

【特許請求の範囲】

【請求項 1】

マルチビットトライ検索エンジンにおいて、
マルチビットトライルーチングテーブルに関するプレフィックス検索動作期間中各々がストライドを処理する一連のハードウェアパイプラインユニット、
を有しており、少なくとも 1 つのパイプラインユニットが、与えられたプレフィックス検索動作に対して前記少なくとも 1 つのパイプラインユニットからの検索結果が次のパイプラインユニットへ継続するか又はリタイアされることを可能とする制御メカニズムを包含している、
ことを特徴とするマルチビットトライ検索エンジン。

10

【請求項 2】

請求項 1 において、前記少なくとも 1 つのパイプラインユニットが、前記少なくとも 1 つのパイプラインユニットにより処理されるストライドを超えてプレフィックス検索動作のためのプレフィックス内にストライドが残存しているか否かを決定することを特徴とするマルチビットトライ検索エンジン。

【請求項 3】

請求項 2 において、前記少なくとも 1 つのパイプラインユニットが、又、プレフィックス検索動作が爾後のパイプラインユニットにおいて処理中であるか否かを決定することを特徴とするマルチビットトライ検索エンジン。

【請求項 4】

請求項 1 において、前記少なくとも 1 つのパイプラインユニットが、前記少なくとも 1 つのパイプラインユニット内の処理がプレフィックス検索動作を完了したか否か及びいずれかのプレフィックス検索動作が爾後のパイプラインユニットにおいて処理中であるか否かを決定することを特徴とするマルチビットトライ検索エンジン。

20

【請求項 5】

請求項 1 において、前記少なくとも 1 つのパイプラインユニットが、爾後のパイプラインユニットにおいて処理中のいずれかのプレフィックス検索動作の完了前にプレフィックス検索動作に対する検索結果をリタイアさせることを特徴とするマルチビットトライ検索エンジン。

【請求項 6】

請求項 1 において、2 つ又はそれ以上のパイプラインユニットの各々が、夫々のパイプラインユニットからの検索結果が次のパイプラインユニットへ継続させるか又はリタイアさせることを可能とする制御メカニズムを包含していることを特徴とするマルチビットトライ検索エンジン。

30

【請求項 7】

請求項 6 において、前記 2 つ又はそれ以上のパイプラインユニットが、2 つ又はそれ以上の所定の最大プレフィックス長の各々に対する最後のストライドに対応する一連のパイプラインユニット内の位置に位置されていることを特徴とするマルチビットトライ検索エンジン。

【請求項 8】

マルチビットトライ検索エンジンにおいて、
一連のハードウェアパイプラインユニット、
を有しており、各ユニットが、
入力又は先行するパイプラインユニットからインデックスを受取るバッファ、
マルチビットトライルーチングテーブルに関してのプレフィックス検索動作期間中にストライドを処理する論理、
を包含しており、

40

前記一連における少なくとも 1 つのパイプラインユニットが、
与えられたプレフィックス検索動作に対する前記少なくとも 1 つのパイプラインユニット内の処理論理から爾後のパイプラインユニット内のバッファか又は前記ルーチングテ

50

ーブル内の次のホップ情報へアクセスする制御器のいずれかへ検索結果を送信するデマルチプレクサ、

前記デマルチプレクサを制御する制御論理、
を有していることを特徴とするマルチビットトライ検索エンジン。

【請求項 9】

請求項 8 において、前記少なくとも 1 つのパイプラインユニット内の制御論理が、いずれかのストライドが、前記少なくとも 1 つのパイプラインユニットにより処理されるストライドを超えてプレフィックス検索動作に対するプレフィックス内に止まるか否かを決定することを特徴とするマルチビットトライ検索エンジン。

【請求項 10】

請求項 9 において、前記少なくとも 1 つのパイプラインユニット内の制御論理が、いずれかのプレフィックス検索動作が爾後のパイプラインユニットにおいて処理中であるか否かを決定することを特徴とするマルチビットトライ検索エンジン。

