

(19) 日本国特許庁(JP)

## (12) 特許公報(B2)

(11) 特許番号

特許第4666866号  
(P4666866)

(45) 発行日 平成23年4月6日(2011.4.6)

(24) 登録日 平成23年1月21日(2011.1.21)

(51) Int.Cl.

H04L 12/56 (2006.01)

F 1

H04L 12/56 100Z

請求項の数 17 (全 18 頁)

(21) 出願番号 特願2001-548968 (P2001-548968)  
 (86) (22) 出願日 平成12年12月22日 (2000.12.22)  
 (65) 公表番号 特表2003-518873 (P2003-518873A)  
 (43) 公表日 平成15年6月10日 (2003.6.10)  
 (86) 國際出願番号 PCT/CA2000/001515  
 (87) 國際公開番号 WO2001/048975  
 (87) 國際公開日 平成13年7月5日 (2001.7.5)  
 審査請求日 平成19年12月13日 (2007.12.13)  
 (31) 優先権主張番号 09/473,074  
 (32) 優先日 平成11年12月28日 (1999.12.28)  
 (33) 優先権主張国 米国(US)

(73) 特許権者 508034325  
 モサイド・テクノロジーズ・インコーポレーテッド  
 カナダ・オンタリオ・K2K・2X1・オタワ・ハインズ・ロード・11・スイート  
 ・203  
 (74) 代理人 100064746  
 弁理士 深見 久郎  
 (74) 代理人 100085132  
 弁理士 森田 俊雄  
 (74) 代理人 100083703  
 弁理士 仲村 義平  
 (74) 代理人 100091409  
 弁理士 伊藤 英彦

最終頁に続く

(54) 【発明の名称】パケットスイッチにおける転送オーバライド

## (57) 【特許請求の範囲】

## 【請求項1】

スイッチであって、  
 少なくとも1つの転送用エントリを含むメモリを含み、該転送用エントリは転送オーバライドを含み、さらに、  
 連続的なレイヤに対して該メモリを連続的にサーチして、受信されたデータパケットの複数の部分に合致する複数の転送用エントリを捜すための手段と、  
該転送用エントリに含まれるネクストフィールドの状態に応じて、該メモリの別のサーチを行なうかどうかを判断するための手段と、

合致する1つの転送用エントリから該転送オーバライドを抽出するための手段とを含む、スイッチ。

## 【請求項2】

該メモリは連想記憶メモリである、請求項1に記載のスイッチ。

## 【請求項3】

最終のサーチに対応する該抽出された転送オーバライドに応じて、該受信データパケットを転送するための手段をさらに含む、請求項1または2に記載のスイッチ。

## 【請求項4】

該転送用エントリは優先度フィールドをさらに含み、該優先度フィールドは該受信データパケットに割当てられた優先度を含む、請求項1～3のいずれか1項に記載のスイッチ。

**【請求項 5】**

該転送用エントリはホストトラフィックラベルフィールドをさらに含み、該ホストトラフィックラベルフィールドは該受信データパケットのための分類データを含む、請求項1～4のいずれか1項に記載のスイッチ。

**【請求項 6】**

該受信データパケットの該部分はプロトコルを特定する、請求項1～5のいずれか1項に記載のスイッチ。

**【請求項 7】**

該転送オーバライドは、受信されたデータパケットがホストに転送されるべきかどうかを示す、請求項1～6のいずれか1項に記載のスイッチ。 10

**【請求項 8】**

先のサーチにおいて抽出された該転送オーバライドが、現時点におけるサーチで抽出された該転送オーバライドで上書きされる、請求項1～7のいずれか1項に記載のスイッチ。

**【請求項 9】**

スイッチにおいて受信されたデータパケットのための転送オーバライドを選択する方法であって、

メモリを提供するステップを含み、該メモリは少なくとも1つの転送用エントリを含み、該転送用エントリは転送オーバライドを含み、さらに、

連続的なレイヤに対して該メモリを連続的にサーチして、受信されたデータパケットの複数の部分に合致する複数の転送用エントリを捜すステップと、 20

現時点における転送用エントリにおけるネクストフィールドの状態に応じて、該メモリの別のサーチを行なうかどうかを判断するステップと、

合致する1つの転送用エントリから該転送オーバライドを抽出するステップとを含む、方法。

**【請求項 10】**

該メモリは連想記憶メモリである、請求項9に記載の方法。

**【請求項 11】**

該連想記憶メモリはレジスタを含む、請求項10に記載の方法。

**【請求項 12】**

最終のサーチに対応する該抽出された転送オーバライドに応じて該受信データパケットを転送するステップをさらに含む、請求項9～11のいずれか1項に記載の方法。 30

**【請求項 13】**

該転送用エントリは優先度フィールドをさらに含み、該優先度フィールドは該受信データパケットに割当てられた優先度を含む、請求項9～12のいずれか1項に記載の方法。

**【請求項 14】**

該転送用エントリはホストトラフィックラベルフィールドをさらに含み、該ホストトラフィックラベルフィールドは該受信データパケットのための分類データを含む、請求項9～13のいずれか1項に記載の方法。

**【請求項 15】**

該受信データパケットの該部分はプロトコルを特定する、請求項9～14のいずれか1項に記載の方法。

**【請求項 16】**

該転送オーバライドは、受信されたデータパケットがホストに転送されるべきかどうかを示す、請求項9～15のいずれか1項に記載の方法。

**【請求項 17】**

先のサーチにおいて抽出された該転送オーバライドが、現時点におけるサーチで抽出された該転送オーバライドで上書きされる、請求項9～16のいずれか1項に記載の方法。

**【発明の詳細な説明】****【0001】**

**【発明の背景】**

コンピュータネットワークにおいて、ネットワーキングスイッチは、スイッチに接続された入力ポートにおいてデータパケットを受信し、そのデータパケットをスイッチに接続された出力ポートに転送する（forward）。該スイッチは、入力ポートで受取ったデータパケットのヘッダに含まれる宛先アドレスに応じて、そのデータパケットをどの出力ポートに転送するかを決定する。

