

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

**特許第3735471号  
(P3735471)**

(45) 発行日 平成18年1月18日(2006.1.18)

(24) 登録日 平成17年10月28日(2005.10.28)

(51) Int. Cl.

F I

**H04L 12/56 (2006.01)**

H04L 12/56 100Z

**G06F 13/00 (2006.01)**

G06F 13/00 355

**H04L 13/08 (2006.01)**

H04L 13/08

H04L 12/56 260Z

請求項の数 13 (全 36 頁)

(21) 出願番号 特願平10-282205  
 (22) 出願日 平成10年10月5日(1998.10.5)  
 (65) 公開番号 特開2000-115243(P2000-115243A)  
 (43) 公開日 平成12年4月21日(2000.4.21)  
 審査請求日 平成15年12月24日(2003.12.24)

(73) 特許権者 000005108  
 株式会社日立製作所  
 東京都千代田区丸の内一丁目6番6号  
 (74) 代理人 100075096  
 弁理士 作田 康夫  
 (74) 代理人 100068504  
 弁理士 小川 勝男  
 (72) 発明者 赤羽 真一  
 東京都国分寺市東恋ヶ窪一丁目280番地  
 株式会社日立製作所中央研究  
 所内  
 (72) 発明者 須貝 和雄  
 神奈川県秦野市堀山下1番地  
 株式会社日立製作所汎用コン  
 ピュータ事業部内  
 最終頁に続く

(54) 【発明の名称】 パケット中継装置およびLSI

(57) 【特許請求の範囲】

## 【請求項1】

受信したマルチキャストパケットの宛先アドレスと送信元アドレスを検索のキーにして上記マルチキャストパケットに対する経路情報を検索し、上記経路情報に従い1つ又は複数のポートに上記マルチキャストパケットを転送するパケット中継装置において、宛先アドレスと送信元アドレスとをこの順に連結した経路アドレスのビットパターンに対応した2分木構造の検索木情報を格納する記憶手段と、  
 受信したマルチキャストパケットから宛先アドレスと送信元アドレスとを読み出し、その読み出した宛先アドレスと送信元アドレスとをこの順に連結して経路アドレスを生成し、その生成された経路アドレスと上記記憶手段に格納されている経路アドレスとの一致判定を行う一致判定手段とを有することを特徴とするパケット中継装置。

10

## 【請求項2】

受信したマルチキャストパケットの宛先アドレスと送信元アドレスを検索のキーにして上記マルチキャストパケットに対する経路情報を検索し、上記経路情報に従い1つ又は複数のポートに上記マルチキャストパケットを転送するパケット中継装置において、宛先アドレスと送信元アドレスとを連結した経路アドレスのビットパターンに対応した2分木構造の検索木情報を格納する記憶手段と、  
 受信したマルチキャストパケットから宛先アドレスと送信元アドレスとを読み出し、その読み出した宛先アドレスと上記記憶手段に格納されている経路アドレスとの一致判定を行い、その読み出した送信元アドレスと上記記憶手段に格納されている経路アドレスとの一

20

致判定を行う一致判定手段とを有することを特徴とするパケット中継装置。

【請求項 3】

上記一致判定手段は、受信したマルチキャストパケットから読み出した宛先アドレスと上記記憶手段に格納されている経路アドレスとの一致判定を、受信したマルチキャストパケットから読み出した送信元アドレスと上記記憶手段に格納されている経路アドレスとの一致判定より先に行うことを特徴とする請求項 2 に記載のパケット中継装置。

【請求項 4】

受信したマルチキャストパケットの宛先アドレスと送信元アドレスを検索のキーにして上記マルチキャストパケットに対する経路情報を検索し、上記経路情報に従い 1 つ又は複数のポートに上記マルチキャストパケットを転送するパケット中継装置において、宛先アドレスと送信元アドレスとをこの順に連結した経路アドレスのビットパターンに対応した 2 の  $p$  乗分木構造 ( $p$  は自然数) の検索木情報を格納する記憶手段と、受信したマルチキャストパケットから宛先アドレスと送信元アドレスとを読み出し、その読み出した宛先アドレスと送信元アドレスとをこの順に連結して経路アドレスを生成し、その生成された経路アドレスと上記記憶手段に格納されている経路アドレスとの一致判定を行う一致判定手段とを有することを特徴とするパケット中継装置。

【請求項 5】

複数のネットワークを相互に接続し、パケットを中継するパケット中継装置であって、受信したマルチキャストパケットの宛先アドレスと送信元アドレスを検索のキーにして上記マルチキャストパケットに対する経路情報を検索し、上記経路情報に従い 1 つ又は複数のポートに上記マルチキャストパケットを転送するパケット中継装置において、送信元アドレスと宛先アドレスとを、宛先アドレス、送信元アドレスの順に連結して経路アドレスとし、上記経路アドレスのビットパターンに対応した 2 分木構造に構成された検索木情報を格納する記憶手段と、受信したマルチキャストパケットの宛先アドレスと送信元アドレスをこの順に連結した経路アドレスの上位ビットから 1 ビットずつ検査し、その検査したビットの値によって上記記憶手段に格納された 2 分木構造をもつ検索木を検索する回路とを有することを特徴とするパケット中継装置。

【請求項 6】

複数のネットワークを相互に接続し、パケットを中継するパケット中継装置であって、受信したマルチキャストパケットの宛先アドレスと送信元アドレスを検索のキーにして上記マルチキャストパケットに対する経路情報を検索し、上記経路情報に従い 1 つ又は複数のポートに上記マルチキャストパケットを転送するパケット中継装置において、送信元アドレスと宛先アドレスとを、宛先アドレス、送信元アドレスの順に連結して経路アドレスとし、上記経路アドレスのビットパターンに対応した 2 の  $p$  乗分木構造 ( $p$  は自然数) に構成された検索情報を格納する記憶手段と、受信したマルチキャストパケットの宛先アドレスと送信元アドレスを連結した経路アドレスの上位ビットから  $p$  ビットずつ検査し、その検査したビットの値によって上記記憶手段に記憶された 2 の  $p$  乗分木構造をもつ検索木を検索する回路とを有することを特徴とするパケット中継装置。

【請求項 7】

複数のネットワークを相互に接続し、パケットを中継するパケット中継装置であって、受信したマルチキャストパケットの宛先アドレスと送信元アドレスを検索のキーとして上記マルチキャストパケットに対する経路情報を検索し、上記経路情報に従い 1 つ又は複数のポートに上記マルチキャストパケットを転送するパケット中継装置において、送信元アドレスと宛先アドレスを、宛先アドレス、送信元アドレスの順に連結して経路アドレスとし、この経路アドレスのビットパターンに対応する 2 の  $p$  乗分木構造 ( $p$  は自然数) をとる検索テーブルを格納する記憶手段と、経路アドレスの上位  $m$  ビット ( $m$  は自然数) に対応する 2 の  $m$  乗個の 2 の  $p$  乗分木のノードを、記憶手段上の決まった位置に展開し、それぞれのノードを、経路アドレスの第 0 ビ

10

20

30

40

50

ットから第 $(m - 1)$ ビットまでが取りうる値に1対1に対応させ、検索時には、経路アドレスの第0ビットから第 $(m - 1)$ ビットの値に従って、展開されたノードの一つを選択し、第 $m$ ビット以降は経路アドレスを $p$ ビットずつ検査し、その検査したビットの値によって上記記憶手段に記憶された2の $p$ 乗分木構造をもつ検索木の枝を検索する回路を有することを特徴とするパケット中継装置。

【請求項8】

複数のネットワークを相互に接続し、パケットを中継するパケット中継装置であって、受信したマルチキャストパケットの宛先アドレスと送信元アドレスを検索のキーにして上記マルチキャストパケットに対する経路情報を検索し、上記経路情報に従い1つ又は複数のポートに上記マルチキャストパケットを転送するパケット中継装置において、

10

送信元アドレスと宛先アドレスとを、宛先アドレス、送信元アドレスの順に連結して経路アドレスとし、上記経路アドレスのビットパターンに対応した2分木構造に構成された検索木情報を格納する記憶手段と、

受信したマルチキャストパケットの宛先アドレスと送信元アドレスをこの順に連結した経路アドレスの上位ビットから1ビットずつ検査し、その検査したビットの値によって上記記憶手段に格納された2分木構造をもつ検索木を検索する回路とを有することを特徴とするパケット中継装置。

【請求項9】

請求項8に記載のパケット中継装置において、

検索木情報を、1つの2分木ノードと、その直下につながる $p - 1$ 段分( $p$ は自然数)の合計(2の $p$ 乗 $- 1$ )個分の2分木ノードを一つの2の $p$ 乗分木ノードにまとめ、まとめられる最下段の2の $(p - 1)$ 乗個の2分木ノードに、それより上段のノードに割り付けられていたエントリデータを埋め込み、2の $p$ 乗分木ノードを、2分木を2の $(p - 1)$ 乗個分併せた形で構成することを特徴とするパケット中継装置。

20

【請求項10】

請求項8に記載のパケット中継装置において、

一つの2分木ノードと、その直下につながる $p - 1$ 段分( $p$ は自然数)の合計(2の $p$ 乗 $- 1$ )個分の2分木ノードを一つの2の $p$ 乗分木ノードにまとめ、まとめられる最下段の2の $(p - 1)$ 乗個の2分木ノードに、それより上段のノードに割り付けられていたエントリデータを埋め込むことにより、2の $p$ 乗分木ノードを、2分木を2の $(p - 1)$ 乗個分併せた形で構成することに加え、2分木を複数個併せるときに、一つだけ持てば良い要素を一つだけ保持するようにすることを特徴とする特徴とするパケット中継装置。

30

【請求項11】

請求項9又は請求項10に記載のパケット中継装置において、

2の $p$ 乗分木ノードを検索のために読むときにノード全てを読まずに、2の $p$ 乗分木ノードを作成するときに併せた2の $(p - 1)$ 乗個の2分木ノードの内の、いずれか一つに対応するデータのみを読むようにすることを特徴とするパケット中継装置。

【請求項12】

請求項10に記載のパケット中継装置において、

各ノードにそのノード自身のマスク長を格納するのではなく、そのノードの直ぐ下に繋がるノードのマスク長を格納することにより、ノードのマスク長を、そのノードのデータを読む前に知り、経路アドレスの、ノードのマスク長で示されるビット位置から、そのビット位置 $+ p - 1$ までの値に従い、ノードのデータの内、読み込む部分を選択することを特徴とするパケット中継装置。

40

【請求項13】

マルチキャストパケットのヘッダ情報を入力され、上記マルチキャストパケットに対する送信ポート等の転送情報を検索して出力する専用LSIにおいて、

マルチキャストパケットの宛先アドレスと送信元アドレスをこの順に連結した経路アドレスの上位ビットから $p$ ビット( $p$ は自然数)ずつ検査し、上記検査ビットの値によって記憶手段に記憶された2の $p$ 乗分木構造をもつ検索木を検索する回路を有することを特徴と

50

する専用 L S I。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、複数のネットワークを相互に接続し、パケットを中継するパケット中継装置、特にマルチキャストパケットの次転送先検索方式に関する。

【0002】

【従来の技術】

ユーザの増加に伴い、インターネットを流れるトラフィック(パケット)が急増し、インターネットが大規模化・高速化している。また、現在のインターネット(Internet Protocol 10  
によるパケット通信網：以下 IP ネットワーク)では、従来のデータ系の通信だけでなく、インターネット電話やインターネット放送などのリアルタイム系アプリケーション(用途)も登場し、音声通信機能の取り込みや、放送機能の取り込みが進んでいる。このような状況下において、IP ネットワークにおける IP マルチキャスト技術は、インターネットにおける動画や音声やコンテンツなどのマルチメディアデータの配信に有効な技術として期待されている。また、IP ネットワークを構成するパケット中継装置(ルータ)における IP マルチキャスト技術のサポート、IP マルチキャスト技術の高速化が課題となっている。

【0003】

ルータは、ある端末から特定の一つの端末にパケットを送信するユニキャスト通信の場合 20  
、受信パケットのヘッダ内の宛先 IP アドレスに対応する経路情報(次に転送すべきルータあるいは端末の IP アドレス、およびルータ内の送信ポート番号)をルーティングテーブルから検索し、パケットを送信する。以下、ユニキャストの経路検索について簡単に説明する。

【0004】

ルーティングテーブルは、宛先 IP アドレスに対応するサブネットワークアドレスとサブネットワークマスク長および、上記の経路情報から構成される情報群を複数保持している。以下では、これらの情報群をエントリと呼ぶ。ここでサブネットワークとは、例えば宛先の企業網の様な端末の部分集合を示し、サブネットワークアドレスは部分集合の IP アドレスをアドレス情報として結合したアドレス群である。また、サブネットワークマスク長は IP アド 30  
レスのうち上位何ビットが上記のサブネットワークの識別子であることを示す値である。ルータは、受信パケットの宛先 IP アドレスにエントリ内のサブネットワークマスク長だけ上位ビットから有効なマスクをかけたものとエントリのサブネットワークアドレスを比較して、一致するエントリの経路情報を検索結果とする。このようにサブネットワーク単位でエントリを構成することによりルーティングテーブルのエントリ数を大幅に縮小し、検索処理の効率化を図っている。複数のエントリに一致した場合は、サブネットワークマスク長が最長のエントリの経路情報を検索結果とする。以下では、この検索方式を最長一致検索と呼ぶ。

【0005】

ある端末から特定の複数の端末にパケットを送信するマルチキャスト通信の場合、ルーティングテーブルの各エントリは、送信元サブネットワークアドレスとアドレスマスクとマルチキャストグループアドレス、および経路情報から構成される。マルチキャスト・グループアドレスとは、送信元がパケットを送信すべき複数の宛先の集合(以下、マルチキャストグループと呼ぶ。)に対して割り当てられる識別子である。ルータは、受信パケットのヘッダ内の送信元 IP アドレスと受信パケットのヘッダ内の宛先 IP アドレス部に入っているマルチキャスト・グループアドレスをキーにして、ルーティングテーブルを検索する。マルチキャストの場合、検索の結果一致したエントリの経路情報は複数の送信ポート番号から構成される。ルータはこの送信ポート番号に従い、受信パケットをコピー処理して、上記特定のマルチキャストグループ向けに出力する。これらの検索処理やコピー処理の負荷が重く、マルチキャスト通信の性能が低いと、マルチキャストパケットを転送処理するルータ全体の性能低下の要因となってしまう。このため、通常のユニキャストパケット中継 50

処理と同様に高速化が要求されている。

【 0 0 0 6 】

ルータの負荷分散処理方式による高速化に関しては、例えば特開平 6 - 1 9 7 1 1 1 号公報（以下「従来技術 1」という。）に言及がある。従来技術 1 では、負荷分散処理方式により、高速中継可能なルータを実現することを目的としている。パケット中継を行うパケット中継モジュールがバスにより複数接続され、各々が同じくバスに接続された管理部からのルーティングテーブル情報に基づいてパケット中継機能を行う。パケット中継モジュールの増設により、性能向上ができる。さらに、従来技術 1 のパケット中継モジュールは、受信パケットのヘッダを抽出し、パケット転送先の検索処理を行うルーティング処理部と、受信パケットをメモリに格納し、上記解析結果に基づいて対応する他のパケット中継モジュールに受信パケットを転送する転送処理部を備え、別々の処理部で機能分担処理させることでパケット中継処理の高速化を図っている。しかし、従来技術 1 では、マルチキャストパケットの中継処理の高速化に関しては記載がない。

10

【 0 0 0 7 】

ルータにおけるマルチキャストパケットの宛先の検索（以下、マルチキャスト経路検索と呼ぶ）に関しては、例えば、（以下「従来技術 2」という。）に記載されている。従来技術 2 では、ルーティングテーブル検索をハッシュ検索方式を用いることにより高速化している。ルーティングテーブル情報をそのまま検索すると、対象となる宛先 IP アドレスが増加しテーブルのエントリが増加するに従って、検索処理にかかる検索時間が著しく増加してしまう。従来技術 2 は、ハッシュ値を計算するためのキーとして受信パケット内の送信元 IP アドレスを使用しており、ハッシュ値が等しい宛先 IP アドレスに対するエントリをグループ化したルーティングテーブル群を備える。