【請求項 11】

請求項 8 において、前記少なくとも 1 つのパイプラインユニット内の制御論理が、前記少なくとも 1 つのパイプラインユニット内の処理がプレフィックス検索動作に対し不一致又は端部ノード一致のいずれかを発生したか否か及びいずれかのプレフィックス検索動作が爾後のパイプラインユニットにおいて処理中であるか否かを決定することを特徴とするマルチビットトライ検索エンジン。

【請求項 12】

請求項 8 において、前記少なくとも 1 つのパイプラインユニット内の制御論理が、爾後のパイプラインユニットにおいて処理中のいずれかのプレフィックス検索動作の完了前にプレフィックス検索動作に対する検索結果をリタイアすることを特徴とするマルチビットトライ検索エンジン。

【請求項 13】

請求項 8 において、2 つ又はそれ以上のパイプラインユニットの各々が、更に、
与えられたプレフィックス検索動作に対する少なくとも 1 つのパイプラインユニット内の処理論理からの検索結果を爾後のパイプラインユニット内のバッファか又は前記ルーチングテーブル内の次のホップ情報へアクセスするための制御器のいずれかへ送信するデマルチプレクサ、

前記デマルチプレクサを制御する制御論理、
を有していることを特徴とするマルチビットトライ検索エンジン。

【請求項 14】

請求項 13 において、前記 2 つ又はそれ以上のパイプラインユニットが 2 つ又はそれ以上の所定の最大プレフィックス長の各々に対する最後のストライドに対応する一連のパイプラインユニット内の位置に位置されていることを特徴とするマルチビットトライ検索エンジン。

【請求項 15】

マルチビットトライ検索エンジンの動作方法において、
マルチビットトライルーチングテーブルに関するプレフィックス検索動作期間中に一連のハードウェアパイプラインユニットの各々におけるストライドを処理し、
少なくとも 1 つのパイプラインユニット内において、与えられたプレフィックス検索動作に対する少なくとも 1 つのパイプラインユニットからの検索結果を次のパイプラインユニットへ継続させるか又はリタイアさせるかのいずれかをイネーブルさせる、
ことを包含していることを特徴とする方法。

【請求項 16】

請求項 15 において、更に、
いずれかのストライドが、前記少なくとも 1 つのパイプラインユニットにより処理されたストライドを超えてプレフィックス検索動作に対するプレフィックス内に止まるか否かを決定する、

10

20

30

40

50

ことを包含していることを特徴とする方法。

【請求項 17】

請求項 16 において、更に、

いずれかのプレフィックス検索動作が爾後のパイプラインユニットにおいて処理中であるか否かを決定する、

ことを包含していることを特徴とする方法。

【請求項 18】

請求項 15 において、更に、

前記少なくとも 1 つのパイプラインユニット内の処理がプレフィックス検索動作を完了したか否か及びいずれかのプレフィックス検索動作が爾後のパイプラインユニットにおいて処理中であるか否かを決定する、

ことを包含していることを特徴とする方法。

【請求項 19】

請求項 15 において、更に、

爾後のパイプラインユニットにおいて処理中のいずれかのプレフィックス検索動作の完了前にプレフィックス検索動作に対する検索結果をリタイアさせる、

ことを包含していることを特徴とする方法。

【請求項 20】

請求項 15 において、更に、

2 つ又はそれ以上のパイプラインユニット内において、夫々のパイプラインユニットからの検索結果を次のパイプラインユニットへ継続させるか又はリタイアさせるかのいずれかをイネーブルさせる、