**【0002】**

該スイッチは、典型的に、入力エンジンの転送用ロジック内に通常実現される転送用テーブルを含む。この転送用テーブルがサーチされ、データパケットに含まれる宛先アドレスに応じてそのデータパケットが転送されるべき 1 または複数の出力ポートが検索される。しかし、未知の宛先アドレスに対するデータパケットが受け取られた場合、該スイッチの転送用テーブルには合致するエントリが存在しない。したがって、入力ポートで受取られたそのようなデータパケットはスイッチ内のホストプロセッサに送られる。たとえば、転送用テーブル内に合致するエントリがないデータパケットは、ホストが転送用テーブルを更新することができるように転送される。

10

**【0003】**

どのデータパケットをホストプロセッサに転送するかを決定するためのある公知の技術は、宛先アドレスのみに基づいて転送を決定するものであり、たとえば、すべてのブロードキャストデータパケットを転送する。すべてのブロードキャストデータパケットがホストプロセッサに転送される場合、ホストプロセッサはホストが必要とする情報を含まないかも知れないデータパケットも処理することになる。また、ホストプロセッサは必要とするがホストには転送されない非ブロードキャストデータパケットもあり得る。したがって、すべてのデータパケットを転送することは必要にホストプロセッサの帯域幅を費やし、ホストが必要とし得るすべてのデータパケットを提供することにはならない。

20

**【0004】**

データパケットをホストプロセッサに転送するための別の公知の技術は、データパケットのヘッダに含まれるデータタイプ、たとえばデータリンクヘッダに記憶されたタイプに基づいて、転送すべきデータパケットを選択するものである。ホストに転送すべきパケットタイプはホスト転送リスト内に記憶される。データパケットは、そのヘッダに含まれるパケットタイプがホスト転送リストの構成要素であった場合にホストに転送される。ホスト転送リストはスイッチ内にハードコードされる。開放型システム相互接続（「OSI」）モデルにおいては、L2 パケットタイプがデータリンクレイヤヘッダに含まれる。

30

**【0005】**

たとえば、この L2 パケットタイプは、アドレスリゾリューションタイプ（Address Resolution Type 「ARP」）ブリッジプロトコルデータユニット（Bridge Protocol Data Unit）、またはインターネットプロトコル（Internet Protocol 「IP」）であり得る。この技術は、転送されるデータパケットを特定のタイプのものに限定することにより、ホストプロセッサに転送されるデータパケットの数を減じる。たとえば、データリンクヘッダに記憶された L2 データタイプを使用して、ブリッジプロトコルデータユニットデータパケットをホストプロセッサのみに転送することも可能であり、ARP データパケットをホストプロセッサと出力ポートの両方に転送することも可能であり、また、IP データパケットをホストと出力ポートの両方に転送することが可能である。

40

**【0006】**

この技術の欠点は、新しい L2 データタイプを有するネットワーキングプロトコルがホストに転送される場合に、新しいデータパケットタイプがホスト転送リストに追加されねばならず、ハードウェアの更新が必要となることである。この技術はまた、たとえば IP データパケット等の、ある特定の L2 データタイプのすべてのデータパケットがホストによる処理に必要ではないかも知れないため、ホストプロセッサの帯域幅を必要に費やしてしまうおそれがある。

**【0007】**

50

### 【発明の概要】

ネットワークスイッチにおけるメモリは、転送オーバライドを有する少なくとも1つの転送用エントリを含む。該スイッチ内の転送オーバライドロジックは、該メモリを連続的にサーチして、受信したデータパケットの複数の部分に合致する複数の転送用エントリを探し出す。該転送オーバライドロジックは、それら合致したエントリから該転送オーバライドを抽出する。該データパケットの合致する部分はヘッダフィールドであり得、複数レイヤでプロトコルを特定し得る。

### 【0008】

該転送用エントリはまた、ネクストフィールドを含む。該転送オーバライドロジックは、該ネクストフィールドの状態に応じて該メモリのサーチを別に行なうかどうかを判断する。該メモリは連想記憶メモリであり得る。

10

### 【0009】

サーチの完了後、該転送オーバライドロジックは転送用エントリから受信したデータパケットのための転送オーバライドを抽出する。該スイッチ内の転送選択ロジックは、該抽出された転送オーバライドに応じて受信データパケットを転送する。

### 【0010】

該転送用エントリは、受信データパケットの内容に応じた優先度フィールドおよび、トラフィックラベルフィールドを含み得る。該トラフィックラベルフィールドおよび優先度フィールドは、ホストによって処理されるようにホストに転送され得る。

### 【0011】

本発明の上述および他の目的、特徴ならびに利点は、添付の図面に示される本発明の好ましい実施例の以下のより特定的な説明から明らかとなるであろう。複数の図面を通じて同一の参照符号は同一の部分を示す。図面は必ずしも一定の比率で描かれておらず、本発明の原理を示すのに強調されているところもある。

20

### 【0012】

#### 【詳細な説明】

図1は、ネットワークスイッチ100を示す。これは、スイッチ100に接続された入力ポート102で受取られたデータパケットを転送するためのホスト転送用ロジック128を含む。入力ポート0-n 102のいずれか1つで受取られたデータパケットは、入力ポートエンジン104で処理される。この入力ポートエンジン104内のホスト転送用ロジック128は、受信データパケットのヘッダにエンコードされたネットワーキングプロトコルから、その受信データパケットをどこに転送するかを判断する。この受信データパケットは、入力データ経路116上をセグメントバッファメモリ108へ、および/または、ホストデータ経路114上をホストインターフェイスロジック106へと転送され得る。

30

### 【0013】

ホスト転送用ロジック128は、データパケットをセグメントバッファメモリ108とホストインターフェイス106の両方に転送するか、または、セグメントバッファメモリ108もしくはホストインターフェイス106のいずれかに転送することが可能であるが、ホスト転送用ロジック128はデータパケットを転送せずに、廃棄することもある。データパケットのヘッダにエンコードされたネットワーキングプロトコルに応じてデータパケットに転送用決定(forwarding decision)を割当てることにより、ホストによって必要とされるデータパケットのみがホストに転送される。また、新しいネットワーキングプロトコルが導入された場合には、ホストプロセッサにより、ホストインターフェイスロジック106およびホストデータ経路114を通じて、新しい転送用決定をホスト転送用ロジック128に追加することができる。

40

### 【0014】

入力データ経路116上を転送されるデータパケットは、セグメントバッファメモリ108内に記憶され、出力データ経路118上を出力ポートエンジン110へと転送される。出力ポートエンジン110はそのデータパケットを、そのデータパケットに含まれるヘッ

