

(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.

F I

H O 4 L 12/56 (2006.01)

H O 4 L 12/56 1 O O Z

請求項の数 17 (全 18 頁)

(21) 出願番号	特願2001-548968 (P2001-548968)	(73) 特許権者	508034325
(86) (22) 出願日	平成12年12月22日 (2000.12.22)		モサイド・テクノロジーズ・インコーポレ
(65) 公表番号	特表2003-518873 (P2003-518873A)		ーテッド
(43) 公表日	平成15年6月10日 (2003.6.10)		カナダ・オンタリオ・K 2 K ・ 2 X 1 ・ オ
(86) 国際出願番号	PCT/CA2000/001515		タワ・ハインズ・ロード・1 1 ・ スイート
(87) 国際公開番号	W02001/048975		・ 2 0 3
(87) 国際公開日	平成13年7月5日 (2001.7.5)	(74) 代理人	100064746
審査請求日	平成19年12月13日 (2007.12.13)		弁理士 深見 久郎
(31) 優先権主張番号	09/473,074	(74) 代理人	100085132
(32) 優先日	平成11年12月28日 (1999.12.28)		弁理士 森田 俊雄
(33) 優先権主張国	米国 (US)	(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ーライド

(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ーライドを含み、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ー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ーライドに応じて、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。

10

20

【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】

該転送オーバーライドは、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ーライドが、現時点におけるサーチで抽出された該転送オーバー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ー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ーライドを含み、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ー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】

30

最終のサーチに対応する該抽出された転送オーバー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。

【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】

40

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】

該転送オーバーライドは、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ーライドが、現時点におけるサーチで抽出された該転送オーバー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】

50

【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捜し出される。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不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不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、転送オーバー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ーライドロジックは、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捜し出す。該転送オーバーライドロジックは、それら合致したエントリから該転送オーバーライドを抽出する。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。該転送オーバーライドロジックは、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、該転送オーバー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ーライドを抽出する。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ー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の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ク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ク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、新しいネットワーク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」)、転送オーバーライドロジック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内の転送オーバーライドロジック200に転送される。

【0017】

転送オーバーライドロジック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ーライド206を捜し出す。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CAM202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、転送オーバーライドロジック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が転送オーバーライドロジック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は転送オーバーライド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。転送オーバーライド206は、そのデータパケット

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 3 0 0 (図 3) 内の転送オーバーライド 2 0 6 が転送選択ロジック 2 0 4 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ーライド 2 0 6 の状態に応じて、入力データ経路 1 1 6 および / またはホストデータ経路 1 1 4 上を転送される。

【 0 0 1 9 】

【 表 1 】

転送オーバーライド	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 C A M 2 0 2 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込アクセス権を有する。転送選択ロジック 2 0 4 は、入力ポート 1 0 2 で受信されたデータパケットを、転送オーバーライドロジック 2 0 0 から転送された転送オーバーライド 2 0 6 の状態に応じて、入力データ 1 1 6 および / またはホストデータ 1 1 4 へと転送する。転送選択ロジック 2 0 4 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 1 1 4 であり、入力バッファの出力は入力データ 1 1 6 である。転送オーバーライド 2 0 6 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、この転送オーバーライド 2 0 6 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 1 1 4 へ、および / または、入力バッファを介して入力データ 1 1 6 へと転送されるべきかどうか決定される。

【 0 0 2 0 】

図 3 は、プロトコル C A M 2 0 2 (図 2) 内の転送用エントリ 3 0 0 を示す。転送用エントリ 3 0 0 は、サーチキー 2 0 8 および関連データ 2 1 0 を含む。関連データ 2 1 0 は、ネクストフィールド 3 0 2 と、選択的なホストトラフィックラベル 3 0 4 と、選択的なサービスの質 (Quality of Service 「 Q o S 」) または E タイプ識別子 (Etype Identifier 「 E I D 」) フィールド 3 0 6 と、転送オーバーライドコード 2 0 6 とを含む。転送用エントリ 3 0 0 のコピーが転送オーバーライドロジック 2 0 0 (図 2) 内に記憶され、転送用エントリ 3 0 0 のコピー内に記憶されている転送オーバーライドコード 2 0 6 は、以後のプロトコル C A M 2 0 2 (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ーライドコード 2 0 6 でオーバーライドされる。

