或いは L E S S T H A N のような、実行されることになる比較演算を定義する。判定基準メモリエントリは、比較の一方の辺の被比較数として用いられる1つ或いは複数の値を含み、受信したパケットからの対応する値は、比較の他の辺に現れる。例えば、判定基準メモリからの1つの被比較数は、ソースアドレスを表す。この値は、受信したパケットのソースアドレスフィールド内に現れる値と比較される。

20

**【 0 0 0 8 】**

制御ロジックはパケット分類要求にตอบสนองして、規則メモリからの規則メモリエントリを検索し、規則メモリエントリにおいて判定基準メモリポインタによって識別された判定基準メモリエントリを検索し、判定基準メモリエントリの値及び分類要求に含まれる対応する値において、規則メモリエントリの演算子によって規定される演算を実行する。この手順は、一定の終了条件が満たされるまで、一連の規則メモリエントリに対して繰返し行われ、その結果、その分類演算の結果を反映して、パケット分類結果が生成される。この結果はパケットプロセッサに提供され、その分類結果に基づいて適当な動作が行われる。

30

**【 0 0 0 9 】****【 発明の実施の形態 】**

本発明の他の態様、特徴及び利点は、以下の詳細な説明において開示される。

**【 0 0 1 0 】**

図1には、ネットワーク交換器10が、各ネットワークセグメント14に接続される多数の回線インタフェース12を備えるものとして示される。回線インタフェース12は、パケット転送のための回線インタフェース12の間の接続部を提供するために用いられる交換器構造部16に接続される。交換器構造部16の動的な構成を含む、交換器10の全動作は、交換器制御部18によって制御される。一般に、種々のネットワークセグメント14は異なるタイプからなる場合がある。例えば、ネットワークセグメント14のうちのあるものは、O C - 3 / S T M - 1 及び O C - 1 2 / S T M - 4 のような種々の標準的なシグナリング速度のうち任意の速度で動作する光学的なリンクであり得る。ネットワークセグメント14の他のものは、例えば同軸ケーブルを用いて、種々のフォーマットの信号を搬送する非光学的なリンクであってもよい。

40

50

## 【 0 0 1 1 】

言うまでもなく、各回線インタフェース 1 2 は、接続される特定のタイプのネットワークセグメント 1 4 で動作するために設計される。各回線インタフェース 1 2 の主なタスクは、取り付けられたネットワークセグメント 1 4 から受信したパケット或いはフレームを、他の回線インタフェース 1 2 に取り付けられたネットワークセグメント 1 4 上で転送するために、交換器構造部 1 6 を介して別の回線インタフェース 1 2 に転送することと、取り付けられたネットワークセグメント 1 4 上で転送するために、交換器構造部 1 6 を介して他の回線インタフェース 1 2 からパケットを受信することである。

## 【 0 0 1 2 】

図 2 は、1つのタイプの回線インタフェース 1 2 の構造を示す。このインタフェースは、4つの個別の光インタフェースポートを含み、各ポートは、物理的な入力/出力及びフレーム処理回路 2 0 と、転送エンジン 2 2 とを備える。転送エンジン 2 2 は、交換器構造部インタフェースロジック 2 4 に接続され、インタフェースロジック 2 4 は図 1 の交換器構造部 1 6 とのインタフェースを構成する。また転送エンジンは、回線インタフェース I / O プロセッサ ( I O P ) 2 6 とのインタフェースも構成する。タイミング制御ロジック 2 8 及び D C パワー回路 3 0 も備えられる。

10

## 【 0 0 1 3 】

各転送エンジン 2 2 は、接続される物理 I / O ブロック 2 0 と交換器構造部インタフェース 2 4 との間の双方向データバスを提供する。受信されたパケットは、図 1 の交換器構造部 1 6 を介して別の回線インタフェース 1 2 に転送するために、多数の固定サイズの A T M 型のセルにセグメント化される。交換器構造部インタフェース 2 4 を介して交換器構造部 1 6 から受信されたセルは、接続される物理 I / O ブロック 2 0 への転送を行うために、パケットに再構成される。

20

## 【 0 0 1 4 】

I O P 2 6 は、バックグラウンド機能、すなわちパケット毎には実行されない、パケットの転送をサポートする機能を実行する汎用プロセッサである。I O P 2 6 によって実行される1つの機能は、図 1 の交換器制御部 1 8 からパケット転送情報及びパケットフィルタリング情報を受信し、転送エンジン 2 2 にその情報を分配することである。この処理は以下に記載される。

## 【 0 0 1 5 】

図 3 は、転送エンジン 2 2 のブロック図を示す。着信セグメント化 - 再構成 ( S A R ) 論理ブロック 4 0 は、図 2 の物理 I / O ブロック 2 0 から交換器構造部 1 6 へのデータバスを提供し、発信 S A R 論理ブロック 4 2 は、交換器構造部 1 6 から各物理 I / O ブロック 2 0 へのデータバスを提供する。各 S A R 4 0、4 2 は、セグメント化機能及び再構成機能を実行する際に用いられる各制御メモリ 4 4、4 6 と、各パケットメモリ 4 8、5 0 に接続される。

30

## 【 0 0 1 6 】

S A R 装置 4 0、4 2 はともに、64ビット P C I バス 5 4 を介して、パケットヘッダ分配器 ( P H D ) 特定用途向け集積回路 ( A S I C ) 5 2 に接続される。以下に詳細に記載するように、P H D A S I C 5 2 は、P C I バス 5 4 と個別の64ビットバス 5 6 との間の F I F O キューインタフェースを提供する。バス 5 6 は、P H D A S I C 5 2 を転送プロセッサ ( F P ) 5 8 及び転送プロセッサメモリ 6 0 と接続する。また P H D A S I C 5 2 は、個別のバス 6 2 によって図 2 の I O P 2 6 にも接続される。

40

## 【 0 0 1 7 】

図 4 は、図 3 の P H D 5 2 の構造を示す。1組の受信キュー、すなわち R X キュー 6 4 は、F P 5 8 に向かうパケットヘッダ及び他の情報を一時的にバッファリングするために用いられる。図のように、R X キュー 6 4 は4つあり、2つが高優先度のトラフィック用であり、残りの2つは低優先度のトラフィック用である。高優先度トラフィックの一例は、約束された速度 ( committed rate ) のような、高品質のサービス ( Q O S ) 保証 ( High Q uality of Service guarantee ) を有するトラフィックである。低優先度のトラフィック

50

は、最善努力 (best-efforts) のような、低い Q O S 保証しかもたないか、或いは全く保証がないトラフィックである。各優先度レベルの場合に、着信 S A R 4 0 から生じるトラフィック用の 1 つのキュー (「 0 」を付される) と、発信 S A R 4 2 から生じるトラフィック用の別のキュー (「 1 」を付される) とが存在する。1 組の送信キュー、すなわち T X キュー 6 6 は、F P 5 8 から S A R 4 0、4 2 に向かうパケットヘッダ及び他の情報を一時的にバッファリングするために用いられる。ルーティング及び分類エンジン 6 8 は、F P 5 8 に代わって、経路探索及び種々のパケットフィルタリング検査を実行する。パケットフィルタリング動作は以下に記載される。ルーティング及び分類エンジン 6 8 は、信号線 6 9 を介してキュー 6 4、6 6 から状態情報を受信し、以下のようにして、F P 5 8 においてこの情報を利用できるようにする。

