

(19)日本国特許庁 (J P)

(12) 特 許 公 報 (B 1)

(11)特許番号

第2896509号

(45)発行日 平成11年(1999) 5月31日

(24)登録日 平成11年(1999) 3月12日

(51)Int.Cl.<sup>6</sup>

識別記号

F I

H 0 4 L 12/56

H 0 4 L 11/20

1 0 2 Z

請求項の数1 (全 7 頁)

(21)出願番号 特願平10-71814

(22)出願日 平成10年(1998) 3月20日

審査請求日 平成10年(1998) 3月20日

特許法第30条第1項適用申請有り 1997年9月24日発行  
社団法人情報処理学会 第55回全国大会講演論文集  
(3) p. 804~805 6U-9「高速大容量IPスイッチングに適した多段接続方式」

(73)特許権者 391012327

東京大学長

東京都文京区本郷7丁目3番1号

(72)発明者 斉藤 忠夫

神奈川県横浜市神奈川区松が丘55-4

(72)発明者 相田 仁

神奈川県川崎市宮前区宮崎2-12-1

宮崎台プラザビル203号

(72)発明者 青木 輝勝

埼玉県上福岡市北野1-7-2

(74)代理人 弁理士 杉村 暁秀 (外8名)

審査官 江嶋 清仁

(56)参考文献 特開 平4-223692 (J P, A)

1997年電子情報通信学会総合大会 B

-6-6 (1997年3月6日発行)

最終頁に続く

(54)【発明の名称】 大容量可変長パケット交換に適したパケットスイッチ方式

(57)【特許請求の範囲】

【請求項1】 複数m個の入力ポートから複数n個の出力ポートへパケットをスイッチングするパケットスイッチ方式において、複数k個のバニヤン網を少なくともn本の並列ラインによってリング型に接続し、前記入力ポートを入力モジュールを介して各バニヤン網の入力部に分散して接続し、前記出力ポートを出力モジュールを介して各バニヤン網の対応する出力部に設けたミスルートタグチェック部に接続し、前記バニヤン網を構成する単位スイッチを、パケットが希望する出力部に出せない場合にはミスルートタグをセットし、このパケットを空いている出力部に出すように構成し、前記ミスルートタグチェック部を、入力されるパケットのミスルートタグをチェックし、これがセットされている場合には次段のバニヤン網側へ、セットされていない場合には前記出力モ

ジュール側へパケットのスイッチングを行うように構成し、前記入力モジュールを、前記入力ポートからのパケットおよびミスルートタグチェック部からのパケットを格納し、これらのパケットを選択的に次段のバニヤン網に入力させるように構成したことを特徴とするパケットスイッチ方式。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、複数の入力ポートから複数の出力ポートへパケットをスイッチングするパケットスイッチ方式に係るものである。

【0002】

【従来の技術】共通バス方式はIPルータとして現在最も主流となっている方式であり、ほとんどの商用IPルータで用いられている。この共通バス方式のパケットス

イチの簡略図を図1に示す。各入力ポートはラウンドロビン方式によって共通バスの使用权を獲得し、この共通バス上にパケットを送り出す(ブロードキャストする)。出力ポートは、このブロードキャストされたパケットのうち自分宛のものだけを受け取る。したがって、ルータ内のパケット損失を0にするためには、共通バスに(各ポートの伝送速度)×(ポート数)以上の動作速度が必要になる。

【0003】共通バス方式は、可変長パケットの交換が容易であり、ルータを構成する回路素子数を少なくすることが可能であるなどの利点があるものの、入出力ポート数に比例して共通バスの伝送速度を大きくしなければならぬ。したがって、この共通バス方式で交換容量が1Tbps(Tera bit per second)を越えるような大容量スイッチを実現することは、素子の物理的限界から非常に困難である。

【0004】パニヤン網は図2に示すように単位スイッチを多段接続することにより大容量スイッチを実現する方式であり、各入力ポートからすべての出力ポートに対して1本ずつの経路が存在している。