20

【 0 0 0 8 】

マルチキャストパケット受信時には、ルータ内部の経路検索処理部は、受信パケットの送信元 IP アドレスからハッシュ値を計算し、該ハッシュ値に対応する上記ルーティングテーブルを検索する。ハッシュ値に従ってルーティングテーブルの検索範囲を限定することで経路検索の高速化を図っている。

【 0 0 0 9 】

一方、経路検索方式としてRadishアルゴリズムが知られている。Radishアルゴリズムに関しては、例えば"A technical memo of WIDE project, Kazuhiko Yamamoto, Akira Kato and Akira Watanabe,Radish - A Simple Table Structure for CIDR"（以下「従来技術 3」という。）に記載されている。

30

【 0 0 1 0 】

このRadish方式のルーティングテーブルは、エントリを 2 分木構造に構成し、検索を高速化している。具体的には、IP アドレス形式の最上位ビットをルート（木構造の根）側とする左右にポインタを持つ複数の頂点（ノード）をポインタでつないだ 2 分木構造の各ノードにエントリを割り付けて構成している。この 2 分木構造を検索する際には、受信パケットの宛先 IP アドレスを上位ビットから 1 ビットずつ検査し、検査されたビットの値（0 または 1）に従って各ノードの左右のどちらかのポインタを選択して次のノードに移動する。このような検索により目的のエントリが割り付けられたノードにたどり着くことができる。

40

【 0 0 1 1 】

従来技術 3 は、宛先 IP アドレスを 1 ビットずつ検査して 2 分木を辿るため、ルーティングテーブルのエントリが多くなっても、高々宛先 IP アドレスのビット数回の検査を行う（2 分木ノードを辿る）ことで検索を終了することができる。

【 0 0 1 2 】

上記のユニキャストの経路検索の説明で述べた、マスク長の異なる複数のエントリにマッチする場合、従来技術 3 のRadish方式のルーティングテーブルの検索においては、木を辿る途中で複数個のエントリにマッチすることになる。エントリがマッチする度にそのエントリ内の経路情報の候補を保持しておき、新たにエントリにマッチした場合には、以前の

50

候補を新しい候補に更新してゆくことで最長一致検索仕様を満たすことができる。

【 0 0 1 3 】

また、従来技術 3 のRadish方式を適用したユニキャスト経路検索を高速に行う技術として、「IPルーティングテーブルのハードウェアによる高速検索方式」 1998年電子情報通信学会通信ソサイエティ大会（以下「従来技術 4」という。）がある。従来技術 4 は次ノード選択回路、ネットワークアドレス比較回路、検索終了判定回路からなる。ここでネットワークアドレス比較回路は、Radish方式で構成した 2 分木から不要な枝を取り除いた場合、ノード自体が保持するネットワークアドレスと、受信パケットの宛先 IP アドレスにノードが保持するマスクをかけたものを比較することによって、削除された枝において検査されるべきビット（1 または複数の連続したビット）の値を一度に検査するための回路であり、この比較の結果が一致の場合は正しいノードに到達したことを意味し、不一致の場合は正しくないノードに到達したことを意味する。

10

【 0 0 1 4 】

従来技術 4 では、次ノードの選択と上記で説明したネットワークアドレス比較処理を並列処理し、検索処理の高速化を図っている。

【 0 0 1 5 】

また、Radish方式と同様なPatricia Trie検索アルゴリズムによるユニキャスト経路検索を高速に行う技術として、「高速 IP アドレス解決 H/W エンジンの開発」 1998年電子情報通信学会通信ソサイエティ大会（以下「従来技術 5」という。）がある。従来技術 5 も従来技術 4 と同様にして木構造の検索の各処理を並列化して高速化を図っている。

20

【 0 0 1 6 】

従来技術 5 のPatricia Trie検索アルゴリズムによるユニキャスト経路検索をさらに高速化する技術として、「マルチウェイ化ツリーによる IP ルーティングテーブルの高速検索方式」 1998年電子情報通信学会通信ソサイエティ大会（従来技術 6）がある。従来技術 6 ではPatricia Trieアルゴリズムによる 2 分木構造を 2 の N 乗分木に構成することにより、高速化を図る旨が記載されている。

【 0 0 1 7 】

また、上記のユニキャスト経路検索において最長一致検索を高速に行う技術として、特開平 10 - 2 2 2 5 3 5 号公報（以下「従来技術 7」という。）がある。従来技術 7 では、従来技術 3 のRadish方式とは異なる方法で 2 分木を構成し、その 2 分木を検索するハードウェアで構成されたデータ検索回路を用いて最長一致検索を高速に行っている。従来技術 7 の 2 分木の具体的な検索方法は、受信パケットの宛先 IP アドレスにノード内に保持されているサブネットマスクをかけたものと、ノードのサブネットワークアドレスを比較し、比較の結果一致すればこのノードの経路情報を検索結果として検索を終了し、不一致の場合は、マスクされた宛先アドレスとノードのサブネットワークアドレスを 0 以上の整数とみなして大小比較を行い、その大小に従って 2 分木の左右の枝を辿るというものである。この際、マスク長が長いエントリから順に 2 分木の上部に配置しておくこと、その順に検索が行われるため、最長一致検索が実現できる旨が記載されている。なお、上記では、マスク長は上位何ビットがサブネットワークアドレスとして有効かを示す値である、という定義に従って従来技術 7 の内容を説明したが、従来技術 7 の明細書中では、マスク長を、IP アドレスのサブネットワークアドレス部分ではない下位のビット数の長さとして定義しており、また、最長一致検索方式をベストマッチ方式と呼んでおり、上記ベストマッチ方式は、マスク長が最短のものを検索結果として採用する。

30

40

【 0 0 1 8 】

【 発明が解決しようとする課題 】

現在、IP マルチキャスト通信は、インターネット上に仮想的に構築されたマルチキャストネットワークにおいて限定的に利用されているにすぎない。そのネットワーク規模もそれほど小さくなく、マルチキャスト通信の検索処理や転送処理は、ルータの転送性能劣化の原因にはなっていない。しかし、今後、インターネットにおけるマルチキャスト機能を利用したアプリケーションが広く普及すると、それに伴い、ルータが保持するマルチキャスト

50

ストルーティングテーブルのエントリ数が多くなる。従って、ルーティングテーブルの検索処理の高速化をエントリ数の多い条件下で実現することが求められている。

【0019】

従来技術2ではハッシュ方式のマルチキャスト経路検索方式が記載されているが、ハッシュ方式の場合、ハッシュ値の算出法はエントリの数にはよらず一定であり、エントリの増加に応じて柔軟にルーティングテーブル群の数を増やすことができない。従って、エントリが増加した場合、分割されるルーティングテーブル自体が長くなり、平均検索時間が長くなるという問題がある。

【0020】

この点、上述のRadishアルゴリズムを用いてユニキャストの経路検索を行う方法によれば、ハッシュ方式に比べて、ルーティングテーブルのエントリ数が増大しても、検索処理時間が増大しないという利点がある。

10

【0021】

しかし、従来技術3乃至従来技術5には、Radishアルゴリズム(あるいはPatricia Trieアルゴリズム)を用いたマルチキャストの経路検索に関する記載はない。

【0022】

また、従来技術6では、2分木構造を2のN乗分木に構成することにより、最大検索時間を1/Nに短縮する旨が記載されているが、マルチキャスト経路検索に関しては言及していない。

【0023】

20

また、従来技術7にもマルチキャストの経路検索を高速化する方式については言及されていない。さらに、従来技術3の2分木の構造が、登録されるエントリによって一意に決定し、その最大検索時間は検索に用いるキーのビット数によって決まるのに対し、従来技術7の2分木の構造は、その作り方によっては、ノードが一方方向に長くつながる木ができてしまい、平均検索時間が長くなる場合がある。このような木が構成されるのを防ぐには、2分木が均等に広がるように構成するアルゴリズムが必要であるが、そのアルゴリズムについては述べられていない。また、従来技術7の2分木構造を構成する際、マスク長が長いエントリから順に2分木の上部に配置しておくこと、その順に検索が行われるため、最長一致検索処理が実現できると述べているが、マスク長が異なるノードのつながり方に関しては詳細は述べられていない。また、従来技術3のように検査ビットの値に従って2分木の左右の枝を辿るのではなく、ノードにおける大小比較の結果、その大小に従って2分木の左右の枝を辿るため、2のp乗分木方式を実現することができない。

30

【0024】

そこで、本発明の第1の目的は、マルチキャストパケットの中継処理に関し、マルチキャストパケット経路検索処理の高速化の手段を提案することである。

【0025】

また、本発明の第2の目的は、Radishアルゴリズムによる2分木検索方式をマルチキャスト経路検索に適用する手段を提案することである。

【0026】

また、本発明の第3の目的は、検索のキーのビット数が大きい場合でも、ノードを辿る回数がビット数より少なく、最大検索時間が少ない経路検索方式を、マルチキャスト経路検索方式に適用する手段を提案することである。

40

【0027】

本発明の第4の目的は、Radishアルゴリズムによるマルチキャストルーティングテーブルの2分木検索方式に従って経路検索処理を高速に行う回路の構成を提案することである。

【0028】

【課題を解決するための手段】

上記の目的を達成するために、本発明では、以下に述べる方式および手段を持つ。

【0029】

パケットの送信元の端末のIPアドレス(送信元IPアドレス)と、パケットを送信すべき

50

端末が属するマルチキャスト・グループのグループアドレス(宛先マルチキャスト・グループアドレス)とを探索のキーとして用いるマルチキャスト経路探索において、送信元IPアドレスと宛先マルチキャスト・グループアドレスを連結して一つの経路アドレスとして定義し、この経路アドレスのビットパターンに従い、マルチキャストルーティングテーブルを2分木に構成し、受信したマルチキャストパケットの送信元IPアドレスと宛先マルチキャスト・グループアドレスとを連結した経路アドレスの上位ビットから1ビットずつ検査し、上記検査ビットの値によって2分木構造をもつ探索木を検索する。

【0030】

また、上記の経路アドレスに関し、送信元IPアドレスと宛先マルチキャスト・グループアドレスの連結順を、宛先マルチキャスト・グループアドレス、送信元IPアドレスの順にして経路アドレスを定義し、この経路アドレスのビットパターンに従い、マルチキャストルーティングテーブルを2分木に構成する。

10

【0031】

また、2分木方式の探索木を構成する各ノードを、2分木ノードから4分木、8分木、あるいは一般に2のp乗と、枝別れの数に2のべき乗で増やすことで、マルチキャストルーティングテーブルを2のp乗分木構造に構成し、受信したマルチキャストパケットの宛先マルチキャスト・グループアドレスと送信元IPアドレスをこの順に連結した経路アドレスを、一つのノードで1ビットではなく、連続する2ビット、3ビット、あるいは一般にpビットを同時に検査し、上記連続するビットの値によって2のp乗分木構造を持つマルチキャストルーティングテーブルを検索する。

20

【0032】

また、上記の経路アドレスとして、マルチキャスト・グループアドレス自体ではなく、マルチキャスト・グループアドレスの下位28ビットのマルチキャスト・グループIDと送信元IPアドレスをこの順に連結したものを採用し、この経路アドレスのビットパターンに従い、マルチキャストルーティングテーブルを2分木あるいは2のp乗分木に構成する。

【0033】

また、探索木を構成するメモリ量を減らすために、2のp乗分木を構成する際、一つの2分木ノードと、その直下につながる(p-1)段分の合計(2のp乗-1)個分の2分木ノードを一つの2のp乗分木ノードにまとめ、まとめられる最下段の2の(p-1)乗個の2分木ノードに、それより上段のノードに割り付けられていたエントリデータを埋め込むことにより、2のp乗分木ノードを2分木ノードを2の(p-1)乗個分併せた形で構成するようにし、さらに、2分木を複数個併せるときに、共通なノードデータの要素を一つだけもつようにする。

30

【0034】

また、2分木ノードを2のp乗分木にまとめる際、2のp乗分木ノードのメモリ量削減と、各2のp乗分木ノードのメモリ量の統一のために、送信ポート情報を探索木のノードから分離して記憶手段の別領域に格納し、探索木のノードには上記送信ポート情報の格納領域へのポインタを保持することで、エントリが割り当てられている2分木ノードと、エントリが割り当てられていない2分木ノードのデータ量を同じにし、これらの2分木ノードをまとめて2のp乗分木ノードを構成する。

40

【0035】

また、経路アドレスの検査すべきビット位置を示すマスク長に関し、各2のp乗分木ノードにそのノード自身のマスク長を格納するのではなく、そのノードの直下につながるノードのマスク長と格納することにより、経路探索処理を行うために探索木データが格納されている記憶手段からノードデータを読み込む際、ノードデータ全てを読み込むのではなく、直前に読み込んだノードに格納されているマスク長を用いて、ノードデータのうち必要な部分のみを選択して読み込む。

【0036】

また、各ノードの先頭に、そのノードにエントリが割り付けられているか否かを示すフラ

50



グを設け、最初に、このフラクを読み込み、エントリが割り付けられていないノードでは、経路情報を読み込まないようにする。

【0037】

また、経路アドレスの上位 $m$ ビットに対応する2の $m$ 乗個の2の $p$ 乗分木ノードを、記憶手段上の決まった位置に展開し、展開されたそれぞれのノードを、経路アドレスの第0ビットから第 $(m-1)$ ビットまでが取りうる値に1対1に対応させ、検索時には、経路アドレスの第0ビットから第 $(m-1)$ ビットの値に従って、該ノードを選択する。

【0038】

また、宛先アドレスと送信元アドレスをこの順に連結して経路アドレスとし、上記経路アドレスの上位 $m$ ビットに対応する2の $m$ 乗個の2の $p$ 乗分木ノードを決まった位置に格納し、それ以降の木は2の $p$ 乗分木構造に構成して格納する記憶手段を持ち、経路検索時には、受信したマルチキャストパケットの宛先アドレスと送信元アドレスをこの順に連結した経路アドレスの第0ビットから第 $(m-1)$ ビットの値に従って、メモリ上の決まった位置に展開されたノードの一つを選択し、経路アドレスの第 $m$ ビット以降は経路アドレスを $p$ ビットずつ検査して2の $p$ 乗分木に構成されたマルチキャストルーティングテーブルを検索する回路をもつ。

【0039】

その他の本願が解決しようとする課題、及びその解決手段は、後述の「発明の実施の形態」の欄、及び本願図面で明らかにされる。

【0040】

【発明の実施の形態】

以下、本発明の実施例を図面を用いて説明する。

【0041】

まず、一般的なネットワークの構成およびそのネットワークにおけるパケット中継動作を図2、3、4を用いて説明する。

【0042】

図2に示したネットワークは、パケットを送受信する端末( $T11$ 、 $T12$ 、 $T21$ 、 $T22$ 、 $T31$ 、 $T32$ 、 $T41$ 、 $T42$ 、 $T51$ 、 $T52$ )と、端末を相互に接続するサブネットワーク( $SN1 \sim SN5$ )と、該サブネットワークを相互に接続するルータ( $R1 \sim R3$ )から構成される。また、ルータ $R1$ はポート $P11$ 、 $P12$ 、 $P13$ を持ち、ルータ $R2$ はポート $P21$ 、 $P22$ 、 $P23$ 、 $P24$ を持ち、ルータ $R3$ はポート $P31$ 、 $P32$ 、 $P34$ 、 $P35$ を持つ。各端末には識別子としてアドレスが割り当てられており、例えばIPネットワーク上の各端末には32ビットのIPアドレスが割り当てられている。このIPアドレスについて図3(a)を用いて説明する。図3(a)に示したIPアドレス(1)は、サブネットワークアドレス(2)とホストアドレス(3)から構成される。

【0043】

サブネットワークアドレスは、サブネットワークの識別子であり、ホストアドレスはサブネットワーク内の端末の識別子である。サブネットワークアドレスのビット数はサブネットマスク長で示されており、このサブネットマスク長だけ上位ビットから有効なマスク(サブネットマスク)とIPアドレスのビットごとの論理積をとることにより、IPアドレスからサブネットワークアドレスを得ることができる。