ことを包含していることを特徴とする方法。

【請求項 21】

請求項 20 において、前記 2 つ又はそれ以上のパイプラインユニットが 2 つ又はそれ以上の所定の最大プレフィックス長の各々に対する最後のストライドに対応する一連のパイプラインユニット内の位置に位置されていることを特徴とする方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、大略、マルチビットトライを基礎としたネットワーク検索エンジンに関するものであって、更に詳細には、ハードウェアパイプライン型マルチビットトライを基礎としたネットワーク検索エンジンの性能を改良する技術に関するものである。

【背景技術】

【0002】

インターネットプロトコル (IP) 等のパケットを基礎とした通信プロトコル用のネットワークルーターは、入ってくる情報を、そのパケットに対して意図されたデスティネーション即ち宛先へのルート即ち経路に沿って次の隣りのものへ指向させる。これを行うために、典型的に、該ルートに沿っての各ルーターは、デスティネーション IP プレフィックスに対する適宜の次のホップアドレスを決定するためにプレフィックス (又はルーチング) テーブルに関してアドレスプレフィックス (通常、単に「プレフィックス」と呼称される) ルックアップ動作を実施せねばならない。このような動作は埋込型ネットワークプロセッサによるか、又は、より一般的には、別個のネットワーク検索エンジンのいずれかにより実施される。

【0003】

元々、ネットワーク検索エンジン用のハードウェアは内容アドレス可能メモリ (CAM) を使用しており、それはビット比較器と 2 つのメモリ要素、即ちデータを格納するために 1 つと比較マスクを格納するために別の 1 つとからなる 1 つのタイプのメモリである。CAM は比較結果を「常に一致」 (即ち、「don't care」) へオーバーライドさせるべくプログラムすることが可能なマスク値の制御下において入力して来るデータを

10

20

30

40

50

データメモリ内に格納されている値と比較する。動作について説明すると、CAMを基礎としたネットワーク検索エンジンは、特定の優先付けした順番でCAMアレイ内のルーチングテーブルの全てのプレフィックスを格納することにより機能し、各プレフィックスの関連する次のホップ情報は別のメモリ内の対応する（リンクした）位置に格納されている。プレフィックスルックアップ期間中、キーがCAMアレイのコンパランド（比較オペランド）上に配置され且つメモリ内の全てのプレフィックスに対して比較される。全ての比較からの一致結果のアレイが優先度論理ユニットを介して送られて最も高い優先度の一致を決定し、勝者の一致が、それから対応する次のホップ情報が読取られ且つリターンされる次のホップメモリへアドレスするために使用される。

【0004】

10

より最近になって、汎用プロセッサ及び通常のメモリを使用するソフトウェアを基礎としたネットワーク検索エンジンが開発されている。このような装置においては、プロセッサは一連のメモリ読取及び比較動作でプレフィックス検索を実施する。ルーチングテーブルプレフィックス及び次のホップ情報が、典型的に、ルックアップ期間中にメモリアクセスの数及びルーチングテーブルを格納する場合におけるメモリ使用を減少させるために開発された種々のソフトウェアアルゴリズムのうちの1つに従って構築されたデータ構造におけるメモリ内に格納される。これらの目的のために、マルチビットトライ及び対応するアルゴリズムは検索動作のための有界数のメモリアクセスで最良のデータ圧縮を達成するデータ構造及びアルゴリズムの範疇に属するものである。

【0005】

20

トライ（「trie」、即ち「retrieve」（検索）の中間の4個の文字）は二進ストリングを表わすために構築されたツリーを基礎としたデータ構造であり、該ストリングにおける各ビット又はグループのビットはそのツリー内の分岐中において採用される方向を決定する。二進（ユニビット即ち単一ビット）トライはビット毎に進行し且つ各ノードから高々2つの分岐を有しており、一方マルチビットは一度に複数のビットを消費し且つ各ノードにおいて幾つかの分岐を有しており、各分岐は次のレベルへ通じている。各ノードにおける分岐選択期間中に消費又は検査されるビットの数はストライドと呼称される。一様幅ストライドトライは、ストライド幅により割算された後にプレフィックス長の残りとするのが可能な可能性として最後のストライドを除いて、同一の幅を有する全てのストライドを具備するトライである。