50

ダの内容に応じて、1または複数の出力ポート112に転送する。セグメントバッファメモリ108におけるデータパケットの記憶に関しては、デイビッドA.ブラウン(David A. Brown)による「インタリーブされたノンブロッキングパケットバッファのための方法および装置("Method and apparatus for an Interleaved Non-Blocking Packet Buffer")」と題された1999年8月31日出願の同時係属中の米国特許出願連続番号第09/386,589号に記載されており、その教示全体がここに引用により援用される。

#### 【0015】

ホストデータ経路114上をホストインターフェイスロジック106へと転送されたデータパケットは、ホストインターフェイスロジック106によりホストポートデータ経路120上をホストまたはマイクロプロセッサ(図示せず)へと転送される。ホストポートインターフェイスロジック106は、周辺制御インターフェイス(Peripheral Control Interface 「PCI」)ホストポートインターフェイスである。ホストポートインターフェイスロジック106はPCIに限定されることはなく、当業者によく知られている他のホストインターフェイスロジックであってもよい。ホストインターフェイスロジック106は、ホストインターフェイスロジック106において実現されるホストインターフェイスプロトコルに応じて、ホストデータ経路114をホストポートデータ経路120に変換する。

#### 【0016】

図2は、図1に示されるホスト転送用ロジック128を示す。ホスト転送用ロジック128は、プロトコル連想記憶メモリ(Content Addressable Memory 「CAM」)、転送オーバライドロジック200および転送選択ロジック204を含む。入力ポートエンジン104(図1)において入力ポート102(図1)でデータパケットが受取られると、そのデータパケットのヘッダ102aがホスト転送用ロジック128内の転送オーバライドロジック200に転送される。

#### 【0017】

転送オーバライドロジック200は、受信データパケットのヘッダ102aの内容に応じてプロトコルCAM202を連続的にサーチして、転送用エントリ300(図3)内に記憶されている転送オーバライド206を探し出す。転送用エントリ300(図3)については、図3に関連して説明する。プロトコルCAM200の複数回のサーチは、データパケットのヘッダ102a内の異なるフィールドの内容を使用して行なわれ得る。データパケットのヘッダ102aは、少なくとも1つのレイヤヘッダを含み、レイヤヘッダは少なくとも1つのフィールドを含む。転送用エントリ300(図3)のサーチは、最下位のレイヤヘッダから開始され、以後、次に高いレイヤヘッダが順にサーチされる。たとえば、転送用エントリの最初のサーチは、データリンクレイヤ(L2)ヘッダのタイプフィールドの内容を用いて行なわれ得る。転送用エントリ300の2回目のサーチは、L2ヘッダのサーチ結果に応じて、ネットワークリンクレイヤヘッダ(L3)内のタイプフィールドの内容を用いて行なわれ得る。転送用エントリ300の3回目のサーチは、L3のサーチ結果に応じて、トランスポートレイヤヘッダ(L4)のソースアドレスフィールドの内容を用いて行なわれ得る。別のレイヤのサーチを続けるかどうかは、各転送用エントリにおけるネクストフィールドの値によって判断される。

#### 【0018】

各サーチの開始時、転送オーバライドロジック200は、データパケットの複数のレイヤヘッダのうち、1つのレイヤヘッダの内容の中から選択することによってサーチキー208を生成し、そのサーチキー208をプロトコル連想記憶メモリ('CAM')202へと転送する。プロトコルCAM202内にこのサーチキー208に合致する転送用エントリ300(図3)があれば、その転送用エントリ300(図3)内に記憶されている関連データ(associated data)210が転送オーバライドロジック200へと転送され、そこにその転送用エントリの現時点の(current)コピーが記憶される。関連データ210は転送オーバライド206を含むが、これは、この実施例においては表1に示すように2ビットの二進コードとして実現される。転送オーバライド206は、そのデータパケット

10

20

30

40

50

をどこに転送すべきかを示す。最後のサーチが行なわれた後、最後のサーチで選択された転送用エントリ 300 (図 3) 内の転送オーバライド 206 が転送選択ロジック 204 へと転送される。データパケットは、表 1 に示される転送オーバライド 206 の状態に応じて、入力データ経路 116 および / またはホストデータ経路 114 上を転送される。

【0019】

【表 1】

転送オーバライド	転送用決定
00	入力データのみ
01	入力データおよびホストデータ
10	ホストデータのみ
11	転送せず (廃棄)

表 1

プロトコル CAM 202 は小さなテーブルである。これは論理ゲートを用いてレジスタ内に実現することが可能であり、ホストプロセッサはそれらレジスタに対する読出および書き込みアクセス権を有する。転送選択ロジック 204 は、入力ポート 102 で受信されたデータパケットを、転送オーバライドロジック 200 から転送された転送オーバライド 206 の状態に応じて、入力データ 116 および / またはホストデータ 114 へと転送する。転送選択ロジック 204 はホストバッファ (図示せず) および入力バッファ (図示せず) を含む。ホストバッファおよび入力バッファへの入力は、受信されたバッファである。ホストバッファの出力はホストデータ 114 であり、入力バッファの出力は入力データ 116 である。転送オーバライド 206 はバッファイネーブルロジック (図示せず) においてエンコードされ、ホストバッファイネーブル信号および入力バッファイネーブル信号が与えられる。したがって、この転送オーバライド 206 により、受信データパケットがホストバッファを介してホストデータ 114 へ、および / または、入力バッファを介して入力データ 116 へと転送されるべきかどうかが決定される。

【0020】

図 3 は、プロトコル CAM 202 (図 2) 内の転送用エントリ 300 を示す。転送用エントリ 300 は、サーチキー 208 および関連データ 210 を含む。関連データ 210 は、ネクストフィールド 302 と、選択的なホストトラフィックラベル 304 と、選択的なサービスの質 (Quality of Service 「QoS」) または E タイプ識別子 (Etype Identifier 「EID」) フィールド 306 と、転送オーバライドコード 206 とを含む。転送用エントリ 300 のコピーが転送オーバライドロジック 200 (図 2) 内に記憶され、転送用エントリ 300 のコピー内に記憶されている転送オーバライドコード 206 は、以後のプロトコル CAM 202 (図 2) のサーチにおいて選択された別の転送用エントリに記憶されている転送オーバライドコード 206 でオーバライドされる。