【0044】

次に、ある端末から特定の一つの端末にパケットを送信するユニキャスト通信について、図2を用いてパケット中継動作を説明する。一例として、端末 $T11$ から端末 $T22$ へユニキャストパケットを送信する場合について説明する。送信されるパケットのヘッダフォーマットを図3(c)に示す。図3(c)の送信元IPアドレスフィールド(7)には送信元端末 $T11$ のIPアドレスが格納され、宛先IPアドレスフィールド(8)には宛先端末 $T22$ のIPアドレスが格納される。端末 $T11$ からのユニキャストパケットを受信したルータ $R1$ は、パケットヘッダ内の宛先IPアドレスフィールドを検査し、宛先端末 $T22$ がサブネットワーク $SN2$ 上にあり、かつ、サブネットワーク $SN2$ がルータ $R2$ 経由で接

10

20

30

40

50

続されていることを認識し、パケットを送信すべきルータ R 2 の IP アドレス(ネクストホップ IP アドレスと呼ぶ)および送信すべきポート P 1 2 を決定し、パケットを出力する。ルータ R 1 からのユニキャストパケットを受信したルータ R 2 は、パケットヘッダ内の宛先 IP アドレスフィールドを検査し、宛先端末 T 2 2 がサブネットワーク S N 2 上にあり、かつ、サブネットワーク S N 2 がルータ R 2 に直接接続されていることを認識し、ネクストホップ IP アドレス(今の場合、端末 2 2 の IP アドレス)および送信すべきポート P 2 2 を決定し、パケットを出力する。

#### 【 0 0 4 5 】

次に、ルータの構成およびルータのユニキャストパケットの中継動作を図 4 を用いて説明する。図 4 のルータ 1 0 は、入力ポート 1 1 -i(= 1 ~ N)、出力ポート 1 2 -i(= 1 ~ N)、ルーティング処理部 1 3 -i(= 1 ~ N)、スイッチ 1 4、通信制御部 1 5 -i(= 1 ~ N)、管理部 1 6 から構成される。上記で説明したユニキャストパケットが入力ポート 1 1 -i から入力されると、ルーティング処理部 1 3 -i がパケットのヘッダ内の宛先 IP アドレスをキーにしてユニキャストルーティングテーブルを検索する。この検索により、ネクストホップ IP アドレスと、出力すべき出力ポート 1 2 -i の番号(送信ポート番号)を決定し、上記ネクストホップ IP アドレスと送信ポート番号をパケットに付加してスイッチ 1 4 に送出する。スイッチ 1 4 は上記送信ポート番号に対応する出力ポート 1 2 -i を保持する通信制御部 1 5 -i にパケットをスイッチングする。スイッチングされたパケットを受信した通信制御部 1 5 -i は、パケットに付加されているネクストホップ IP アドレスに対応するデータリンク層の M A C (Media Access Control) アドレスを対応テーブルに従って決定してパケットの宛先 M A C アドレスとして付加して、送信ポート番号に対応する出力ポート 1 2 -i にパケットを送出する。

#### 【 0 0 4 6 】

なお、上記のユニキャストルーティングテーブルは、管理部 1 6 が他のルータと接続情報をやり取りして作成し、各ルーティング処理部 1 3 -i に配布する。また、管理部 1 6 は、他のルータおよび端末とデータリンク層の情報のやり取りをし、ネクストホップ IP アドレスと、その IP アドレスを持つルータおよび端末の M A C アドレスの対応テーブルを作成して各通信制御部 1 5 -i に配布する。

#### 【 0 0 4 7 】

次に、ある端末から特定の複数の端末にパケットを送信するマルチキャスト通信について、図 2 を用いてパケット中継動作を説明する。一例として、端末 T 1 1 からある特定のマルチキャストグループに属する端末(T 2 1、T 3 1、T 3 2、T 4 1、T 4 2、T 5 1)へマルチキャストパケットを送信する場合について説明する。マルチキャストグループは、クラス D の IP アドレス(マルチキャスト・グループアドレス)によって識別される。このクラス D の IP アドレスの構造を図 3 (b)に示す。クラス D の IP アドレス(4)は、クラス D であることを示す "1 1 1 0" という上位 4 ビット(5)と 2 8 ビットのマルチキャスト・グループ ID (以下、G I D)(6)から構成される。上記のマルチキャスト・グループアドレスによって識別されるマルチキャストグループは、複数のサブネットワークにまたがるのが可能である。また、グループに属する端末のメンバー構成は動的で、端末は自由にマルチキャストグループに参加、離脱することができる。

#### 【 0 0 4 8 】

端末 T 1 1 から、あるマルチキャストグループ宛に送信されたマルチキャストパケットのヘッダ内の送信元 IP アドレスフィールド(図 3 (c)の 7)には送信元端末 T 1 1 の IP アドレスが格納され、宛先 IP アドレスフィールド(図 3 (c)の 8)には、ユニキャストパケットの場合とは異なり、宛先のマルチキャスト・グループアドレスが格納される。端末 T 1 1 からのマルチキャストパケットを受信したルータ R 1 は、ユニキャストパケットの場合とは異なり、パケットヘッダ内の送信元 IP アドレスフィールドと宛先 IP アドレスフィールドの 2 つのフィールドを検査し、このマルチキャストパケットの送信元端末 T 1 1 がサブネットワーク S N 1 上にあることを認識し、さらに、このサブネットワーク S N 1 から、宛先マルチキャストアドレスに属する端末の全てにパケットを送信するためにルー

10

20

30

40

50

タ R 1 がパケットを送信すべき送信ポート P 1 2、P 1 3 を決定し、上記ポートにパケットをコピーして出力する。ルータ R 1 からのマルチキャストパケットを受信したルータ R 2 は、R 1 と同様な検査を行い、このマルチキャストパケットの送信元端末 T 1 1 がサブネットワーク S N 1 上にあることを認識し、さらに、宛先マルチキャストアドレスに属する端末が存在するサブネットワーク S N 2、S N 3 にパケットを送信するためにルータ R 2 がパケットを送信すべき送信ポート P 2 2、P 2 3 を決定し、上記ポートにパケットをコピーして出力する。ルータ R 3 も同様にして R 1 から受信したパケットを送信すべきポート P 3 4、P 3 5 を決定し、上記ポートにパケットをコピーして出力する。

#### 【 0 0 4 9 】

次に、ルータのマルチキャストパケットの中継動作を図 4 を用いて説明する。上記で説明したマルチキャストパケットが入力ポート 1 1 -i から入力されると、ルーティング処理部 1 3 -i がパケットのヘッダ内の送信元 I P アドレスと宛先 I P アドレス(宛先マルチキャストアドレス)をキーにしてマルチキャストルーティングテーブルを検索する。この検索により、出力すべき 1 つまたは複数の出力ポート 1 2 -i の番号(送信ポート番号)を決定し、受信したパケットをコピーし、各送信ポート番号をパケットに付加してスイッチ 1 4 に送出する。スイッチ 1 4 は上記送信ポート番号に対応する出力ポート 1 2 -i を保持する通信制御部 1 5 -i にパケットをスイッチングする。スイッチングされたパケットを受信した通信制御部 1 5 -i は、ユニキャストパケットの場合とは異なり、マルチキャストパケットのヘッダ内の宛先マルチキャストアドレスを用いてデータリンク層のマルチキャスト用 M A C アドレスを生成し、パケットの宛先 M A C アドレスとして付加して、送信ポート番号 20 に対応する出力ポート 1 2 -i にパケットを送出する。

#### 【 0 0 5 0 】

なお、上記のマルチキャストルーティングテーブルは、ユニキャストルーティングテーブルと同様にして管理部 1 6 が他のルータと接続情報をやり取りして作成し、各ルーティング処理部 1 3 -i に配布する。

#### 【 0 0 5 1 】

次に、本発明の一実施例として、ルータの全体構成例を図 1 を用いて説明する。

#### 【 0 0 5 2 】

まず、ルータ 7 0 0 の構成を図 1 を用いて説明する。5 0 5 はスイッチであり、図 4 のスイッチ 1 4 に対応する。上記スイッチ 5 0 5 に、装置全体の管理機能と経路情報の収集・分配等の機能をもつ管理部 6 0 1 が接続されている。管理部 6 0 1 は図 4 の管理部 1 6 に対応する。また、スイッチ 5 0 5 には、高速にパケット中継処理を行う機能を有するパケット処理部 6 0 0 を複数接続することができる。パケット処理部 6 0 0 は、図 4 のルーティング処理部 1 3 -i に対応する。管理部 6 0 1 は、それぞれのパケット処理部 6 0 0 にルーティングテーブルを配布し、それぞれのパケット処理部 6 0 0 が上記ルーティングテーブルに基づいてパケット中継処理を行うことによりルータの機能を実現する。 30

#### 【 0 0 5 3 】

さらに、各パケット処理部 6 0 0 には、ネットワークを接続するための通信ポートを備える通信制御部 5 0 1、6 0 2、6 0 3 (図 4 の通信制御部 1 5 -i に対応)が接続される。通信制御部は、各種のネットワークに接続することができる。また、そのネットワークの通信速度に応じて、高速通信(例えば、図 2 の 5 0 0)ならば 1 ポート、低速通信(例えば、図 2 の 6 0 4、6 0 5)ならば複数ポートを設け、1 つまたは複数のネットワークと接続することができる。 40

#### 【 0 0 5 4 】

次に、上記のパケット処理部 6 0 0 の構成について、図 1 を用いて説明する。図 1 において、パケット処理部 6 0 0 は、パケット転送処理を行う転送処理部 5 0 4 と、パケットを格納するパケットバッファを備えるメモリ 5 0 3 と、メモリ 5 0 3 に格納されたパケットのヘッダ情報を用いて経路検索を行う経路検索処理部 4 0 0 と、転送処理部 5 0 4 および通信制御部 5 0 1 を制御する制御回路 5 0 2 とを有する。経路検索処理部 4 0 0 は、ルーティングテーブルを格納するメモリ 3 0 0 と、そのルーティングテーブルを検索し、検索 50

結果を転送処理部 504 へ出力する経路検索処理回路 200 から構成される。

【0055】

次に、パケット処理部 600 における、マルチキャストパケットの中継処理の動作を説明する。なお、パケット中継処理を行う前に、管理部 601 が、スイッチ 505 に接続している全パケット処理部 600 に、それぞれのマルチキャストルーティングテーブルを配布しており、各パケット処理部 600 はメモリ 300 にルーティングテーブルを格納している状態とする。通信制御部 501 がネットワーク 500 から受信したマルチキャストパケットをメモリ 503 へ転送し格納する。転送処理部 504 はメモリ 503 に格納されたパケットデータの内、パケットヘッダ情報を抽出して経路検索処理部 400 へ転送する。経路検索処理回路 200 は、受信パケットのヘッダ内の送信元アドレスと宛先マルチキャスト・グループアドレスを用いてメモリ 300 内に格納されているマルチキャストルーティングテーブルの検索を行い、検索結果として、パケットを送信すべき N 個 (N は 1 以上の整数) の送信ポート番号を転送処理部 504 へ転送する。

10

【0056】

上記の検索結果を受信した転送処理部 504 は、先にメモリ 503 に格納されているパケットに対してパケットヘッダの書換などの処理を施し、パケットに上記 N 個の送信ポートの内の 1 つの番号を付加してスイッチ 505 に転送する。転送処理部 504 は上記の処理を N 回繰り返し、1 つの受信パケットに対する転送処理を終了する。また、送信ポート番号が自パケット処理部に接続されている通信制御部のポート番号である場合は、自パケット処理部に接続されている通信制御部にパケットを送信する。

20

【0057】

スイッチ 505 は転送処理部 504 により転送されたパケットを、付加されている送信ポート番号に従い、各パケット処理部にスイッチングする。上記スイッチングされたパケットを受信したパケット処理部 600 は、上記パケットを通信制御部 (501、602、603) へ転送する。通信制御部は、上記パケットのヘッダ内の宛先マルチキャストアドレスからマルチキャスト用 MAC アドレスを生成し、宛先 MAC アドレスとしてパケットに付加して各ネットワークに送信する。

【0058】

本発明は、上記の経路検索処理部 400 の高速化に関するものである。以下では、この経路検索処理部 400 に関し、まず、方式 1 として、マルチキャストルーティングテーブルを 2 分木に構成し、その 2 分木構造を検索する検索方式 (以下、2 分木検索方式と呼ぶ) について説明する。方式 2 として、方式 1 を改良し、さらに検索を高速に行うために、マルチキャストルーティングテーブルを 2 の p 乗分木に構成し、その 2 の p 乗分木構造を検索する検索方式 (以下、2 の p 乗分木検索方式と呼ぶ) について説明する。また、その方式 2 の説明の中で、マスク長 m ビットのノードをメモリ上に展開して方式 1、2 のノードを渡す回数を削減する方式 (方式 3) について説明する。最後に、本発明の一実施例として、方式 2 および 3 の方式を適用した経路検索処理を実現するハードウェア構成について説明する。

30

【0059】

まず、方式 1 の 2 分木検索方式について説明する。

40

【0060】

従来の技術で説明したように、マルチキャスト経路検索を行う際、受信パケットのヘッダ内の送信元 IP アドレスと、宛先マルチキャスト・グループアドレスを検索のキーとしてマルチキャストルーティングテーブルを検索する。すなわち、この 2 つの情報の組み合わせにより、そのマルチキャストパケットの送信元から宛先までの経路が識別される。

【0061】

従って、送信元 IP アドレスとマルチキャスト・グループアドレスを連結して、これを経路アドレスとして定義し、この経路アドレスと経路情報の組をエントリとしてマルチキャストルーティングテーブルを構成することにより、この経路アドレスをキーとして、マルチキャストルーティングテーブルを検索することができる。

50

## 【 0 0 6 2 】

また、上記の 2 つのアドレスのうち、マルチキャスト・グループアドレスに関しては、マルチキャスト・グループ一つに対して一つのアドレスが割り当てられているので、ルーティングテーブル検索時は、エントリに保持されたマルチキャスト・グループアドレスとの一致比較を行えばよい。

## 【 0 0 6 3 】

しかし、上記の 2 つのアドレスのうち、送信元 IP アドレスに関しては、受信パケットの送信元 IP アドレスに、エントリに保持されているサブネットマスク長だけ上位ビットから有効なマスクをかけたものとエントリに保持されている送信元サブネットワークアドレスに一致するエントリの経路情報を検索結果とするが、この場合、ユニキャスト経路検索と同様に最長一致検索を行う必要がある。

10

## 【 0 0 6 4 】

以上のことから、送信元 IP アドレスとマルチキャスト・グループアドレスの連結順を、マルチキャスト・グループアドレス、送信元 IP アドレスの順にして、これを経路アドレスとして定義し、この経路アドレスのビットパターンに従ってマルチキャストルーティングテーブルを 2 分木構造に構成することにより、従来技術 3 に述べられている Radish 方式を用いたユニキャスト経路検索方式と同様にして、経路アドレスに関して最長一致検索を行うことができる。この場合、経路アドレスの定義から、マルチキャスト・グループアドレスについては一致比較、送信元 IP アドレスに関しては最長一致検索を行うことができる。

20

## 【 0 0 6 5 】

図 5 に 2 分木で構成されたマルチキャストルーティングテーブルを検索する際に用いる経路アドレスを示す。なお、この経路アドレスは後述する 2 の p 乗分木検索方式にも採用する。図 5 (a) の経路アドレス 3 2 は、マルチキャスト・グループアドレス (3 0) と送信元 IP アドレス (3 1) とを連結したものであり、ビット数は 6 4 ビットである。図 5 (b) の経路アドレス 3 3 は、マルチキャスト・グループアドレス (3 0) から "1 1 1 0" という IP アドレスのクラス D を示す上位 4 ビットの値を除いたマルチキャスト・グループ ID (3 4) と、送信元 IP アドレス (3 1) とを連結したものであり、ビット数は 6 0 ビットである。木構造を構成するための値 (今の場合は検索アドレス) のビット数が少ない方が木の構造が浅くなり、検索時に迎える最大ノード数が減るため、以下では、ビット数の少ない図 5 (b) の経路アドレス 3 3 を採用して説明する。

30

## 【 0 0 6 6 】

次に、上記の経路アドレスを用いてマルチキャストルーティングテーブルを 2 分木に構成する方法について説明する。木構造の考え方は経路アドレスのビット数には依存しないので、簡単のため、マルチキャスト・グループ ID は 1 ビット、送信元 IP アドレスは 2 ビットと仮定して説明する。この場合経路アドレスは 3 ビットとなる。

