

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3591420号
(P3591420)

(45) 発行日 平成16年11月17日(2004.11.17)

(24) 登録日 平成16年9月3日(2004.9.3)

(51) Int. Cl.⁷

F I

G06F 13/00	G06F 13/00	353Q
G06F 12/08	G06F 12/08	U
G06F 12/12	G06F 12/12	A
H04L 12/56	H04L 11/20	102Z

請求項の数 13 (全 18 頁)

(21) 出願番号 特願2000-106744 (P2000-106744)
 (22) 出願日 平成12年4月7日(2000.4.7)
 (65) 公開番号 特開2001-290718 (P2001-290718A)
 (43) 公開日 平成13年10月19日(2001.10.19)
 審査請求日 平成13年3月9日(2001.3.9)

(73) 特許権者 000004237
 日本電気株式会社
 東京都港区芝五丁目7番1号
 (74) 代理人 100111729
 弁理士 佐藤 勝春
 (72) 発明者 村瀬 勉
 東京都港区芝五丁目7番1号
 日本電気株式会社内
 審査官 石井 茂和

最終頁に続く

(54) 【発明の名称】 ルータにおけるキャッシュテーブル管理装置およびプログラム記録媒体

(57) 【特許請求の範囲】

【請求項1】

全エントリに対して最長一致検索を行うためのフォワーディングテーブルと、
 全エントリの一部のエントリに対して最長一致検索を行うためのキャッシュテーブルとを
 備え、

前記フォワーディングテーブルにおいてエントリを前記キャッシュテーブルに入れるとき
 には、そのエントリを親とする全エントリを、また前記キャッシュテーブルからエントリ
 を出すときには、そのエントリの親となる全エントリを、当該エントリの使用状況の評価
 値に対する閾値および削除あるいは移動されるエントリの領域対応の優先度に応じて選択
 することを特徴とするルータにおけるキャッシュテーブル管理装置。

10

【請求項2】

全エントリに対して最長一致検索を行うためのフォワーディングテーブルあるいは全エン
 トリの一部のエントリに対して最長一致検索を行うためのキャッシュテーブル内に仮想的
 な複数の領域を設定する手段と、

該領域内では、それぞれ固有に、削除あるいは移動対象を決めるためのエントリ群の使用
 状況の評価値と、少なくとも削除対象に選択される可否を決める閾値と選択された場合に
 実際に削除あるいは移動されるエントリの該領域に応じて決まる優先度の一つを持つ手段
 と、

該閾値および優先度に応じて入れ替え対象のエントリ群を選択する手段とを持つことを特
 徴とするルータにおけるキャッシュテーブル管理装置。

20

【請求項3】

各エントリあるいは各エントリ群と各仮想的な領域を関連付ける手段、または各エントリあるいは各エントリ群にそれぞれ優先度を関連付ける手段、または各エントリあるいは各エントリ群にそれぞれ閾値を関連付ける手段の内の少なくとも一つの手段を備えたことを特徴とする請求項2に記載のルータにおけるキャッシュテーブル管理装置。

【請求項4】

特定の時間区間において、エントリ群を構成する各エントリが持つ使用状況の評価値に対して、該使用状況を全てのエントリに対してあるいは一部のエントリに対して組み合わせた値あるいは統計値を各エントリ群の使用状況の評価値として用いる手段を備えたことを特徴とする請求項2に記載のルータにおけるキャッシュテーブル管理装置。

10

【請求項5】

プレフィクスビット列およびプレフィックス長から成るアドレスの集合を示す情報と、このアドレスの集合に対するパケットの出方路を示す情報と、優先度情報との組が登録されたエントリを複数有し最長一致検索されるフォワーディングテーブルと、
入れ替えるべきエントリを親とする前記フォワーディングテーブルの全エントリが書き込まれ、またエントリが削除・移動されるときには、その親も削除・移動されるキャッシュテーブルと、

前記フォワーディングテーブルの全エントリについて、その内容に加えてヒット情報が登録され、前記フォワーディングテーブルまたは前記キャッシュテーブルがヒットしたときにそのエントリの内容が更新されるヒット状況データベースと、

20

パケットが入力されるとそのパケットヘッダから宛先ネットワークアドレスを取り出し、このアドレスをキーにして前記キャッシュテーブルまたは前記フォワーディングテーブルを検索し、得られた出方路にパケットを転送するパケット処理回路と、

前記キャッシュテーブルに登録されている全エントリの前記優先度情報を含み、前記エントリ群入れ替えを行うときには、該優先度情報と前記ヒット状況データベースからの情報とを考慮して入れ替えるべきエントリ群を選択するエントリ選択装置と備えたことを特徴とするルータにおけるキャッシュテーブル管理装置。

【請求項6】

前記エントリ選択装置は、

前記ヒット状況データベースからエントリ群全ての情報を受け取り、この全エントリから、エントリ固有の優先度である領域、エントリ使用状況の評価値および閾値の代表値を決めるエントリ群代表値回路と、

30

前記キャッシュテーブルに既に存在するエントリ群に関して、そのエントリ群とその代表値を組にして表に持っており、その表の各エントリ群の代表値と、前記エントリ群代表値回路から送られてくる代表値とを比較して、キャッシュテーブルに既に存在するエントリ群がどのような優先順位にあるかを送出する前記領域、エントリ使用状況の評価値および閾値対応の比較回路と、

該比較回路から送られてきた優先順位データを基にエントリの最終的な入替え順位を決定する調停回路と、

前記キャッシュテーブルの空き状況を管理しており、当該エントリ群を追加することのできる空きがあるかどうかを調べ、空きが無い場合には、前記調停回路が決定した順位の最も低いエントリが所属するエントリ群から前記キャッシュテーブルのエントリを消去して十分な空きを作り、また削除したエントリ群を前記表からの削除のために前記比較回路に送るエントリ決定送付回路とから成ることを特徴とする請求項5に記載のルータにおけるキャッシュテーブル管理装置。

40

【請求項7】

前記エントリ使用状況の評価値として、ヒット回数またはヒット時刻またはその両者を採用することを特徴とする請求項2、請求項4または請求項6に記載のルータにおけるキャッシュテーブル管理装置。

【請求項8】

50

前記各代表値は、閾値と領域についてはエン트리中の最も高い値、ヒット回数は平均値、ヒット時刻は最近のものとする特徴とする請求項 7 に記載のルータにおけるキャッシュテーブル管理装置。

【請求項 9】

前記調停回路は、0 ~ 1 の乱数を発生し、その乱数を 2 等分して、1 / 2 づつをヒット回数とヒット時刻に割り当て、割り当てられた属性の順位を最終的な順位とする特徴とする請求項 7 または請求項 8 に記載のルータにおけるキャッシュテーブル管理装置。

【請求項 10】

前記エントリの優先度を受信パケットから得られる情報に応じて設定する手段を設けたことを特徴とする請求項 2 ないし請求項 9 のいずれかに記載のルータにおけるキャッシュテーブル管理装置。

10

【請求項 11】

前記フォーディングテーブル内の優先度は、ポリシーサーバにより変更可能としたことを特徴とする請求項 5 ないし請求項 10 に記載のルータにおけるキャッシュテーブル管理装置。

【請求項 12】

プレフィクスビット列およびプレフィックス長から成るアドレスの集合を示す情報と、このアドレスの集合に対するパケットの出方路を示す情報と、優先度情報との組が登録されたエントリを複数有し最長一致検索されるフォーディングテーブルと、

入れ替えるべきエントリを親とする前記フォーディングテーブルの全エントリが書き込まれ、またエントリが削除・移動されるときには、その親も削除・移動されるキャッシュテーブルと、