【0021】

QoS フィールド 306 は、受信データパケットの優先度を提供する。ホストトラフィックラベル 304 は、受信データパケットをどのように分類するかに関する情報を提供するタグである。ホストトラフィックラベル 306 および QoS または EID フィールド 306 は、ホストによって処理されるように、ホストインターフェイスロジック 106 を通じてホストプロセッサに転送することが可能である。転送用エントリ 300 に含まれるサーチキー 208 および関連データ 210 の内容は、転送オーバライドロジック 200 (図 2) によって処理されるヘッダの内容に依存する。新しいサーチキー 208 が、各サーチにつき、異なるレイヤヘッダの内容を選択することによって生成される。転送用エントリ 300 については、図 4A から図 4E に関連して説明する。

10

20

30

40

50

**【0022】**

転送オーバライドロジック200(図2)は、ネクストフィールド302の状態から、データパケットに含まれる次に高いレイヤヘッダの内容を用いて転送用エントリ300の別のサーチを行なうかどうかを判断する。現時点におけるサーチからの転送オーバライドコード206が、転送オーバライドロジック200(図2)内の転送用エントリ(図示せず)のコピーに記憶されている先のサーチからの転送オーバライドコード206に優先する override)。したがって、最後のサーチにおいて選択された転送オーバライドコード206が、そのデータパケットのための転送用決定となる。

**【0023】**

たとえば、L2ヘッダに記憶されているタイプがARPである場合、すべてのARPデータパケットが入力データ経路116上およびホストデータ経路114上を転送されるので、さらなるサーチは不要である。したがって、転送用エントリ300内のネクストフィールド302は「0」に設定される。しかし、L3レイヤヘッダに記憶されているタイプがIPである場合には、IPデータパケットのタイプを判断するのに、L3レイヤヘッダにおいてさらなるサーチが必要である。L3レイヤヘッダのタイプフィールドに記憶されているIPタイプがインターネットグループ管理プロトコル(Internet Group Management Protocol 「IGMP」)である場合、ネクストフィールド302はさらなるサーチが不要であることを示す「0」に設定され、そのデータパケットは入力データ経路116(図1)上およびホストデータ経路114(図1)上を転送される。

**【0024】**

IPタイプデータパケットのデフォルトの転送用決定は、入力データ経路116(図1)上の転送に設定され得る。この転送用決定は、より高いレベルのレイヤヘッダの内容を用いた以後のサーチにおいて転送オーバライドコード206(図2)が選択されると、データパケットをホストデータ経路114(図1)上を転送するようにオーバライドされる。CAM内に転送用エントリ300を記憶することにより、サーチは迅速に行なわれ、転送用決定もまた、データパケットをスイッチ100(図1)を通じてワイヤ速度(wire-speed)で転送できるように迅速に決定される。

**【0025】**

図4Aは、空または無効のエントリ300aを示す。空のサーチキー208aの3つの最上位ビット(Most Significant Bits 「MSBs」)は「000」に設定され、無効な関連データ210aが記憶される。

**【0026】**

図4Bは、Eタイプ識別子('EID')エントリ300bを示す。EIDサーチキー208bのMSB410は「1」に設定される。EIDサーチキー208bはL2ペイロードタイプ414およびカスタムOUIインジケータ412を含む。EID関連データ210bは、ネクストフィールド402b、ホストトラフィックラベル(Host Traffic Label 「HTL」)404b、EID406および転送オーバライド(Forward Override 「FO」)408bを含む。

**【0027】**

図4Cは、IPv4タイプエントリ300cを示す。IPタイプサーチキー208cの3つのMSBs426は「001」に設定される。IPタイプサーチキー208cはIPv4タイプ428を含む。IPタイプ関連データ210cは、ネクストフィールド402c、HTL404c、LNQoS有効ビット420、レイヤ3QoS422、およびFO408cを含む。

**【0028】**

図4Dは、IPバージョンエントリ300dを示す。IPバージョンサーチキー208dの3つのMSBs416は「001」に設定される。IPバージョンサーチキー208dはIPバージョン418を含む。IPバージョン関連データ210dは、ネクストフィールド402d、HTL404d、LNQoS有効ビット420、レイヤ3QoS422、およびFO408dを含む。

10

20

30

40

50

**【0029】**

図4Eは、L4ポートエントリ300eを示す。L4サーチキー208eの2つのMSBs430は「01」に設定される。L4ポートサーチキー208eはL4ポート番号432を含む。L4ポート関連データ210eは、HTL404e、LNQoS有効フィールド420、L4QoS424、およびFO408eを含む。

**【0030】**

図5Aは、スイッチ100(図1)に接続された入力ポート102(図1)で受取られ得る先行技術のデータパケットを示す。図5Bから図5Eは、図5Aに示される先行技術のデータパケット内に含まれ得る先行技術のデータパケットヘッダの内容を示す。

**【0031】**

図5Aは、データフィールド510およびネットワーキングレイヤのためのヘッダ502、504、506、508を含む、先行技術のデータパケット500を示す。開放型システム相互接続(Open Systems Interconnection 「OSI」)参照モデル(Reference Model)における複数レイヤのうち4つのレイヤのヘッダが示される。これらは、物理レイヤ(L1)ヘッダ502、データリンクレイヤ(L2)ヘッダ504、ネットワーキングレイヤ(L3)ヘッダ506およびトランスポートレイヤ(L4)ヘッダ508を含む。たとえば、データリンク(L2)レイヤ504はイーサネット(R)(Ethernet)、ネットワーキングレイヤ(L3)ヘッダ506はインターネットプロトコル(Internet Protocol 「IP」)、トランスポートレイヤ(L4)ヘッダ508は伝送制御プロトコル(Transmission Control Protocol 「TCP」)またはユニキャストデータグラムプロトコル(Unicast Datagram Protocol 「UDP」)であり得る。

10

**【0032】**

図5Bは、イーサネット(R)データリンク(L2)ヘッダ504のフォーマットを示す。イーサネット(R)データリンク(L2)ヘッダ504は、宛先アドレス514、ソースアドレス516、選択的なバーチャルローカルエリアネットワーク特定(Virtual Local Area Network Identification 「VLAN ID」)フィールド518、および長さ/タイプフィールド520を含む。