【0005】パニヤン網は網内の各単位スイッチが並列的に動作するため、高速大容量化を実現する方法として適している。しかしその反面、図3に示すように、パケットの同一入線によって内部ブロッキングが発生してしまう問題点がある。なお、図2、3では単位スイッチとして2×2単位スイッチを用いた場合を取り上げたが、4×4、8×8、16×16などの単位スイッチを用いても同様の接続が可能である。

【0006】タンデムパニヤン網はパニヤン網における内部ブロッキングを低減するための一手法であり、図4に示すように複数のパニヤン網を一行に配置したスイッチである。タンデムパニヤン網ではパケットをすべて1段目のパニヤン網(図4の一番左のパニヤン網)に入力させるが、もしパニヤン網で内部ブロッキングが発生した場合には、そのパケットのミスルートタグをセットし、空いている出力ポートに出力させる。そして1段目のパニヤン網を通過後に続けて2段目のパニヤン網に入力させる。一方、1段目のパニヤン網で無事に希望する出力ポートに出られたパケットは、2段目以降のパニヤン網を通らずにそのまま出力バッファに入れられる。

【0007】以上のように、内部ブロッキングが発生したパケットは希望する出力ポートに出られるまで次段のパニヤン網に入力することができるため、パニヤン網で発生する大部分の内部ブロッキングを回避することが可能である(タンデムパニヤン網において内部ブロッキングが発生するのは1段目から最終段まですべてのパニヤン網でミスルートタグがセットされた場合のみである)。

【0008】

【発明が解決しようとする課題】タンデムパニヤン網方

式は、パニヤン網における内部ブロッキングを低減するための手段として有効である。しかし、タンデムパニヤン網方式では、はじめすべてのセルを1段目のパニヤン網に入力させ、1段目で内部ブロッキングが発生したパケットは引き続き2段目以降のパニヤン網に入力させてゆく。したがって入力トラヒックが大きい場合には、1段目のパニヤン網での内部ブロッキングの発生率が高いため、2段目以降のパニヤン網に入力するパケットの確率がかなり高くなってしまい、パケットのスイッチング遅延および遅延ゆらぎの増加を引き起こす。さらに、遅延ゆらぎの増加に伴い、パケットの順序狂い率も増加してしまうという問題がある。

【0009】本発明の目的は、上述した従来のタンデムパニヤン網方式における問題を解決し、内部ブロッキングの発生率を抑えることにより、パケットのスイッチング遅延および遅延揺らぎが減少するように適切に構成した、大容量可変長パケット交換に適したパケットスイッチ方式を提供することにある。

【0010】

【課題を解決するための手段】上記の目的を達成するため、本発明によるパケットスイッチ方式は、複数k個のパニヤン網を少なくともn本の並列ラインによってリング型に接続し、前記入力ポートを入力モジュールを介して各パニヤン網の入力部に分散して接続し、前記出力ポートを出力モジュールを介して各パニヤン網の対応する出力部に設けたミスルートタグチェック部に接続し、前記パニヤン網を構成する単位スイッチを、パケットが希望する出力部に出せない場合にはミスルートタグをセットし、このパケットを空いている出力部に出すように構成し、前記ミスルートタグチェック部を、入力されるパケットのミスルートタグをチェックし、これがセットされている場合には次段のパニヤン網側へ、セットされていない場合には前記出力モジュール側へパケットのスイッチングを行うように構成し、前記入力モジュールを、前記入力ポートからのパケットおよびミスルートタグチェック部からのパケットを格納し、これらのパケットを選択的に次段のパニヤン網に入力させるように構成したことを特徴とする。