20

前記フォーディングテーブルの全エントリについて、その内容に加えてヒット情報が登録され、前記フォーディングテーブルまたは前記キャッシュテーブルがヒットしたときにそのエントリの内容が更新されるヒット状況データベースとを備えたルータを、

パケットが入力されるとそのパケットヘッダから宛先ネットワークアドレスを取り出し、このアドレスをキーにして前記キャッシュテーブルまたは前記フォーディングテーブルを検索し、得られた出方路にパケットを転送するパケット処理回路と、

前記キャッシュテーブルに登録されている全エントリの前記優先度情報を含み、前記エントリ群入れ替えを行うときには、該優先度情報と前記ヒット状況データベースからの優先度情報とを比較して入れ替えるべきエントリ群を選択するエントリ選択装置として機能させるためのプログラムを記録したコンピュータ読み込み可能なプログラム記録媒体。

30

【請求項 13】

前記エントリ選択装置は、

前記ヒット状況データベースからエントリ群全ての情報を受け取り、この全エントリから、エントリ固有の優先度である領域、エントリ使用状況の評価値およびこれらの各々についての閾値の代表値を決めるエントリ群代表値手段と、

前記キャッシュテーブルに既に存在するエントリ群に関して、そのエントリ群とその代表値を組にして表に持っており、その表の各エントリ群の代表値と、前記エントリ群代表値回路から送られてくる代表値とを比較して、キャッシュテーブルに既に存在するエントリ群がどのような優先順位にあるかを送出する前記領域、エントリ使用状況の評価値および閾値対応の比較手段と、

40

該比較回路から送られてきた優先順位データを基にエントリの最終的な入替え順位を決定する調停手段と、

前記キャッシュテーブルの空き状況を管理しており、当該エントリ群を追加することのできる空きがあるかどうかを調べ、空きが無い場合には、前記調停回路が決定した順位の最も低いエントリが所属するエントリ群から前記キャッシュテーブルのエントリを消去して十分な空きを作り、また削除したエントリ群を前記表からの削除のために前記比較回路に送るエントリ決定送付手段とから成ることを特徴とする請求項 12 に記載のプログラム記録媒体。

50

【発明の詳細な説明】**【0001】****【発明の属する技術分野】**

本発明は、ルータにおけるキャッシュテーブル管理、特にフォーディングテーブルとキャッシュテーブルとの間でエントリを群単位で出し入れするときのキャッシュテーブルの管理に関する。

【0002】**【従来の技術】**

複数のネットワークを接続するルータは、異なるネットワークを越えて行われる通信を媒介するため、通信情報の固まり（パケット）の転送（フォーディング）を行わなければならない。ルータは、接続されたネットワークから入力されたパケットをネットワークインタフェース（以下、「インタフェース」と記す）で受信し、そのパケットのヘッダ部分に含まれる送信先アドレスを基に、各ルータの持つフォーディングテーブルを検索して、パケットを送信すべきインタフェースの番号を決定して、パケットのフォーディングを行う。

10

【0003】

この場合、ネットワークの数やネットワークに接続される端末の数が多いと、フォーディングテーブルの規模も大きくなり、フォーディングが低速化する。そこで、従来から、キャッシュテーブルを使用して、フォーディング、ひいてはパケット処理を高速化した技術が知られている。

20

【0004】

例えば、特開平6 - 261078号公報に記載された「テーブル検索方法及びルータ装置」では、各インタフェースはルーティングテーブル及びルーティングテーブル検索用のキャッシュメモリを備える。ルーティングテーブルには、特定のネットワークがそのルータのインタフェースを介して到達可能であるか、また、そのインタフェースを介してパケットを送出したとき、送出ネットワーク内での宛先をどこにすればよいかの情報が書かれている。すなわち、ルーティングテーブルの各エントリは宛先ネットワークアドレス、宛先ネットワークアドレスに対するマスク、送出ネットワーク内での宛先ネットワークアドレス及び送出インタフェース番号を格納する。

【0005】

キャッシュメモリはCAM (Content Addressable Memory) と対応テーブルとから成る。CAMには、ルーティングテーブル上の該当エントリへのポインタが得られた宛先ネットワークアドレスが登録され、対応テーブルは、CAMのエントリにリンクしているエントリを持ち、そこにルーティングテーブルの該当エントリへのポインタが登録される。そして、同じネットワークアドレスに対して再度ルーティングテーブルを検索する場合には、まず、CAMに対して宛先ネットワークアドレスの照合を行い、該当アドレスが登録されている場合は、CAMの該当ネットワークアドレスが登録されているエントリにリンクしている対応テーブルのエントリからルーティングテーブルへのポインタを得る。この結果、検索回数の多い宛先ネットワークアドレスがキャッシュメモリに登録されることになるのでテーブル検索時の高速化が図れ、また、対応テーブルを設けるのでCAMのエントリサイズを小さくすることができる。

30

40

【0006】

しかし、この従来技術ではキャッシュメモリにおけるテーブル検索は、検索キーとしてホストアドレスを用いた完全一致検索方式を採用しているため、検索に時間がかかりすぎたり、ヒット率を上げようとすると大容量のキャッシュメモリを必要とするという問題点がある。

【0007】

そこで、小林ほかにより特願平11 - 033287において、最長一致検索のできるフォーディングテーブル（以下、「フルテーブル」と記す）と、最長一致検索のできるキャッシュテーブルを設けた「キャッシュを有するルータの方路決定装置及びプログラムを記

50

録した機械読み取り可能な記録媒体」が提案されている。

【0008】

この技術について、先ず、それぞれのテーブルの構成と検索方法について説明する。第1の構成例においては、フルテーブルには、ルーティングに必要な全ての経路を含むエントリを入れ、キャッシュテーブルにはフルテーブルの一部のエントリを入れておく。また、第2の構成例においては、キャッシュテーブルにはフルテーブルの一部のエントリを入れておき、フルテーブルにはキャッシュテーブルにあるエントリを除いた全ての必要なエントリを入れておく。

【0009】

キャッシュヒット率を高めるためには、キャッシュテーブルとフルテーブルでエントリを入れたり出したりの操作を行う必要があるが、第1の構成例においては、フルテーブルからキャッシュテーブルへはエントリをコピーし、キャッシュテーブルからフルテーブルへは、データを廃棄（削除）することになる。一方、構成例2においては、フルテーブルからキャッシュテーブルへはエントリを移動し、キャッシュテーブルからフルテーブルへも、やはりデータを移動することになる。以下の説明では、第1の構成例について、コピーと削除という動作で説明するが、第2の構成例の場合には、第1の構成例におけるコピーが移動に、第1の構成例における削除が移動という操作に相当する。なお、コピーすることを単に「入れる」と呼ぶ。

【0010】

フルテーブルは大量のデータを検索するので検索時間は長い。キャッシュテーブルは容量は小さいが高速に検索できる。先ず、検索要求があれば、まずキャッシュを最長一致検索で検索する。キャッシュでエントリが見つければ（ヒットするという）、検索はここで終了する。しかし、見つからない場合（ミスヒット）には、引き続きフルテーブルを最長一致検索する。フルテーブルには全てのエントリが入っているので、必ずヒットし検索は終了する。最長一致検索とは、特願平11-033287にも記載されているように、プレフィクス（Prefix）が一致するものを候補とし、その候補のエントリから最も長いプレフィクス長をもつエントリをヒットエントリとする検索をいう。