【 0 0 2 1 】

Q o S フィールド 3 0 6 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 3 0 4 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 3 0 6 および Q o S または E I D フィールド 3 0 6 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ースロジック 1 0 6 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 3 0 0 に含まれるサーチキー 2 0 8 および関連データ 2 1 0 の内容は、転送オーバーライドロジック 2 0 0 (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 2 0 8 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 3 0 0 については、図 4 A から図 4 E に関連して説明する。

【 0 0 2 2 】

転送オーバーライドロジック 2 0 0 (図 2) は、ネクストフィールド 3 0 2 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 3 0 0 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ーライドコード 2 0 6 が、転送オーバーライドロジック 2 0 0 (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ーライドコード 2 0 6 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ーライドコード 2 0 6 が、そのデータパケットのための転送用決定となる。

【 0 0 2 3 】

たとえば、L 2 ヘッダに記憶されているタイプが A R P である場合、すべての A R P データパケットが入力データ経路 1 1 6 上およびホストデータ経路 1 1 4 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 3 0 0 内のネクストフィールド 3 0 2 は「 0 」に設定される。しかし、L 2 ヘッダに記憶されているタイプが I P である場合には、I P データパケットのタイプを判断するのに、L 3 レイヤヘッダにおいてさらなるサーチが必要である。L 3 レイヤヘッダのタイプフィールドに記憶されている I P 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 「 I G M P 」) である場合、ネクストフィールド 3 0 2 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 1 1 6 (図 1) 上およびホストデータ経路 1 1 4 (図 1) 上を転送される。

【 0 0 2 4 】

I P 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 1 1 6 (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ーライドコード 2 0 6 (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 1 1 4 (図 1) 上を転送するようにオーバーライドされる。C A M 内に転送用エントリ 3 0 0 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 1 0 0 (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。

【 0 0 2 5 】

図 4 A は、空または無効のエントリ 3 0 0 a を示す。空のサーチキー 2 0 8 a の 3 つの最上位ビット (Most Significant Bits 「 M S B s 」) は「 0 0 0 」に設定され、無効な関連データ 2 1 0 a が記憶される。

【 0 0 2 6 】

図 4 B は、E タイプ識別子 (「 E I D 」) エントリ 3 0 0 b を示す。E I D サーチキー 2 0 8 b の M S B 4 1 0 は「 1 」に設定される。E I D サーチキー 2 0 8 b は L 2 ペイロードタイプ 4 1 4 およびカスタム O U I インジケータ 4 1 2 を含む。E I D 関連データ 2 1 0 b は、ネクストフィールド 4 0 2 b、ホストトラフィックラベル (Host Traffic Label 「 H T L 」) 4 0 4 b、E I D 4 0 6 および転送オーバーライド (Forward Override 「 F O 」) 4 0 8 b を含む。

【 0 0 2 7 】

図 4 C は、I P v 4 タイプエントリ 3 0 0 c を示す。I P タイプサーチキー 2 0 8 c の 3 つの M S B s 4 2 6 は「 0 0 1 」に設定される。I P タイプサーチキー 2 0 8 c は I P v 4 タイプ 4 2 8 を含む。I P タイプ関連データ 2 1 0 c は、ネクストフィールド 4 0 2 c、H T L 4 0 4 c、L N Q o S 有効ビット 4 2 0、レイヤ 3 Q o S 4 2 2、および F O 4 0 8 c を含む。

【 0 0 2 8 】

図 4 D は、I P バージョンエントリ 3 0 0 d を示す。I P バージョンサーチキー 2 0 8 d の 3 つの M S B s 4 1 6 は「 0 0 1 」に設定される。I P バージョンサーチキー 2 0 8 d は I P バージョン 4 1 8 を含む。I P バージョン関連データ 2 1 0 d は、ネクストフィールド 4 0 2 d、H T L 4 0 4 d、L N Q o S 有効ビット 4 2 0、レイヤ 3 Q o S 4 2 2、および F O 4 0 8 d を含む。