**【0033】**

図5Cは、IPネットワークレイヤ(L3)ヘッダ506のフォーマットを示す。IPネットワークレイヤ(L3)ヘッダ506は、ソースIPアドレス544、宛先IPアドレス546、IPバージョンフィールド522、IP長さフィールド524、サービスのタイプ(Type of Service 「TOS」)526、全長(Total Length)528、アイデンティフィケーション530、フラグ532、フラグメントオフセット534、寿命(Time to Live 「TTL」)536、プロトコルタイプ540、ヘッダチェックサム542、オプション548およびパッド550を含む。

30

**【0034】**

図5Dは、TCPトранspoートレイヤ(L4)ヘッダ508aのフォーマットを示す。TCPトランspoートレイヤ(L4)ヘッダ508aは以下のフィールドを含む：TCPソースポート552a、TCP宛先ポート554a、連続番号556、肯定応答(acknowledgment)番号558、TCPオフセット560、リザーブドフィールド562、TCPフラグ564、ウインドウ566、TCPヘッダチェックサム568、緊急ポインタ(URGent Pointer)570、オプション572、およびTCPパッド574。

40

**【0035】**

図5Eは、UDPトランspoートレイヤ(L4)ヘッダ508bのフォーマットを示す。UDPトランspoートレイヤ(L4)ヘッダ508bは以下のフィールドを含む：UDPソースポート552b、UDP宛先ポート554b、UDPメッセージ長582、およびUDPヘッダチェックサム584。

**【0036】**

図6は、受信データパケットのヘッダにエンコードされているネットワークプロトコルに応じてその受信データパケットのための転送オーバライド206(図2)を選択するため

50

に、転送オーバライドロジック 200（図2）において行なわれる工程を示すフローチャートである。図6について、図2、図4A～4Eおよび図5A～5Eに関連して以下に説明する。

#### 【0037】

ステップ600において、転送オーバライドロジック200（図2）は、入力ポート102（図1）からデータパケットヘッダを受取る。転送オーバライドロジック200（図2）は、転送オーバライド206（図2）を初期化し、かつ、転送オーバライドロジック200内のEIDレジスタ（図示せず）を初期化する。処理はステップ602に続く。

#### 【0038】

ステップ602において、転送オーバライドロジック200（図2）は、EIDサーチキー208b（図4B）をプロトコルCAM202（図1）へと転送する。EIDサーチキー208b（図4B）のMSBは「1」に設定され、受信データパケットに含まれるイーサネット（R）データリンク（L2）ヘッダ504（図5B）内の長さ／タイプフィールド520（図5B）からのL2ペイロードタイプ414（図4B）を含む。もし合致するEIDエントリ300b（図4B）があれば、処理はステップ604に続く。もしなければ、受信データパケットの処理は完了し、転送オーバライド206（図2）は転送オーバライドロジック202（図2）内のUnknown\_EID\_Default\_Overrideレジスタ（図示せず）に記憶されているデフォルト値に設定される。転送オーバライドのデフォルト値は、そのデータパケットがセグメントバッファメモリ108（図1）に転送されるように、ホストプロセッサによって選択され得る。

#### 【0039】

ステップ604において、転送オーバライド206（図2）は、EID関連データ210b（図4B）内の転送オーバライドフィールド408b（図4B）に記憶されているL2転送オーバライド値に設定される。処理はステップ606に続く。

#### 【0040】

ステップ606において、転送オーバライドロジック200（図2）は、EIDエントリ300b（図4B）内のネクストフィールド402b（図4B）を調べる。もしネクストフィールド402b（図4B）がプロトコルCAM202（図2）のさらなるサーチが必要であることを示していれば、処理はステップ608に続く。そうでなければ、受信データパケットの処理は完了する。

#### 【0041】

ステップ608において、転送オーバライドロジック200（図2）は、IPバージョンサーチキー208d（図4D）をプロトコルCAM202（図2）に転送する。IPバージョンサーチキーのMSB416は「001」に設定され、IPバージョンサーチキー208d（図4D）は、入力データパケット（図5A）のIPネットワーキングレイヤ（L3）ヘッダ506（図5C）のIPバージョンフィールド522（図5C）から抽出されたIPバージョン432（図4D）を含む。もしプロトコルCAM202（図2）において合致が見つかれば、IPバージョン関連データ210dが転送オーバライドロジック200に転送されて、処理はステップ610に続く。もし見つからなければ、受信データパケットの処理は完了する。

#### 【0042】

ステップ610において、転送オーバライドロジック200（図2）は、転送オーバライド206（図2）を、IPバージョンタイプエントリ300d（図4D）内のIPバージョン関連データ210d（図4D）内に記憶されている転送オーバライド値408d（図4D）に設定する。処理はステップ612に続く。

#### 【0043】

ステップ612において、転送オーバライドロジック200（図2）は、IPバージョンエントリ300c内のネクストフィールド402c（図4C）を調べる。もしネクストフィールド402d（図4D）がプロトコルCAM202（図2）のさらなるサーチが必要であることを示していれば、処理はステップ614に続く。そうでなければ、受信データ

10

20

30

40

50

パケットの処理は完了する。

#### 【0044】

ステップ614において、転送オーバライドロジック200(図2)は、IPタイプサーチキー208c(図4C)をプロトコルCAM202(図2)に転送する。IPタイプサーチキー208c(図4C)のMSB426(図4C)は「001」に設定され、IPタイプサーチキー208cは、入力データパケット(図5A)のネットワーキングレイヤ(L3)ヘッダ506(図5C)内のプロトコルタイプフィールド540(図5C)から抽出されたIPv4タイプ428(図4C)を含む。もしプロトコルCAM202(図2)においてIPタイプサーチキー208cに対する合致が見つかれば、IPタイプ関連データ210c(図4C)が転送オーバライドロジック200(図2)に転送されて、処理はステップ616に続く。もし見つからなければ、受信データパケットの処理は完了する。  
10

#### 【0045】

ステップ616において、転送オーバライドロジック200(図2)は、転送オーバライド206(図2)をIPタイプエントリ300cに記憶されている転送オーバライド値408cに設定する。処理はステップ618に続く。

#### 【0046】

ステップ618において、転送オーバライドロジック200(図2)は、IPタイプエントリ300c内のネクストフィールド402c(図4C)を調べる。もしネクストフィールド402c(図4C)がプロトコルCAM202(図2)のさらなるサーチが必要であることを示していれば、処理はステップ620に続く。そうでなければ、受信データパケットの処理は完了する。  
20