【0011】

例えば、（A）100.120.0.0/16と（B）100.0.0.0/8と（C）100.120.140.0/24と（D）100.120.180.0/24というプレフィクスで表わされるネットワークアドレスを有する4つのエントリがあるときに、検索キーとして、ネットワークアドレス「100.120.140.5」が与えられたとする。なお、100.120.0.0等はプレフィクスビットと呼ばれ、ドットで区切られた数字はそれぞれ8ビットで10進数を表現する。この8ビットで構成される10進数が0のときは任意数（don't care）を意味する。また、/の後の数はプレフィクス長を意味する。与えられたネットワークアドレス「100.120.140.5」は、（A）、（B）、（C）のいずれにもプレフィクスが一致するため、（A）、（B）、（C）の3つとも検索結果の候補になる。しかし、最長一致検索では、候補の中から最も長いプレフィクス長のエントリを選択することになるため、この場合、プレフィクス長=24が最も長いので、検索の結果、（C）100.120.140.0/24のエントリがヒットする。

【0012】

次に、キャッシュテーブルに入れるエントリについて述べる。この特願平11-033287に提案の発明では、キャッシュテーブルに出し入れするエントリは、注目している単独のエントリのコピーや削除ではなく、所定の規則を持って複数あるいは場合によっては単独のエントリで移動することになる。すなわち、同じプレフィクスを持つエントリのプレフィクス長の短いものを親、長いものを子と呼ぶときに、キャッシュインのときは自分の全ての子を必ずいっしょにキャッシュに入れ、キャッシュアウトのときは自分の全ての親を必ずいっしょにキャッシュから取り出すのである。以下の説明では、この規則に従って移動すべき複数のエントリをエントリ群と呼ぶ。

10

20

30

40

50

【 0 0 1 3 】

このエントリ群の作り方は、具体例で示せば次のようである。例えば、前記 (A) , (B) , (C) および (D) の 4 つのエントリにおいて、(A) 1 0 0 . 1 2 0 . 0 . 0 / 1 6 というエントリをフルテーブルからキャッシュに移動させるときは、(A) , (C) および (D) がエントリ群になる。また、(A) をキャッシュテーブルからフルテーブルに移動させるときには、(A) のみがエントリ群になる。

【 0 0 1 4 】

また、キャッシュからフルテーブルへの方にエントリを移動する場合のエントリ選択方法として L R U (L e a s t R e c e n t l y U s e d) という方法が記述されている。L R U とは、エントリが参照された順序において最も古いものを対象とする方法である。これは次のような理由による。上述の (A) 1 0 0 . 1 2 0 . 0 . 0 / 1 6 と (C) 1 0 0 . 1 2 0 . 1 4 0 . 0 / 2 4 と (D) 1 0 0 . 1 2 0 . 1 8 0 . 0 / 2 4 とでは出方路としてのインタフェース番号が異なっているときに、もし、(A) 1 0 0 . 1 2 0 . 0 . 0 / 1 6 だけをキャッシュインしておく、「1 0 0 . 1 2 0 . 1 4 0 . 5」が検索キーとして与えられると、最長一致検索の結果として (C) 対応のインタフェース番号が得られるべきところ、それとは異なる (A) 対応のインタフェース番号がキャッシュにおいて得られることになるからである。また、(C) をキャッシュアウトするとき (A) をキャッシュに残しておく、「1 0 0 . 1 2 0 . 1 4 0 . 5」が検索キーとして与えられると、(C) 対応のインタフェース番号と異なる (A) 対応のインタフェース番号がキャッシュヒットするからである。

【 0 0 1 5 】

【 発明が解決しようとする課題 】

ところで、キャッシュシステムでは、キャッシュに入れるべきデータはフルテーブル全体の一部分であるため、フルテーブルからエントリを選択してキャッシュに入れる必要がある。このとき、高速検索を行うためにはキャッシュでヒットする確率を上げなければならないので、キャッシュに入れるべきエントリの選択が重要である。また、キャッシュのエントリ内容を動的に変更する場合には、キャッシュから出すべきエントリの選択も重要である。

【 0 0 1 6 】

このような選択方法としては、上述の 2 つの先例におけるフォーワーディング技術では L R U 方式が採用されている。しかしながら、このような L R U は、単独のエントリを出し入れする場合にあっては効果的であっても、特願平 1 1 - 0 3 3 2 8 7 におけるようにエントリ単独ではなく、エントリ群、すなわち関連の複数のエントリを束にしてキャッシュに出し入れするような制約の下では効果的とはいえないという第 1 の問題点がある。例えば、図 1 2 に例示するように、使用時系列「3 : 0 0」「2 : 0 0」のプレフィクス (A) 1 0 0 . 1 2 0 . 0 . 0 / 1 6 と、使用時系列「1 : 0 0」, 「0 : 3 0」のプレフィクス (C) 1 0 0 . 1 2 0 . 1 4 0 . 0 / 2 4 を含むキャッシュから (C) を出すときには、その親である (A) も出すことになるが、そうすると (A) よりも長い間使用されていない (D) がキャッシュに残ることになるのである。

【 0 0 1 7 】

また、L R U のようなアルゴリズムを採用すると、データの選択が参照度に従って機械的になされてしまい、意図的にデータの重要度、具体的には、ポリシー上または料金体系上から回線接続の容易性の度合をつけることができないという第 2 の問題点もある。例えば、エントリ別に重要度が異なるときにおいても、従来の L R U では、キャッシュへのエントリ出し入れを全てを平等に扱ってしまい、データの重要度に応じた扱いができない。

【 0 0 1 8 】

本発明は、このような問題を解決するために成されたものであって、その目的は、キャッシュを有するルータにおいて、キャッシュのヒット確率を上げることによって、キャッシュ検索より時間のかかるフルテーブル検索を行う確率を下げ、ルータのパケット処理時間を減らし、ルータのパケット処理能力を向上させることのできるキャッシュに入出力するエ

10

20

30

40

50

ントリ群の管理装置およびプログラム記録媒体を提供することにある。

【0019】

本発明の他の目的は、エントリの重要度に応じて、検索処理に重み付けをするためのエントリ群選択、つまりキャッシュに入出力するエントリ群の選択方法を提供することにある。

【0020】

【課題を解決するための手段】

第1の本発明のルータにおけるキャッシュ管理装置は、全エントリに対して最長一致検索を行うためのフルテーブルと、全エントリの一部のエントリに対して最長一致検索を行うためのキャッシュとを備え、前記フォワーディングテーブルにおいてエントリを前記キャッシュテーブルに入れるときには、そのエントリを親とする全エントリを、また前記キャッシュテーブルからエントリを出すときには、そのエントリの親となる全エントリを、当該エントリの使用状況の評価値に対する閾値および削除あるいは移動されるエントリの領域対応の優先度に応じて選択することを特徴とする。

10

【0021】

第2の本発明のルータにおけるキャッシュ管理装置は、全エントリに対して最長一致検索を行うためのフルテーブルあるいは全エントリの一部のエントリに対して最長一致検索を行うためのキャッシュ内に仮想的な複数の領域を設定する手段と、該領域内では、それぞれ固有に、削除あるいは移動対象を決めるためのエントリ群の使用状況の評価値と、少なくとも削除対象に選択される可否を決める閾値と選択された場合に実際に削除あるいは移動されるエントリの該領域に応じて決まる優先度の一つを持つ手段と、該閾値および優先度に応じて入れ替え対象のエントリ群を選択する手段とを持つことを特徴とする。

20

【0022】

この場合、各エントリあるいは各エントリ群と各仮想的な領域を関連付ける手段、または各エントリあるいは各エントリ群にそれぞれ優先度を関連付ける手段、または各エントリあるいは各エントリ群にそれぞれ閾値を関連付ける手段の内の少なくとも一つの手段を備えるようにしてもよい。