【0011】本発明によるパケットスイッチ方式においては、図5に示すように複数のパニヤン網をリング状に接続し、入力ポートを各パニヤン網に分散している。これにより、入力トラヒックが分散され、各パニヤン網への見かけ上の入力トラヒック量を低減することができる。例えばパニヤン網の平面数がNであるとすると、タンデムパニヤン網方式では総トラヒック量が1段目のパニヤン網に入力するのに対し、本方式では各パニヤン網にその1/Nのトラヒックを入力させればよいことになる。この分散効果により、各パニヤン網での入力トラヒック量が見かけ上減少するため、1段目のパニヤン網での内部ブロッキング発生率がタンデムパニヤン網と比較

して減少し、パケットのスイッチング遅延および遅延ゆらぎを小さくすることができる

【0012】

【発明の実施の形態】図6は、本発明のパケットスイッチ方式によるパケットスイッチの一実施形態を示すブロック図である。このパケットスイッチ1は、入力ポート $I_h$  ( $h = 1, 2, \dots, m$ )に結合した入力モジュール $IM_h$ と、バニヤン網 $B_i$  ( $i = 1, 2, \dots, k$ )4と、ミスルートタグチェック部 $MT_{h,i}$ と、出力ポート $O_j$  ( $j = 1, 2, \dots, n$ )に結合した出力モジュール $OM_j$ とを具える。この図に示す実施形態においては、 $m = n = 8$ 、 $k = 4$ としているが、他の数としてもよいことはもちろんである。

【0013】バニヤン網 $B_i$ は、すでに図2に示したように、単位スイッチを多段接続することにより大容量スイッチを実現する。各単位スイッチは、通常のパニヤン網とは異なり、パケットが希望する出力ポート $O_j$ に通じる出力部に出せない場合にはミスルートタグをセットし、空いている出力部にこのパケットを出す。図2では、 $2 \times 2$ 単位スイッチを用いた場合のパニヤン網を示しているが、 $4 \times 4$ 、 $8 \times 8$ 、 $16 \times 16$ 、または $32 \times 32$ 単位スイッチを用いても同様の多段接続が可能である。バニヤン網 $B_i$ の入力部および出力部の数は、少なくとも出力ポートの数 $n$ だけなければならない。これらのバニヤン網を少なくとも $n$ 本のラインでリング状に接続する。この図においてパケットは、リング状に接続されたバニヤン網を反時計方向に進む。

【0014】ミスルートタグチェック部 $MT_{h,i}$ は、 $1 \times 2$ スイッチの機能を持つ。入力パケットのミスルートタグをチェックし、これがセットされている場合には次段のパニヤン網 $B_i$ 側へ、セットされていない場合には出力モジュール $OM_j$ 側へパケットのスイッチングを行う。

【0015】入力モジュール $IM_h$ は、図8に示すように2つのFIFO(First In First Out)バッファ11および12とセクタ13とを具える。2つのFIFOバッファ11および12はそれぞれ、入力ポート $I_h$ またはミスルートタグチェック部 $MT_{h,i}$ から転送されてきたパケットを格納する。セクタ13は、2つのFIFOバッファ11および12の一方からパケットを選択し、次段のパニヤン網 $B_i$ に入力させる。本実施形態では、入力ポート $I_h$ からのパケットを低優先、ミスルートタグチェック部 $MT_{h,i}$ からのパケットを高優先と定義しているため、セクタ13は、ミスルートタグチェック部 $MT_{h,i}$ からのパケットがFIFOバッファ11に存在する場合にはそのパケットを優先的に次段のパニヤン網 $B_i$ に入力させる。

【0016】また、入力ポート $I_h$ 、ミスルートタグチェック部 $MT_{h,i}$ の双方から100%の入力トラフィックが存在する場合には、FIFOバッファ11および12

がオーバーフローを起こす可能性がある。したがって、入力モジュール $IM_h$ をはじめとするバニヤン網のリング内部のすべての装置を、入力ポート $I_h$ の2倍の伝送速度で動作させる。