#### 【0047】

ステップ620において、転送オーバライドロジック200(図2)は、「01」に設定されたMSB430(図4E)とL4ヘッダ508(図5D、5E)内のソースポートフィールド552(図5D～5E)から抽出されたL4ソースポート番号とを含むL4ポートサーチキー208e(図4E)を、プロトコルCAM202(図2)に転送する。もしプロトコルCAM202(図2)内にソースポートの合致が見つかれば、L4ポート関連データ210e(図4E)のコピーが転送オーバライドロジック200(図2)に転送され、処理はステップ624に続く。もし見つからなければ、処理はステップ622に続く。  
30

#### 【0048】

ステップ624において、転送オーバライドロジック200(図2)は、転送オーバライド206(図2)を、L4ポートエントリ300e(図4E)内のL4ポート関連データ210e(図4E)に記憶されている転送オーバライド値408e(図4E)に設定する。受信データパケットの処理はこれで完了する。

#### 【0049】

ステップ622において、転送オーバライドロジック200(図2)は、「01」に設定されたMSB430(図4E)および入力データパケット(図5A)のL4ヘッダ508(図5D、5E)に記憶されている宛先ポート554(図5D、5E)からのL4宛先ポート番号を含むL4ポートサーチキー208e(図4E)を、プロトコルCAM202(図2)に転送する。もしプロトコルCAM202(図2)において宛先ポートの合致が見つかれば、L4ポート関連データ210e(図4E)のコピーが転送オーバライドロジック200(図2)に転送され、処理はステップ626に続く。  
40

#### 【0050】

ステップ626において、転送オーバライドロジック200(図2)は、転送オーバライド206(図2)を、L4ポートエントリ300e(図4E)内のL4ポート関連データ200e(図4E)に記憶されている転送オーバライド値408e(図4E)に設定する。受信データパケットの処理はこれで完了する。

#### 【0051】

したがって、受信データパケットのヘッダにエンコードされているネットワークプロトコルに応じて転送オーバライド206を選択する目的で、プロトコルCAM202(図2)において最高5回の連続したサーチが、転送オーバライドロジック200(図2)により、受信データパケットのヘッダの内容を用いて行なわれ得る。

#### 【0052】

図7は、ルーティング情報プロトコル(Routing Information Protocol「RIP」)データパケット206(図2)のための転送オーバライドを選択するためにプロトコルCAM202においてサーチされる4つのプロトコルCAM300エントリを示す。これらのエントリは、EIDエントリ300ba、IPv4タイプエントリ300ca、IPバージョンエントリ300daおよびL4ポートエントリ300eaを含む。RIPデータパケットのための転送オーバライドを選択するためのプロトコルCAM202(図2)のサーチについて、図6に関連して説明する。  
10

#### 【0053】

ステップ602(図6)において、RIPデータパケットに含まれるイーサネット(R)データリンク(L2)ヘッダ504(図5B)の内容を用いて、サーチが行われる。MSBが「1」に設定されL2ペイロードタイプ414がIPに設定されたサーチキー208ba、すなわち、RIPデータパケットに含まれるイーサネット(R)データリンク(L2)ヘッダ504(図5B)内の長さ/タイプフィールド520(図5B)の内容が、転送オーバライドロジック200(図2)からプロトコルCAMへと転送される。プロトコルCAM202内の合致する転送用エントリ300baは、関連データ210baを含む。この関連データ210baは、「NEXT」に設定されたネクストフィールド402baおよび、「入力データのみ」に設定された転送オーバライドフィールド408baを含む。  
20

#### 【0054】

ステップ606(図6)において、プロトコルCAM202(図2)の別のサーチが行なわれる。これは、ネクストフィールド402baが「NEXT」に設定されているためである。このサーチは、RIPデータパケットに含まれるIPネットワークレイヤ(L3)ヘッダ506(図5C)の内容を用いて行なわれる。MSBが「001」に設定されかつIPバージョン418aがRIPデータパケットに含まれるIPネットワークレイヤ(L3)ヘッダ506(図5C)内のIPバージョンフィールド522の内容に設定されたサーチキー208daが、転送オーバライドロジック200(図2)からプロトコルCAM202(図2)に転送される。プロトコルCAM202内の合致する転送用エントリ300daは、IPバージョン関連データ210daを含む。IPバージョン関連データ210daは、「NEXT」に設定されたネクストフィールド402daおよび、「入力データのみ」に設定された転送オーバライドフィールド408daを含む。  
30

#### 【0055】

ネクストフィールド402daが「NEXT」に設定されているために、ステップ612(図6)において、プロトコルCAM202(図2)の別のサーチが行なわれる。このサーチは、RIPデータパケットに含まれるIPネットワークレイヤ(L3)ヘッダ506(図5C)の内容を用いて行なわれる。MSBが「001」に設定されかつIPタイプ428(図4C)がUDPに設定されたIPタイプサーチキー208ca、すなわち、RIPデータパケットに含まれるIPネットワークレイヤ(L3)ヘッダ506(図5C)内のIPプロトコルタイプフィールド540(図5C)の内容が、転送オーバライドロジック200(図2)からプロトコルCAM202へと転送される。プロトコルCAM202内の合致する転送用エントリ300caは、関連データ210caを含む。この関連データ210caは、「NEXT」に設定されたネクストフィールド402caおよび、「入力データのみ」に設定された転送オーバライドフィールド408caを含む。  
40

#### 【0056】

ネクストフィールド402caが「NEXT」に設定されているので、ステップ618(図6)においてプロトコルCAM202の別のサーチが行なわれる。このサーチは、受信  
50

されたRIPデータパケットに含まれるUDPトランSPORTレイヤ(L4)ヘッダ508b(図5E)の内容を用いて行なわれる。MSBが「001」に設定されかつL4ポート番号がUDPトランSPORTレイヤ(L4)ヘッダ508b(図5E)内のUDPソースポートフィールド552b(図5E)の内容に設定されたサーチキー208eaが、転送オーバライドロジック200(図2)からプロトコルCAM202(図2)へと転送される。プロトコルCAM202(図2)内の合致する転送用エントリ300eaは、L4ポート関連データ210eaを含む。このL4ポート関連データ210eaは、「STO P」に設定されたネクストフィールド402eaおよび、「入力データおよびホストデータ」に設定された転送オーバライドフィールド408eaを含む。転送オーバライド206(図2)は、転送オーバライド408eaの内容に従って設定される。RIPデータパケットは、ホストインターフェイスロジック106(図1)およびセグメントバッファメモリ108(図1)の両方に転送される。  
10