10

**【 0 0 1 8 】**

転送エンジン 2 2 の全動作が、図 3 及び図 4 を参照して記載されるであろう。パケットは、図 2 の関連する物理層回路 2 0 から着信 S A R 4 0 によって受信され、パケットメモリ 4 8 に記憶される。着信 S A R 4 0 は、パケットヘッダを、P H D 5 2 内の適当な R X キュー 6 4 の 1 つに転送する。F P 5 8 は P H D 5 2 をポーリングしてキュー状態を判定するとともに、適宜 R X キュー 6 4 からのパケットヘッダを検索する。ヘッダ処理の一部として、F P 5 8 は、ルーティング及び分類要求において、各ヘッダからの一定の情報要素を、ルーティング及び分類エンジン 6 8 に送出する。ルーティング及び分類エンジン 6 8 は、その要求内のヘッダ要素に対する経路探索及び種々のパケットフィルタリング検査を実行し、これらの検査の結果を、結果キュー (以下に記載する) に入れる。F P 5 8 は、結果キューから経路探索及び分類結果を取得し、これらの結果を用いて、そのパケットのための新しいヘッダを作成する。その新しいヘッダは、セグメント化した後にパケットが転送されることになる内部回路を特定する情報とともに、T X キュー 6 6 の 1 つを介して P H D 5 2 に返送される。着信 S A R 4 0 は新しいヘッダを検索し、受信したパケットのペイロード部分とともにそのヘッダをパケットメモリ 4 8 に入れ、新しいパケットをセグメント化し、その結果として生成されたセルを、F P 5 8 によって特定された内部回路上にある図 1 の交換器構造部 1 6 に転送する。

20

**【 0 0 1 9 】**

発信方向では、発信 S A R 4 2 が図 1 の交換器構造部 1 6 からパケットを受信し、これらのパケットを再構成してパケットメモリ 5 0 に入れる。パケットヘッダは P H D 5 2 に送出され、F P 5 8 によって P H D 5 2 から検索される。大部分のパケットの場合、経路探索及びフィルタリング検査は、着信処理中に既に行われているため、これらの動作は繰返し行われず。しかしながらプロトコルによっては、着信及び発信の両方の場合に探索及びフィルタリングを必要とするため、これらの動作は、ルーティング及び分類エンジン 6 8 とともに、F P 5 8 によって任意選択的に実行される。妥当な場合には、F P 5 8 は、セグメント化された発信パケットが受信される内部回路の同一性に一部基づいて、そのパケットのための新しいヘッダを作成する。この新しいヘッダは、送信回路情報とともに、P H D 5 2 に書き込まれる。P H D 5 2 は、新しいヘッダを、発信 S A R 4 2 に転送する。発信 S A R 4 2 は、この新しいヘッダをパケットペイロードとともにパケットメモリ 5 0 に入れ、そのパケットを、図 2 の関連する物理層インタフェース 2 0 に送信する。

30

40

**【 0 0 2 0 】**

図 5 はルーティング及び分類エンジン 6 8 の構造を示す。図 3 の F P 5 8 からの要求は、1 つの要求キュー 7 0 に入れられ、1 つの結果キュー 7 2 において結果が返送される。各キュー 7 0 及び 7 2 は、1 6 個までの要求 / 結果エントリを保持する。ルート探索エンジン (R L E) 7 4 は、典型的にはヘッダに含まれる行先アドレス (D A) に基づいて、経路探索を実行する。パケット分類エンジン (P C E) 7 6 は、パケットヘッダに含まれる特定の情報に基づいて、パケットフィルタリング検査を実行する。P C E 7 6 の動作が以下に詳細に記載される。入力 F I F O バッファ 7 8 は、要求キュー 7 0 と、R L E 7 4 及び P C E 7 6 それぞれとの間に配置され、出力 F I F O バッファ 8 0 は、R L E 7 4 及び P C E 7 6 それぞれと、結果キュー 7 2 との間に配置される。F I F O 7 8 及び 8 0 は、

50

R L E 7 4 によって実行される処理と、P C E 7 6 によって実行される処理との間を分割する手段を提供する。マルチプレクサ 8 1 によって、F P 5 8 は、結果キュー 7 2、或いは要求キュー 7 0、結果キュー 7 2 及び図 4 の信号線 6 9 上に現れる状態情報からの状態を含む状態情報のいずれかを読み出すことができるようになる。これらのエントリの構造が以下に記載される。

【 0 0 2 1 】

図 6 は、図 5 の要求キュー 7 0 を介して P C E 7 6 及び R L E 7 4 に渡されるルーティング及び分類要求の構造を示す。その要求のサイズは、4 つの 6 4 ビットワードである。種々のフィールドが以下に定義される。

【表 1】

フィールド名	説明
Type	R L E エントリタイプ：0 = ノード、1 = リーフ (葉)
I n d .	R L E インダイレクトルート： 1 = 間接、0 = 直接
R e s .	未使用の予約ビット
O r d e r	R L E ポインタに付加される D A ビットの数
R L E P t r .	D A が (O r d e r に基づいて) 付加される R L E エントリの基底アドレス
P C T ルート 0	P C E 規則 0 のための開始アドレス
P C T ルート 1	P C E 規則 1 のための開始アドレス
0	0 に設定され、アライメント検査のために用いられる。
リクエスト I D	要求識別子、要求と一致するようになる結果にコピーされる。
I P T O S	受信したパケットのサービスタイプ (T O S) フィールドの内容
I P プロトコル	受信したパケットのプロトコルフィールドの内容
T C P フラグ	受信したパケットの T C P フラグフィールドの内容
I P ソースアドレス	受信したパケットの I P ソースアドレス
I P 行先アドレス	受信したパケットの I P 行先アドレス
T C P / U D P ソースポート	パケットを受信した T C P / U D P ポートの識別子
T C P / U D P 行先ポート	受信したパケットの行き先となる T C P / U D P ポートの識別子

【 0 0 2 2 】

上記の表に示されるように、個別の 2 組の分類検査のための規定があり、1 つは「P C E R o o t 0」を付されるアドレスで開始し、もう 1 つは「P C E R o o t 1」で開始する。これらの個別の開始アドレスの重要性が以下に記載される。

【 0 0 2 3 】

上記のように、その要求の適切なフィールドが、図 5 の R L E 7 4 及び P C E 7 6 のための各入力 F I F O 7 8 に供給される。R e q . I D 及び I P . D e s t . A d d r のようなフィールドのうちいくつかは、R L E 7 4 及び P C E 7 6 の両方に供給される。他のフィールドは一方にのみ供給される。特に P C E にルーティングされるフィールドの使用法が以下に記載される。

【 0 0 2 4 】

図 7 及び図 8 は、図 4 のルーティング及び分類エンジン 6 8 から読み出される 2 つの異なるタイプのエントリの各構造を示す。図 7 は、図 5 の結果キュー 7 2 から得られ、分類サーチの結果を搬送する結果エントリを示す。図 8 は、図 3 の F P 5 8 に状態情報を搬送するために用いられる状態エントリを示す。

【 0 0 2 5 】

図 7 に示される結果エントリのフィールドは以下のように定義される。

【表 2】

フィールド名	説明
T	タイプ：0 = PCE 結果、1 = PCE 状態
要求 ID	(要求からの) 要求識別子
P	PCE 不一致検出：0 = 一致検出、1 = 不一致検出
I	RLE インダイレクトルート：0 = 正常、1 = 間接
L	PLE ロングサーチ：0 = ショート、1 = ロング
E	エラー指示子：0 = 正常、1 = エラー
Z	0 パディング
R 1 - M	PCE ルート 1 における一致 (P = 0 の場合にのみ有効)：0 = ルート 0 において一致、1 = ルート 1 において一致
深さ	ルート探索サーチの深さ
PCE マッチアドレス	PCE において検査された最後の規則のアドレス
RLE フラグ	RLE テーブルエントリからのフラグ
RLE ネクストホップポインタ	RLE テーブルエントリからのポインタ

10

20

【 0 0 2 6 】

図 8 に示される状態エントリのフィールドは以下のように定義される。

【表 3】

フィールド名	説明
0	未使用、0 に設定される
TX メッセージ	プロセッサから IOP メッセージキューへの転送時に残る空間
RCE 結果	結果キュー 7 2 の未決定エントリの数、キューが空きのときのみ状態が挿入されるため、通常は 0
RCE 要求	要求キュー 7 0 内の空きエントリの数
T x - 0	TX キュー 6 6 内の空きエントリの数
T x - 1	TX キュー 6 6 内の空きエントリの数
H i - 0	RX キュー 6 4 内の空きエントリの数
H i - 1	RX キュー 6 4 内の空きエントリの数
L o - 0	RX キュー 6 4 内の空きエントリの数
L o - 1	RX キュー 6 4 内の空きエントリの数