## 【 0 0 6 7 】

図 6 に 2 分木の構造の例を示す。図 6 に示すように、各ノードは、経路アドレスモ a a a モと、経路アドレスマスク長 n、および、そのノードの下につく二つのノードへのポインタを持つ。ここで、経路アドレスマスク長とは、そのノードが持つ経路アドレスの上位何ビットが有効であることを示す値であり、また、そのノードでの受信パケットの経路アドレスの検査ビット位置も示す。また、経路アドレスと経路アドレスマスクの組を表記法モ a a a : n モで表す。この表記法により、マスク付き経路アドレスを表すことにする。

40

## 【 0 0 6 8 】

各ノードを木の上から順に、経路アドレスマスク長 0 ビット、1 ビット、2 ビット、3 ビットのノードと呼ぶ。

## 【 0 0 6 9 】

経路アドレスマスク長 0 ビットのノード (4 0) では、経路アドレスマスク長の値で示されている経路アドレスの第 0 ビットを検査し、その値が 0 か 1 かに従い左/右のポインタを辿ることにより経路アドレスマスク長 1 ビットのノード (4 1、4 2) に移り、経路ア

50

ドレスマスク長1ビットのノードでは、経路アドレスの第1ビットが0か1かに従い左/右のポインタを辿ることにより経路アドレスマスク長2ビットのノード(43、44、45、46)に移り、経路アドレスマスク長2ビットのノードでは、経路アドレスの第2ビットが0か1かに従い左/右のポインタを辿ることにより経路アドレスマスク長3ビットのノード(47、48、49、50、51、52、53、54)に移る。

#### 【0070】

検索したい経路アドレスについて、この木の経路アドレスマスク長0ビットのノード(40)から順に、各ビットが0か1かに従いポインタを辿った場合、経路アドレスマスク長0ビットのノードは経路アドレスがどの値をとる場合にも通過し、経路アドレスマスク長1ビットのノード(41、42)は左から順に経路アドレスの各ビットが0xx、1xxの場合に通過し、経路アドレスマスク長2ビットのノード(43、44、45、46)は左から順に経路アドレスの各ビットが00x、01x、10x、11xの場合に通過し、経路アドレスマスク長3ビットのノード(47、48、49、50、51、52、53、54)は左から順に経路アドレスの各ビットが000、001、010、011、100、101、110、111の場合に通過する。ここで、xは、そのビット値が0または1のどちらでも良いことを示す。

#### 【0071】

従って、経路アドレスマスク長0ビットのノード(40)は、経路アドレスがマスク付き経路アドレス 000:0に属する場合に通過し、経路アドレスマスク長1ビットのノード(41、42)は、経路アドレスがそれぞれマスク付き経路アドレス000:1、100:1に属する場合に通過し、経路アドレスマスク長2ビットのノード(43、44、45、46)は、経路アドレスがそれぞれマスク付き経路アドレス000:2、010:2、100:2、110:2に属する場合に通過し、経路アドレスマスク長3ビットのノード(47、48、49、50、51、52、53、54)は、経路アドレスがそれぞれマスク付き経路アドレス000:3、001:3、...、111:3に属する場合に通過する。

#### 【0072】

上記の通り、この木の各ノードは、経路アドレスと経路アドレスマスク長が異なる全マスク付き経路アドレスに1対1に対応している。

#### 【0073】

上記の2分木において、図7に示すマルチキャストルーティングテーブルの各エントリに対応するノード40、48、44、42、53に該エントリを割り付ける。具体的には、図12を用いて後述するように、ノードのデータ構造内にエントリが割り付けられているか否かを示すフラグを設け、エントリが割り付けられたノード内の上記フラグの値を1に設定し、かつ送信ポート情報あるいは送信ポート情報へのポインタをノード内に格納する。本明細書では、ノードに\*を付記することにより、ノードにエントリが割り付けられていることを表すことにする。

#### 【0074】

受信パケットの経路アドレスが011(図6の60)の場合、この木の上から各ビットが0か1かに従いポインタを辿ると、"\*"が付記されたノード40、44が、マスク付きの検索で一致することが分かる。そこで、ルーティングテーブルエントリが複数一致した場合は、送信元IPアドレスに関して最長一致検索を行うことを考慮すると、一致した"\*"付きノード40、44の内、上記で定義した経路アドレスマスク長が最も長いノード、すなわち最も末端に近いノード44に割り付けられたエントリの経路情報を、ルーティングテーブルの検索結果とする。なお、図7のルーティングテーブルの例では、エントリ内の経路情報は省略している。

#### 【0075】

上記検索方法から分かるように、"\*"が付記されておらず、かつ"\*"付きのノードにたどり着くための途中経路にもなっていないノード47、49、50、51、52、54、45は木から取り除いても、検索結果には影響しない。むしろ、最下のノードが、"\*"が付

10

20

30

40

50

記されていないノードの場合は最下まで移動せずに検索が終了するために効率的である。そこで、"\*"が付記されておらず、かつ"\*"付きのノードにたどり着くための途中経路にもなっていないノードを木から取り除くと図8のようになる。

#### 【0076】

上記の方法を用い、経路アドレス長が32ビットと仮定した場合、図9に示す経路テーブルに対応する2分木を構成すると、図10に示す分岐も"\*"も無いノードの長い列61ができる。なお、図9のルーティングテーブルの例では、図7と同様にエントリ内の経路情報は省略している。また、図10の0xNNNNNNNNという表記法は、NNNNNNNNが16進数表示であることを示す。このように、左右の片方のポインタだけに次のノードがつながり、かつエントリが割り当てられていないノードを取り除くことにより、検索時に辿るノード数を減らすことができる。

10

#### 【0077】

図10の例においては、分岐も"\*"も無いノード列61を取り除き、直ぐ上のノード62の分岐方向(図10では右側)に、取り除かれたノード列61の直ぐ下のノード63を付ける。その結果、図11に示す形となる。このように途中のノード列を取り除くことを、以後、木の縮退と呼ぶ。

#### 【0078】

次に縮退した2分木の検索法を説明する。

#### 【0079】

図11に示す例では、経路アドレスマスク長0ビットのノード62で第0ビットの検査を行った後、経路アドレスマスク長15ビットのノード63に跳ぶので、経路アドレスマスク長15ビットのノード63で第15ビットだけを検査したのでは、途中のビット、即ち第1ビットから第14ビットが検査できない。そこで、第1ビットから第15ビットの検査を一回の処理で行う為に、受信パケットの経路アドレスの第1ビットから第15ビットとノード63の経路アドレス0x85040000の第1ビットから第15ビットの一致比較を行う。比較結果が一致すれば正しいノードにたどり着いたこと、即ち、縮退しない木で1ビットずつ比較しても、このノードにたどり着いたことを意味し、一致しなければ、正しくないノードにたどり着いたこと、即ち縮退しない木では、行き先が無いことを意味する。

20

#### 【0080】

ここで、図11に示す例では、第0ビットは、既にテストされ、第0ビットが等しくなる方の分岐が選択されている為、常に一致する。一般に、あるノードにたどり着く毎に、正しいノードにたどり着いたか否かを検査していれば、第0ビットから、そのノードの経路アドレスマスク長までのビットに関しては、受信パケットの経路アドレスと、ノードの経路アドレスとが等しいことが保証されているので、次のノードにたどり着いたときに、前にどのビットまで検査したかに関らず、第0ビットからノードの経路アドレスマスク長までのビットに関して、受信パケットの経路アドレスとノードの経路アドレスとが等しいか否かを調べて良い。

30

#### 【0081】

次に、上記で説明した2分木を構成するノードのデータ構造について、図12を用いて説明する。図12(a)において、ワード110、111内の次のノードのマスク長0、1は、このノード自身の経路アドレスマスク長ではなく、このノードの直下のノードの経路アドレスマスク長である。このように、自分自身でなく直下のノードのマスク長を設定する理由は高速化のためであり、その説明は図20を用いて後述する。ワード110、111内のフラグ0、フラグ1は、このノードがエントリが割り付けられているノードか否か、即ち、図6、8、10、11で示した木の例において、このノードが、"\*"が付記されているノードか否かを示すフラグ(以下、エントリ有りフラグと呼ぶ)や、このノードに一致するパケットに関する属性フラグなどが設定されるフィールドである。フラグ0とフラグ1には、同じ値を設定する。これは、ワード110とワード111の一つだけを読めば良いようにするためであり、このように、ノードの全てを読むのではなく、一部分を読むことによる高速化については図20を用いて後述する。ワード110、111内の次のノ

40

50

ドへのポインタ 0、1 は、経路アドレスの、このノードの経路アドレスマスク長で示されるビット位置の値が、それぞれ 0、1 のときに次に辿るノードへのポインタである。ワード 1 1 2、1 1 3 の経路アドレスは、このノードに対応する経路アドレスであり、例えばワード 1 1 2 にはノードの経路アドレスの内、上位 2 8 ビットを設定し、ワード 1 1 3 にはノードの経路アドレスの内、下位 3 2 ビットを設定する。

#### 【0082】

図 1 2 (a) では、2 分木ノードのデータ構造の中に送信ポート情報 1 1 4 を保持する例を示している。図 1 2 (b) では、送信ポート情報 1 1 4 を 2 分木ノード構造から分離し、分離した送信ポート情報へのポインタ 1 1 5 を保持する例を示している。送信ポート情報は、受信マルチキャストパケットを転送すべき 1 つまたは複数の送信ポート番号を示し、ルータのポート数が多い場合は数ワード必要になる。従って、図 1 2 (a) のように、2 分木ノード構造内に送信ポート情報 1 1 4 を保持する場合、エントリが割り付けられている 2 分木ノードと、エントリが割り付けられていない 2 分木ノードで、その大きさが大きく異なることになる。このようにノードの大きさが異なると、ノード内の次ノードへのポインタに必要なビット数が多くなることや、そのポインタから実際のメモリのアドレスを計算する処理が複雑になり、特にハードウェア構成が複雑になるという問題が考えられる。

10

#### 【0083】

図 1 2 (b) のように送信ポート情報をノードから分離し、ノードにはその送信ポート情報へのポインタ 1 1 5 を保持する場合、エントリが割り付けられている / いらないに関わらずノードの大きさは一定になり、ハードウェア構成を単純にする利点がある。

20

#### 【0084】

次に、方式 2 の 2 の p 乗分木検索方式について説明する。

#### 【0085】

上記で説明した 2 分木方式では、一つのノードに 2 つの分岐先があり 1 ビットずつ検索していた。検索のキーのビット数は、検索時に辿るノードの数 (の最大値) に一致する。したがって、最大の検索時間は (式 1) で表される。

#### 【0086】

(最大検索時間) = (一つのノードの処理時間) × (検索のキーのビット数) …… (式 1)  
すなわち、最大検索時間は検索のキーのビット数に比例する。この最大検索時間をさらに大幅に短縮し、高速化することが求められる。

30

#### 【0087】

そこで、以下で説明する 2 の p 乗分木方式では、一つのノードに 2 の p 乗の分岐先を設け、同時に p ビットずつ検索することにより、従来に比べ、検索時間を 1 / p に短縮するものである。以後、一つのノードに 2 の p 乗の分岐先があるノードのことを 2 の p 乗分木ノードと呼ぶ。

#### 【0088】

2 の p 乗分木ノードは、2 分木方式で説明した 2 分木ノードから構成される木を変形することにより作成する。木の変形の方法は、経路アドレスマスク長 n の 2 分木ノード一つと、この 2 分木ノードの下に存在する、経路アドレスマスク長 (n + 1) ビットから (n + p - 1) ビットの 2 分木ノードを、一つの 2 の p 乗分木ノードに対応させるものである。また、以下では、上記のように対応させてまとめた 2 の p 乗分木ノードに関し、まとめる前に頂点にあった 2 分木ノードのマスク長 n を引き継いで、この 2 の p 乗分木ノードのマスク長を n と定義する。

40

#### 【0089】

上記の変形法の例として、2 分木から 8 分木への変形法を図 1 3 に示す。

#### 【0090】

経路アドレスのビット数が 6 0 ビットの場合で、8 分木を構成する場合を考える。この場合、一つの 8 分木に対応させる 2 分木ノードの選び方として、マスク長 0 ~ 2、3 ~ 5、…、5 7 ~ 5 9、6 0 ビットの 2 分木ノードを、それぞれ一つの 8 分木ノードとする場合 (図 1 3 (a)) と、マスク長 1 ~ 3、4 ~ 6、…、5 5 ~ 5 7、5 8 ~ 6 0 ビットの 2

50



分木ノードを、それぞれ一つの8分木ノードとする場合(図13(b))と、第2~4、5~7、...、56~58、59~60ビットの2分木ノードを、それぞれ一つの8分木ノードとする場合(図13(c))の3通りがあり、どの区切り方でも、構成可能だが、木全体では、エントリの追加、削除を容易に行えるように、上記3通りのビット位置の区切り方の内、一つを使用する。

#### 【0091】

上記3通りのビット位置の区切り方の内、最初の区切り方以外では、マスク長が0ビットから始まっていないので最初のビットの検査を別に行う必要がある。この検査には、マスク長 $m$ ビットのノードをメモリ上の決まった位置に展開する方法(以下、方式3と呼ぶ)を使用する。

10

#### 【0092】

以下、方式3について説明する。方式3は、2の $p$ 乗分木構造のマスク長 $m$ ビットのノードを、ノードがある場合も無い場合も区別せず全て予め用意し、メモリ上の決まった位置に展開しておき、経路検索開始時に、受信パケットの経路アドレスの上位 $m$ ビットの値に従って上記の展開されたノードの内の一つを選択してメモリから読み込む。この方式により上位 $m$ ビット分の検索時間を省くことができる。

#### 【0093】

上記の方式3を適用した例を、図13(b)、(c)に示した区切り位置の場合について図14(a)、(b)に示す。

#### 【0094】

20

図14(a)に示す構成では、マスク長1ビットの8分木ノード90、91をメモリ上の決まった位置に並べ、それぞれを、受信パケットの経路アドレスの第0ビットが0か1かに従い選択することにより、最初の第0ビットの検査を行ったことになる。図14(b)に示す構成では、マスク長2ビットの8分木ノード92、93、94、95をメモリ上の決まった位置に並べ、それぞれを、受信パケットの経路アドレスの第0~1ビットの値が00か01か10か11かに従い選択することにより、最初の第0、1ビットの検査を行ったことになる。

#### 【0095】

さらに、図13(a)、(b)、(c)に示す各ビットの区切り位置の場合に、それぞれ最初に並べる8分木ノード数を1、2、4個ではなく、これらの8倍である8、16、32個、8の2乗倍である64、128、256個、或いは一般に8の $q$ 乗倍個にし、最初の1回、2回、或いは一般に $q$ 回の8分木ノードの検索時間を無くすることも可能である。この場合、メモリ上の決まった位置に展開する8分木ノードのマスク長 $m$ は、図13(a)、(b)、(c)に示す各ビットの区切り位置の場合、それぞれ、 $m = 3 \times q$ 、 $1 + 3 \times q$ 、 $2 + 3 \times q$ となり、この上位 $m$ ビット分の検索時間を省くことができる。 $q = 1$ の場合、即ち8分木ノードの1回のノードの検索時間を無くす場合で、図13(a)、(b)、(c)に示す3通りの各ビットの区切り位置の場合についてのメモリ上へのノードの展開法を、図15(a)、(b)、(c)にそれぞれ示す。

30

#### 【0096】

図15(a)では、マスク長 $m = 3$ の8分木ノードを8個メモリ上の決まった位置に展開し、受信パケットの経路アドレスの第0~2ビットの計3ビットの値に従って、展開された8分木ノードの内の一つを選択する。図15(b)では、マスク長 $m = 1 + 3 = 4$ の8分木ノードを16個メモリ上の決まった位置に展開し、受信パケットの経路アドレスの第0~3ビットの計4ビットの値に従って、展開された8分木ノードの内の一つを選択する。図15(c)では、マスク長 $m = 2 + 3 = 5$ の8分木ノードを32個メモリ上の決まった位置に展開し、受信パケットの経路アドレスの第0~4ビットの計5ビットの値に従って、展開された8分木ノードの内の一つを選択する。

40

#### 【0097】

以上、8分木を例にして方式3について説明したが、同様に、マスク長 $m$ ビットの2の $p$ 乗分木ノードをメモリ上の決まった位置に展開し、上位 $m$ ビット分の検索時間を省く