#### 【0057】

プロトコルCAM202(図2)のサーチは、ネクストフィールド408eaが「STO P」に設定されているので完了する。したがって、RIPデータパケットのための転送オーバライド206(図2)を選択するために、プロトコルCAM202(図2)のサーチは4回行なわれる。

#### 【0058】

図4Bから図4Eに示すように、サービスの質(QoS)フィールド422、424もまた、プロトコルCAM202(図2)の関連データ210内に記憶され得る。QoSフィールド306(図3)は、受信データパケットに優先度を割当てるのに使用され得る。割当てられた優先度は、出力ポートエンジン110(図1)により、出力ポート112(図1)に転送されるパケットの順番を選択するのに用いられる。たとえば、QoSフィールド306(図3)の内容は、テルネット(Telnet)データパケットに最高の優先順位を割当て、ハイパーテキストマークアップ言語(HyperText Markup language「HTTP」)データパケットに最低の優先順位を割当てる等、異なる優先度を割当てるのに使用され得る。スイッチ100は、各QoS値につき別個のポートキューを含み得、データパケットのメモリ内の場所を示すポインタが、そのデータパケットタイプのための転送用エントリ300(図3)におけるQoSフィールド306(図3)の内容に応じて、1または複数のポートキューに記憶される。  
20  
30

#### 【0059】

以上に本発明をローカルエリアネットワーク(Local Area Network「LAN」)について説明したが、本発明はLANに限定されるものではなく、ワイドエリアネットワーク(Wide Area Network)においてデータパケットのための転送オーバライドを選択するにも実現され得る。

#### 【0060】

本発明をその好ましい実施例に関連して特定的に図示し説明したが、本発明の形状および詳細に前掲の請求項に包含される本発明の範囲から離れることなく種々の変更が加えられ得ることは、当業者には理解されるであろう。

#### 【図面の簡単な説明】

【図1】 本発明の原理に従ったネットワークスイッチであって、スイッチに接続された入力ポートで受取られたデータパケットを転送するためのホスト転送用ロジックを含むネットワークスイッチを示す。

【図2】 図1に示されるホスト転送用ロジックを示す。

【図3】 図2に示されるプロトコル連想記憶メモリ(「CAM」)のエントリのフォーマットを示す。

【図4A】 図2に示されるプロトコルCAMの空のエントリを示す。

【図4B】 図2に示されるプロトコルCAMのレイヤ2タイプエントリを示す。

【図4C】 図2に示されるプロトコルCAMのIPv4タイプエントリを示す。

【図4D】 図2に示されるプロトコルCAMのIPバージョンエントリを示す。  
40  
50

【図4 E】 図2に示されるプロトコルCAMのレイヤ4エントリを示す。

【図5 A】 スイッチに接続された入力ポートで受取られ得る先行技術のデータパケットを示す。

【図5 B】 図5 Aに示されるデータパケットに含まれ得る先行技術のイーサネット(R)(データリンクレイヤ(L2))ヘッダを示す。

【図5 C】 図5 Aに示されるデータパケットに含まれ得る先行技術のインターネットプロトコル(ネットワークレイヤ(L3))ヘッダを示す。

【図5 D】 図5 Aに示されるデータパケットに含まれ得る先行技術の传送制御プロトコル('TCP') (トランスポートレイヤ(L4))ヘッダを示す。

【図5 E】 図5 Aに示されるデータパケットに含まれ得る先行技術のユーザデータグラムプロトコル(User Datagram Protocol 'UDP') (トランスポートレイヤ(L4))ヘッダを示す。 10