30

【0006】

通常、マルチビットトライアルゴリズムは、一様ストライド幅トライ内にプレフィックスを格納し且つ検索し、同一の親（次により高いストライドレベル）ストライド値を有し同一レベル内の全ての分岐をトライテーブルと呼称されるテーブル内へグループ化することにより動作する。各レベルにおいて、対応するストライド値は次のレベルへ到達するのに必要な情報を包含するトライテーブルエントリ内へのインデックスを与える。マルチビットトライは、共通の高次ビット（ストライド）を有するプレフィックスは同一の親トライテーブルを共用するという利点を有しており、プレフィックスを格納するのに必要なメモリを減少させる。

【0007】

40

長さ1のプレフィックスが各々が n ビットからなる m ストライドへ分割する場合に、次のレベルのトライテーブル内のエントリの最大可能数は 2^n である。該アルゴリズムは同一の親からの全ての次のレベルのストライド値を、トライノードと呼称されるデータ構造で次のレベル（子）トライテーブルのベースアドレスを包含するポインターと共に、親トライテーブル内のエントリ内に格納される 2^n ビットデータフィールドへエンコードさせる。最大寸法 2^n の代わりに、存在する実際の数のテーブルエントリに対してメモリを割り当てることによりテーブル圧縮が達成される。各プレフィックスの最後のストライドに対して、端部ノードと呼称される同様のタイプのデータ構造が使用されるが、この場合には、ポインターは次のレベルのトライテーブルの代わりに次のホップ情報を包含するテーブルを指し示すことが異なる。

50

【0008】

ルーチングテーブルルックアップも同一幅ストライドで実施され、入力検索器（典型的に、32又は64ビットのIPアドレス）からの次のレベルのストライドの値が、そのストライド値の親のテーブルエントリにおける関連するデータフィールドと共に、デコードされ且つ処理される。同一のプレフィックスストライド値を有する格納されているルートがそのトライ内に存在するものと決定されると、親テーブル内の情報を使用してインデックスが計算され、次いでテーブルポインター及び計算されたインデックスを使用して検索が継続して行われ次のレベルのトライテーブルエントリへ通じるアドレスを形成する。一致が見つからない場合には、その検索は成功せずに終了する。検索が端部ノードに到達し且つ一致が見つかった場合には、その検索は成功であり且つ関連する次のホップ情報が次のホップテーブルから読取られる。 10

【0009】

マルチビットトライにおいてルックアップを成功裡に完了することは複数のレベルのトライをトラバースすることを必要とする。各レベルにおいて、ツリーノードがメモリからアクセスされ且つストライド値と共に使用されて次のレベルにおけるトライノードのアドレスを決定する。

【0010】

然しながら、ハードウェアマルチビットトライネットワーク検索エンジン内のレイテンシー即ち待ち時間は、ネットワークルーター内、従ってネットワーク自身における全体的なレイテンシーに影響を与える場合がある。 20

【発明の開示】

【発明が解決しようとする課題】

【0011】

本発明は、以上の点に鑑みなされたものであって、上述した如き従来技術の欠点を解消し、改良したマルチビットトライネットワーク検索エンジン及びその動作方法を提供することを目的とする。