10

20

30

40

50

【 0 0 2 9 】

図 4 E は、L 4 ポートエントリ 3 0 0 e を示す。L 4 サーチキー 2 0 8 e の 2 つの M S B s 4 3 0 は「 0 1 」に設定される。L 4 ポートサーチキー 2 0 8 e は L 4 ポート番号 4 3 2 を含む。L 4 ポート関連データ 2 1 0 e は、H T L 4 0 4 e、L N Q o S 有効フィールド 4 2 0、L 4 Q o S 4 2 4、および F O 4 0 8 e を含む。

【 0 0 3 0 】

図 5 A は、スイッチ 1 0 0 (図 1) に接続された入力ポート 1 0 2 (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 5 B から図 5 E は、図 5 A 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。

【 0 0 3 1 】

図 5 A は、データフィールド 5 1 0 およびネットワークレイヤのためのヘッダ 5 0 2、5 0 4、5 0 6、5 0 8 を含む、先行技術のデータパケット 5 0 0 を示す。開放型システム相互接続 (Open Systems Interconnection 「O S I」) 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 (L 1) ヘッダ 5 0 2、データリンクレイヤ (L 2) ヘッダ 5 0 4、ネットワークレイヤ (L 3) ヘッダ 5 0 6 およびトランスポートレイヤ (L 4) ヘッダ 5 0 8 を含む。たとえば、データリンク (L 2) レイヤ 5 0 4 はイーサネット (R) (Ethernet)、ネットワークレイヤ (L 3) ヘッダ 5 0 6 はインターネットプロトコル (Internet Protocol 「I P」)、トランスポートレイヤ (L 4) ヘッダ 5 0 8 は伝送制御プロトコル (Transmission Control Protocol 「T C P」) またはユニキャストデータグラムプロトコル (Unicast Datagram Protocol 「U D P」) であり得る。

【 0 0 3 2 】

図 5 B は、イーサネット (R) データリンク (L 2) ヘッダ 5 0 4 のフォーマットを示す。イーサネット (R) データリンク (L 2) ヘッダ 5 0 4 は、宛先アドレス 5 1 4、ソースアドレス 5 1 6、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 「V L A N I D」) フィールド 5 1 8、および長さ / タイプフィールド 5 2 0 を含む。

【 0 0 3 3 】

図 5 C は、I P ネットワークレイヤ (L 3) ヘッダ 5 0 6 のフォーマットを示す。I P ネットワークレイヤ (L 3) ヘッダ 5 0 6 は、ソース I P アドレス 5 4 4、宛先 I P アドレス 5 4 6、I P バージョンフィールド 5 2 2、I P 長さフィールド 5 2 4、サービスのタイプ (Type of Service 「T O S」) 5 2 6、全長 (Total Length) 5 2 8、アイデンティフィケーション 5 3 0、フラグ 5 3 2、フラグメントオフセット 5 3 4、寿命 (Time to Live 「T T L」) 5 3 6、プロトコルタイプ 5 4 0、ヘッダチェックサム 5 4 2、オプション 5 4 8 およびパッド 5 5 0 を含む。

【 0 0 3 4 】

図 5 D は、T C P トランスポートレイヤ (L 4) ヘッダ 5 0 8 a のフォーマットを示す。T C P トランスポートレイヤ (L 4) ヘッダ 5 0 8 a は以下のフィールドを含む：T C P ソースポート 5 5 2 a、T C P 宛先ポート 5 5 4 a、連続番号 5 5 6、肯定応答 (acknowledgment) 番号 5 5 8、T C P オフセット 5 6 0、リザーブドフィールド 5 6 2、T C P フラグ 5 6 4、ウィンドウ 5 6 6、T C P ヘッダチェックサム 5 6 8、緊急ポインタ (Urgent Pointer) 5 7 0、オプション 5 7 2、および T C P パッド 5 7 4。

【 0 0 3 5 】

図 5 E は、U D P トランスポートレイヤ (L 4) ヘッダ 5 0 8 b のフォーマットを示す。U D P トランスポートレイヤ (L 4) ヘッダ 5 0 8 b は以下のフィールドを含む：U D P ソースポート 5 5 2 b、U D P 宛先ポート 5 5 4 b、U D P メッセージ長 5 8 2、および U D P ヘッダチェックサム 5 8 4。

【 0 0 3 6 】

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ーライド 2 0 6 (図 2) を選択するため

10

20

30

40

50

に、転送オーバーライドロジック 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 4 A ~ 4 E および図 5 A ~ 5 E に関連して以下に説明する。