【図6】 受信データパケットのための転送オーバライドを選択するために図2に示される転送オーバライドロジックにおいて行なわれる工程を示すフローチャートである。

【図7】 プロトコルCAM内のエントリを示す。

【図1】

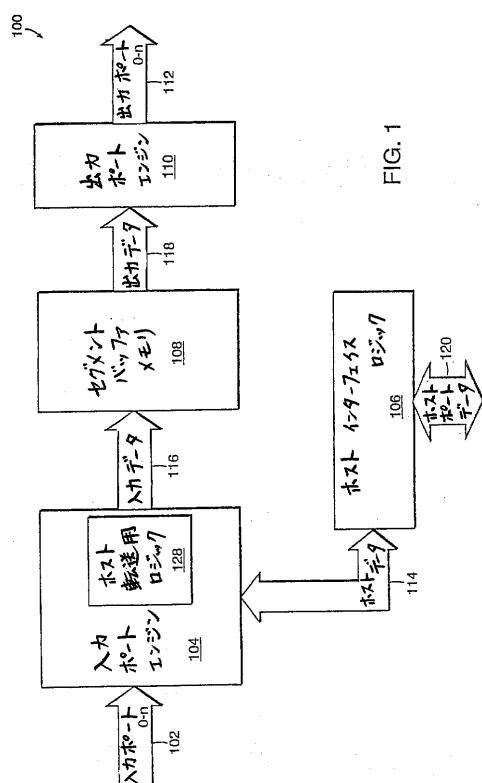


FIG. 1

【図2】

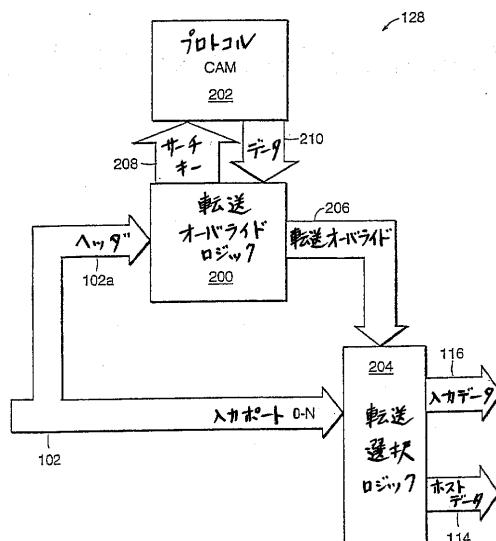


FIG. 2

【図3】

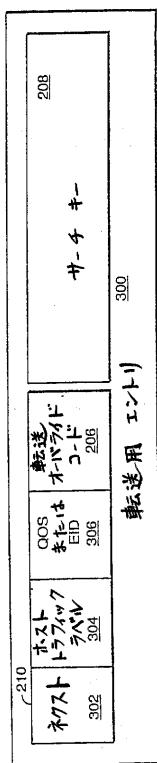


FIG. 3

【図4A】

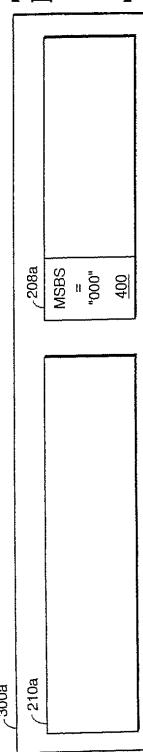


FIG. 4A

【図4B】

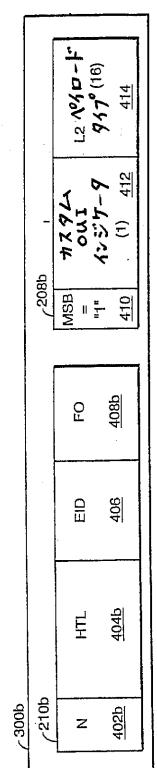


FIG. 4B

【図4C】

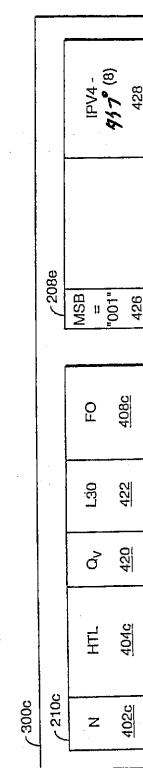


FIG. 4C

【図 4 D】

300d			
210d	208d	208d	208d
N	HTL	Q <sub>v</sub>	L3Q
402d	404d	420d	422
508	510	408d	408d

FIG. 4D

【図 4 E】

300e			
210e	208e	208e	208e
N	HTL	Q <sub>v</sub>	L4Q
401e	420e	424	408e
510	512	430	432

FIG. 4E

【図 5 A】

500			
物理 L14	物理 L12	物理 L3	物理 L4
504	504	506	508
502	502	506	508

PRIOR ART  
FIG. 5A

【図 5 B】

504			
光アダプタ 514	光アダプタ 516	選択的 VLAN ID 518	長さ/タイム 520
514	516	518	520

PRIOR ART  
FIG. 5B

【図 5 C】

IP サン		IP 長さ	サービス番号	全長
506		524	526	528
IP サン				534
530			532	534
アイデンティティ-ション			7ラグ	7ラグ
寿命	プロトコル	9ラグ	ヘッダ	チエックサム
536	540	540	542	542
506	IP アドレス	544		
	宛先 IP アドレス	546		
	オプション	548	10, 12	550

PRIOR ART  
FIG. 5C

【図 5 D】

7ラスポート (TCP) 552a		宛先ポート (TCP) 554a	
連続番号 556			
肯定応答番号 558			
TCP オフセット 560	リザード 562	TCP 7ラグ 564	ウイング 566
TCP ヘッダ チエックサム 568		緊急ポインタ 570	
オプション 572		TCP パンド 574	

FIG. 5D  
PRIOR ART

【図 5 E】

(UDP) 7ラスポート 552b	(UDP) 宛先ポート 554b
UDP メッセージ長 582	UDP ヘッダ チエックサム 584

FIG. 5E  
PRIOR ART

【図 6】

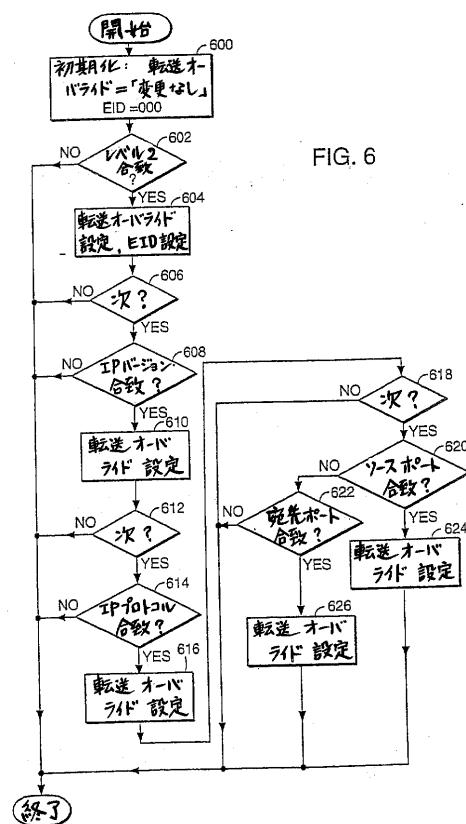


FIG. 6

【図7】

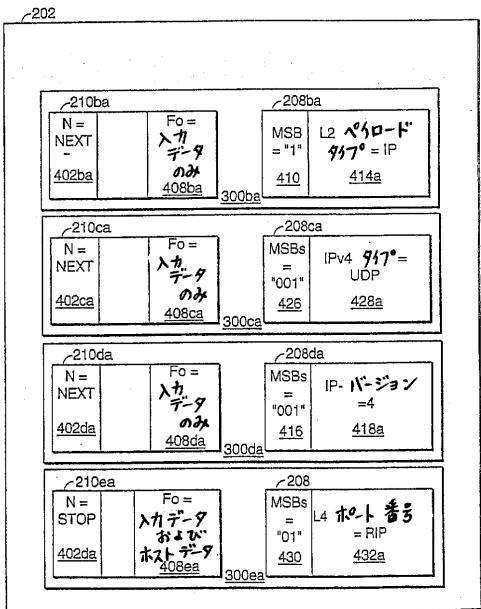


FIG. 7

---

フロントページの続き

(74)代理人 100096781

弁理士 堀井 豊

(74)代理人 100096792

弁理士 森下 八郎

(72)発明者 ブラウン, デイビッド・エイ

アメリカ合衆国、ケイ・0・エイ 1・エル・0 オンタリオ州、カーブ、アリシア・クレッセント、110

審査官 衣鳩 文彦

(56)参考文献 国際公開第99/000948 (WO, A1)

特開平10-257066 (JP, A)

特表2002-507362 (JP, A)

特開平09-162922 (JP, A)

特開昭62-137799 (JP, A)

(58)調査した分野(Int.Cl., DB名)

H04L 12/56