40

【 0 0 2 7 】

50

ルーティング及び分類エンジン 68 の全体的な動作が、図 5 ~ 図 8 を参照して以下に記載される。図 3 の F P 5 8 は、探索及び分類要求を、要求キュー 70 に書き込む。要求が要求キュー 70 の先頭に達するとき、その要求からの種々の情報要素は、R L E 7 4 及び P C E 7 6 のための各入力 F I F O 7 8 に同時に書き込まれる。R L E 7 4 及び P C E 7 6 は独立して各要求の個別の部分において動作し、一般に、異なる時間で、所与の要求に対する各処理動作を終了する。これらの動作の結果が、出力 F I F O 8 0 に書き込まれる。所与のバケットに対する両方の結果の組が、出力 F I F O 8 0 の先頭に達したとき、1 つの組み合わせられた結果が、結果キュー 72 に書き込まれる。その組み合わせられた結果は F P 5 8 によって読み出され、それを用いて F P 5 8 は、上記のように、図 3 の S A R 4 0 及び 4 2 のための新しいバケットヘッダ及び回路情報を作成する。

10

## 【 0 0 2 8 】

より詳細には、F P 5 8 は、バッチ形式でルーティング及び分類エンジン 68 を用いる。要求キュー 70 内に十分な空間があるとき、一連の要求、1) 4 つの結果エン트리、2) 3 つの結果エン트리と後続する 1 つの状態エン트리、3) 2 つの結果エン트리と後続する 2 つの状態エン트리、4) 1 つの結果エン트리と後続する 3 つの状態エン트리、5) 4 つの状態エントリが存在する。

## 【 0 0 2 9 】

一般に F P 5 8 は、結果キュー 72 が空になるまで、読出しコマンドを発行し、1 つ或いは複数の状態エントリが結果ブロックに含まれるときには必ず、空であると推定される。その後 F P 5 8 は、ルーティング及び分類エンジン 68 が次の要求のバッチを処理している間に、その結果を用いる。F P 5 8 は、R L E 7 4 及び P C E 7 6 がビジーで保持され、キュー 70 及び 72 と、F I F O 7 8 及び 80 とがオーバーフローするのを防ぐように、状態情報を用いて要求の流れを管理する。

20

## 【 0 0 3 0 】

例示される実施形態では、読み出すことができる状態エントリは 1 つしか存在せず、結果ブロック内に多数の状態エントリが存在する場合には、この 1 つのエントリに対して何度も読出しが行われたことを表すことに留意されたい。別の実施形態では、例えば統計情報を収集するために、或いは他のバックグラウンド処理を行うために、第 2 ~ 第 4 の状態エントリにおいて付加的な低優先度の情報を供給することが有用な場合もある。

## 【 0 0 3 1 】

結果に状態情報を付加する 1 つの重要な利点は、F P バス 56 を用いる際の効率が改善されることである。F P 5 8 が結果の読出しコマンドを発行するときにはいつでも、有効な結果情報或いは有効な状態情報のいずれかが返送される。さらに、結果ブロックはバースト形式で返送され、読出しに関するオーバーヘッドを低減できるようにする。また F P 5 8 は、1 つの読出しトランザクションにおいて R L E 7 4 及び P C E 7 6 キューのついでの情報と、R X キュー 64 及び T X キュー 66 についての情報とを取得する。

30

## 【 0 0 3 2 】

図 9 は、図 5 の P C E 7 6 の構造を示す。フィルタ及び結合子 (以下に記載する) を表すデータは、規則メモリ (R M) 82 及び判定基準メモリ (C M) 84 に記憶される。C M 84 は、3 つの一般的にアドレス指定されたメモリ C M 0 86、C M 1 88 及び C M 2 90 を備える。3 つの比較論理ブロック 92、94 及び 96 は、判定基準メモリ 86、88 及び 90 にそれぞれ関連する。アドレス指定及び制御ロジック 98 は、図 5 の要求キュー 70 から受信した要求を復号化して、R M 82 及び C M 84 のためのアドレスを生成し、各要求によって要求されるような多数の規則を通して配列され、図 5 の結果キュー 72 に渡される結果を生成する。またそのアドレス指定及び制御ロジック 98 は、図 3 の I O P 26 に対するインタフェースも形成し、I O P 26 によって R M 82 及び C M 84 を読出し及び書込みできるようにする。バストランシーバ 100 は、I O P 26 と R M 82 及び C M 84 との間の必要なデータバスを提供する。A N D ゲート 102 は、比較論理ブロック 92、94 及び 96 からの対応する M A T C H n 出力が全て真であるとき、1 つの M A T C H 信号を与える。

40

50

## 【 0 0 3 3 】

パケットフィルタリングのための規則の組は典型的には、ネットワーク管理ステーション ( N M S ) によって発信されるが、特定された流れに基づいて F P 5 8 によって動的に割り当てられることもできる。以下の情報の一部或いは全部が、フィルタのために N M S 或いは F P 5 8 によって供給される。その情報は、マスクを有する I P 行先アドレス、マスクを有する I P ソースアドレス、 I P プロトコル識別子、 T C P / U D P ソースポート及び行先ポート識別子、サービス識別子の I P タイプ及びマスク、種々のフラグである。そのパケットがフィルタ判定基準と一致するか否かを判定するために、フィルタからの種々の情報要素が、受信された各パケットからの対応する要素と比較される。一致する場合には、パケットがその行先に向けて進めるようにする動作のような、いくつかのデフォルト動作が典型的には行われる。

10

## 【 0 0 3 4 】

従来通りに、パケットフィルタは、線形にサーチされる比較の組からなる順序付けされたリストとして表される。 P C E 7 6 では、フィルタ要素は、判定基準 ( 比較値 ) 及び規則 ( リスト自体及び各比較のために用いられる演算子 ) に分割される。この規則と判定基準の分離は、個別の規則メモリ ( R M ) 8 2 及び判定基準メモリ ( C M ) 8 4 を使用する際に反映される。メモリ 8 2 及び 8 4 は各機能のために個別に最適化され、その結果、効率及び性能が改善される。また、 C M 8 4 内のエントリは、 R M 8 2 内の多数の規則によって参照されることができ、さらに記憶効率が高められる。

## 【 0 0 3 5 】

R M 8 2 は、規則メモリエントリのアレイを含み、各エントリは 2 つのタイプのうちの 1 つであり得る。第 1 のタイプは、 1 組の演算子と、対応するフィルタのための被比較数を記憶する C M 8 4 の行へのポインタとを含む。第 2 のタイプは、別の規則メモリエントリへのポインタを含む。これらのエントリを用いて、順次サーチされる 1 組の規則において、不連続のセグメント間でジャンプを行う。例示される実施形態では、 R M 8 2 は 1 6 K までのエントリを含むことができる。

20

## 【 0 0 3 6 】

C M 8 4 は、 3 つの個別のメモリ C M 0 8 6、 C M 1 8 8 及び C M 2 9 0 にセグメント化され、例示される実施形態では、各メモリは 4 K までのエントリを含むことができる。 C M 8 4 の機構は、 I P パケット分類において固有の階層を利用する。一定のフィールド上でのフィルタリングは通常、同様に他のフィールドに基づくフィルタリングを伴うので、個別のメモリ C M 0、 C M 1 及び C M 2 内にどのフィールドが記憶されるかを限定することが望ましい。これらの限定によって、さらに記憶効率が高められる。大部分の一般的にフィルタリングされたフィールド、ソースアドレス及び行先アドレスは、 3 つの全てのメモリ C M 0 8 6、 C M 1 8 8 及び C M 2 9 0 においてサポートされる。以下に記載するように、他のフィールドは、 C M 1 8 8 及び / または C M 2 9 0 においてのみサポートされる。このアーキテクチャは、 C M 8 4 内の空間を割り当てることができる自由度を最大限にすると同時に、効果的な並列サーチを可能にする。 C M 8 4 の構造及び使用法は、より詳細に以下に記載される。