【0023】

また、特定の時間区間において、エントリ群を構成する各エントリが持つ使用状況の評価値に対して、該使用状況を全てのエントリに対してあるいは一部のエントリに対して組み合わせた値あるいは統計値を各エントリ群の使用状況の評価値として用いる手段を備えるようにしてもよい。

30

【0024】

第3の本発明のルータにおけるキャッシュ管理装置は、プレフィクスビット列およびプレフィクス長から成るアドレスの集合を示す情報と、このアドレスの集合に対するパケットの出方路を示す情報と、優先度情報との組が登録されたエントリを複数有し最長一致検索されるフルテーブルと、入れ替えるべきエントリを親とする前記フルテーブルの全エントリが書き込まれ、またエントリが削除・移動されるときには、その親も削除・移動されるキャッシュと、前記フルテーブルの全エントリについて、その内容に加えてヒット情報が登録され、前記フルテーブルまたは前記キャッシュがヒットしたときにそのエントリの内容が更新されるヒット状況データベースと、パケットが入力されるとそのパケットヘッダから宛先ネットワークアドレスを取り出し、このアドレスをキーにして前記キャッシュテーブルまたは前記フォワーディングテーブルを検索し、得られた出方路にパケットを転送するパケット処理回路と、前記キャッシュに登録されている全エントリの前記優先度情報を含み、前記エントリ群入れ替えを行うときには、該優先度情報と前記ヒット状況データベースからの優先度情報とを比較して入れ替えるべきエントリ群を選択するエントリ選択装置と備えたことを特徴とする。

40

【0025】

エントリ選択装置は、前記ヒット状況データベースからエントリ群全ての情報を受け取り、この全エントリから、エントリ固有の優先度である領域、エントリ使用状況の評価値お

50

よび閾値の代表値を決めるエン트리群代表値回路と、前記キャッシュに既に存在するエン트리群に関して、そのエン트리群とその代表値を組にして表に持っており、その表の各エン트리群の代表値と、前記エン트리群代表値回路から送られてくる代表値とを比較して、キャッシュに既に存在するエン트리群がどのような優先順位にあるかを送出する前記領域、エン트리使用状況の評価値および閾値対応の比較回路と、該比較回路から送られてきた優先順位データを基にエントリの最終的な入替え順位を決定する調停回路と、前記キャッシュの空き状況を管理しており、当該エン트리群を追加することのできる空きがあるかどうかを調べ、空きが無い場合には、前記調停回路が決定した順位の最も低いエントリが所属するエントリ群から前記キャッシュのエントリを消去して十分な空きを作り、また削除したエントリ群を前記表からの削除のために前記比較回路に送るエントリ決定送付回路とで構成してもよい。

10

【0026】

より詳しくは、前記エン트리使用状況の評価値として、ヒット回数またはヒット時刻またはその両者を採用してもよく、前記各代表値は、閾値と領域についてはエントリ中の最も高い値、ヒット回数は平均値、ヒット時刻は最近のものとする事もできる。

【0027】

また、前記調停回路は、0～1の乱数を発生し、その乱数を2等分して、1/2づつをヒット回数とヒット時刻に割り当て、割り当てられた属性の順位を最終的な順位とするように構成してもよい。

【0028】

また、前記エントリの優先度を受信パケットから得られる情報に応じて設定する手段を設けてもよいし、ポリシーサーバにより変更可能としてもよい。

20

【0029】

本発明によれば、ヒットする可能性の高いエントリをキャッシュにおくことができるため、キャッシュミスヒット率の低い検索が可能であり、それゆえ全体の検索処理速度を向上させることが可能となる。すなわち、ヒット数の高いエントリが知られているときに、該エントリを優先的にキャッシュに入れておくことで、ヒットする可能性の高いキャッシュ内容となる。逆に、ヒット数の低いエントリはできるだけキャッシュに存在しないように、その優先度を下げることによって、キャッシュを有効に使用できる。

【0030】

また、エン트리群の平均、最大値、最小値といった統計数値を用いてエン트리群の代表値を得、その代表値をヒットする可能性を示す値として用いることにより、ヒット率を上げることができる。エン트리群は非常に多数のエントリから成るものもあり、まためったに使われないことがわかっている場合など様々な状況が考えられるため、そのエントリ群の一部のエントリだけをを用いて前記代表値を決めることは、そのエントリ群の真の代表値により近くなるのである。

30

【0031】

また、ポリシー上または料金体系上から重要度に応じて、エントリあるいはキャッシュの領域に優先度を設けておくことにより、重要度に応じて検索速度を調整することが可能となる。

40

【0032】**【発明の実施の形態】**

次に、本発明の実施の形態について図面を参照して詳細に説明する。

【0033】

図1は本発明のルータの第1実施形態の全体構成図であり、パケット処理回路101、キャッシュテーブル102、フルテーブル103、ヒット状況データベース104およびエントリ選択装置105から構成される。先ず、各構成要素の概要を説明する。ここでは、前述の第1の構成例を想定している。すなわち、フルテーブル103には、ルーティングに必要な全ての経路を含むエントリを入れておき、キャッシュテーブル102には、フルテーブル103の一部のエントリを入れておく。

50

【 0 0 3 4 】

図 1 において、パケット処理回路 1 0 1 はキャッシュテーブル 1 0 2 とフルテーブル 1 0 3 に接続されており、キャッシュテーブル 1 0 2 はフルテーブル 1 0 3 とヒット状況データベース 1 0 4 とエントリ選択装置 1 0 5 とに接続されている。ヒット状況データベース 1 0 4 は、さらにフルテーブル 1 0 3 とエントリ選択装置 1 0 5 と接続されている。エントリ選択装置 1 0 5 は、さらにキャッシュテーブル 1 0 2 とフルテーブル 1 0 3 と接続されている。以下、断りがない限り、最長一致検索のことを単に検索と記す。

【 0 0 3 5 】

フルテーブル 1 0 3 は、プレフィクスビット列とプレフィクス長とから成るアドレスの集合を示す情報と、このアドレスの集合に対するパケットの出方路を示す情報と、フルテーブル 1 0 3 とキャッシュテーブル 1 0 2 のエントリを入れ替えるときにエントリを選択するための優先度情報との組が登録されたエントリを複数有する。また、キャッシュテーブル 1 0 2 は、受信したパケットの送信先アドレスをキーにしてフルテーブル 1 0 3 を検索したときにヒットしたエントリを「親」とするフルテーブル 1 0 3 の全エントリ（エントリ群）の内容が登録される。なお、ここで、フルテーブル、キャッシュテーブルといった場合、テーブル自体の他、それらの検索機構をも含んでいるものとする。

【 0 0 3 6 】

パケット処理回路 1 0 1 は、パケットが入力されるとそのパケットヘッダから宛先ネットワークアドレス（以下、「アドレス」と記す）を取り出し、このアドレスをキーにしてキャッシュテーブル 1 0 2 を検索する。キャッシュテーブル 1 0 2 はこのキーで検索され、ヒットするエントリがあれば、そのヒットした旨をヒットしたエントリのあるキャッシュテーブル 1 0 2 のアドレスと共にパケット処理回路 1 0 1 に通知する。一方、キャッシュテーブル 1 0 2 にヒットしなかった場合には、ミスヒットした旨をパケット処理回路 1 0 1 に通知する。パケット処理回路 1 0 1 はヒットアドレスを通知されたときには検索処理を完了し、ミスヒットを通知されたときにはフルテーブル 1 0 3 の検索を行う。

【 0 0 3 7 】

フルテーブル 1 0 3 の検索は、キーをフルテーブル 1 0 3 に入力して行う。フルテーブル 1 0 3 は、入力されたキーを基に検索され、ヒットしたアドレスをパケット処理回路 1 0 1 に通知する。