【課題を解決するための手段】

【0012】

本発明は、ネットワークルーター用のマルチビットトライネットワーク検索エンジンを提供しており、該検索エンジンは一連のハードウェアパイプラインユニットを有しており、各ユニットはマルチビットトライに関するプレフィックス検索動作期間中にストライドを処理し且つ、最後のパイプラインユニット以外の少なくとも1つのパイプラインユニット内において、検索結果を残りのパイプラインユニットを介して通過させるのではなく夫々のパイプラインユニットからの検索結果をリタイアさせるメカニズムを包含している。早期のリタイアは、パイプライン型検索動作の順番外完了をサポートしないシステムに対する爾後のパイプラインユニットにおけるアクティブな検索の動作の不存在と共に、処理すべき爾後のストライドの不存在か又は現在のストライドにおける検索の完了（不一致又は端部ノード一致）のいずれかによりトリガさせることが可能である。早期リタイアメカニズムは、そのパイプラインより短い最大プレフィックス長に対する最後のストライドに対応するパイプラインユニット（例えば、64ビットプレフィックスではなく20又は32ビットプレフィックス）において、何等かのその他の基準に基づいて選択したパイプラインユニットにおいて、又は全てのパイプラインユニットにおいて包含させることが可能である。プレフィックス検索動作に対する最悪及び／又は平均のレイテンシー即ち待ち時間が減少される。 30 40

【発明を実施するための最良の形態】

【0013】

以下に説明する図1乃至図4A-4B、及び本明細書における本発明の原理を説明するために使用する種々の実施例は単に例示的なものであって、本発明の範囲を制限するような態様で理解すべきものではない。当業者によって理解されるように、本発明の原理は任意の適宜に構成した装置において実現することが可能なものである。 50

【0014】

図1は本発明の1実施例に基づくコンフィギュラブル (c o n f i g u r a b l e) 即ち形態特定可能な多段パイプラインマルチビットトライネットワーク検索エンジンを使用する処理システムを示している。処理システム100はインターネットプロトコル (I P) ネットワークルーターの一部を実現しており、且つ外部メモリ103へ結合されているネットワークパケット検索エンジン (N P S E) 102へ結合しているシステム制御器又はネットワーク処理ユニット (N P U) 101を包含している。N P S E 102は次のホップアドレスに対するプレフィックスのみならずシステム制御器101からアドレス及び／又はコマンドを受取り、且つ次のホップアドレスをシステム制御器101へリターンする。N P S E 102は、プレフィックステーブル104から決定したデータテーブル内の次のホップアドレスへのデータポインターと共に、外部メモリに対するデータテーブルメモリアドレスを外部メモリ103へ送信する。N P S E 102は外部メモリ103からアドレス及びデータポインターによって識別されるデータテーブル105内のテーブル及びエントリからの次のホップアドレスを受取る。システム制御器101はバックプレーン106へ結合しており且つネットワークインターフェース107を介してネットワークへ結合している。

【0015】

N P S E 102は、好適には、高速32ビットネットワークプロセッサユニット (N P U) インターフェースによりシステム制御器101へ結合しており、且つ高速32ビットスタチックランダムアクセスメモリ (S R A M) インターフェースにより外部メモリ103へ結合している。各インターフェースは、クワッドデータレート (Q D R) アーキテクチャにより実現されており、システム制御器101がN P Uインターフェースに対するマスターとして且つ外部メモリ103がS R A Mインターフェースに対するスレーブとして設けられており、4.5ギガバイト／秒 (G B / s) のピーク集合帯域幅に対する250メガヘルツ (M H z) の最大クロック周波数でクロック当たり2個のワードで32ビット同時的データ入力／出力 (I / O) を供給する。当業者によって理解されるように、N P U及びS R A Mインターフェースのその他のコンフィギュレーション即ち形態のものを使用することも可能である。

【0016】

図2は本発明の1実施例に基づく形態特定可能な多段パイプライン型マルチビットトライネットワーク検索エンジンに対するマイクロアーキテクチャの概略図である。N P S E 102は、N P S E 102をシステム制御器101へ結合させるN P Uインターフェース200、アービター (仲裁器) 201、C P U 202により実行されるプログラムを包含する関連するメモリ (S R A M) 203を具備する中央処理装置 (C P U) 202、N P S E 102を外部メモリ103へ結合させるS R A M制御器204、パイプライン論理ユニット205a-205n及び一連の仮想メモリバンク206a-206nを形成する対応する組 (この例示的实施例においては16) の形態特定可能なメモリブロックのアレイ (この例示的实施例においては16)、とを包含しており、パイプライン論理ユニット205a-205n及びメモリバンク206a-206nは仮想バンク形態特定をイネーブル即ち可能とさせるクロスバー207aによって結合されている。N P S E 102により実施される主要なタスクはルートの挿入、削除及び検索であり、そのことは、N P S E 102が、内部及び外部メモリ内にルートを挿入し且つそれからルートを除去することによりルーチングテーブルを維持し、且つ提示されたプレフィックスに関してのルート検索を実施し且つ関連する次のホップ情報をリターンする。