30

## 【 0 0 3 7 】

図 1 0 は、規則メモリエントリとも呼ばれる、図 9 の R M 8 2 内のエントリの構造を示す。各 3 9 ビットエントリは、 1 ビットのタイプフィールドを有する。このフィールドが 1 の場合には、エントリのビット 1 3 ~ 0 は、 R M 8 2 内の別の位置へのポインタ、すなわち別の規則メモリエントリへのポインタを含む。このフィールドが 0 である場合には、そのエントリはフィルタ検査を実行するための情報を含む。この場合には、ビット 1 1 ~ 0 は、検査のためのオペランドが見出される、 C M 8 4 の行のアドレスを含んでおり、ビット 3 5 ~ 1 2 は、その要求からの、各オペランド及びフィールドにおいて実行される動作の符号化ビットを含む。これらの動作は以下に詳細に記載される。ビット 3 6 は、例えば範囲検査を実行するための複合規則を形成するために用いられるキャリービットである。キャリービットが 0 である場合には、その規則はそれ単独で評価される。キャリービット

40

50

が1である場合には、その規則は、次の規則も真と評価する場合にのみ、真と評価する。ビット37は、要求の一部として検査される規則の列の最後に到達したことを示す終了(Done)ビットである。

【0038】

判定基準演算子フィールドは、8つの3ビット論理演算子コードを含む。各演算子コードは、その要求及び判定基準メモリエントリから選択される対応する被比較数において実行される演算を規定する。判定基準メモリエントリのフィールドが以下に記載される。被比較数に対する判定基準演算子ビットの割当ては、以下の通りである。

35 ~ 33	CM0	SA/DAフィールド	
32 ~ 30	CM1	プロトコルフィールド	10
29 ~ 27	CM1	ソースポートフィールド	
26 ~ 24	CM1	SA/DA 或いは DP フィールド	
23 ~ 21	CM2	プロトコルフィールド	
20 ~ 18	CM2	TOS 或いは マスクフィールドを有する TOS	
17 ~ 15	CM2	ソースポート 或いは マスクフィールドを有する フラグ	
14 ~ 12	CM1	SA/DA 或いは SP 或いは DP フィールド	

【0039】

演算子コードは実行される比較を規定しており、その場合、要求からの被比較数は演算子の左側にあり、判定基準メモリエントリからの被比較数は右側にある。例えば、演算子が「>」である場合には、評価される式は、(要求データ > 判定基準データ)である。演算子コードは以下の通りである。

000	Greater than	(~より大きい)	
001	Less than	(~より小さい)	
010	Equal	(~に等しい)	
011	Not Equal	(~に等しくない)	
1xx	Don't care	(未定義、すなわち被比較数値に関わらず強制的に真)	20

【0040】

判定基準演算子を用いて、以下に記載するように、比較論理ブロック92、94及び96内のロジックを構成する。

【0041】

図11は、図9のCM0 86におけるエントリの構造を示す。各エントリは、38ビット長である。1つのビット、すなわち、ビット37を用いて、そのエントリのための2つの可能な構成を、32ビットソースアドレス(SA)か、或いは32ビット行先アドレス(DA)のいずれかとして区別する。ビット31~0は、対応するフィルタによって必要とされるSA 或いはDA値を含む。ビット36~32は、エントリ内のSA/DAと、その要求のSA/DAとの間の比較の範囲を制限するために用いられる5ビットの符号化されたマスク値を含む。マスクの使用法は以下に詳細に記載される。

【0042】

図12は、図9のCM1 88内のエントリの構造を示す。各エントリは47ビット長である。ビット46~45に示されるように、4つの異なる構成を取り得る。PTCLフィールドは、4つの全ての構成においてIPプロトコルを特定する。構成2及び3の16ビットのSP及びDPはそれぞれ、ソースポート識別子及び行先ポート識別子を表す。ビット36~32の内容は、構成2及び3では未定義である。

【0043】

図13は、図9のCM2 90内のエントリの構造を示す。各エントリは51ビット長である。ビット50~48に示されるように、8つの異なる構成を取り得る。構成2~7のTOSフィールドは、サービスのIPタイプを特定する。構成3~7では、TOSマスクフィールドは、以下のように、TOS比較の範囲を限定するために用いられる8ビットマスクを含む。8ビットFLAGフィールドは、TCP/UDPパケットからの対応する

10

20

30

40

50

ビットに対して比較されるフラグ値を含む。8ビットFLAG\_MSKフィールドは、以下のように、FLAG\_S比較の範囲を限定するために用いられる。

#### 【0044】

図14は、比較論理ブロック92、94及び96の全般的な構造を示す。2つ以上の比較器ロジック104-1、...104-nのブロックを用いて、多数の比較を平行して実行し、その場合に、各比較は、要求の所与のフィールドと判定基準メモリエントリの対応するフィールドとの間で行われる。CM0\_86の場合の比較ロジック92では、例えば、2つの比較器論理ブロック104が用いられ、1つは要求のソースアドレスフィールド用であり、もう1つは要求の行先アドレスフィールド用である。CM1\_88の場合の比較ロジック94は、ソースアドレス、行先アドレス、IPプロトコル、ソースポート及び行先ポートのための比較器論理ブロック104を含む。CM2\_90の場合の比較ロジック96は、ソースアドレス、行先アドレス、IPプロトコル、ソースポート、行先ポート、マスクのないサービスタイプ、マスクのあるサービスタイプ及びフラグのための比較器論理ブロック104を含む。

10

#### 【0045】

比較器論理ブロック104からの出力は、NOT\_EQUAL(~と等しくない: <math>\neq</math>)、EQUAL(~と等しい: <math>=</math>)、LESS\_THAN(~より小さい: <math><</math>)及びGREATER\_THAN(~より大きい: <math>></math>)のための指示を含む。これらの信号は、DON'T\_CARE機能を実施するために用いられる論理「1」とともに、各セクタ106-1、...106-nの入力に供給される。セクタ106は、制御入力として、演算子タイプ規則メモリエントリから演算子を受信する。これらの演算子は、上記のように、規則メモリエントリのビット35~12内に存在する。

20

#### 【0046】

セクタ106の各出力は別のセクタ108に供給され、セクタは、判定基準メモリエントリからの構成ビットに基づいて、セクタ106の出力の種々の組み合わせの中から選択を行う。例えば、CM0の場合の比較ロジック92では、構成セクタ108は、判定基準メモリエントリのビット37の値に基づいて、SA比較結果とDA比較結果との間で選択を行う。他の比較論理ブロック94及び96内の構成セクタ108も同様に動作する。構成セクタ108からの出力信号MATCHは、その要求内のデータが、各判定基準メモリ86、88或いは90からの判定基準を満足するか否かを示す。図9に示されるように、比較ブロック92、94及び96からのMATCH出力はANDゲート10によって互いにANDを取られ、分類動作を制御するために、アドレス指定及び制御ロジック98に1つのMATCH指示を与える。

30

#### 【0047】

図15は、比較器論理ブロック104の全般的な構造を示す。EQUAL比較器110は、2つの被比較数が等しいか否かを判定し、LESS\_THAN比較器112は、被比較数の一方が他方の被比較数より小さいか否かを判定し、GREATER\_THAN比較器114は、被比較数の一方が他方の被比較数より大きいか否かを判定する。EQUAL比較110からの出力は、インバータ116によって反転され、NOT\_EQUAL指示が得られる。