【 0 0 3 8 】

このとき、キャッシュテーブル 1 0 2 とフルテーブル 1 0 3 のうちのいずれかヒットした方のテーブルは、ヒット状況データベース 1 0 4 を更新する。ヒット状況データベース 1 0 4 には、フルテーブルの全エントリについて、エントリ番号、キャッシュテーブル 1 0 2 でのヒットアドレス、フルテーブル 1 0 3 でのヒットアドレス、各エントリの優先度、ヒットがあった時刻およびどのエントリがどのエントリと同じエントリ群に属するかの情報が書かれている。ここで、エントリの優先度とは、フルテーブル 1 0 3 のエントリのうちのどのエントリをキャッシュテーブル 1 0 2 に入れるべきか、またどのエントリを出すべきかの優先度合をいい、後述のように、フルテーブルにも登録されているエントリ固有の閾値、エントリ固有の領域、ヒット回数と、ヒット状況データベースにのみ登録されるヒット時刻等の優先度情報で定められる。

【 0 0 3 9 】

フルテーブル 1 0 3 は、エントリ選択装置 1 0 5 の指示に従って、フルテーブル 1 0 3 のエントリ内容をキャッシュテーブル 1 0 2 にコピーする。キャッシュテーブル 1 0 2 には、エントリ選択装置 1 0 5 の指示に従ってフルテーブル 1 0 3 から入力されたエントリが登録され（書き込まれ）、エントリ選択装置 1 0 5 の指示に従って、エントリ内容が消去（削除）される。

【 0 0 4 0 】

エントリ選択装置 1 0 5 は、ヒット状況データベース 1 0 4 を参照して、各エントリの優先度に基づいて、キャッシュテーブル 1 0 2 に存在させるエントリを決定し、キャッシュテーブル 1 0 2 とフルテーブル 1 0 3 にデータの移動を指示する。

10

20

30

40

50

【 0 0 4 1 】

次に、図 2 , 図 3 , 図 4 および図 5 を参照して、上述の各構成要素 1 0 1 ~ 1 0 5 を詳細に説明する。

【 0 0 4 2 】

図 2 はパケット処理回路 1 0 1 の詳細を示す。図 2 を参照すると、このパケット処理回路 1 0 1 は、ヘッダコピー回路 1 0 1 1 がパケットを受け取り、このパケットから IP アドレスを含むヘッダをコピーしてアドレス抽出回路 1 0 1 2 に送る。また、パケット自体は出方路情報付加回路 1 0 1 4 に送る。アドレス抽出回路 1 0 1 2 は、コピーされたヘッダから IP アドレスを抽出して、キャッシュテーブル 1 0 2 とフルテーブル 1 0 3 に送る。ヒットアドレス・出方路情報変換回路 1 0 1 3 は、キャッシュテーブル 1 0 2 とフルテーブル 1 0 3 のいずれかからヒットアドレス情報を受け取り、このヒットアドレス情報を出方路情報に変換して出方路情報付加回路 1 0 1 4 に送る。出方路情報付加回路 1 0 1 4 は、ヘッダコピー回路 1 0 1 1 からのパケットと出方路情報付加回路 1 0 1 4 からの出方路情報を受け取り、それらを結合する。ヒットアドレス・出方路情報変換回路 1 0 1 3 における変換は、一覧表を検索して行う。一覧表はヒットアドレスと出方路インタフェース番号の組が表になったものである。

10

【 0 0 4 3 】

図 3 はキャッシュテーブル 1 0 2 の詳細を示す。図 3 を参照すると、キャッシュテーブル 1 0 2 では、検索キーレジスタ 1 0 2 1 がパケット処理回路 1 0 1 におけるアドレス抽出回路 1 0 1 2 からの IP アドレスを受け取る。最長一致検索回路 1 0 2 2 は、この IP アドレスとビットマスク付メモリ 1 0 2 5 の各エントリの内容とを比較して、最長一致となるエントリを決定する。最長一致検索回路 1 0 2 2 は、決定したエントリのアドレス（メモリのアドレス）、つまりヒットアドレスあるいは一致しなかったという情報をパケット処理回路 1 0 1 および結果情報回路 1 0 2 3 に送る。結果情報回路 1 0 2 3 は、ヒットアドレスにヒットした時刻、つまり現時刻の情報を付加してヒット状況データベース 1 0 4 に送る。エントリ操作回路 1 0 2 4 は、エントリ選択装置 1 0 5 およびフルテーブル 1 0 3 から送られてくる情報を受け取り、ビットマスク付メモリ 1 0 2 5 の各エントリの削除、更新、読出しまたは書込みを行う。

20

【 0 0 4 4 】

フルテーブル 1 0 3 は、当然にキャッシュテーブル 1 0 2 より大規模であるが、その構成は、図 3 に示したキャッシュテーブル 1 0 2 のブロック図と同様である。

30

【 0 0 4 5 】

図 4 は、ヒット状況データベース 1 0 4 の詳細を示す。図 4 を参照すると、書込み読出し制御回路 1 0 4 1 が、キャッシュテーブル 1 0 2 またはフルテーブル 1 0 3 からヒットアドレスとその時刻情報を受け取り、状況テーブル 1 0 4 2 のうちのヒットアドレスと一致するエントリヒット時刻情報を書き込む。また、フルテーブル 1 0 3 からヒット情報を受け取った場合には、書込み読出し制御回路 1 0 4 1 は、そのエントリ群の全エントリ情報をエントリ選択装置 1 0 5 に送る。

【 0 0 4 6 】

状況テーブル 1 0 4 2 は、エントリ毎に、そのエントリの番号、キャッシュテーブル 1 0 2 でのヒットアドレス、フルテーブル 1 0 3 でのヒットアドレス、キャッシュテーブル 1 0 2 におけるそのエントリの領域、キャッシュテーブル 1 0 2 における閾値、フルテーブル 1 0 3 における閾値、キャッシュヒット回数、ヒット時刻、キャッシュテーブル 1 0 2 にエントリ群を入れる時のそのエントリ番号およびフルテーブル 1 0 3 にエントリ群を入れる時のそのエントリ番号を格納する。

40

【 0 0 4 7 】

図 5 は、エントリ選択装置 1 0 5 の詳細を示す。図 5 を参照すると、エントリ群代表値回路 1 0 5 7 はヒット状況データベース 1 0 4 から上述のエントリ群全ての情報（N 個）を受け取る。エントリ群代表値回路 1 0 5 7 は、図 6 にその処理フローチャートを示すように、この全エントリから、閾値、領域、ヒット回数およびヒット時刻の各々について、最

50

大値を持つエン트리と最小値を持つエントリとを除外し、(N - 2)個のエントリを抽出する。そして、抽出したエントリから閾値、領域、ヒット回数、ヒット時刻の4つの代表値を決める。各代表値は、図6に示すように、閾値と領域についてはエントリ中の最も高い値、ヒット回数は平均値、ヒット時刻は最近のものとする。各代表値は、閾値保持回路1053、領域保持回路1054、ヒット回数比較回路1055およびヒット時系列比較回路1056に送る。

【0048】

ここで、閾値とは、エントリ群選択の優先度を定める他の属性、すなわち、領域、ヒット回数およびヒット時刻に対して比較の対象となる基準値であり、。また、領域とは、そのエントリ固有の優先度をいい、ポリシーあるいは料金体系から決められ得る。