50

ことができる。p、qの値を大きくすると、経路検索時間を短縮することができるが、多くのメモリを必要とするので、p、qの値は、メモリ効率と性能のトレードオフから決めるようにする。

【0098】

以上、方式3について、2のp乗分木ノードの場合について述べたが、同様にして、方式1の2分木検索方式にこの方式3を適用することも可能である。

【0099】

次に、4分木ノード、8分木ノード、16分木ノード、或いは一般に2のべき乗分木のノードの構成法を図16を用いて説明する。

【0100】

図16は4分木で、ある1つの2分木ノードA、B、C、D、Eと、その直下の2個の2分木ノードA0、A1、B0、B1、C0、C1、D0、D1、E0、E1の、各々合計3個の2分木ノードをまとめて一つの4分木ノード100、101、102、103、104にする場合の例であり、合計3個の2分木ノードをつぶして、下の方の2分木ノードだけの大きさにする。つぶし方は、経路アドレスマスク長に関する最長一致検索の仕様に従い、経路検索を行った場合に、2分木の場合と4分木の場合とで、経路検索結果が同じになる、という条件を満たすように行う。

【0101】

4分木の場合の、このノードのつぶし方を図17に示す。図17において、\*Aという表記は、ノードAに割り付けられているエントリ内の経路情報を表すことにする。また、\*Aという表記がないノードにはエントリが割り付けられていないことを意味する。

【0102】

2分木構造において、3つのノードが全部ある場合(図17(a))、全ノードにエントリが割り付けられていたら、下のノードの経路情報\*A0、\*A1を残し、上のノードは削除する。これは、ノードAのエントリが一致したらノードA0かノードA1のどちらかのエントリが必ず一致するので、最長一致検索を行うことから、ノードAの経路情報\*Aが使用されることが無いからである。

【0103】

上のノードAにエントリが割り付けられており、下のノードA0、A1の内A1にだけエントリが割り付けられていない場合(図17(b))、ノードA1に、Aの経路情報\*Aを格納する。下のノードA0、A1の内A0だけエントリが割り付けられていない場合も同様である。

【0104】

下のノードA0、A1の両方にエントリが割り付けられていない場合(図17(c))には、A0、A1の両方のノード内に、Aの経路情報\*Aを格納する。

【0105】

下のノードA1が無い場合(図17(d))、A1を補い、A1には、上のノードAの経路情報\*Aを格納し、さらに、ノードA1の下にはノードが繋がっていないので、ノードA1の下の方のノードへのポインタはNULLにする。下のノードA0、A1の内A0だけエントリが割り付けられていない場合も同様である。

【0106】

下のノードA0、A1の両方が無い場合(図17(e))、両方を補い、両方のノードに、Aの経路情報\*Aを格納し、両方のノードの下の方のノードへのポインタはNULLにする。

【0107】

上のノードAにエントリが割り付けられていない場合(図17(f))、ノードAをただ単に削除する。

【0108】

上のノードAに経路が割り付けられておらず、下のノードA0、A1の内、A1にもエントリが割り付けられていない場合(図17(g))、4分木にした場合もA1の経路情報は無い。下のノードA0、A1の内、A0にエントリが割り付けられていない場合も同様であ

10

20

30

40

50

る。

#### 【 0 1 0 9 】

上のノード A にエントリが割り付けられてなく、下のノード A 0、A 1 の両方にエントリが割り付けられていない場合(図 1 7 (h))、4 分木にした場合も A 0、A 1 の両方のノードの経路情報は無い。

#### 【 0 1 1 0 】

下のノード A 0 だけしかない場合(図 1 7 (I))には、下のノード A 1 を補う。下のノード A 1 だけしかない場合も同様である。

#### 【 0 1 1 1 】

8 分木の場合も同様にして、一つにまとめる 7 個の 2 分木ノードをつぶして、一番下の 4 個のノードだけの大きさにする。上の方の 3 つの 2 分木ノードのつぶし方の 2 つの例を図 1 8 に示す。

#### 【 0 1 1 2 】

図 1 8 (a) は一つにまとめる 7 個の 2 分木ノードが全てあるが、その内のいくつかにはしかエントリが割り付けられていない場合の例であり、最下の 4 つのノードの内、エントリが割り付けられていないノード A 0 1、A 1 0 には、そのノードの上方につながっているノードの内、エントリが割り付けられている最も下、即ち最も経路アドレスマスク長が長いノード(それぞれ、A、A 1)の経路情報 \* A、\* A 1 を格納する。

#### 【 0 1 1 3 】

図 1 8 (b) は一つにまとめる 7 個の 2 分木ノードの内のいくつかしかノードが存在しない場合の例であり、存在しないノード A 0 1、A 1 0 を、まずエントリが割り付けられていないノードとして補い、図 1 8 (a) と同じ規則で、経路情報を格納する。最下の 4 つのノード A 0 0、A 0 1、A 1 0、A 1 1 の内、補ったノード A 0 1、A 1 0 の下には、ノードが繋がっていないので、下のノードへのポインタは NULL にする。

#### 【 0 1 1 4 】

一般に 2 の p 乗分木の場合も同様にして、一つにまとめる (2 の p 乗 - 1) 個の 2 分木ノードをつぶして、一番下の 2 の (p - 1) 乗個のノードだけの大きさにする。

#### 【 0 1 1 5 】

また、4、8、16、... 分木ノードでは、2 分木ノードを 2、4、8、... 個まとめて扱うので、1 つの 4、8、16、... 分木ノードにまとめられる 2 分木ノード間で共通な要素は、一つだけ持てばよく、これにより、4、8、16、... 分木ノードのメモリ量を小さくできる。1 つの 4、8、16、... 分木ノードにまとめられる 2 分木ノード間で共通な要素には、ノードが持つ経路アドレス、及び、経路アドレスマスク長があるが、経路アドレスマスク長については、後述するように、このノード自身の経路アドレスマスク長ではなく、このノードの直下のノードの経路アドレスマスク長を持つようにするので、メモリ量を小さくする効果は無い。

#### 【 0 1 1 6 】

次に、2 の p 乗分木ノードの一つの例として、4 分木ノードのデータ構造について図 1 9 を用いて説明する。4 分木ノードを構成する際、上記 2 分木ノードを 2 つまとめ、まとめられる 2 つの 2 分木ノードに共通のデータを一つだけ保持するようにする。この共通のデータは、そのノードの経路アドレスだけである(ワード 1 2 4、1 2 5)。また、まとめられる 2 分木ノードが、経路が割り付けられているか否かによりその大きさが異なると、それらをまとめて構成する 4 分木ノードの大きさも異なってしまう。従って、図 1 2 (b) で示したように、送信ポート情報をノードから分離し、ノード内にはこの送信ポート情報へのポインタを保持する 2 分木ノードをまとめる。

#### 【 0 1 1 7 】

図 1 9 において、ワード 1 2 0、1 2 1、1 2 2、1 2 3 内の次ノードのマスク長 t t、フラグ t t、次ノードへのポインタ t t はそれぞれ、受信パケットの経路アドレスの、このノードの経路アドレスマスク長 m で示された第 m ビット、および第 m + 1 ビットの値 t t に対応する値である。ここで t t は 0 0、0 1、1 0、1 1 の値を取りうる。2 分木の

10

20

30

40

50

場合と同様に、フラグ 0 0 とフラグ 0 1 は同じ値を設定し、フラグ 1 0 とフラグ 1 1 も同じ値を設定する。

【 0 1 1 8 】

図 1 2 ( b ) に示す例では、2 分木ノードは、2 のべき乗の大きさである 1 6 バイトにわずかに入りきらない大きさになっているが、4 分木ノードにし、1 ノード内に経路アドレスを一つしか持たないようにすることで、2 のべき乗の大きさである 3 2 バイトに丁度収まるようになっている。8 分木ノードにし、経路アドレスを 1 ノードで一つだけ保持するようにすれば、6 4 バイトの大きさに収まる上に、4 バイトの余裕ができ、この領域は他の情報を入れるのに使用できる。さらに 1 つにまとめる 2 分木の数を増やせば(すなわち、2 の p 乗分木の p の値を大きくとれば)、2 のべき乗の大きさに対し、1 つのノードの大きさをさらに小さくできる。

10

【 0 1 1 9 】

このように、1 ノードの大きさを 2 のべき乗の大きさに収めることで、ハードウェアの構成を非常に簡単にできる。ハードウェアの構成を簡単にできる例を以下に示す。

【 0 1 2 0 】

1 つ目の例として、4 分木ノードが 3 2 バイトに収まる場合、メモリを複数バンクで構成していた場合でも 1 ノードのメモリ領域がバンク境界にまたがることのないこと、メモリにダイナミック R A M を使用していた場合でも 1 ノードのメモリ領域が R o w アドレス境界にまたがることのないこと。

【 0 1 2 1 】

20

2 つ目の例として、ノード内の各要素のメモリ上のアドレス(以下、メモリアドレスと呼ぶ)を求めるときに、そのノードへのポインタとそのポインタからのオフセットの足し算でなく、メモリアドレスの上位ビットはポインタの値にし、下位ビットをオフセットにすればよいこと。例えば 4 分木ノードが 3 2 バイトに収まる場合、あるノード内のある要素のメモリアドレスを生成するには、そのノードへのポインタをメモリアドレスの 2 の 5 乗ビット以上に割り付け、ノード内のその要素へのオフセットをメモリアドレスの 2 の 0 乗ビットから 2 の 4 乗ビットに割り付ければよい。

【 0 1 2 2 】

3 つ目の例として、例えば 4 分木ノードが 3 2 バイトに収まる場合、各ノードに保持する次のノードへのポインタとして、(次のノードの先頭のバイトアドレス) ÷ 3 2、という値を保持すればよく、1 ノード内で 1 ポインタあたり 5 ビットづつデータ量を減らせることが挙げられる。

30

【 0 1 2 3 】

次に、例えば、図 1 の経路検索処理回路 2 0 0 を専用 L S I で構成する場合のノードの読み込み方式について説明する。4、8、1 6、... 分木ノードにした場合に 1 つのノードが大きくなり、検索処理時にテーブルを保持するメモリから専用 L S I 内に 1 つのノードを全て読み込むと、ノードを大きくするに従い読み込み時間が伸び、性能低下要因となる、という問題があるが、この問題は、ノードを大きくしたときに、1 つのノード全てを読み込まず一部だけを読み込む、という方で回避する。この方法について、図 2 0 を用いて説明する。

40

【 0 1 2 4 】

図 2 0 に 4 分木の場合の例を示す。既に図 1 9 を用いて説明したように、経路アドレスマスク長 m ビットの 4 分木ノードは、図 1 2 ( b ) で示した、経路アドレスの第 m ビット目の値が 0 の場合に対応する 2 分木ノードと、1 の場合に対応する 2 分木ノードを併せた構造になっている。従って、受信パケットの経路アドレスの第 m ビット目の値にしたがって、対応する方の 2 分木ノードの部分だけを読み込むことにより、ノードの大きさが大きくなっても 2 分木ノードの場合と同じデータ量を読み込むようにする。このとき、図 1 9 で示した、1 つにまとめられる 2 分木ノード間で一つだけ保持する要素であるノードの経路アドレス(ワード 1 2 4、1 2 5)は、受信パケットの経路アドレスの m ビット目の値に係わらず読み込むようにする。

50

## 【0125】

さらに、受信パケットの経路アドレスの第 $(m+1)$ ビット目の値に従い、2分木ノードで2つ存在した次ノードへのポインタの内、一方だけを読み込むようにすることにより、読み込むデータ量をさらに少なくする。

## 【0126】

この方法は2分木検索方式にも使用できる。例えば第 $m$ ビット目の2分木ノードの場合には、受信パケットの経路アドレスの第 $m$ ビット目の値に従い、2つの次ノードへのポインタの内、一方だけを読み込むようにする。

## 【0127】

上記方法を全て行い、結局、このノードの経路アドレスマスク長を $m$ とした場合、宛先アドレスの第 $m$ 、 $m+1$ ビットの値が00か、01か、10か、11かに従い、それぞれ(ワード120 ワード124 ワード125 ワード126)、(ワード121 ワード124 ワード125 ワード126)、(ワード122 ワード124 ワード125 ワード127)、(ワード123 ワード124 ワード125 ワード127)の順にデータを読み込むようにする(図20(b))。

10

## 【0128】

このように、あるノードの一部分だけを読み込むためには、このノードの経路アドレスマスク長 $m$ を知る必要があり、このノードの経路アドレスマスク長 $m$ は、1ノードのデータ読み込みの最初に読み込むか、この情報を1つ前のノードに移し、一つ前のノードのデータ読み込み時に読み込む必要がある。ノードのマスク長 $m$ を1ノードのデータ読み込みの最初に読み込む方法は、例えば、図1の経路検索処理回路200を専用LSIで構成する場合、宛先の第 $m$ ビット目の値の抽出のための専用LSI内のゲートディレイ、及び、次に読み込む部分のメモリアドレスをメモリに出力してからメモリからのデータを専用LSI内に読み込むまでの時間であるメモリリードレイテンシだけ、マスク長 $m$ を読み込んでから次に読み込む部分を選択して読み込むまで時間が空いてしまうので、ノードの一部だけを読むことによる性能向上効果が少ない。ノードのマスク長 $m$ を1つ前のノードに移し、一つ前のノードのデータの読み込み時に読み込む方が性能向上効果がある。

20

## 【0129】

さらに、ノードの経路アドレスマスク長 $m$ を1つ前のノードに移す場合、1ノードのデータを読み込む順序を、1番目に次のノードの経路アドレスマスク長 $m$ 、次のノードへのポインタ、次にノードの経路アドレス、送信ポート情報へのポインタの順にすることにより、次のノードの最初に読み込む部分のメモリアドレスが最も早く計算できるようにする。

30

## 【0130】

次のノードへのポインタは、次のノードのメモリ領域の先頭部分を指しており、次のノードの先頭から最初に読み込む部分までのメモリアドレスのオフセットは、次のノードのマスク長 $m$ を読み込み、受信パケットの経路アドレスの該当ビット位置の値を検査することにより、得られる。

## 【0131】

次に、1ノード内で、条件によっては、読み込む必要が無い要素を、その条件に従って読み込まないようにすることで、読み込みの時間を削減し、高速化を図る方法について図21を用いて説明する。

40

## 【0132】

図21に4分木の場合の例を示す。2分木構造の説明で述べたように一般に木構造においては、全ノードにエントリが割り付けられているわけではなく、枝の分岐の個所では、エントリが割り付けられていなくてもノードを設ける必要がある。図21に示すように、ノードデータの最初に読み込むワード120、121、122、123内のフラグ内のエントリ有りフラグの値に従い、エントリが割り付けられていないノードでは、送信ポート情報へのポインタを読み込まないようにすることで、読み込み時間の短縮を図ることができる。エントリ有りフラグは1ビットで表現できるので、この情報を読み込むことによる読み込み時間の増大は小さい。

50

## 【0133】

この方法では、このノードのマスク長を $m$ とすると、宛先アドレスの第 $m$ 、 $m+1$ ビットの値が00で、ワード120を読み、フラグ00から、4分木を構成する0番目の2分木に経路情報が無いと判った場合、ワード124、125だけを読めば良く、経路情報が有ると判った場合にだけ、図21(b)に示すようにワード124　ワード125　ワード126の順に読めば良い。宛先アドレスの第 $m$ 、 $m+1$ ビットの値が01、10、11の場合も同様である。

## 【0134】

以上、2分木検索方式(方式1)、2の $p$ 乗分木検索方式(方式2)、およびマスク長 $m$ ビットのノードをメモリ上に展開する方法(方式3)という各検索方式について説明した。次に、図1の経路検索処理回路200が上記の方式を用いて経路検索処理を行う際のフローチャートについて図22を用いて説明する。図22のフローチャートでは、方式2と方式3とを組み合わせた場合の例を示す。この例では、経路アドレスの第0ビットから第 $(m-1)$ ビットの値に従って、メモリ上の決まった位置に展開された2の $p$ 乗分木ノードの一つを選択する。以下ではこのように選択し、検索の最初に読込む2の $p$ 乗分木ノードを初段ノードと呼ぶ。第 $m$ ビット以降は経路アドレスを $p$ ビットずつ検索し、2の $p$ 乗分木を検索する。図1のメモリ300には、上記の方式2および方式3の検索方式に従った2の $p$ 乗分木ノードデータ、および送信ポート情報が格納されているとする。