【0017】出力モジュール $OM_j$ は図9に示すように、タイミング調整バッファ21、共通バス22、制御部23および出力バッファ24から構成されている。タイミング調整バッファ21は、各バニヤン網 $B_i$ から出力したパケットを格納する。制御部23は共通バス22を管理し、各タイミング調整バッファ21にラウンドロビン方式によって共通バス22の使用権を与える。タイミング調整バッファ21に格納されたパケットは、この使用権を獲得した時に出力バッファ24に移される。

【0018】パケットを入力ポート $I_h$ から入力し、スイッチングして、希望する出力ポート $O_j$ に出力するまでの流れは次の通りである。

【0019】まずパケットは入力ポート $I_h$ から入力され、入力モジュール $IM_h$ 内のFIFOバッファ12に格納される。入力モジュール $IM_h$ では、入力ポート $I_h$ からのパケットよりもミスルートタグチェック部 $MT_{h,i}$ からのパケットを優先するため、このパケットはミスルートタグチェック部 $MT_{h,i}$ からのパケットがFIFOバッファ11内に存在しない場合のみセクタ13を通してバニヤン網 $B_i$ に入力される。

【0020】バニヤン網 $B_i$ に入力したパケットは、単位スイッチ間をスイッチングしてゆくが、もし希望する出力ポート $O_j$ に通じる出力部が使用中の場合には、パケットヘッダ内のミスルートタグをセットされ、使用されていない出力部に出される。バニヤン網 $B_i$ を通過したパケットは続いてミスルートタグチェック部 $MT_{h,i}$ に渡される。

【0021】ミスルートタグチェック部 $MT_{h,i}$ では、パケットにミスルートタグがセットされているかどうかをチェックし、セットされている場合には引き続き次段のパニヤン網 $B_i$ に入力させる。一方、ミスルートタグがセットされていない場合には、出力モジュール $OM_j$ に転送される。

【0022】出力モジュール $OM_j$ に入ったパケットはタイミング調整バッファ21に格納され、共通バス22の使用権を獲得するまで待機する。そして共通バス使用権が回ってきたら、パケットは共通バス22を通じて出力バッファ24に入れられ、出力ポート $O_j$ を経て本パケットスイッチ1から出力される。

【0023】従来のタンデムバニヤン網方式のパケットスイッチと本方式のパケットスイッチとをシミュレーションにより評価した結果を示す。ここで、使用するバニヤン網を、図10に示すような $4 \times 4$ 単位スイッチを用いた $64 \times 64$ バニヤン網とし、ポート伝送速度を155.52Mbpsとし、入力負荷を90%とし、IPデータグラムを21~1500octetの可変長とし

た。

【0024】図11にバニヤン網の平面数 - パケット損失率特性を示す。図11では、バニヤン網の段数の増加に伴ってタンデムバニヤン網方式、本方式ともにパケット損失率が減少してゆく様子が示されているが、特に本方式でははじめからトラヒックの分散が行われているため、タンデムバニヤン網方式と比較してパケット損失率が小さくなっている。また、本方式はタンデムバニヤン網方式と比較してパケット損失率の収束が早いいため、少ないバニヤン網の段数で高性能化を図ることができる。これは設計の立場に立った場合、大きな長所となるであろう。

【0025】一方、図12では、6段のバニヤン網を用いたタンデムバニヤン網方式と本方式に対して、入力パケットが何段目のバニヤン網で出力できたかを示している。

【0026】図12では、本方式ではほとんどのパケットが3段目以内のバニヤン網でスイッチングを完了しているのに対し、タンデムバニヤン網方式では5段目のバニヤン網に入力するパケットが10%近くもあり、本方式と比較して、遅延および遅延ゆらぎが大きくなっていることがわかる。