10

【0049】

比較回路1055～1056は、キャッシュテーブル102に既に存在するエントリ群に関して、そのエントリ群とその代表値を組にして表に持っており、その表の各エントリ群の代表値と、エントリ群代表値回路1057から送られた値とを比較する。また、その際に、領域と閾値を考慮して比較を行う。本実施例では、ヒット回数、ヒット時刻を閾値と比較し、閾値を超えたエントリ群のみをまず選択し、さらに、比較回路1055～1056に保持されている全エントリ群のヒット回数、ヒット時刻に、優先度を掛け算して算出した値の大きい順に優先度をつける。比較の結果、送られた代表値と、キャッシュテーブル102に既に存在するエントリ群とが、どのような順位にあるかを調停回路1052に送る。調停回路1052では、それぞれの回路から送られてきた順位データを元に、それらに差異がある場合には、それらを参考にして、図7にその処理フローチャートを示すように、最終的な順位を決定する。図7を参照すると、0～1の乱数を発生し、その乱数を2等分して、1/2づつをヒット回数、ヒット時刻に割り当て、割り当てられた属性の順位を最終的な順位としてエントリ決定送付回路1051に送る。

20

【0050】

エントリ決定送付回路1051は、キャッシュテーブル102の空き状況を管理しており、N個のエントリを追加することのできる空きがあるかどうかを調べ、空きが有るときには、そのままN個のエントリを書き込み、また空きが無いには、順位の最も低いエントリが所属するエントリ群からキャッシュテーブル102のエントリを消去し、空きが十分になるまでこの手順を繰り返す。最後に、削除したエントリ群を各回路1053～1056

30

【0051】

なお、優先度は、上の説明では、閾値、領域、ヒット回数およびヒット時刻の4つの属性から定めるとしているが、本発明ではこれにとられることなく、優先度を定めるのにこれらの属性のうち少なくとも一つを使用することも可能である。

【0052】

次に、本発明の第2の実施例として、キャッシュテーブル102の内部を2つの領域に分け、それぞれA、Bとした例について説明する。この例ではヒット状況データベース104には、ヒット回数をエントリ毎に記録する。また、エントリ選択装置105は、領域Aにおいては、エントリ群のヒット回数の代表値が閾値M以下のものをキャッシュテーブル102からの消去対象とし、領域Bにおいては、エントリ群のヒット回数の代表値が閾値Mの0.5倍以下のものを消去対象とし、さらに消去対象となったエントリ群に対して領域Aと領域Bに2対1の重みをつけて、キャッシュテーブル102から削除するエントリ群を選択する。例えば、閾値を領域Aは10、領域Bは5とし、領域Aと領域Bそれぞれからこれらの閾値を下回るエントリ群を選び出す。さらに、この選択されたエントリ群に対して、ある確率変数Nに基づいて削除するエントリ群を選択する。重み付けをして領域Aのエントリ群はN、領域Bのエントリ群は2Nとする。この結果、最も確率の高いエントリ群を消去対象として選択する。さらに削除が必要であれば、この操作を繰り返す。

40

【0053】

50

また、Aの重み付けを0（ゼロ）として、Aを確率によらずに削除対象から外すことも可能である。逆に閾値を無限大として、全てのエントリのヒット回数の代表値が閾値以下とする、つまり閾値を事実上使用しないことも可能である。

【0054】

また、ヒット回数を用いない方法も可能である。上記例では、閾値とヒット回数を代表値として用いることを領域A、Bで行ったが、これ以外の任意の組合せを領域A、Bについて行うことも可能である。例えば、閾値はA、Bという領域に異なる値を割り付、このA、Bとは全く独立なa、bという領域ではそれぞれヒット回数の平均値とヒット回数の最大値といった別の代表値を用いることも可能である。

【0055】

また、エントリ選択装置105は、上述のキャッシュテーブル102に対する操作と同様の操作をフルテーブル103に対しても行い、フルテーブル103からキャッシュテーブル102に入れるエントリを選択する。

【0056】

次に、第3の実施例について説明する。本例では、各エントリは、その優先度を定めるための属性として、フルテーブルのどの領域に入れることができるか（どの領域に属することができるか）という領域情報と閾値とを持つ。このような属性は、本実施例では固定とするが、動的に変更することが可能である。動的に変更する場合、例えば、属性の変更はポリシーサーバの指示に応じて行われる。ポリシーサーバは網の運用のためのルールが網の管理者によって記述されているものである。例えば、あるエントリは昼間は優先度が高いが夜は優先度が低いといったルールがポリシーサーバに記述されていると、ポリシーサーバはそれに応じて昼と夜で優先度を切り替える必要があることを知り、各エントリの優先度の切り替えを行う。

【0057】

図8は、そのようなポリシーサーバ301が設けられた本発明の第2の実施形態のルータの全体構成図を示す。ポリシーサーバ301は、ヒット状況データベース104とエントリ選択装置105に対して、各エントリの領域、閾値の値を定義し、これをマッピング回路1043とマッピング回路1058を通じて優先制御に反映させる。

【0058】

すなわち、ポリシーサーバ301は、ユーザがフルテーブルの各エントリに対して定義した領域、閾値の値を端末302を通じて受け取り、ポリシーサーバ301内に保持する。さらに、ポリシーサーバ301は、保持している情報に変更があったときに、その情報をパケット処理回路101内のヘッダコピー回路1011とアドレス抽出回路1012を経由してヒット状況データベース104とエントリ選択装置105に送る（図9参照）。

【0059】

ヒット状況データベース104では、マッピング回路1043が書込み読み出し制御回路1041を通じて、状況テーブル1042に書き込み（図10参照）、また、エントリ選択装置105では、マッピング回路1058が閾値保持回路1053および領域保持回路1054に存在する各エントリの閾値、領域を変更する（図11参照）。

【0060】

次に、本発明の第4の実施例について説明する。この例では、各エントリが持つ属性は、入力されたパケットにより変更され得る。パケットは、そのヘッダに属性を決める値を持っており、この値に従って、ヒットしたエントリの属性を変更する。

【0061】

アドレス抽出回路1012は、コピーされたパケットヘッダからアドレスを抽出して、そのアドレスをキャッシュテーブル102とフルテーブル103に送るとともに、パケットヘッダのアドレス、あるいはそれ以外のヘッダ部分、ここでは、IPバージョン4パケットのヘッダのTime To Live（TTL）の値によって領域の値、閾値を決定し、マッピング回路1043とマッピング回路1058に送る。領域の値、閾値は、例えば、TTLが32以上のときには領域 = 低優先、閾値 = 50とし、TTLが32未満のとき

10

20

30

40

50

には領域 = 高優先、閾値 = 100とする。マッピング回路1043とマッピング回路1058の動作は第3実施例における動作と同様である。

【0062】

次に、本発明の第5の実施例について説明する。この例では、エン트리選択装置105は、エン트리群の代表値を評価する際に、エン트리群を構成する全てのエントリのなかから所定の規則に従ってエントリを選んで、その選ばれたエントリ（計算対象エントリ）の値に基づいて求める。例えば、エン트리群を構成するエントリの最大値と最小値を除いた全ての値の平均値を代表値とする。あるいは、エン트리群を構成するエントリのヒット時刻の最近（最も近い過去）の時刻を持つエントリを除いたエントリの時刻の平均値を代表値とする。

10

【0063】

次に、本発明の第6の実施例では、エン트리選択装置105はエン트리群の代表値を評価する際に、計算対象となるエントリの全ての値のなかから所定の規則に従って選ばれた値を用いる。例えば、エントリのヒット時刻の系列において、最近（最も近い過去）の時刻を代表値とする。

【0064】

次に、以上に説明したキャッシュの管理方法を実行するプログラムを半導体メモリ、CD-ROM、フロッピーディスク等の記録媒体に記録し、それをルータに読み込ませて実行するようにしてもよい。そのルータは、上述のフルテーブル、キャッシュおよびヒット状況データベースを備えており、また、読み込んだプログラムは、ルータに上述の packets 処理回路およびエン트리選択装置と同様な機能を発揮させる。