【0037】

ステップ 600 において、転送オーバーライドロジック 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。転送オーバーライドロジック 200 (図 2) は、転送オーバーライド 206 (図 2) を初期化し、かつ、転送オーバーライドロジック 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、転送オーバーライドロジック 200 (図 2) は、EID サーチキー 208b (図 4 B) をプロトコル CAM 202 (図 1) へと転送する。EID サーチキー 208b (図 4 B) 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 5 B) 内の長さ / タイプフィールド 520 (図 5 B) からの L2 ペイロードタイプ 414 (図 4 B) を含む。もし合致する EID エントリ 300b (図 4 B) 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、転送オーバーライド 206 (図 2) は転送オーバーライドロジック 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。転送オーバー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、転送オーバーライド 206 (図 2) は、EID 関連データ 210b (図 4 B) 内の転送オーバーライドフィールド 408b (図 4 B) に記憶されている L2 転送オーバー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、転送オーバーライドロジック 200 (図 2) は、EID エントリ 300b (図 4 B) 内のネクストフィールド 402b (図 4 B) を調べる。もしネクストフィールド 402b (図 4 B) がプロトコル CAM 202 (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、転送オーバーライドロジック 200 (図 2) は、IP バージョンサーチキー 208d (図 4 D) をプロトコル CAM 202 (図 2) に転送する。IP バージョンサーチキーの MSB 416 は「001」に設定され、IP バージョンサーチキー 208d (図 4 D) は、入力データパケット (図 5 A) の IP ネットワーキングレイヤ (L3) ヘッダ 506 (図 5 C) の IP バージョンフィールド 522 (図 5 C) から抽出された IP バージョン 432 (図 4 D) を含む。もしプロトコル CAM 202 (図 2) において合致が見つければ、IP バージョン関連データ 210d が転送オーバーライドロジック 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、転送オーバーライドロジック 200 (図 2) は、転送オーバーライド 206 (図 2) を、IP バージョンタイプエントリ 300d (図 4 D) 内の IP バージョン関連データ 210d (図 4 D) 内に記憶されている転送オーバーライド値 408d (図 4 D) 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、転送オーバーライドロジック 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 4 C) を調べる。もしネクストフィールド 402d (図 4 D) がプロトコル CAM 202 (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。

【 0 0 4 4 】

ステップ 6 1 4 において、転送オーバーライドロジック 2 0 0 (図 2) は、IP タイプサーチキー 2 0 8 c (図 4 C) をプロトコル C A M 2 0 2 (図 2) に転送する。IP タイプサーチキー 2 0 8 c (図 4 C) の M S B 4 2 6 (図 4 C) は「 0 0 1 」に設定され、IP タイプサーチキー 2 0 8 c は、入力データパケット (図 5 A) のネットワーキングレイヤ (L 3) ヘッダ 5 0 6 (図 5 C) 内のプロトコルタイプフィールド 5 4 0 (図 5 C) から抽出された I P v 4 タイプ 4 2 8 (図 4 C) を含む。もしプロトコル C A M 2 0 2 (図 2) において IP タイプサーチキー 2 0 8 c に対する合致が見つければ、IP タイプ関連データ 2 1 0 c (図 4 C) が転送オーバーライドロジック 2 0 0 (図 2) に転送されて、処理はステップ 6 1 6 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

【 0 0 4 5 】

ステップ 6 1 6 において、転送オーバーライドロジック 2 0 0 (図 2) は、転送オーバーライド 2 0 6 (図 2) を IP タイプエン트리 3 0 0 c に記憶されている転送オーバーライド値 4 0 8 c に設定する。処理はステップ 6 1 8 に続く。

【 0 0 4 6 】

ステップ 6 1 8 において、転送オーバーライドロジック 2 0 0 (図 2) は、IP タイプエン트리 3 0 0 c 内のネクストフィールド 4 0 2 c (図 4 C) を調べる。もしネクストフィールド 4 0 2 c (図 4 C) がプロトコル C A M 2 0 2 (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 6 2 0 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