10

## 【0135】

なお、以下では方式2と方式3とを組み合わせた例について説明するが、方式1と方式3とを組み合わせた場合でも同様なフローチャートに従うことで実現可能である。また、このフローチャートに従うことにより、ソフトウェアでもハードウェアでも経路検索処理を実現することができる。ソフトウェアで実現する場合、図1の経路検索処理回路200にはCPUを用いればよい。またハードウェアで実現する場合、図1の経路検索処理回路200を専用LSIで構成すればよい。

20

## 【0136】

図22の処理810は木構造検索処理であり、また、処理811は送信ポート情報処理である。まず、木構造検索処理810について説明する。

## 【0137】

図1の経路検索処理回路200は、受信パケットの宛先マルチキャスト・グループIDおよび送信元IPアドレスを受け取ると、これらの値から図5を用いて説明した経路アドレスを生成し、この経路アドレスと初段ノードのマスク長 $m$ の値から初段ノードへのポイントを生成し、このポイントと、経路アドレスの第 $m$ ビットから第 $(m+p-1)$ ビットの値(以下、検査ビット値と呼ぶ)に従ってメモリ300に格納されている初段ノードの読み込みアドレスを生成し、メモリ300から該初段ノードの一部を読み込む(図22の800)。

30

## 【0138】

次に、図1の経路検索処理回路200は、受信パケットの経路アドレスにノードのマスク長だけ上位ビットから有効とするマスクをかけたものと、ノードの経路アドレスと比較し(図22の801)、不一致の場合は木構造検索処理810を終了する(図22の809)。一致する場合は図22の処理802に進む。

40

## 【0139】

次に、図22の処理802、および803について説明する。これらの処理は、最長一致検索を実現するための経路情報の更新処理である。経路情報としては、図1のメモリ300が出力するノードデータの内、エントリ有りフラグと、送信ポート情報へのポイントがある。図1の経路検索処理回路200は、ノードデータ内のエントリ有りフラグを検査し、エントリ有りフラグの値が1の場合(図22の812)のみ、読み込んだノード内の新たな経路情報をレジスタに保持する(図22の803)。エントリ有りフラグの値が0の場合は更新処理を行わない(図22の813)。

## 【0140】

50

次に、図 1 の経路検索処理回路 200 は、図 1 のメモリ 300 が出力するノードデータの内の次ノードへのポインタが NULL かどうかを判定し(図 22 の 804)、NULL の場合は木構造検索処理 810 を終了する。NULL でない場合はそのポインタと、新たな検査ビットの値に従ってメモリ 300 に格納されている初段ノードの読み込みアドレスを生成し、メモリ 300 から該ノードデータを読み込む(図 22 の 805)。

#### 【0141】

以下、上記の処理を繰り返すことにより、2 の p 乗分木方式の経路検索を行うことができる。

#### 【0142】

次に、図 22 の送信ポート情報処理 811 について説明する。木構造検索処理の結果、図 1 の経路検索処理回路 200 内のレジスタには、経路情報として、上記のエントリ有りフラグと送信ポート情報へのポインタが保持されている。図 1 の経路検索処理回路 200 は、まず、レジスタに保持されているエントリ有りフラグを調べ(図 22 の 806)、その値が 0 の場合は経路検索処理を終了し、転送処理部 504 へ検索結果無しという通知をする。エントリ有りフラグの値が 1 の場合は、検索の結果、あるエントリに一致したことになるため、送信ポート情報へのポインタを用いてメモリ 300 から送信ポート情報を読み出す(図 22 の 807)。この送信ポート情報からパケットを送信すべき 1 つまたは複数のポート番号を生成して図 1 の転送処理部 504 へ送信し(図 22 の 808)、すべてのポート番号を送信し終わると経路検索処理を終了する。

#### 【0143】

次に、本発明の一実施例として、図 22 で説明した検索方式をハードウェアで実現する場合の構成例について、図 23、図 24 を用いて説明する。

#### 【0144】

図 23 に経路検索処理部 400 をハードウェアで構成した場合の構成例を示す。経路検索処理回路 200 は、木構造検索回路 201 と、読み込みアドレス生成回路 202 と、メモリ制御回路 206 と、送信ポート情報処理回路 203 と、経路検索処理制御回路 204 とからなる。

#### 【0145】

木構造検索回路 201 は、メモリ 300 に格納された 2 の p 乗分木構造を検索し、次に読み込むべきノードのポインタの生成、受信パケットの経路アドレスの検査ビット値の抽出、木構造検索の終了判定、検索結果である経路情報の候補の更新を行う。また、読み込みアドレス生成回路 202 は、木構造検索処理時には、木構造検索回路 201 から出力される読み込むべきノードへのポインタ、および検査ビット値、および経路検索処理制御回路 204 から出力されるタイミング信号に従い、図 20 を用いて説明した順序でノード内の読み込みワードのメモリアドレスを生成してメモリ制御回路 206 に送信する。また、送信ポート情報処理時には、木構造検索回路 201 内のレジスタに検索結果として保持されている送信ポート情報へのポインタから送信ポート情報のメモリアドレスを生成する。また、メモリ制御回路 206 は、上記のメモリアドレスと、経路検索処理制御回路 204 から出力されるタイミング信号に従い、メモリ制御信号を生成する。また、送信ポート情報処理回路 203 は、メモリ 300 から読み出された送信ポート情報からパケットを送信すべき 1 つまたは複数のポート番号を生成して転送処理部 504 へ送信する。また、経路検索処理制御回路 204 は、経路検索処理回路 200 全体の制御(各回路の動作タイミングおよび動作状態管理など)を行う。

#### 【0146】

上記の経路検索処理回路 200 を専用 LSI で実現する場合、メモリ 300 は上記専用 LSI の外部に備えても良いし、専用 LSI 内の内蔵メモリを使用しても良い。メモリ 300 として専用 LSI の内蔵メモリ使用すると、外部メモリを使用する場合に比べてメモリアクセス時間が短くなり、メモリ 300 内に格納されたルーティングテーブルの検索処理を高速に行うことができる。

#### 【0147】

10

20

30

40

50

次に、経路検索処理部 400 の動作について図 23 を用いて説明する。また、木構造検索回路 201 の詳細動作については、図 24 を用いて後述する。

【0148】

木構造検索回路 201 は、転送処理部 504 から受信パケットの宛先マルチキャスト・グループ ID および送信元 IP アドレスを受け取ると、これらの値から経路アドレスを生成し、この経路アドレスとノードのマスク長の値から次ノードへのポインタを生成して、読み込みアドレス生成回路 202 に転送する。また、木構造検索回路 201 は、ノードのマスク長で示される経路アドレスの検査ビット位置の値(検査ビット値)を抽出して、読み込みアドレス生成回路 202 に転送する。

【0149】

読み込みアドレス生成回路 202 はこのノードへのポインタと、検査ビット値と、経路検索処理制御回路 204 からのタイミング信号を用いて、図 20 を用いて説明した順序でノード内の読み込みワードのメモリアドレスを生成し、メモリ制御回路 206 に送信する。メモリ制御回路 206 は上記メモリアドレスと経路検索処理制御回路からのタイミング信号を用いてメモリ制御信号を生成し、メモリ 300 へ出力する。上記のメモリ制御信号を受信したメモリ 300 は、対応するノードデータを信号線 214 を用いて木構造検索回路へ転送する。

【0150】

木構造検索回路 201 はこのノードデータを用いて、図 22 の処理 801、802、803、804、805 を行う。これらの処理の詳細は図 24 で後述する。図 22 の処理 801 および 804 に対応する判定処理において木構造検索を終了すると判定した場合は、木構造検索終了信号を経路検索処理制御回路 204 へ出力し、経路検索処理制御回路 204 は送信ポート情報読み込み処理を開始する。終了と判定されない場合は、終了と判定されるまで図 22 の処理 801、802、803、804、805 を繰り返す。

【0151】

次に、経路検索処理制御回路 204 は、木構造検索回路 201 内に保持された経路情報(エントリ有りフラグ、送信ポート情報へのポインタ)のうちのエントリ有りフラグを調べ、その値が 0 の場合は経路検索処理を終了し、転送処理部 504 へ検索結果無しという通知をする。エントリ有りフラグの値が 1 の場合は、送信ポート情報読み込みおよび送信ポート生成処理制御を開始し、木構造検索回路 201 を制御して、送信ポート情報へのポインタを読み込みアドレス生成回路 202 へ出力させる。読み込みアドレス生成回路 202 はこの送信ポート情報へのポインタと、経路検索処理回路 204 からのタイミング信号を用いて、読み出すべき送信ポート情報が格納されているメモリアドレスを生成し、メモリ制御回路 206 へ送信し、メモリ制御回路 206 は、上記のメモリアドレスと、経路検索処理制御回路 204 から出力されるタイミング信号に従い、メモリ制御信号を生成してメモリ 300 へ出力し、この制御信号を受信したメモリ 300 から信号線 214 上に送信ポート情報が出力される。送信ポート情報処理回路 203 はこの送信ポート情報を取り込み、パケットを送信すべき 1 つまたは複数のポート番号を生成して転送処理部 504 へ送信する。

【0152】

送信ポート情報の構造については、複数の送信ポート番号をポインタでつないだリスト構造や、送信ポート番号のビットマップ化などが考えられるが、リスト構造の場合、送信ポート番号が多いとメモリアクセス回数が増加し、結果として検索処理が遅くなる問題がある。送信ポート番号をビットマップ化する方法は、そのビットマップをメモリ 300 からバーストリードして送信ポート情報処理回路 203 内に保持しておき、そのビットマップをデコードして送信ポート番号を生成することにより、メモリアクセス回数を減らすことができる。

【0153】

すべての送信ポート番号を出力した後で、経路検索処理制御回路 204 は経路検索処理を終了し、次のパケット処理の制御を行う。

10

20

30

40

50



## 【 0 1 5 4 】

次に、図 2 3 の木構造検索回路 2 0 1 の詳細を図 2 4 を用いて説明する。

## 【 0 1 5 5 】

まず、図 2 2 の初段ノードリード処理 8 0 0 および次ノードリード処理 8 0 5 に対応する処理について、図 2 4 を用いて説明する。転送処理部 5 0 4 より信号線 2 0 5 を用いて送信される宛先マルチキャスト・グループアドレスおよび送信元 IP アドレスは、経路アドレス生成回路 2 0 7 内に保持される。経路アドレス生成回路 2 0 7 はこれらの情報から、図 5 (b) で説明した 6 0 ビットの経路アドレス 3 3 を生成し、初段ノードへのポインタ生成回路 2 0 8、検査ビット抽出回路 2 0 9、マスク処理回路 2 1 0 へ出力する。

## 【 0 1 5 6 】

初段ノードへのポインタ生成回路 2 0 8 は、予め初段ノードマスク長レジスタ 2 2 9 に設定されている値  $m$  に従い、受信パケットの経路アドレスの第 0 ビットから第  $(m - 1)$  の上位  $m$  ビットの値を抽出し、この  $m$  ビットの値に従って初段ノードへのポインタを生成してポインタセクタ 2 1 1 へ出力する。ポインタセクタ 2 1 1 は、信号線 2 2 6 によって図 2 3 の経路検索処理制御回路 2 0 4 から送信される初段ノード読み込み / 初段ノード以外のノード読み込み選択信号に従い、初段ノードの読み込み時は、上記の初段ノードへのポインタを選択して図 2 3 の読み込みアドレス生成回路 2 0 2 へ出力する。また、ポインタセクタ 2 1 1 は、初段ノード以外のノードの読み込み時は、次ノードへのポインタレジスタ 2 1 5 に保持されている次ノードへのポインタを選択して読み込みアドレス生成回路 2 0 2 へ出力する。

## 【 0 1 5 7 】

また、上記の処理と並行して、検査ビット抽出回路 2 0 9 は、初段ノード読み込み時には、検査ビット位置セクタ 2 1 2 において選択されて出力される初段ノードマスク長レジスタ 2 2 9 の設定値  $m$  に従い、経路アドレスの第  $m$  ビットから第  $(m + p - 1)$  ビットまでの  $p$  ビットの検査ビット値を抽出して図 2 3 の読み込みアドレス生成回路 2 0 2 へ出力する。また、初段ノード以外のノードの読み込み時は、後述する次ノードマスク長レジスタ 2 1 6 に保持されている次ノードのマスク長  $m_1$  が、検査ビット位置セクタ 2 1 2 において選択されて検査ビット抽出回路 2 0 9 に出力され、検査ビット抽出回路 2 0 9 は、この次ノードのマスク長  $m_1$  に従い、経路アドレスの第  $m_1$  ビットから第  $(m_1 + p - 1)$  ビットまでの  $p$  ビットの検査ビット値を抽出して図 2 3 の読み込みアドレス生成回路 2 0 2 へ出力する。

## 【 0 1 5 8 】

図 2 3 の読み込みアドレス生成回路 2 0 2 およびメモリ制御回路 2 0 6 は、上記のノードへのポインタと  $p$  ビットの検査ビット値を用いて、図 2 0 を用いて説明した順序でノード内の読み込みワードのメモリアドレスおよびメモリ制御信号を生成し、メモリ 3 0 0 は入力されたメモリアドレスに格納されているノード内の読み込みワードを信号線 2 1 4 に出力する。

## 【 0 1 5 9 】

信号線 2 1 4 のビット幅が 3 2 ビットの場合、図 2 0 で説明した方式を採用すると一回のノード (の一部) の読み込みワード数は 4 ワードとなり、これらのデータは図 2 0 の表に示した順番で信号線 2 1 4 に出力され、各レジスタ (次ノードへのポインタレジスタ 2 1 5、次ノードマスク長レジスタ 2 1 6、経路アドレス上位レジスタ 2 1 7、経路アドレス下位レジスタ 2 2 2、フラグレジスタ 2 2 4、送信ポート情報ポインタレジスタ 2 3 0) に保持される。各レジスタの保持タイミングは経路検索処理制御回路 2 0 4 からの制御信号 (図示していない) により制御される。次ノードへのポインタレジスタ 2 1 5 には次ノードへのポインタが保持され、次ノードマスク長レジスタ 2 1 6 には次ノードのマスク長が保持され、経路アドレス上位レジスタ 2 1 7 にはノードの経路アドレスの上位 2 8 ビットが、経路アドレス下位レジスタ 2 2 2 には経路アドレスの下位 3 2 ビットがそれぞれ保持され、フラグレジスタ 2 2 4 にはエントリ有りフラグが保持され、送信ポート情報ポインタレジスタ 2 3 0 には送信ポート情報へのポインタが保持される。

10

20

30

40

50

## 【0160】

次に、図22の経路アドレス一致比較処理801に対応する処理について、図24を用いて説明する。初段ノード読み込み時には、マスク長セクタ228を介して初段ノードマスク長レジスタ229の設定値がマスク処理回路210に入力する。

## 【0161】

初段ノード以外のノード読み込み時には、マスク長セクタ228を介してマスク長レジスタ227に保持されているノードのマスク長がマスク処理回路210に入力する。このマスク長レジスタ227の値は、一つ前に読み込んだノードに格納されている次ノードのマスク長の値であり、次ノードマスク長レジスタ216に保持されていたものである。次ノードマスク長レジスタ216の値は、ノードの読み込み毎に更新されるため、更新される前に、現在読み込んでいるノード(以下、現ノードと呼ぶ)の経路アドレス一致比較処理に使用する現ノードのマスク長をマスク長レジスタ227に保持しておく。

10

## 【0162】

マスク処理回路210は、これらのマスク長だけ上位ビットから有効とするマスクを生成し、このマスクと、経路アドレス生成回路207から出力される受信パケットの経路アドレスとの論理積をとり、その結果(以下、マスクした受信パケットの経路アドレスと呼ぶ)を一致比較回路213に出力する。また、経路アドレス上位レジスタ217、経路アドレス下位レジスタ222に保持されているノードの経路アドレスの上位28ビット、下位32ビットはこの順に連結され、ノードの経路アドレスとして一致比較回路213に入力される。一致比較回路213は、このノードの経路アドレスと、上記で説明した、マスクされた受信パケットの経路アドレスとを比較し(以下経路アドレス比較と呼ぶ)、その結果が不一致となる場合には不一致信号を信号線218を用いて木検索終了判定回路219に転送する。木検索終了判定回路219は上記不一致信号を受信し、木構造検索終了信号を図23の経路検索処理制御回路204に送信する。