20

【0065】

【発明の効果】

本発明の第1の効果は、キャッシュを有するルータにおいて、エントリを群単位でキャッシュから出し入れする場合にも、キャッシュのヒット確率を上げることができることである。その理由は、よく使用されるエントリをキャッシュに置くようにするからである。よく使われるエントリは、使用頻度や、総使用回数などで定める優先度を用いて選ぶ。例えば、ランダムに選んだ場合に比べて使用頻度の高いものを選んだ場合に明らかにキャッシュヒット率が向上する。

【0066】

また、第2の効果は、あるポリシーに基づいて、処理性能を向上させるべきエントリに属性を指定できるため、優先度合いをつけることが可能になることである。例えば、顧客毎に異なる料金体系を採用している場合、高い料金を得ている顧客の持つアドレスを包含するようなエントリの優先度を上げることで、その顧客と通信するパケットの処理を高速化でき、低い料金の顧客との差別化が可能となる。

30

【図面の簡単な説明】

【図1】本発明の第1の実施形態のブロック図

【図2】図1におけるパケット処理回路101の詳細ブロック図

【図3】図1におけるキャッシュテーブル102の詳細ブロック図

【図4】図1におけるヒット状況データベース104の詳細ブロック図

40

【図5】図1におけるエン트리選択装置105の詳細ブロック図

【図6】図5におけるエン트리群代表値回路1057の処理フローチャート

【図7】図5における調停回路1052の詳細ブロック図

【図8】本発明の第2の実施形態のブロック図

【図9】図8におけるパケット処理回路101の詳細ブロック図

【図10】図8におけるヒット状況データベース104の詳細ブロック図

【図11】図8におけるエン트리選択装置105の詳細ブロック図

【図12】本発明を具体的に説明するための図

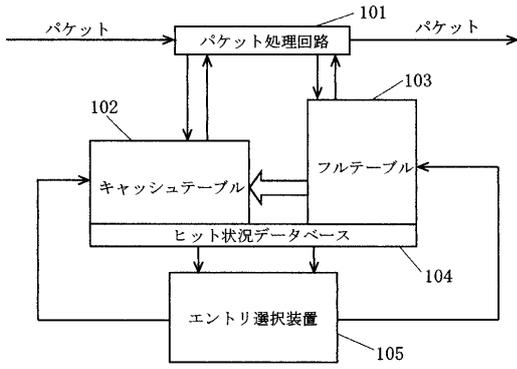
【符号の説明】

101 パケット処理回路

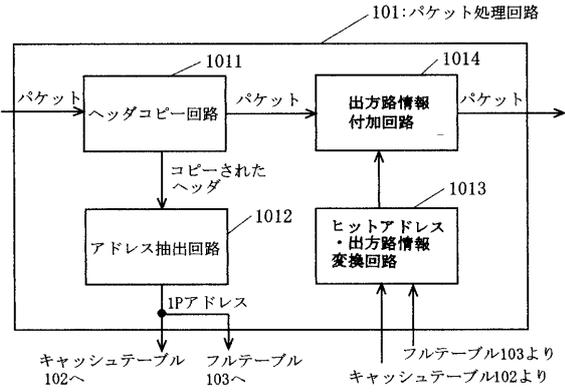
50

1 0 2	キャッシュテーブル	
1 0 3	フルテーブル	
1 0 4	ヒット状況データベース	
1 0 5	エン트리選択装置	
3 0 1	ポリシーサーバ	
3 0 2	端末	
1 0 1 1	ヘッダコピー回路	
1 0 1 2	アドレス抽出回路	
1 0 1 3	ヒットアドレス・出方路情報変換回路	
1 0 1 4	出方路情報付加回路	10
1 0 2 1	検索キーレジスタ	
1 0 2 2	最長一致検索回路	
1 0 2 3	結果情報回路	
1 0 2 4	エン트리操作回路	
1 0 2 5	ビットマスク付メモリ	
1 0 4 1	書込み読出し制御回路	
1 0 4 2	状況テーブル	
1 0 4 3	マッピング回路	
1 0 5 1	エン트리決定回路	
1 0 5 2	調停回路	20
1 0 5 3	閾値保持回路	
1 0 5 4	領域保持回路	
1 0 5 5	ヒット数比較回路	
1 0 5 6	ヒット時系列比較回路	
1 0 5 7	エントリ群代表値回路	
1 0 5 8	マッピング回路	