【 0 0 4 7 】

ステップ 6 2 0 において、転送オーバーライドロジック 2 0 0 (図 2) は、「 0 1 」に設定された M S B 4 3 0 (図 4 E) と L 4 ヘッダ 5 0 8 (図 5 D、5 E) 内のソースポートフィールド 5 5 2 (図 5 D ~ 5 E) から抽出された L 4 ソースポート番号とを含む L 4 ポートサーチキー 2 0 8 e (図 4 E) を、プロトコル C A M 2 0 2 (図 2) に転送する。もしプロトコル C A M 2 0 2 (図 2) 内にソースポートの合致が見つければ、L 4 ポート関連データ 2 1 0 e (図 4 E) のコピーが転送オーバーライドロジック 2 0 0 (図 2) に転送され、処理はステップ 6 2 4 に続く。もし見つからなければ、処理はステップ 6 2 2 に続く。

30

【 0 0 4 8 】

ステップ 6 2 4 において、転送オーバーライドロジック 2 0 0 (図 2) は、転送オーバーライド 2 0 6 (図 2) を、L 4 ポートエン트리 3 0 0 e (図 4 E) 内の L 4 ポート関連データ 2 1 0 e (図 4 E) に記憶されている転送オーバーライド値 4 0 8 e (図 4 E) 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。

【 0 0 4 9 】

ステップ 6 2 2 において、転送オーバーライドロジック 2 0 0 (図 2) は、「 0 1 」に設定された M S B 4 3 0 (図 4 E) および入力データパケット (図 5 A) の L 4 ヘッダ 5 0 8 (図 5 D、5 E) に記憶されている宛先ポート 5 5 4 (図 5 D、5 E) からの L 4 宛先ポート番号を含む L 4 ポートサーチキー 2 0 8 e (図 4 E) を、プロトコル C A M 2 0 2 (図 2) に転送する。もしプロトコル C A M 2 0 2 (図 2) において宛先ポートの合致が見つければ、L 4 ポート関連データ 2 1 0 e (図 4 E) のコピーが転送オーバーライドロジック 2 0 0 (図 2) に転送され、処理はステップ 6 2 6 に続く。

40

【 0 0 5 0 】

ステップ 6 2 6 において、転送オーバーライドロジック 2 0 0 (図 2) は、転送オーバーライド 2 0 6 (図 2) を、L 4 ポートエン트리 3 0 0 e (図 4 E) 内の L 4 ポート関連データ 2 0 0 e (図 4 E) に記憶されている転送オーバーライド値 4 0 8 e (図 4 E) 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。

【 0 0 5 1 】

50

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ーライド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、転送オーバーライドロジック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ー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ー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、転送オーバーライドロジック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ーライドフィールド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、転送オーバーライドロジック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ーライドフィールド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、転送オーバーライドロジック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ーライドフィールド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 R I P データパケットに含まれる U D P トランスポートレイヤ (L 4) ヘッダ 5 0 8 b (図 5 E) の内容を用いて行なわれる。 M S B が「 0 0 1 」に設定されかつ L 4 ポート番号が U D P トランスポートレイヤ (L 4) ヘッダ 5 0 8 b (図 5 E) 内の U D P ソースポートフィールド 5 5 2 b (図 5 E) の内容に設定されたサーチキー 2 0 8 e a が、転送オーバーライドロジック 2 0 0 (図 2) からプロトコル C A M 2 0 2 (図 2) へと転送される。プロトコル C A M 2 0 2 (図 2) 内の合致する転送用エントリ 3 0 0 e a は、 L 4 ポート関連データ 2 1 0 e a を含む。この L 4 ポート関連データ 2 1 0 e a は、「 S T O P 」に設定されたネクストフィールド 4 0 2 e a 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ーライドフィールド 4 0 8 e a を含む。転送オーバーライド 2 0 6 (図 2) は、転送オーバーライド 4 0 8 e a の内容に従って設定される。 R I P データパケットは、ホストインターフェイスロジック 1 0 6 (図 1) およびセグメントバッファメモリ 1 0 8 (図 1) の両方に転送される。

10

【 0 0 5 7 】

プロトコル C A M 2 0 2 (図 2) のサーチは、ネクストフィールド 4 0 8 e a が「 S T O P 」に設定されているので完了する。したがって、 R I P データパケットのための転送オーバーライド 2 0 6 (図 2) を選択するために、プロトコル C A M 2 0 2 (図 2) のサーチは 4 回行なわれる。