40

#### 【0048】

各比較器110、112及び114への入力、CM84からの被比較数(「CM被比較数」として示される)及び要求からのおそらくマスクされた被比較数(REQ被比較数として示される)である。マスク処理ロジックが、関連するマスクを有するフィールドのために用いられる。ANDゲート118は、ビット毎にマスク処理を実施する。Flags\_Maskの場合と同様に、マルチビットマスク(「CMマスク」として示される)が直接用いられる場合があり、或いはSA/DAマスクの場合と同様に、復号化されるか、又はエキスパンダロジック120により拡張される場合もある。エキスパンダロジック120は、5ビットの符号化されたマスク値によって示されるように、多数の後続のビット位置において0を有し、その他の場所では1を有する32ビット値を生成する。例えば、

50

マスク値が、10進の11に等価な2進の01011である場合には、復号化されたマスクは、先行する21ビット位置において1を有し、後続の11ビット位置で0を有する16進のFFFFFF800である。このマスクは、SA/DAの最上位21ビットのみが比較結果に作用することを示す。

#### 【0049】

パケット分類エンジン(PCE)76の動作は、全般に以下のように進む。

1. RM82及びCM84が図2のIOP26によって初期化される。これは、動的な割当て、或いはネットワーク管理ステーション(NMS)(以下に記載する)のいずれかによって、電源投入時、及び動作中に行われる。

2. FP58によって提示されるパケット分類要求が、図5の要求キュー70から検索される。

3. RM82が、その要求のルート0アドレスの内容によって索引付けされ、そのサーチの第1の規則メモリエントリを検索する。そのエントリがポインタタイプである場合には、検索されたエントリ内の規則メモリアドレスに対して、このステップが繰り返される。このステップは、何度も繰り返すことができる。

4. 検索された規則メモリエントリが演算子タイプである場合には、判定基準メモリエントリが、規則メモリエントリ内のCMアドレスにより規定される位置において検索される。CM84からの選択された被比較数が、規則メモリエントリ内の演算子に従って、対応する要求のフィールドと比較される。上記のように、種々のフィールドがマスクされる場合がある。

5. 1に設定されたDONEビットが到達するか、或いは一致条件が見出される(すなわち比較演算の結果が真になる)まで、規則メモリアドレスが1だけインクリメントされる。ある規則は、CARRYビットを設定する場合もあり、それは、一致が宣言される前に、次の規則も真であると評価することを要求している。

6. サーチにおいて遭遇する任意の規則メモリエントリが、ポインタタイプのエントリである場合には、判定基準メモリエントリではなく、別の規則メモリエントリを指している。この場合、逐次的な規則評価が継続され、指示される規則メモリエントリで開始される。

7. 要求のルート0アドレスで一旦開始されると、上記プロセスが実行される。ルート0に関連するフィルタでDONEが到達する場合には、ルート1アドレスで開始して、そのプロセスは繰り返される。一致が見出される場合には、その結果は、その一致がルート0か、ルート1規則を用いて見出されたことを示す。

8. 一致を見出すか、或いはルート1サーチにおいてDONEを見出すことによって、サーチが終了するとき、結果が、フィルタリング検査の結果を示す結果キュー72に書き込まれる。その結果は、最後に検査された規則のアドレスを含み、一致が見出されたか否かを含んでいる。一致が見出された場合には、そのアドレスは、その結果に適した動作を開始する、動作テーブルに索引付けするためのFP58によって用いられる。例えば、その一致が、ある一定の値より小さいDAを有する全てのパケットが入ることになることを示す規則の場合には、その動作テーブルは、パケットが意図的に廃棄されるようになるルーチンを指している。

#### 【0050】

上記のように、CM84は、種々の異なる構成において用いることができる。3つのメモリCM086、CM188及びCM290はそれぞれ、異なる構成を実現するための異なるモードにおいて用いることができる。以下の真理値テーブルは判定基準メモリ86、88及び90の種々の構成モードを用いて実行することができる種々の比較を提供する。「1」は、所与の構成モードを用いて比較を実行できることを示しており、「0」は、比較を実行できないことを示している。

#### 【表4】

10

20

30

40

	SA および マスク	DA および マスク	PTCL	SP	DP	TOS	TOS および マスク	FLAG および マスク
CM 0-0	1	0	0	0	0	0	0	0
CM 0-1	0	1	0	0	0	0	0	0
CM 1-0	1	0	1	0	0	0	0	0
CM 1-1	0	1	1	0	0	0	0	0
CM 1-2	0	0	1	1	1	0	0	0
CM 1-3	0	0	1	1	1	0	0	0
CM 2-0	1	0	1	0	0	0	0	0
CM 2-1	0	1	1	0	0	0	0	0
CM 2-2	0	0	1	1	1	1	0	0
CM 2-3	0	0	0	1	1	1	1	0
CM 2-4	0	0	0	0	1	1	1	1
CM 2-5	0	0	0	1	0	1	1	1
CM 2-6	0	0	0	1	0	1	1	1
CM 2-7	0	0	0	1	0	1	1	1

10

20

30

## 【0051】

従って例えば、SA比較は、CM0-0、CM1-0及びCM2-0の任意のものを用いて実行することができる。FLAG比較は、CM2-4～CM2-7の任意のものを用いて実行することができる。種々の構成モードの任意のものを用いて所与の比較を実行するための能力は、CM84を編成する際に所望の自由度を提供し、それにより効率が高められる。判定基準メモリ空間の割当ては、以下に詳細に記載される。

## 【0052】

判定基準メモリ構成情報を符号化する種々の方法を用いることにより、別の実施形態において、記憶効率を高めることができる場合もある。例示される実施形態では、各判定基準メモリエントリのための構成メモリを格納するために、30ビットが用いられることに留意されたい。これらの30ビットは、規則メモリエントリ内の演算子コードの24ビットと、CM0エントリの1ビットと、CM1エントリの2ビットと、CM2エントリの3ビットとを含む。この方式により、CM0、CM1及びCM2内の復号化が簡単になる。しかしながら、判定基準メモリエントリのための被比較数及び演算子の全ての実現可能な構成の数は、約 $3.3 \times 10^6$ であり、従って22ビットのみを用いて表すことができる。従って例えば、演算子及び被比較数情報が復号化される各規則メモリエントリにおいて、1つの22ビット構成フィールドを用いることができる場合もある。しかしながら、そのような実施形態によって必要とされる復号化は、各構成ビットとCM84の各セクション

40

50

との間の一対一の対応関係の欠如により、例示される実施形態よりも、一般的に、より複雑になる。

【0053】

図16は、パケットフィルタリング情報が図1の交換器10において管理され、利用される方法を示す。一般に、パケットフィルタリング情報の情報源は、ネットワーク管理ステーション(NMS)であり、典型的には図1の交換器10から離れて配置される。NMSは、簡易ネットワーク管理プロトコル(SNMP)のような、ネットワーク管理プロトコルを用いて、図1の交換器制御部18内に存在する中央演算装置(CP)と通信を行う。CPは、NMSからフィルタリング情報を受信し、その情報を分配するために、各回線インタフェース12のIOP26に応答することができる。さらに、CPは不揮発(NV)記憶装置においてその情報を保持し、NMSが利用できないときに、交換器10が動作できるようにする。

10

【0054】

CPからIOP26に送出されるフィルタリング情報は、(1)それぞれ受信したパケットに適用することができる少数の判定基準を規定するフィルタと、(2)結合子、すなわち交換器10内の種々のポート及び/または回路と異なるフィルタのグループを関連付ける情報と、(3)フィルタとの関連を有し、フィルタ判定基準が満たされる場合に実行される動作とを含む。

【0055】

動作時に、IOP26が初期化される時、CPは、NV記憶装置からの現存するフィルタリングテーブルと結合データベースとを検索し、各回線インタフェース12のIOP26にそれをダウンロードする。NMSがフィルタ及び結合子を追加、削除或いは変更するとき、新しい情報をCPに渡すためにSNMP動作要求を発行する。次に、CPはその変更を各IOP26に通知する。