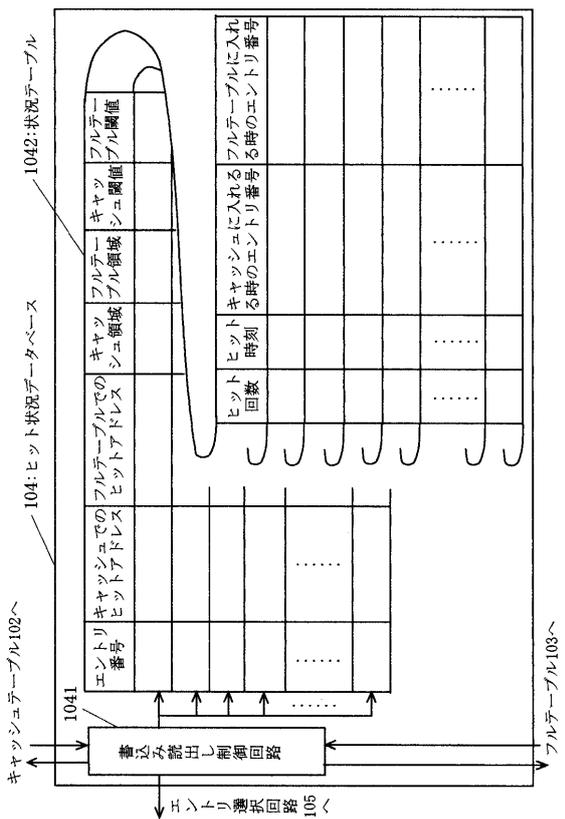
【 図 1 】



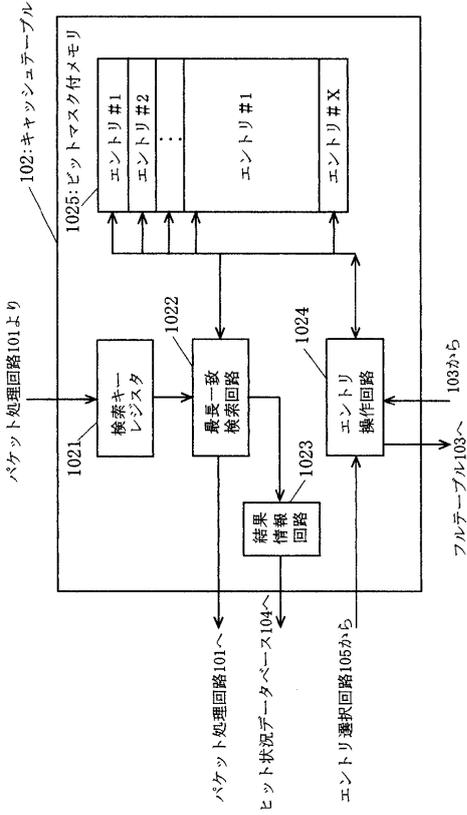
【 図 2 】



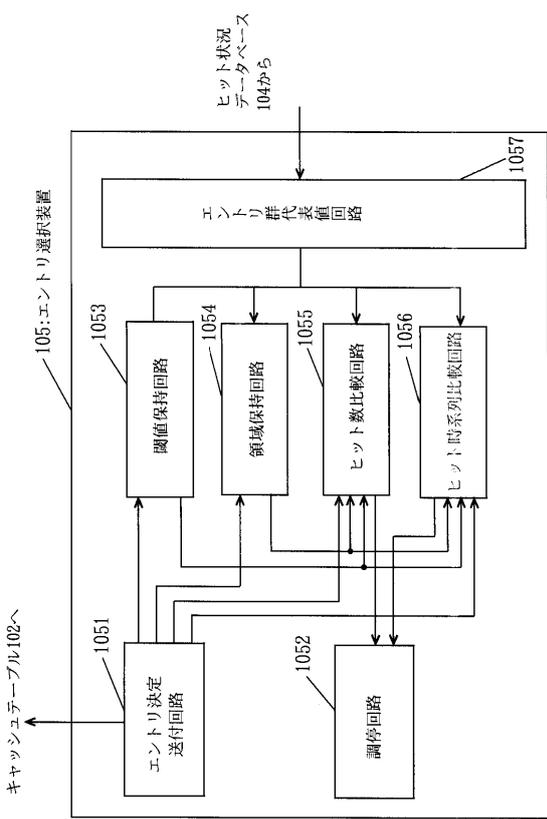
【 図 4 】



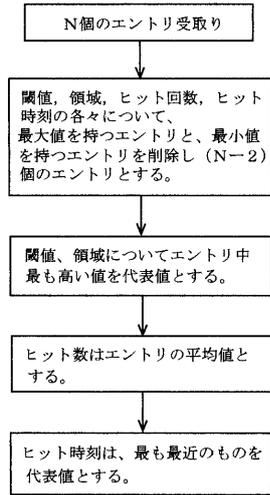
【 図 3 】



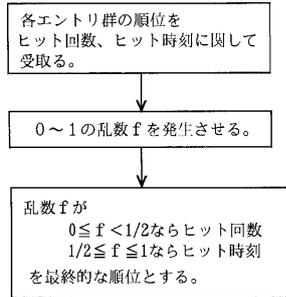
【 図 5 】



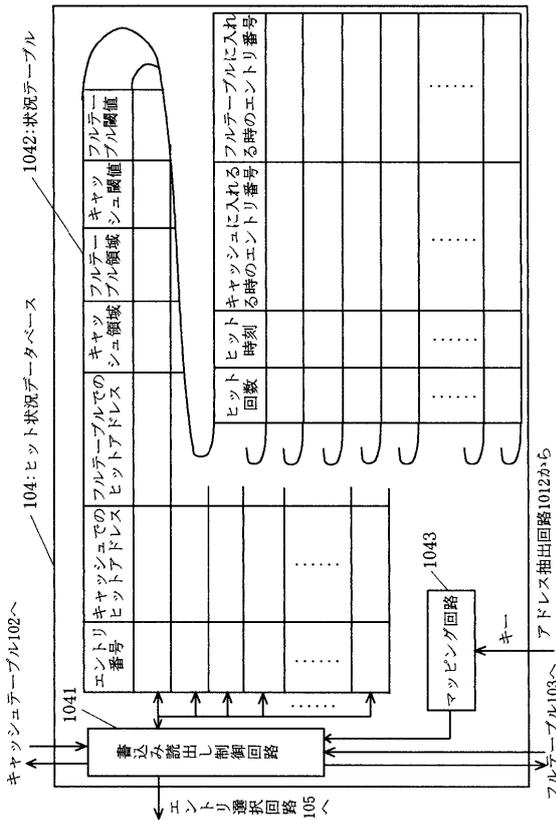
【 図 6 】



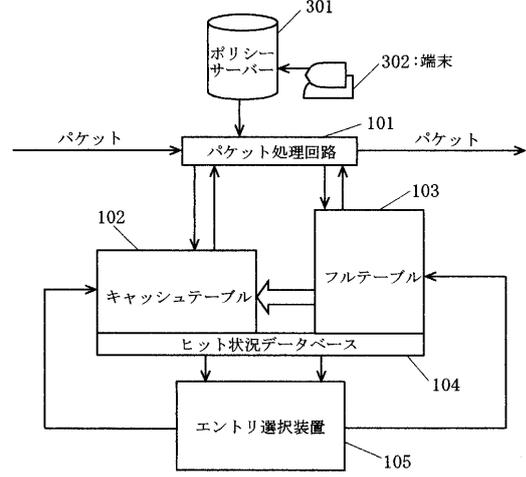
【 図 7 】



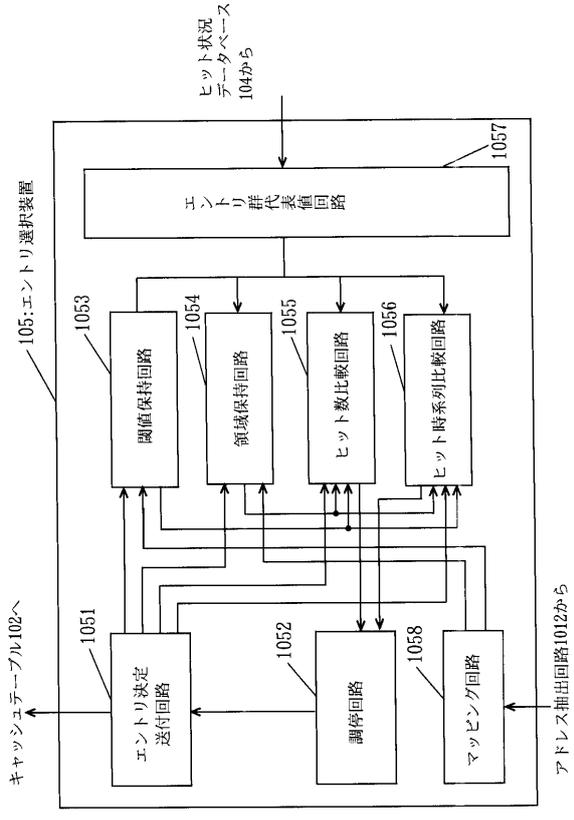
【 図 10 】



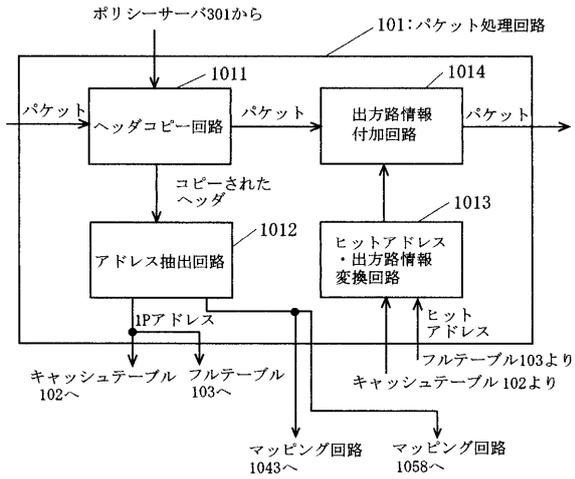
【 図 8 】



【 図 11 】



【 図 9 】



【 図 1 2 】

IPプレフィクス	プレフィクス長	出方路1F番号	使用時刻
100. 120. 0. 0	16	3	3:00, 2:00
100. 0. 0. 0	8	1	1:00, 0:30
100. 120. 140. 0	24	2	5:00
100. 120. 180. 0	24	4	2:00, 1:00

フロントページの続き

- (56)参考文献 特開平11-112564(JP,A)
特開平11-289345(JP,A)
特開平10-290232(JP,A)
特開2000-076168(JP,A)
特開2000-101606(JP,A)
国際公開第99/014906(WO,A1)

(58)調査した分野(Int.Cl.⁷, DB名)

G06F 13/00 353
G06F 12/08
G06F 12/12
H04L 12/56