【 0 0 5 8 】

図 4 B から図 4 E に示すように、サービスの質 (Q o S) フィールド 4 2 2、4 2 4 もまた、プロトコル C A M 2 0 2 (図 2) の関連データ 2 1 0 内に記憶され得る。 Q o S フィールド 3 0 6 (図 3) は、受信データパケットに優先度を割当ててのに使用され得る。割当てられた優先度は、出力ポートエンジン 1 1 0 (図 1) により、出力ポート 1 1 2 (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、 Q o S フィールド 3 0 6 (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 「 H T T P 」) データパケットに最低の優先順位を割当てて等、異なる優先度を割当ててのに使用され得る。スイッチ 1 0 0 は、各 Q o S 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 3 0 0 (図 3) における Q o S フィールド 3 0 6 (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

【 0 0 5 9 】

以上に本発明をローカルエリアネットワーク (Local Area Network 「 L A N 」) について説明したが、本発明は L A N 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ーライドを選択するのにも実現され得る。

【 0 0 6 0 】

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】

40

【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 (「 C A M 」) のエントリのフォーマットを示す。

【図 4 A】 図 2 に示されるプロトコル C A M の空のエントリを示す。

【図 4 B】 図 2 に示されるプロトコル C A M のレイヤ 2 タイプエントリを示す。

【図 4 C】 図 2 に示されるプロトコル C A M の I P v 4 タイプエントリを示す。

【図 4 D】 図 2 に示されるプロトコル C A M の I P バージョンエントリを示す。

50

【図 4 E】 図 2 に示されるプロトコル C A M 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 (L 2)) 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 (L 3)) 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伝送制御プロトコル (「 T C P 」) (トランスポートレイヤ (L 4)) ヘッダを示す。

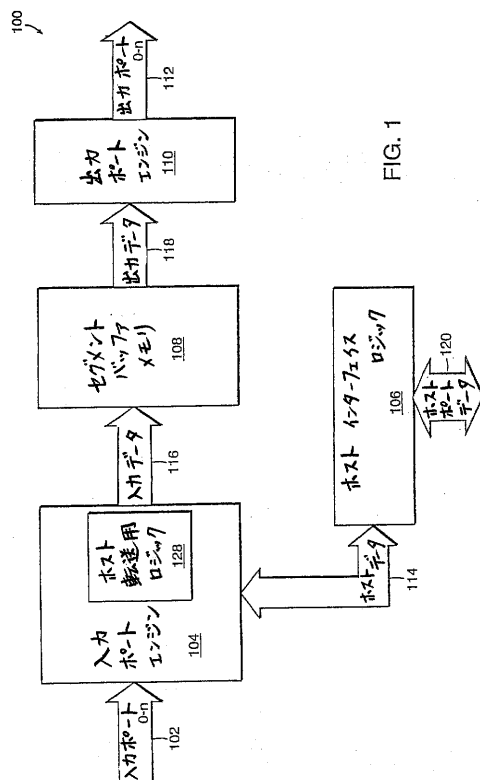
【図 5 E】 図 5 A に示されるデータパケットに含まれ得る先行技術のユーザデータグラムプロトコル (User Datagram Protocol 「 U D P 」) (トランスポートレイヤ (L 4)) ヘッダを示す。

10

【図 6】 受信データパケットのための転送オーバーライドを選択するために図 2 に示される転送オーバー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 C A M 内のエントリを示す。