20

【0056】

IOP26は、CPからフィルタリング情報を検索し、フィルタ、結合子及び動作のローカルコピーを作成してメモリに格納する。IOP26は、CPが新しい情報を送出する場合には必ず、これらのローカルコピーを更新する。IOP26は、種々のフィルタの場合に行うことができる種々の動作のテーブルを用いて、図22の各転送エンジン22内のFPメモリ60をプログラミングする。またIOP26は、フィルタ及び結合子に対応するRMエントリ及びCMエントリを作成し、これらのエントリを用いて、PCE76(図5)のRM82及びCM84(図9)をプログラミングする。IOP26がCPから新しいフィルタリング情報を受信する場合には必ず、必要に応じてRM及びCMエントリが追加、削除或いは変更される。

30

【0057】

FP58は、PCE76の支援によりパケットを処理するための役割を果たす。IOP26によって供給される情報を用いるとき、FP58は、ルート0及びルート1アドレスに対する、それぞれ受信したパケットのポート及び回路識別のマップを作成し、これらのアドレスを用いてPCE要求を作成し、その要求を要求キュー70(図5)を介してPCE76に書き込む。上記のように、一般にFP58は、可能ならば多数の要求からなるバーストを書き込むことによって、バッチ形式でPCE76を動作させようと試みる。PCE76は上記のようにその要求を処理する。FP58は結果を取得するためにPCE76をポーリングし、その結果は上記のようにブロックでPCE76によって返送される。各結果において一致が示される場合、その結果からのPCE一致アドレスは、そのパケットがどの動作を行うかを確認するためにIOP26によって確立される動作テーブルへの索引として用いられる。その後FP58は、指示される動作を実行する。

40

【0058】

上記のように、RM82及びCM84では、性能を高めるために1つのIC上に実装されるメモリは比較的小さい。これらのメモリ内の限られた空間が効率的に管理されることは重要である。IOP26は、フィルタ判定基準のためのCM84内の空間の割当て、及び

50

規則セットのための R M 8 2 内の空間の割当てを行う役割を果たす。これらの動作は順次以下に記載される。

【 0 0 5 9 】

フィルタは、2つのタイプのうちの1つ、すなわち独立型か、複合型のいずれかとすることができる。独立型フィルタは、1つの規則のみを用いて実現することができる。複合型フィルタは多数の規則を必要とする。種々のタイプの複合型フィルタを構成することができるが、例示される実施形態で用いられる複合型フィルタは、範囲フィルタのみである。範囲フィルタは、ある範囲の上限を検査するために1つの規則を、その範囲の下限を検査するための別の規則を必要とする。従って、フィルタを加える際の第1のステップは、そのフィルタが独立型フィルタ、範囲フィルタのいずれであるかを決定することである。そのフィルタが独立型フィルタである場合には、1つの判定基準メモリ構成のみが必要とされるのに対して、範囲フィルタの場合には2つの構成が必要とされる。その後、C M 8 4の内容がサーチ、かつ/または評価され、C M 8 4においてそのフィルタを最もよく表現する方法を決定する。一旦構成が選択されれば、そのフィルタ情報は、C M 8 4への更新情報として追加される。これらのプロセスは、以下に詳細に記載される。

10

【 0 0 6 0 】

所与のフィルタを実現するために用いることができる判定基準メモリエントリの種々のタイプの構成がある。これらは、判定基準メモリ 8 6、8 8 及び 9 0 のうちのいずれが用いられるかによって、7つの C M 構成に編成される、以下の表は、C M 構成に従って編成される、より一般的に用いられる構成タイプのうちのいくつかを示す。

20

【表 5】

タイプ	CM 構成	CM0 モード	CM1 モード	CM2 モード
6	(CM0, CM1, CM2)	0:SA	1:DA	2:PTCL, SP, DP, TOS
5_1	(CM1, CM2)		0:SA	2:PTCL, SP, DP, TOS
5_2			1:DA	2:PTCL, SP, DP, TOS
5_3			0:SA	1:DA, PTCL
4_1	(CM0, CM2)	0:SA		2:PTCL, SP, DP, TOS
4_2		1:DA		2:PTCL, SP, DP, TOS
4_3		0:SA		1:DA, PTCL
3_1	(CM0, CM1)	0:SA	2:PTCL, SP, DP	
3_2		1:DA	2:PTCL, SP, DP	
3_3		0:SA	1:DA, PCTL	
2_1	(CM2)			0:SA, PTCL
2_2				1:DA, PTCL
2_3				2:PCTL, SP, DP, TOS
1_1	(CM1)		0:SA, PTCL	
1_2			1:DA, PTCL	
1_3			2:PTCL, SP, DP	
0_1	(CM0)	0:SA		
0_2		1:DA		

30

40

【 0 0 6 1 】

C M 構成は、リソース消費の観点から、最も高価なものから最も安価なものまでランク付けされる。多くのフィルタの場合、種々の構成のうちの任意の構成を用いることができるが、その目標は、最も安価、すなわち「最低限の」構成を用いて、メモリの使用効率を最大限にすることである。例えば、( S A、S A M a s k ) 比較のみを必要とするフィルタは、任意の C M 構成を用いて実装することができ、その最低限の構成は ( C M 0 ) である。別の例としては、( S A、S P 及び P T C L ) を必要とするフィルタは、4つの構成 ( C M 0、C M 1 )、( C M 0、C M 2 )、( C M 1、C M 2 )、( C M 0、C M 1、C

50

M 2 ) のいずれかを用いて実装することができ、その最低限の構成は ( C M 0 、 C M 1 ) である。

【 0 0 6 2 】

最低限の構成は、その最低限利用可能な構成をサーチする際の開始点として用いられる。最低限の構成が利用可能な場合には、それが用いられる。そうでない場合には、利用可能な構成が見つかるまで、順次、より高価な構成が考慮される。上記の例では、その構成は以下の順序、すなわち ( C M 0 、 C M 1 ) 、 ( C M 0 、 C M 2 ) 、 ( C M 1 、 C M 2 ) 及び ( C M 0 、 C M 1 、 C M 2 ) の順にサーチされる。構成サーチは、一群のリンクされた空き判定基準メモリ位置のリストを利用し、各リストは、1つの行、例えば ( C M 0 ) 、 ( C M 0 、 C M 1 ) 等において特定の空き列の組を表す。フィルタによって必要とされる比較のタイプに基づいて、I O P 2 6 は、利用可能なエントリで最初のものを検索する所定のシーケンスで、全ての適切なリストをサーチする。

10

【 0 0 6 3 】

選択される判定基準メモリ構成が、所与のフィルタのために最低限必要とされる構成より大きい場合には、残りの部分は、他のフィルタによって利用可能である。従って上記の例では、( C M 0 、 C M 1 ) のみが必要とされるときに、構成 ( C M 0 、 C M 1 、 C M 2 ) が用いられる場合には、( C M 2 ) の1ユニットは、他のフィルタによって利用可能である。

【 0 0 6 4 】

一旦ある構成が選択されたなら、フィルタの種々の要素が、C M 8 4 の種々のセクションに割り当てられる。( S A 、 S P 及び P T C L ) を必要とするフィルタの例で説明を続けると、構成 ( C M 0 、 C M 1 ) が選択されるものと仮定すれば、S A 及び S A M a s k は C M 0 に割り当てられ、S P 及び P T C L は C M 1 に適切に割り当てられる。この割当ては、上記テーブルからの構成タイプ 3 \_ 1 に対応する。C M 0 8 6 はモード 0 において構成され、C M 1 8 8 はモード 2 において構成される。一旦割当てが終了すれば、C M 8 4 のためのエントリ ( 或いは範囲フィルタのためのエントリ ) が生成される ( 図 1 1 ~ 図 1 3 及びそれに付随する説明を参照されたい ) 。また、対応する規則メモリエントリに対する判定基準演算子 ( ビット 3 5 ~ 1 2 ) のためのデータも生成される ( 図 1 0 及びそれに付随する説明を参照されたい ) 。規則メモリエントリの残りは、以下に説明するフィルタ結合中に生成される。