【0017】

図3は本発明の1実施例に基づく形態特定可能な多段パイプライン型マルチビットトライネットワーク検索エンジン内のマルチビットトライ検索のハードウェアパイプライン化を示している。マルチビット検索アルゴリズムのハードウェア実現例において、複数のトライレベルを介しての検索は図示した如くにパイプライン化することが可能であり、各処理レベルは与えられた検索器に対し異なるストライドに関して動作し且つ新たな検索器に

対する処理が各サイクル毎に開始される。任意の与えられた時間において、パイプラインの各レベルは、そのレベルにおいてのストライドビット（その検索器に対し）を処理しながら、異なる検索器に関して動作を行う。1つの実現例においては、最後（パイプラインユニット205n）以外のパイプラインユニットのいずれかにおけるミス又は端部ノード一致は、パイプラインユニットの残部を介して更なる処理を行うことなしに通過される結果を発生する。

【0018】

殆どのマルチビットトライアルゴリズムは汎用のプログラム可能プロセッサ上で稼動するソフトウェアプログラムとして実現される。マルチビットトライアルゴリズムのハードウェア実現例が提案される範囲において、このようなハードウェアトライアルを介しての検索に関連するレイテンシー即ち待ち時間についての説明は割愛する。

10

【0019】

マルチビットトライアルゴリズムのハードウェア実現例の輸入成分は、トライテーブルを格納するために必要とされるオンチップ格納部の量である。一方、メモリは、コアルータにおける数十万個のプレフィックスを有する大きなルーチングテーブルを格納するのに十分に大きなものでなければならない。同時に、毎秒数億個の検索の高い処理能力ルックアップ条件に起因して、メモリへのアクセスはパイプライン化せねばならず、そのことは、メモリがコンフィギュラブル即ち形態特定可能なものであり且つパイプラインの各ステージ即ち段又はトライのレベルにおいての同時的アクセス可能性に対して柔軟性のあるものでなければならない。

20

【0020】

オンチップで形態特定可能メモリ構造を有する1つの結果は、各トライレベルにおいてのストライドの処理がより時間がかかるということである。実施例のNPSE102において、16個のパイプラインユニット205a-205nが存在しており、各々は4ビットのストライドを処理する。インターネットプロトコルバージョン6（IPv6）パケットに対する64ビットアドレスプレフィックス（典型的に、IPv6パケットに対する128ビットプレフィックスの64ビットのみが検索において関連性がある）は書込速度において検査することが可能であるが、各レベルにおいて行われる作業はより大きく、形態特定可能オンチップSRAMメモリバンク206a-206nはアクセスされ、トライノードエントリが抽出され、且つ次のトライレベルアドレスが計算されることが必要である。各パイプラインユニット205a-205nにおけるこれらの処理に対して最大で4個のサイクルが必要とされる場合があり、その結果64サイクルのレイテンシー即ち待ち時間となる。