【図 1】



【図 2】

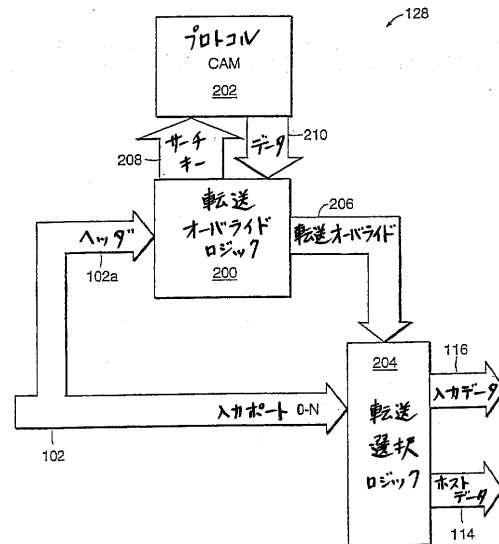


FIG. 2

【図 3】

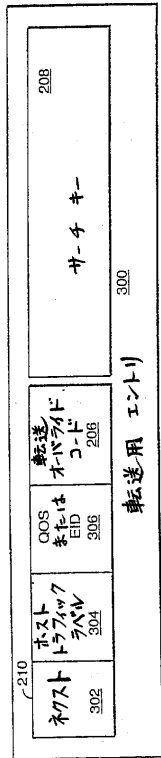


FIG. 3

【図 4 A】

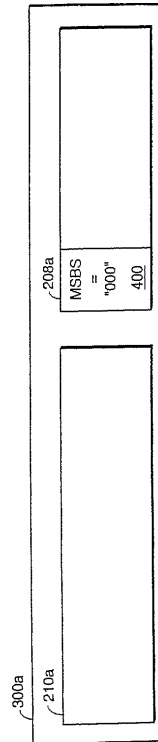


FIG. 4A

【図 4 B】

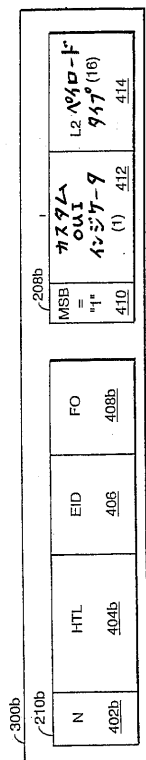


FIG. 4B

【図 4 C】

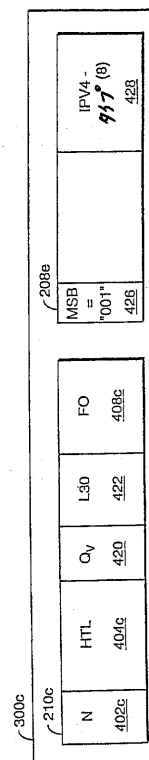
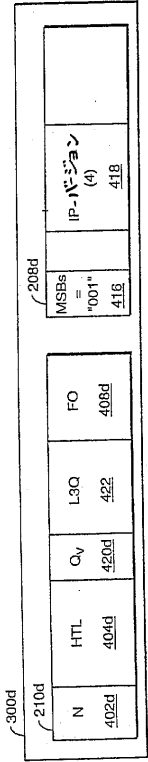
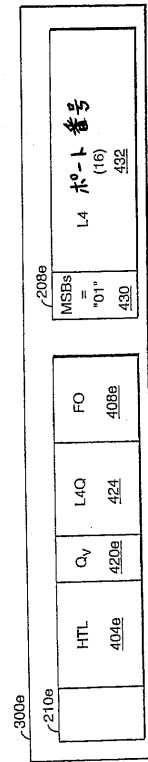