20

30

【 0 0 6 5 】

フィルタが削除されるとき、削除されるフィルタに用いられる C M 構成は再利用できるようになる。より大きな構成にすることができる場合には、利用可能な構成が連結される。これは、上記のように全体として、或いは部分的に後に用いることができる。例えば、1ユニットの ( C M 0 ) が、フィルタの削除によって空きになり、同じ行の C M 8 4 において1ユニットの ( C M 1 ) が利用可能である場合には、1ユニットの ( C M 0 、 C M 1 ) が形成される。

【 0 0 6 6 】

上記のように、I O P 2 6 は、フィルタの「結合」を表す規則の組、すなわち論理ポート或いは回路と1つ或いは複数のフィルタの組との間の関係を保持するための役割も果たす。このプロセスは R M 8 2 の割当て及びプログラミングを含む。結合子が追加されるとき、追加される結合子のサイズが最初に評価される。そのサイズは、論理ポート或いは回路、及び範囲フィルタ及び非範囲フィルタの混合体で用いられるフィルタの数により決定される。一旦結合子のサイズがわかれば、R M 8 2 内の空間が割り当てられる。例示される実施形態では、規則メモリ空間は、そのサイズが 2 の累乗であるセグメントに割り当てられる。結合子に割り当てられるまで、空きセグメントは各空きリスト上に保持され、削除された結合子からのセグメントは、再利用するために空きリストに戻される。個々にリンクされたりリストを用いて、そのセグメントは互いに連結される。例えば、各空きリストは、同じサイズからなる割り当てられていないセグメントの個々にリンクされたりリストである。各結合子は、一般に異なるサイズのセグメントからなる個々にリンクされたりリストで

40

50

ある。

【 0 0 6 7 】

2 1 個の規則メモリエントリを必要とする結合子を考慮する。この結合子の場合、サイズ 1 6、4、2 及び 2 のセグメントが割り当てられることが好ましい。1 6 エントリのセグメントは、1 5 の規則と、4 エントリセグメントへのポインタとを格納する。4 エントリセグメントは、3 つの規則と、2 エントリセグメントの 1 つへのポインタとを格納し、2 エントリセグメントは、1 つの規則と、他の 2 エントリセグメントへのポインタとを格納する。最後のセグメントは 2 つの規則を格納する。割当て中に、所望のサイズのセグメントが利用できない場合には、より大きなセグメントが利用される。セグメント内の未使用の空間は単に未使用のまま残されることができ、或いは別法では、判定基準メモリ構成の割当ての場合に上で説明したのと同様にして、他の結合子に割り当てるために利用することができる。

10

【 0 0 6 8 】

一旦メモリ割当てが終了すれば、R M 8 2 のためのエントリが作成され、R M 8 2 に書き込まれる。このプロセスの間に、各演算子タイプの規則は、その規則で用いられることになる対応する判定基準メモリエントリのアドレスでプログラミングされ、その演算子は、規則 / 判定基準の対によって表されるフィルタに基づく適当な値でプログラミングされる。その規則は、フィルタが検査されることになる所望のシーケンスに従って、R M 8 2 において論理シーケンスに配列される。R M 8 2 のセグメント内では、その規則は逐次的に配列される。多数のセグメントに及ぶ結合子の場合、そのセグメントは、規則が所望のシーケンスで評価されるように、互いに連結される。

20

【 0 0 6 9 】

パケット分類に関係する種々の装置及び方法が記載されてきた。本発明は主にインターネットプロトコル ( I P ) パケット或いはメッセージを参照して記載されてきたが、記載された技術は、他のタイプのメッセージの場合でも用いられることは明らかであろう。また、ここで開示された本発明の概念から逸脱することなく、上記の技術に対する他の変更形態及びその変形形態が実現可能であることは当業者には明らかであろう。従って本発明は、添付の請求の範囲及びその精神によってのみ制限されるものと見なされるべきである。

【 0 0 7 0 】

【 発明の効果 】

上記のように本発明によれば、パケットフィルタリングを行うために高速のパケット分類を提供する、ネットワーク装置内のパケット処理ロジックを実現することができる。

30

【 図面の簡単な説明 】

【 図 1 】 本発明によるパケット分類エンジンを組み込んだネットワーク交換器のブロック図である。

【 図 2 】 図 1 のネットワーク交換器における回線インタフェースのブロック図である。

【 図 3 】 図 2 の回線インタフェース上のパケット転送エンジンのブロック図である。

【 図 4 】 図 3 の転送エンジン内のパケットヘッダ分配器の特定用途向け集積回路 ( A S I C ) のブロック図である。

【 図 5 】 図 4 のパケットヘッダ分配器 A S I C 内のルーティング及び分類エンジンのブロック図である。

40

【 図 6 】 図 5 のルーティング及び分類エンジンに渡されるルーティング及び分類要求の構造図である。

【 図 7 】 図 5 のルーティング及び分類エンジンによって供給されるルーティング及び分類結果の構造図である。

【 図 8 】 図 5 のルーティング及び分類エンジンによって供給される状態指示の構造図である。

【 図 9 】 図 5 のルーティング及び分類エンジン内のパケット分類エンジン ( P C E ) のブロック図である。

【 図 1 0 】 図 9 のパケット分類エンジン内の規則メモリエントリの構造図である。

50

【図 1 1】図 9 のパケット分類エンジン内の第 1 の判定基準メモリエントリの構造図である。

【図 1 2】図 9 のパケット分類エンジン内の第 2 の判定基準メモリエントリの構造図である。

【図 1 3】図 9 のパケット分類エンジン内の第 3 の判定基準メモリエントリの構造図である。

【図 1 4】図 9 のパケット分類エンジン内の判定基準メモリのバンクに対する比較論理ブロックのブロック図である。

【図 1 5】図 1 4 の比較論理ブロックにおいて用いられる比較器論理ブロックのブロック図である。

10

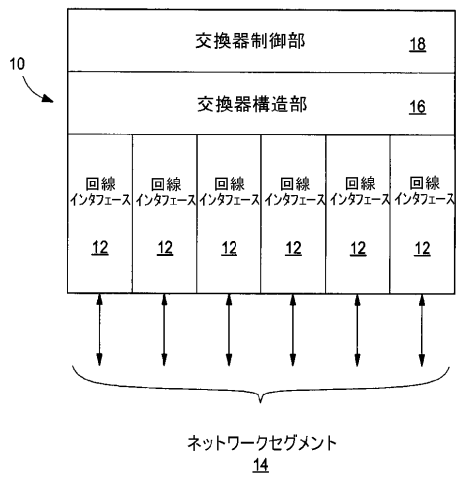
【図 1 6】図 1 の交換器内の種々の処理構成要素によって、パケットフィルタリング情報が如何に作成され、分配され、用いられるかを示す図である。