30

【0021】

同一のNPSE102は各々が異なる最大プレフィックス長を有する複数個のルーチングテーブルを収容することが可能であり、例えばインターネットプロトコルバージョン4（IPv4）パケットに対しての32ビットプレフィックス又は仮想プライベートネットワーク（VPN）パケットに対する20ビットプレフィックスであり、パイプラインユニット205a-205nのレイテンシーは64ビット未満の長さのプレフィックスに対して検索完了時間に不必要な制限を加える。更に、上述したパイプライン化実現例は、例えばプレフィックスルックアップが最初の1つ又は2つのレベルで最も長いプレフィックスの一致を見つけ出す場合であっても64サイクル全体のレイテンシーが無駄にされるという欠点を有しており、従って平均的な場合のルックアップレイテンシーも損失を被る。

40

【0022】

図4A及び4Bは本発明の種々の実施例に基づく形態特定可能多段パイプライン型マルチビットトライネットワーク検索エンジン内の処理レベルに対するパイプラインユニットの構造を詳細に示している。図4Aは与えられたテーブルに対して最悪の場合のレイテンシーをアドレスする1つの実施例に対するパイプラインユニットの詳細を示している。構築され且つNPSE102においてトラバースされる各テーブルは、特定のネットワーク適用例に対してのものであり、即ちIPv4/IPv6、VPN等の適用例であり、それ

50

はテーブル内の最大プレフィックス長 l は通常知られていることを意味している。

【0023】

この最大プレフィックス長は NPSE102 での検索に対するトライレベルの最大数 $L_{\max}$ を決定し、尚 $L_{\max} = \text{ceiling}(l\%4)$ であり、それはそのテーブルにおける検索に対する最悪のレイテンシーを決定する。従って、IPV4 テーブルは 32 ビットの最大プレフィックス長及び 8 の検索レベルの最大数 (4 ビットストライドに対し) を有している。NPSE102 においては、パイプラインユニット 205a-205n 内の各ストライドレベル検索は 4 つのサイクルを必要とし、32 ビット IPV4 検索器に対する最悪のレイテンシーは 32 を超えることのないサイクルである。然しながら、8 番目のパイプラインユニットの結果が上述した如くに残りの 8 個のパイプラインユニットを介して通過せねばならない場合には、完全な 64 個のサイクルが必要とされる。

【0024】

図 4A は、全てのルックアップが順番に取扱われるものと仮定し、最悪のレイテンシーを 64 サイクルから 32 サイクルへ減少させる構造を例示している。一連のパイプラインユニット 205a-205n 内の各パイプラインユニット 205i は、バッファ 402 を介して前の処理レベル $i-1$ からの入力を受取る対応する処理レベル i に対する処理論理 401 を包含している。ルックアップ動作がパイプラインに沿って進行する場合に、ストライドが各レベルにおいて消費される場合に処理されるべき残存する入力検索器内の入力ビットの数が追跡される。少なくとも 1 つのパイプラインユニット内において (IPV4 / IPV6 ルックアップの例においては 8 番目のパイプラインユニット)、信号 403 が発生され、ストライドが残存していない時を表わす。この信号 403 は論理 404 によって使用されて制御ユニット 405 (例えば、デマルチプレクサ) への入力を発生し、検索結果がパイプラインを介して継続するか (即ち、処理レベル $i+1$ 用のバッファへ通過される) 又はリタイアされるか否かを決定する。

【0025】

図 4A の構造の場合には、ルックアップ動作が、入力検索器において処理されるべき更なるビットが存在していないレベルに到達すると、パイプラインにおけるさらに下方においてアクティブな検索が存在しない場合には、その動作の結果はリタイアすることが可能である。そのために、論理 404 は一連 205a-205n、即ち物理的に「前方」であるか又は更にパイプラインチェーンに沿ってのパイプラインユニット内の各爾後のパイプラインユニットからのアクティブな検索信号を受取るが、興味のある時刻において、時間においてより速い検索を処理する (現在のパイプラインユニットにおいて処理中の検索が発行される前に発行される)。然しながら、NPSE102 が最大のサポートされる長さより短い単一長のプレフィックスを処理するに過ぎない場合 (即ち、IPV4 又は IPV6 プレフィックスのいずれかを取扱うべく適合されているが、動作において、NPSE102 が IPV4 プレフィックスのみを取扱うために使用される)、このようなアクティブな信号検索は必要とされない。