【0027】

【発明の効果】本発明によれば、入力ポートをあらかじめ分散しておくことによって、入力トラヒックの分散を図り、各バニヤン網への見かけ上の入力トラヒック量を低減することができる。これにより、例えばバニヤン網の平面数がNであるとすると、タンデムバニヤン網方式では総トラヒック量が1段目のバニヤン網に入力するのに対し、本方式では各バニヤン網にその1/Nのトラヒックを入力させればよいことになる。この分散効果により、各バニヤン網での入力トラヒック量が見かけ上減少するため、1段目のバニヤン網での内部ブロッキング発生率がタンデムバニヤン網と比較して減少し、パケットのスイッチング遅延および遅延ゆらぎを小さくすることができる。したがって本発明によれば、大容量可変長パケット交換に適したパケットスイッチ方式が提供される。

【図面の簡単な説明】

【図1】従来の共通バス方式のパケットスイッチを示す線図である。

【図2】バニヤン網を示す線図である。

【図3】バニヤン網における内部ブロッキングを説明する線図である。

【図4】タンデムバニヤン網を示す線図である。

【図5】本発明による、バニヤン網をリング状に接続し

たパケットスイッチを図式的に示す線図である。

【図6】本発明のパケットスイッチ方式によるパケットスイッチの一実施形態の構成を示すブロック図である。

【図7】図6のパケットスイッチにおけるミスルートタグチェック部の構成を示すブロック図である。

【図8】図6のパケットスイッチにおける入力モジュールの構成を示すブロック図である。

【図9】図6のパケットスイッチにおける出力モジュールの構成を示すブロック図である。

【図10】4×4単位スイッチを用いた64×64バニヤン網を示す線図である。

【図11】バニヤン網の平面数 - パケット損失率特性を示すグラフである。

【図12】入力パケットが何段目のバニヤン網で出力できたかを示すグラフである。

【符号の説明】

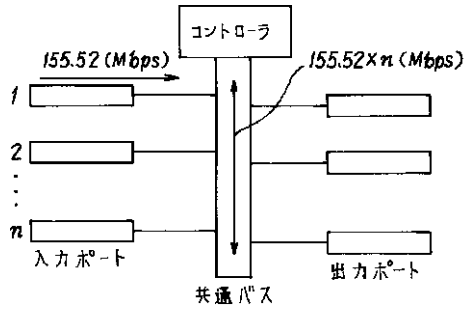
- 1 パケットスイッチ
- 11、12 FIFOバッファ
- 13 セレクタ
- 21 タイミング調整バッファ
- 22 共通バス
- 23 制御部
- 24 出力バッファ
- $I_h$  入力ポート
- $IM_h$  入力モジュール
- $B_i$  バニヤン網
- $MT_{h,i}$  ミスルートタグチェック部
- $OM_j$  出力モジュール
- $O_j$  出力ポート

【要約】

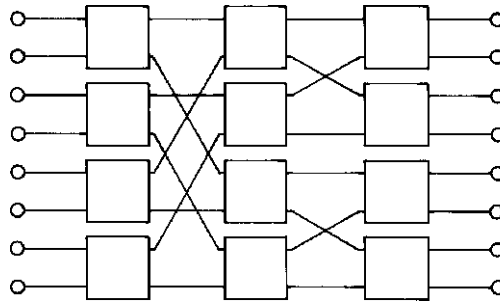
【課題】 パケットのスイッチング遅延および遅延揺らぎが減少するように適切に構成した大容量可変長パケット交換に適したパケットスイッチ方式を提供する。

【解決手段】 複数k個のバニヤン網を少なくともn本の並列ラインによってリング型に接続し、入力ポートを入力モジュールを介して各バニヤン網の入力部に分散して接続し、出力ポートを出力モジュールを介して各バニヤン網の対応する出力部に設けたミスルートタグチェック部に接続し、前記バニヤン網を構成する単位スイッチを、パケットが希望する出力部に出せない場合にはミスルートタグをセットし、このパケットを空いている出力部に出すように構成し、前記ミスルートタグチェック部を、入力されるパケットのミスルートタグをチェックし、これがセットされている場合には次段のバニヤン網側へ、セットされていない場合には前記出力モジュール側へパケットのスイッチングを行うように構成した。