【符号の説明】

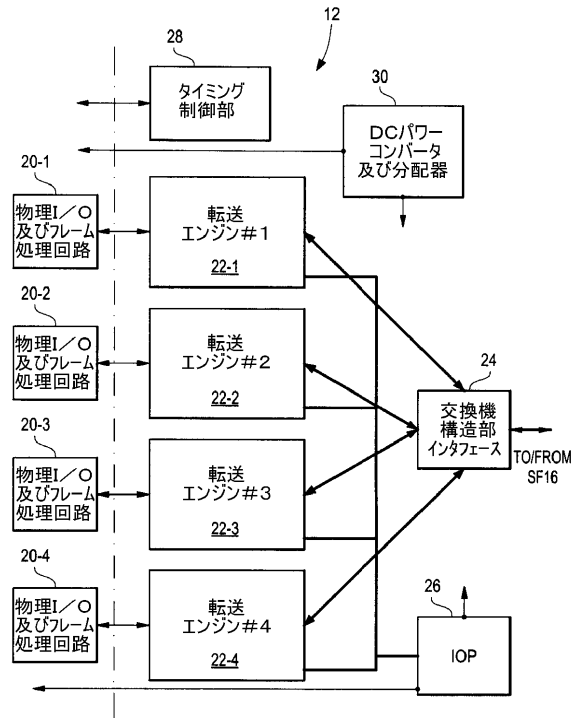
1 0	ネットワーク交換器	
1 2	回線インタフェース	
1 4	ネットワークセグメント	
1 6	交換器構造部	
1 8	交換器制御部	
2 0	物理 I / O 及びフレーム処理回路	
2 2	転送エンジン	20
2 4	交換器構造部インタフェース	
2 6	I O P	
2 8	タイミング制御部	
3 0	D C パワーコンバータ及び分配器	
4 0	着信 S A R	
4 2	発信 S A R	
4 4、4 6	制御メモリ	
4 8、5 0	パケットメモリ	
5 2	パケットヘッダ分配装置 ( P H D )	
5 4	P C I バス	30
5 6	F P バス	
5 8	転送プロセッサ ( F P )	
6 0	F P メモリ	
6 4	R x キュー	
6 6	T x キュー	
6 8	ルーティング及び分類エンジン	
6 9	信号線	
7 0	要求キュー	
7 2	結果キュー	
7 4	ルート探索エンジン ( R C E )	40
7 6	パケット分類エンジン ( P C E )	
7 8	I N F I F O	
8 0	O U T F I F O	
8 2	規則メモリ ( R M )	
8 4、8 6、8 8、9 0	判定基準メモリ ( C M )	
9 2、9 4、9 6	比較論理ブロック	
9 8	アドレス指定及び制御ロジック	
1 0 0	バストランシーバ	
1 0 2	A N D ゲート	
1 0 4	比較論理ブロック	50

- 106 セレクタ
- 108 構成セレクタ
- 110 EQUAL比較器
- 112 LESS THAN比較器
- 114 GREATER THAN比較器
- 116 インバータ
- 118 ANDゲート
- 120 エキスパンダロジック

【図1】

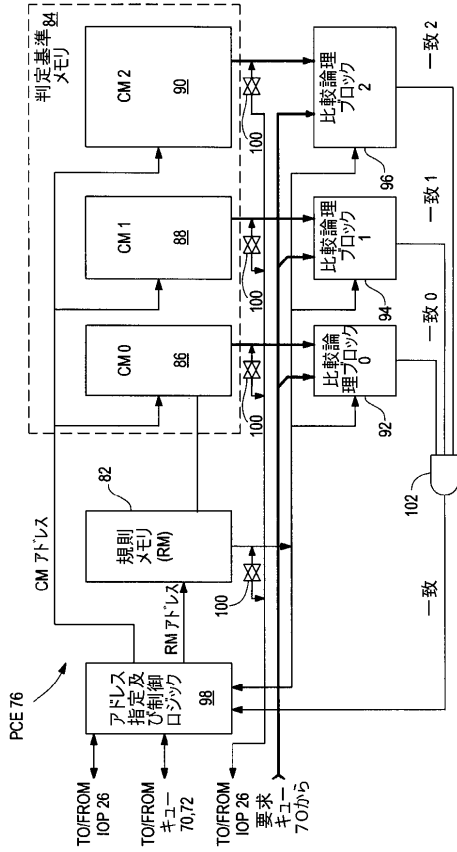


【図2】

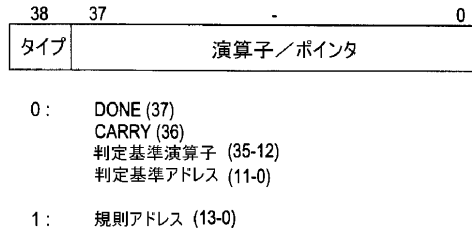




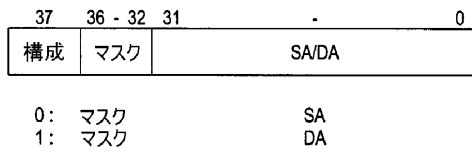
【図 9】



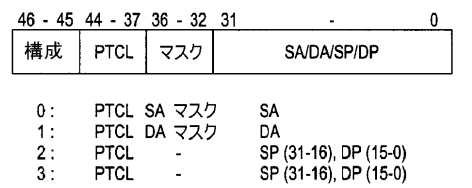
【図 10】



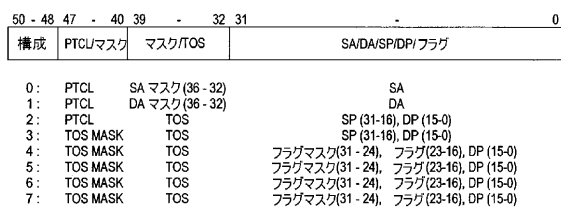
【図 11】



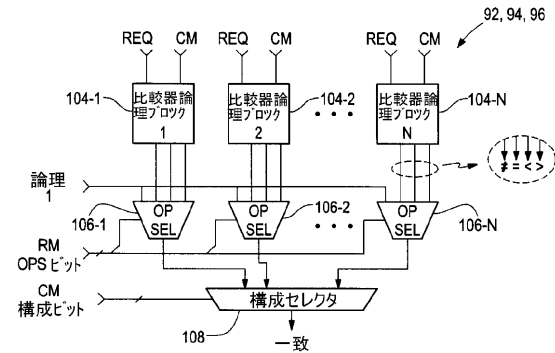
【図 12】



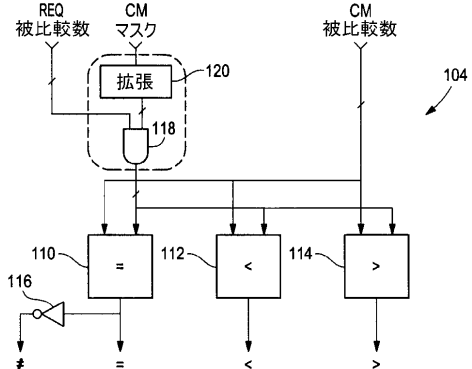
【図 13】



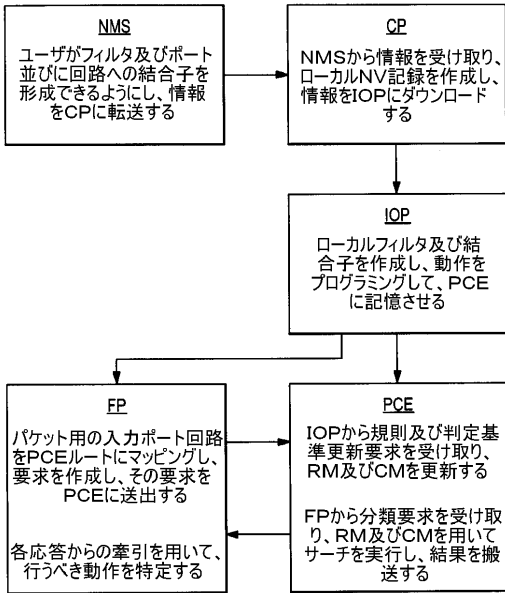
【図 14】



【図 15】



【 図 1 6 】



## フロントページの続き

- (74)代理人 100091889  
弁理士 藤野 育男
- (74)代理人 100101498  
弁理士 越智 隆夫
- (74)代理人 100096688  
弁理士 本宮 照久
- (74)代理人 100102808  
弁理士 高梨 憲通
- (74)代理人 100104352  
弁理士 朝日 伸光
- (74)代理人 100107401  
弁理士 高橋 誠一郎
- (74)代理人 100106183  
弁理士 吉澤 弘司
- (72)発明者 アンドリュー テー . ヘブ  
アメリカ合衆国 0 1 7 4 9 マサチューセッツ , ハドソン , レイクサイド アヴェニュー 6 2
- (72)発明者 サンジェイ ジー . チェリアン  
アメリカ合衆国 0 3 0 3 3 ニューハンプシャー , ブルックリン , マックスウェル ドライヴ  
6

審査官 石井 研一

- (56)参考文献 特開平 1 1 - 3 0 8 2 7 2 ( J P , A )  
特開平 1 0 - 2 4 3 3 9 1 ( J P , A )

- (58)調査した分野(Int.Cl. , DB名)  
H04L 12/56  
H04L 12/28  
H04M 3/00