FIG. 4C

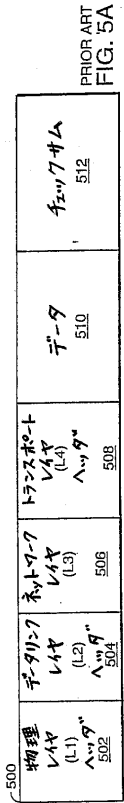
【図 4 D】



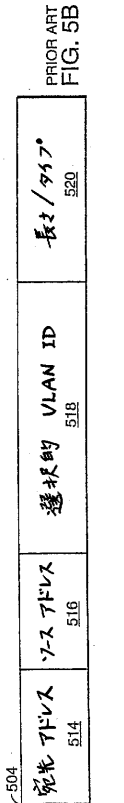
【図 4 E】



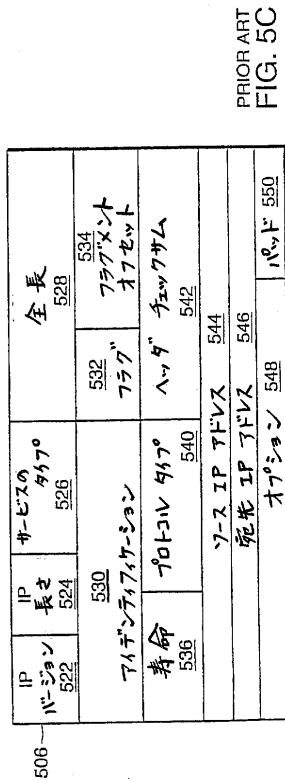
【図 5 A】



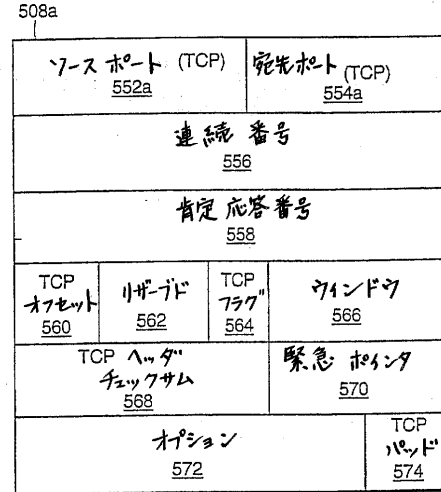
【図 5 B】



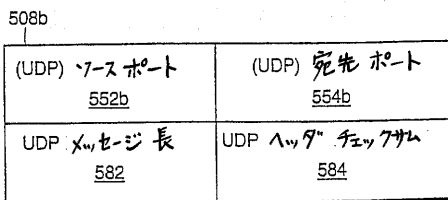
【図 5 C】



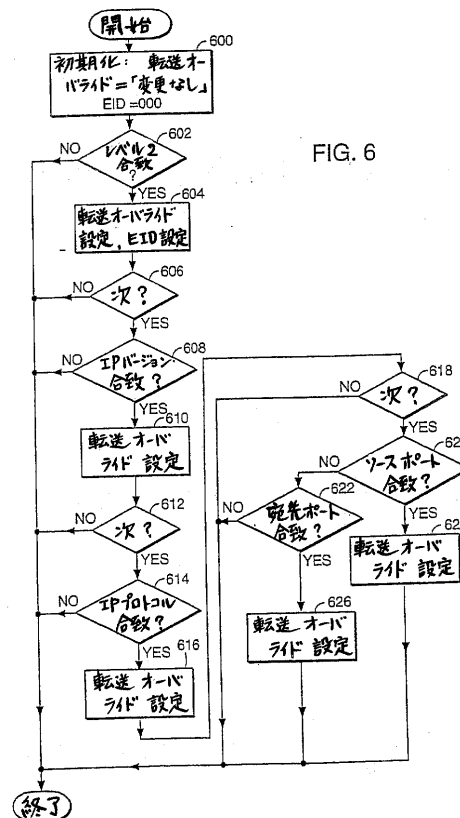
【図 5 D】

FIG. 5D
PRIOR ART

【図 5 E】

FIG. 5E
PRIOR ART

【図 6】



【図 7】

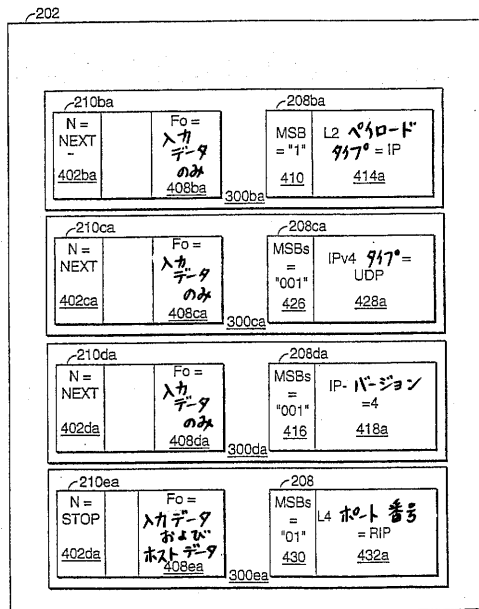


FIG. 7

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、カープ、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 9 9 / 0 0 0 9 4 8 (W O , A 1)

特開平 1 0 - 2 5 7 0 6 6 (J P , A)

特表 2 0 0 2 - 5 0 7 3 6 2 (J P , A)

特開平 0 9 - 1 6 2 9 2 2 (J P , A)

特開昭 6 2 - 1 3 7 7 9 9 (J P , A)

(58)調査した分野(Int.Cl. , D B 名)

H04L 12/56