20

## 【0163】

また、マスク長レジスタ227の値は、上記の経路アドレス比較が行われた後、次ノードマスク長レジスタ216に保持されている次ノードのマスク長の値によって更新され、次のノードの読み込み時の経路アドレス比較に使用される。

## 【0164】

次に、図22の処理802、および803に対応する処理について、図24を用いて説明する。

30

## 【0165】

一致比較回路213における経路アドレス比較の結果一致した場合、一致比較回路213は信号線220を用いて一致信号を更新判定回路221に出力する。この一致信号が更新判定回路221に入力され、かつ、フラグレジスタ224に保持されているエン트리有りフラグの値が1の場合(図25の812)のみ、更新判定回路221は更新信号をフラグレジスタ223および送信ポート情報へのポインタレジスタ225に出力する。更新信号を受信したフラグレジスタ223は、フラグレジスタ224に保持されているエン트리有りフラグを新たに保持し、同じく更新信号を受信した送信ポート情報へのポインタレジスタ225は、送信ポート情報ポインタレジスタ230に保持されている送信ポート情報へのポインタを新たに保持する(図22の経路情報更新処理803)。エン트리有りフラグの値が0の場合は更新判定回路221は更新信号を送信しないので、フラグレジスタ223および送信ポート情報へのポインタレジスタ225は更新処理を行わない(図22の813)。

40

## 【0166】

経路情報として上記のフラグおよび、送信ポート情報へのポインタ以外の情報を追加する必要がある場合は、木構造のノード内にそれらの情報を追加し、それらの情報を保持、更新するレジスタを新たに追加すればよい。

## 【0167】

次に、図22の次ノードへのポインタがNULLかどうかの判定処理804に対応する処

50

理について、図 2 4 を用いて説明する。次ノードへのポインタレジスタ 2 1 5 内に保持されている次ノードへのポインタは木検索終了判定回路 2 1 9 に入力される。この次ノードへのポインタが NULL の場合、木検索終了判定回路 2 1 9 は図 2 3 の経路検索処理制御回路 2 0 4 に木検索終了信号を出力する。

【 0 1 6 8 】

以上、図 2 3 の木構造検索処理回路 2 0 1 の詳細動作を図 2 4 を用いて説明したが、ハードウェアで構成するため、図 2 2 の木構造検索処理の中の 8 0 1、8 0 2、8 0 3、8 0 4、8 0 5 の各処理は逐次処理をする必要はなく、各処理に必要なデータが各レジスタ 2 1 5、2 1 6、2 1 7、2 2 2、2 2 4、2 3 0 に保持された後に、各処理を開始すればよく、上記の各処理を並列処理を行うことにより、高速に木構造の検索を行うことができる。

10

【 0 1 6 9 】

【 発明の効果 】

マルチキャスト経路検索において検索のキーとなる宛先マルチキャスト・グループアドレスと送信元 IP アドレスをこの順に連結して一つの経路アドレスとして定義し、この経路アドレスのビットパターンに従い、マルチキャストルーティングテーブルを 2 分木に構成することにより、エントリ数が増大しても検索時間は増加しない検索処理を行うことができる。

【 0 1 7 0 】

また、検索木を構成する各ノードを、従来の方法である 2 分木ノードから 4 分木、8 分木、あるいはそれ以上と、枝別れの数  $2$  のべき乗で増やすことにより、1 つのノードで 1 ビットでなく、連続する 2 ビット、3 ビット、あるいはそれ以上のビット数を同時に検査でき、検索終了までに辿るノードの数が減り、経路検索処理の高速化を図る効果がある。

20

【 0 1 7 1 】

図 2 5 に、エントリ数に対する従来技術 2 のハッシュ方式のマルチキャスト経路検索の処理性能 9 0 0、および本発明の 2 分木方式のマルチキャスト経路検索の処理性能 9 0 1、2 の  $p$  乗分木方式のマルチキャスト経路検索の処理性能 (例として、4 分木方式の性能 9 0 2、8 分木方式の性能 9 0 3) を示す。ハッシュ方式では、エントリ数に反比例して処理性能が劣化するのに対し、2 分木方式および 2 の  $p$  乗分木方式ではエントリ数に依存せず一定となる。また、2 分木方式に対し、4 分木方式は 2 倍の性能を実現でき、また、8 分木方式では 3 倍の性能を実現できる。

30

【 0 1 7 2 】

さらに、上記の経路アドレスとして、マルチキャスト・グループアドレス自体ではなく、マルチキャスト・グループアドレスの下位 2 8 ビットのマルチキャスト・グループ ID と送信元 IP アドレスをこの順に連結したものを採用し、この経路アドレスのビットパターンに従い、マルチキャストルーティングテーブルを 2 分木あるいは 2 の  $p$  乗分木に構成することにより、検索木の深さを決定する経路アドレスのビット数を 4 ビット減らし、ノードを渡る最大回数を減らすことができる。

【 0 1 7 3 】

また、4 分木、8 分木、あるいは一般に 2 の  $p$  乗分木を構成するときに、1 つの 2 分木ノードと、その直下につながる  $(p - 1)$  段分の合計  $(2$  の  $p$  乗  $- 1)$  個分の 2 分木ノードを一つの 2 の  $p$  乗分木ノードにまとめ、まとめられる最下段の 2 の  $(p - 1)$  乗個の 2 分木ノードに、それより上段のノードに割り付けられていた経路データを埋め込むことにより、まとめる前には 2 分木ノード換算で  $(2$  の  $p$  乗  $- 1)$  個分のメモリ量だったものを、2 の  $(p - 1)$  乗個分のメモリ量に減らす効果がある。

40

【 0 1 7 4 】

さらに、2 分木ノードを 2 の  $p$  乗分木にまとめる際、送信ポート情報を検索木のノードから分離して記憶手段の別領域に格納し、検索木のノードには上記送信ポート情報の格納領域へのポインタを保持することで、エントリが割り当てられている 2 分木ノードと、エントリが割り当てられていない 2 分木ノードのデータ量を同じにし、これらの 2 分木ノード

50

をまとめて2のp乗分木ノードを構成する際に、2のp乗分木ノードのメモリ量の削減と、各2のp乗分木ノードのメモリ量の統一を図ることができる

また、この2分木を複数個併せた形で構成した4分木、8分木、あるいはそれ以上の枝別れ数のノードを、検索のために読むときにノード全てを読むのではなく、必要な部分のみを読むようにすることにより、ノードが大きくなることによるデータの読み込み時間の増大を防ぎ、経路検索処理の高速化を図る効果がある。

【0175】

また、各ノードに、そのノード自体のマスク長ではなく、そのノードのすぐ下につながるノードのマスク長を格納することにより、ノードのマスク長をノードのデータを読む前に知ることができ、ノードのデータを読む前に、ノードのどの部分を読み込めば良いかが分  
10  
かり、必要な部分のみを読むことができるようになり、ノードが大きくなることによるデータの読み込み時間の増大を防ぎ、経路検索処理の高速化を図る効果がある。

【0176】

また、各ノードの先頭に、そのノードにエントリが割り付けられているか否かを示すフラグを設け、最初に、このフラグを読み込み、エントリが割り付けられていないノードでは、経路情報を読み込まないようにすることにより、データの読み込み時間が短縮され、経路検索処理の高速化を図る効果がある。

【0177】

また、経路アドレスの上位数ビット分、ノードをメモリ上の決まった位置に展開し、受信パケットの経路アドレスの上位数ビットに従い、ノードが格納されている位置を直接にア  
20  
クセスすることは、検索処理時間を無くし、経路検索処理の高速化を図る効果がある。

【0178】

また、上記の高速化手段をハードウェアで構成することにより、経路検索処理の高速化を図る効果がある。

【図面の簡単な説明】

【図1】本発明の一実施例であるパケット中継装置の例であるルータのブロック図。

【図2】一般的なネットワークの構成例を示す図。

【図3】IPアドレスの構成、およびクラスDのIPアドレスの構成、およびIPパケットのパケットヘッダフォーマット。

【図4】一般的なルータの構成例を示す図。  
30

【図5】本発明で定義する経路アドレスの構成図。

【図6】経路アドレス長3ビットの場合の全てのノードがある2分木。

【図7】経路アドレス長3ビットの場合のマルチキャストルーティングテーブルの例。

【図8】図6においてエントリが割り付けられておらず、かつ、エントリ付きのノードへの途中経路にもなっていないノードを取り除いた木。

【図9】経路アドレス32ビットの場合のマルチキャストルーティングテーブルの例。

【図10】図9に示したマルチキャストルーティングテーブルに対応する木。

【図11】図10において枝別れもエントリの割り付けも無いノードを取り除いた木。

【図12】2分木ノードのデータ構造を示す図。

【図13】2分木から8分木への変形時に一つの8分木ノードにまとめられる2分木ノードを囲んだ木。  
40

【図14】マスク長が0ビットから始まらないようにビット位置を区切った場合に最初のノードをメモリ上に展開することにより区切り位置までのビットの検索を行う木。

【図15】先頭のさらに多くのビット数の検索時間を省いた木。

【図16】2分木から4分木への変形時に1つの4分木ノードにまとめられる2つの2分木ノードを2つの2分木ノード分につぶした木。

【図17】1つの4分木ノードにまとめられる3つの2分木ノードの、2つの2分木ノードへのつぶし方。

【図18】1つの8分木ノードにまとめられる7つの2分木ノードの、4つの2分木ノードへのつぶし方。  
50

【図 19】4 分木ノードのデータ構造を示す図。

【図 20】ノードの大きさが大きくなったときにノードデータリード時間の増大を防ぐために 1 つのノード全てを読み込まずに一部だけを読み込む方法を示す図。

【図 21】条件によっては読み込む必要が無い要素を、条件に従い読み込まないようにすることで読み込みの時間を削減することで高速化を図る方法を示す図。

【図 22】本発明の一実施例である経路検索処理部の検索処理フローチャート。

【図 23】本発明の一実施例である経路検索処理部のブロック図。

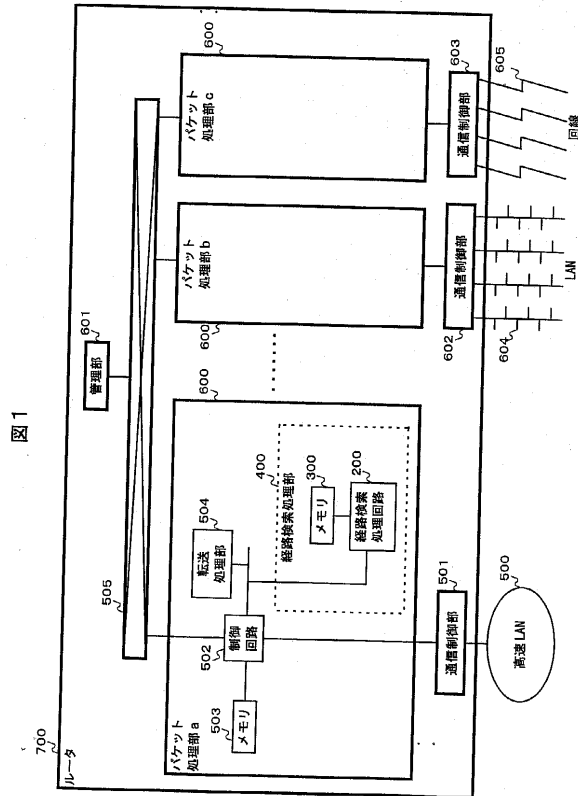
【図 24】本発明の一実施例である、木構造のマルチキャストルーティングテーブル検索を行う木構造検索処理回路のブロック図。

【図 25】ハッシュ方式、2 分木方式、2 の p 乗分木方式の、エントリ数に対する処理性能を示す図。 10

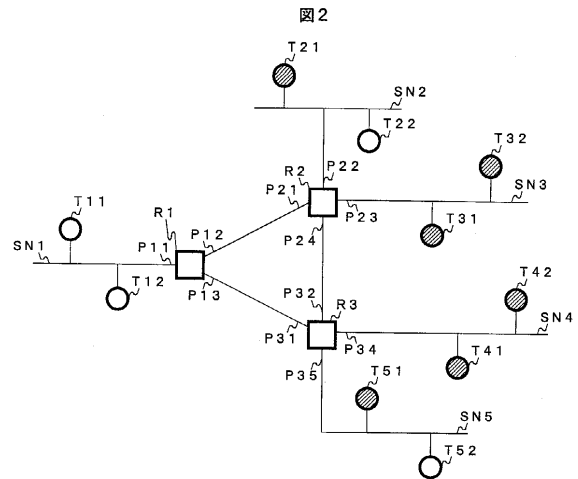
【符号の説明】

T 1 1、T 1 2、T 2 1、T 2 2、T 3 1、T 3 2、T 4 1、T 4 2、T 5 1、T 5 2 . . . ネットワーク上でパケットを送受信する端末、S N 1 ~ S N 5 . . . 各端末が属するサブネットワーク、R 1 ~ R 3 . . . サブネットワークを相互に接続するルータ、P 1 1、P 1 2、P 1 3、P 2 1、P 2 2、P 2 3、P 2 4、P 3 1、P 3 2、P 3 4、P 3 5 . . . 各ルータのポート、1 . . . I P アドレス、2 . . . サブネットワークを識別するサブネットワークアドレス、3 . . . ネットワーク内の端末を識別するホストアドレス、4 . . . クラス D の I P アドレス、5 . . . クラス D であることを示すクラス D の I P アドレスの上位 4 ビット、6 . . . マルチキャスト・グループ I D、7 . . . I P パケットヘッダの送信元アドレスフィールド、8 . . . I P パケットヘッダの宛先アドレスフィールド、1 1 - i ( = 1 ~ N ) . . . ルータの入力ポート、1 2 - i ( = 1 ~ N ) . . . ルータの出力ポート、1 3 - i ( = 1 ~ N ) . . . ルーティング処理部、1 4 . . . スイッチ、1 5 - i ( = 1 ~ N ) . . . 通信制御部、1 6 . . . 管理部、3 2 . . . マルチキャスト・グループアドレス、送信元 I P アドレスの順に連結して生成した経路アドレス、3 3 . . . マルチキャスト・グループ I D、送信元 I P アドレスの順に連結して生成した経路アドレス、6 0 . . . 経路アドレス長 3 ビットの場合の受信パケットの経路アドレスの一例、4 0 ~ 5 4、6 2 ~ 6 5、7 0 ~ 8 4、9 0 ~ 9 5、1 0 0 ~ 1 0 4、1 3 0 ~ 1 4 4 . . . 木を構成するノード、6 1 . . . 枝別れもエントリの割り付けも無いノードの長い列、1 1 0 ~ 1 1 3、1 1 5 . . . 2 分木ノードを構成する 3 2 ビットを単位とするワード、1 1 4 . . . 送信ポート情報、1 2 0 ~ 1 2 7 . . . 4 分木ノードを構成する 3 2 ビットを単位とするワード、2 0 0 . . . 経路検索処理回路、2 0 1 . . . 木構造検索回路、2 0 2 ~ 2 0 4、2 0 6 . . . 経路検索処理回路を構成する各回路、2 0 5 . . . 転送処理部、経路検索処理部間の信号線、2 0 7 ~ 2 1 3、2 1 5 ~ 2 3 0 . . . 木構造検索回路を構成する各回路および信号線、2 2 6 . . . 経路検索処理制御回路からの読み込みノード(初段ノードかそれ以外か)選択制御信号線、3 0 0 . . . マルチキャストルーティングテーブルが格納されるメモリ、4 0 0 . . . 経路検索処理回路とメモリから構成される経路検索処理部、5 0 0、6 0 4、6 0 5 . . . ルータに接続されるネットワークの例、5 0 1、6 0 2、6 0 3 . . . 通信制御部、5 0 2 . . . パケットデータのメモリ格納およびメモリからの読み出しを制御する制御回路、5 0 3 . . . パケットデータが格納されるメモリ、5 0 4 . . . パケット転送処理部、5 0 5 . . . 複数のパケット処理部および管理部が接続されるスイッチ、6 0 0 . . . パケット処理部、6 0 1 . . . ルータ全体の管理を行う管理部、7 0 0 . . . ルータ。 20 30 40