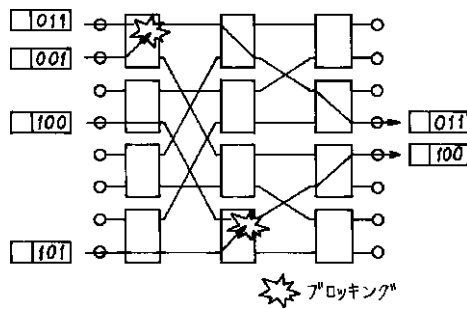
【図1】



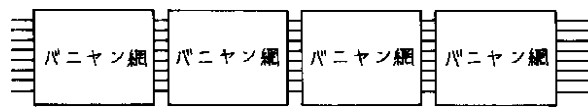
【図2】



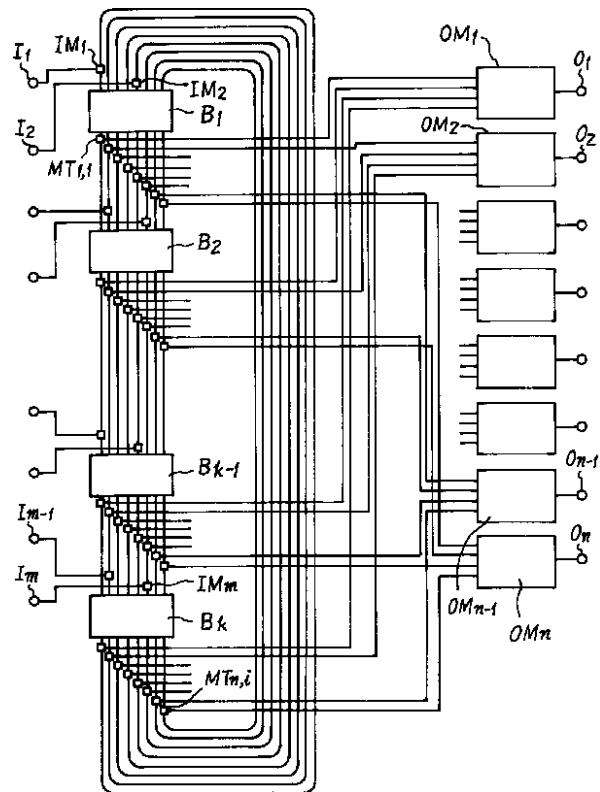
【図3】



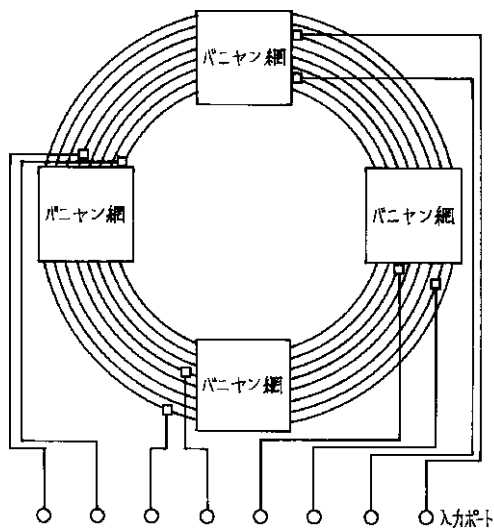
【図4】



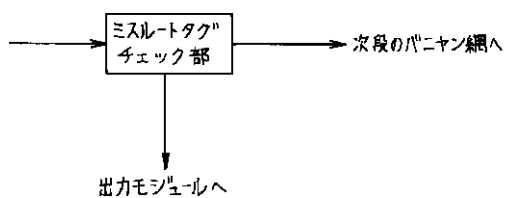
【図6】



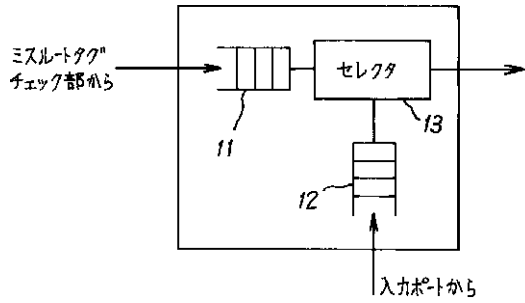
【図5】