【図 1】

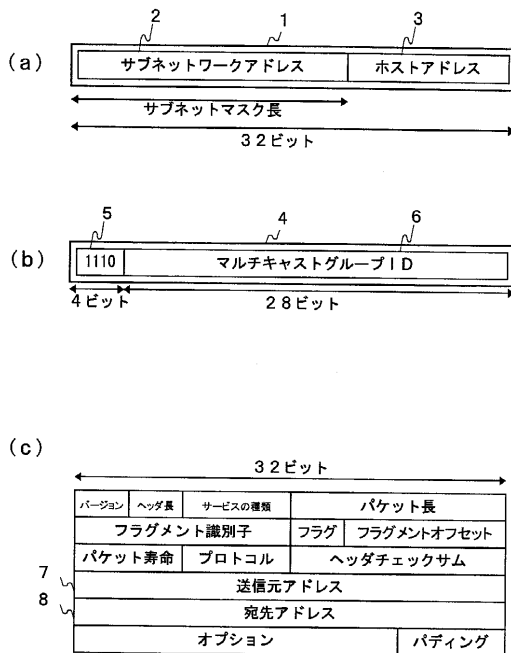


【図 2】

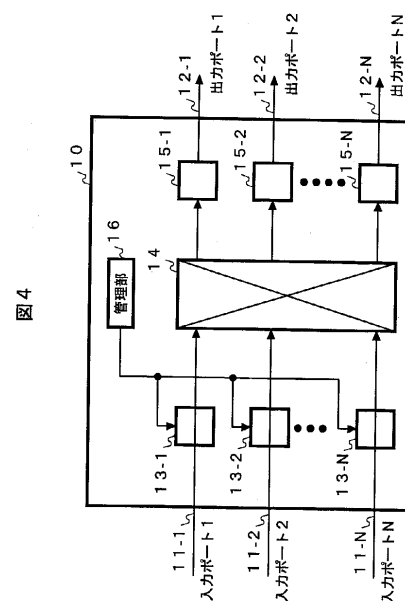


【図 3】

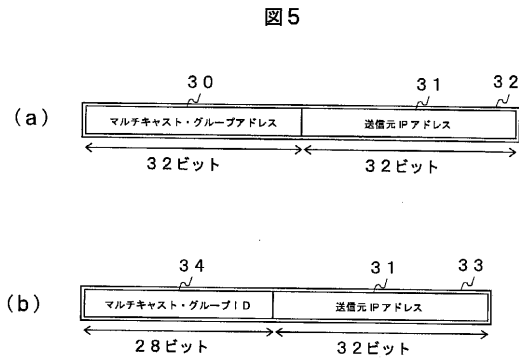
図 3



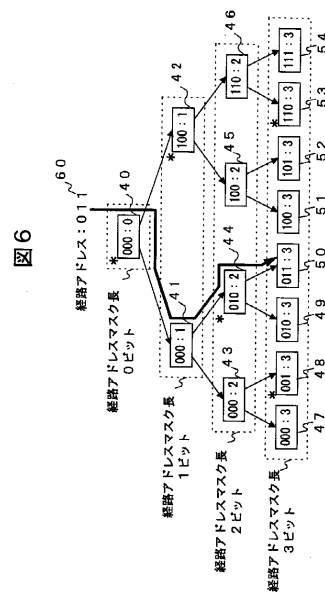
【図 4】



【図 5】



【図 6】

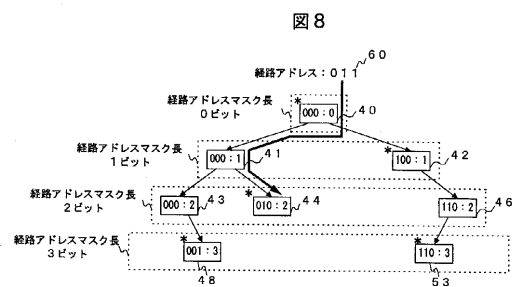


【図 7】

図 7

経路アドレス: 経路アドレスマスク長	
000 :	0
001 :	3
010 :	2
100 :	1
110 :	3

【図 8】



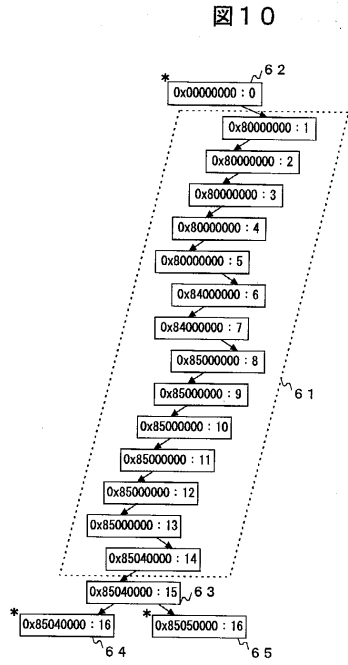
【図 9】

図 9

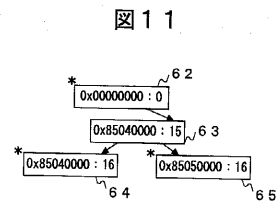
マルチキャストルーティングテーブル

経路アドレス: 経路アドレスマスク長	
0x00000000 :	0
0x85040000 :	16
0x85050000 :	16

【図 10】

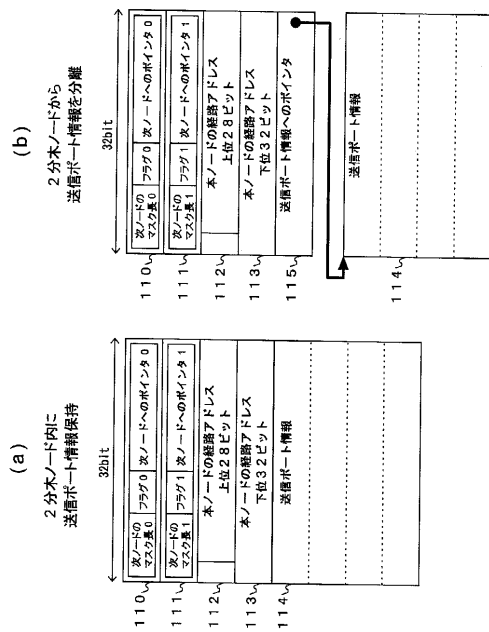


【図 11】

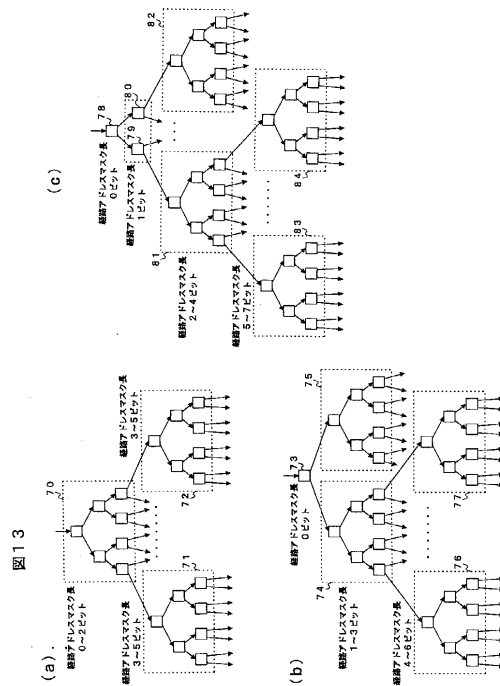


【図 12】

図 12

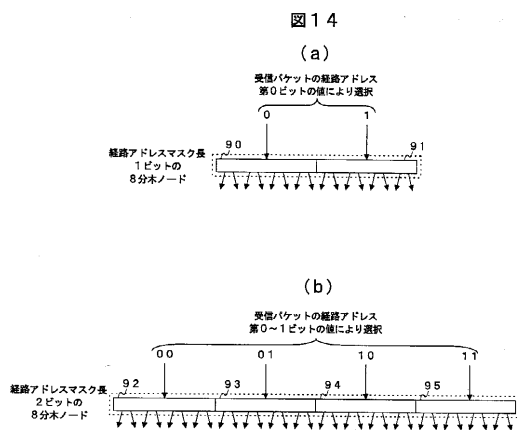


【図 13】

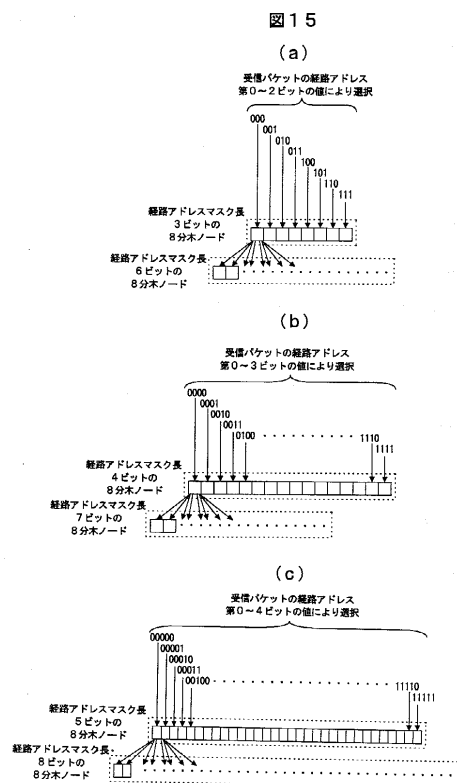




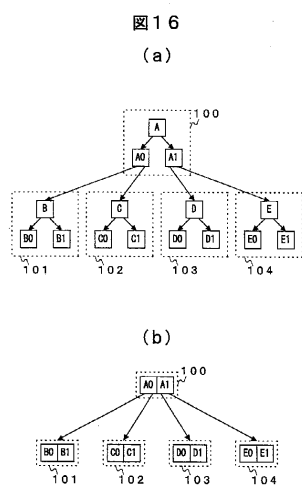
【 図 1 4 】



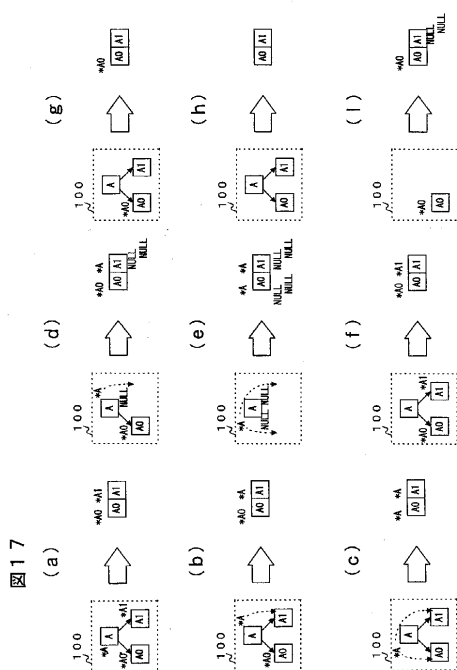
【 図 1 5 】



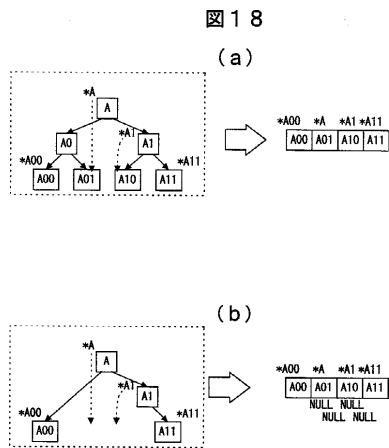
【 図 1 6 】



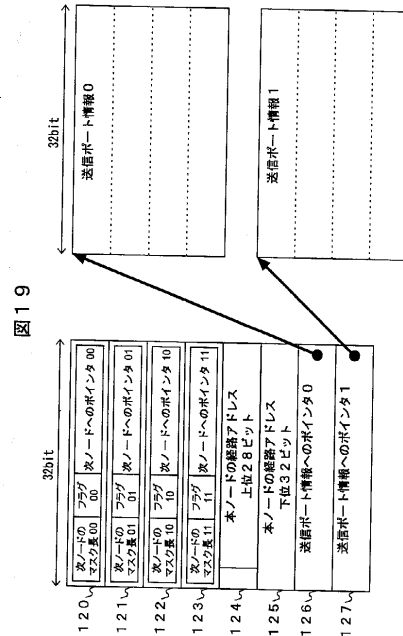
【 図 1 7 】



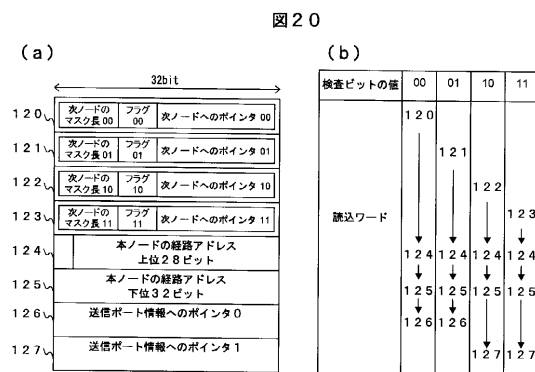
【 図 1 8 】



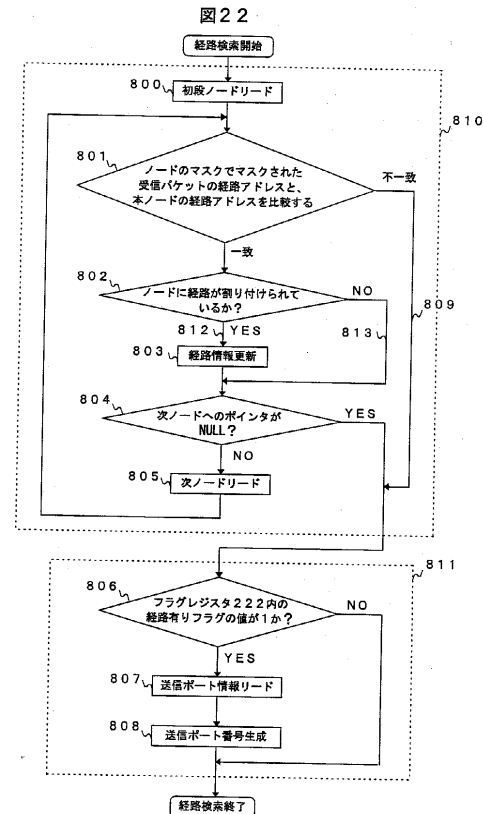
【 図 1 9 】



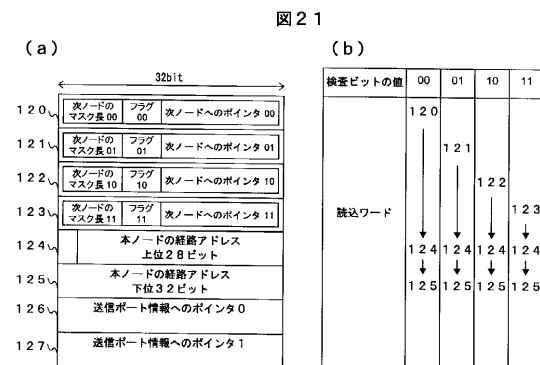
【 図 2 0 】



【 図 2 2 】



【 図 2 1 】





---

フロントページの続き

- (72)発明者 相本 毅  
東京都国分寺市東恋ヶ窪一丁目280番地 株式会社日立製作所中央研究所内
- (72)発明者 松山 信仁  
神奈川県秦野市堀山下1番地 株式会社日立インフォメーションテクノロ  
ジー内
- (72)発明者 左古 義人  
神奈川県秦野市堀山下1番地 株式会社日立製作所汎用コンピュータ事業  
部内
- (72)発明者 関野 浩  
神奈川県秦野市堀山下1番地 株式会社日立インフォメーションテクノロ  
ジー内

審査官 清水 稔

- (56)参考文献 特開2000-083056(JP,A)  
鹿島和幸 他4名,IPルーティングテーブルのハードウェアによる高速検索,電子情報通信学会  
通信ソサイエティ大会講演論文集2,1998年9月7日,B-7-123,p.244  
J.Moy,Multicast Extensions to OSPF,Network Working Group Request for Comments:1584,  
1994年3月

(58)調査した分野(Int.Cl.,DB名)

H04L 12/56  
G06F 13/00 355  
H04L 12/18  
H04L 12/28  
H04L 12/46  
H04L 13/08