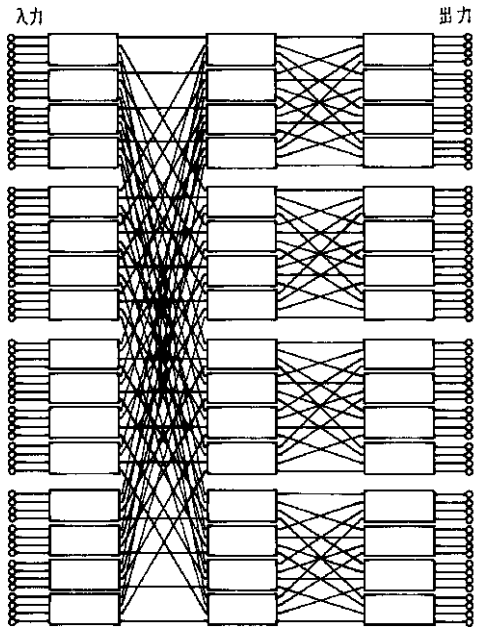
【図7】



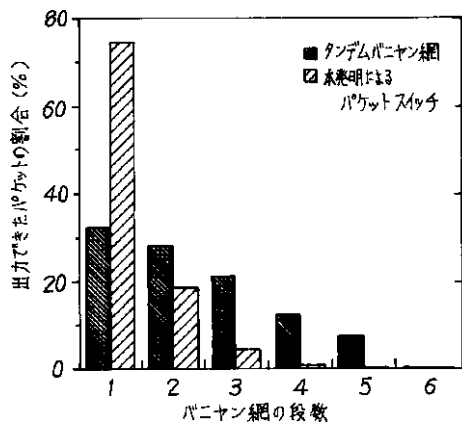
【図8】



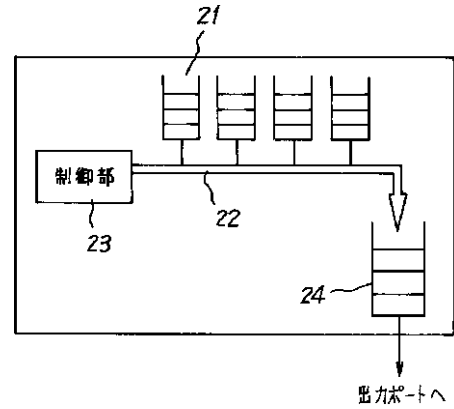
【図10】



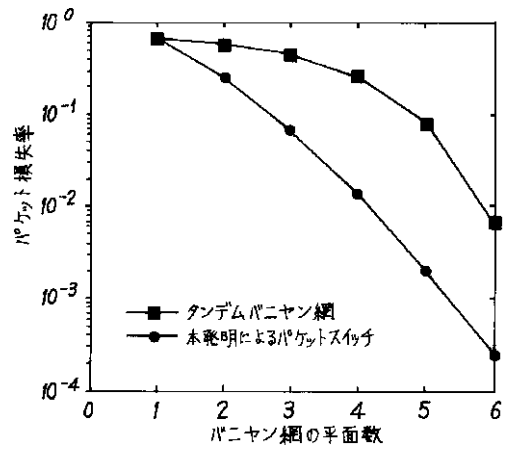
【図12】



【図9】



【図11】



フロントページの続き

(58)調査した分野(Int.Cl.<sup>6</sup>, DB名)

H04L 12/28

H04L 12/56