【0026】

信号 403、論理 404 及び制御ユニット 405 は、32 ビット検索器を有するテーブルに対する最悪の場合のレイテンシーは、8 番目のものより大きな処理レベルにおけるアクティブの未決のルックアップが存在しない場合には、64 サイクルの代わりに 32 サイクルであることを確保する。早期のリタイアの場合、成功したルックアップが次のホップポインター情報と共に外部 SRAM103 へ転送され、一方失敗したルックアップは NPUI01 へ帰還する。

【0027】

更に、図 4A における構造の複数個のインスタンスはパイプラインユニット 205a-205n 内において実現することが可能である。例えば、IPV4 / IPV6 ルックアップを取扱うために 8 番目のパイプラインユニットにおいて実現されることに加えて、図 4A の構造は、又、NPSE102 が異なる長さのプレフィックスに対するルックアップを取扱う場合に 5 番目のものより大きな処理レベルにおいてアクティブな未決のルックアッ

プが存在しない場合には、32又は64サイクルではなく20サイクルで20ビットVPNプレフィックスに関するルックアップの早期のリタイアを可能とするために5番目のパイプラインユニットにおいて実現することが可能である。

【0028】

図4Bは平均的な場合のレイテンシーに対処する場合である。図4Aの構造であっても、8番目のものより速い処理レベルにおいて成功裡に又は不成功裡に完了した32ビットプレフィックスルックアップは、尚且つ、8番目の処理レベルの後にリタイアするに過ぎない（何故ならば、ストライドが処理されるべく残存するからである）。然しながら、全てのルックアップが順番に処理されるものと仮定すると、図4Bの構造は、検索がより速くにパイプラインからリタイアすることを可能とする。処理されるべく残存する更なるストライドが存在するか否かに依存する信号の代わりに、処理論理401は、検索が完了したか否かを表わす信号406を発生する（即ち、ミス又は端部ノードに対する一致のいずれかが決定されたか否か）。

【0029】

図4Bのシステムの場合には、検索動作が成功裡に又は不成功裡に完了すると、検索完了信号406がアサート即ち活性化される。信号406がアサートされ且つパイプラインの更に下側にアクティブな検索が存在しない場合には、この特定の検索は、処理されるべきプレフィックス内にストライドが残存するか否かに拘わらずに、パイプラインからリタイアすることが可能である。検索結果は、最早、パイプラインの残部を介して段階的に進むことは必要ではない。

【0030】

図4Bの検索完了検知システムは、可及的に最も速い点において各検索のリタイアを可能とするために全てのパイプラインユニット205a-205n（即ち、全ての処理レベルにおいて）内に実現することが可能である。然しながら、相互接続ルーチング（経路付け）及びタイミング拘束条件、及びその他の実際的な理由が、平均的な場合のレイテンシーに利点があるようなあるレベル又はプレフィックス長においてのみ図4Bの構造を使用することを正当化する場合がある。図4A及び4Bのメカニズムは、最後のパイプラインユニット205nにおいて必要とされるものではないが、そのパイプラインユニット205nからのアクティブな検索信号が、処理レベルi乃至n-1に対する先行するパイプラインユニットに対して必要とされる場合がある。

【0031】

更に、図2のパイプライン実現例に対してエキストラなレベルの複雑性が付加され且つ全ての検索要求に対するトランザクション識別子でパイプラインにおいて順番外に検索を完了することが許容される場合には、パイプライン内の後の処理レベルが未だにルックアップ要求を処理している場合であっても、検索を可及的速やかにリタイアさせることが可能である。NPU101が順番外の検索結果及びNPU101とNPSE102との間のインターフェースの性質を受付けることが可能であるか否かに依存して、検索応答を順番でリターンするために並び替えバッファを使用することが可能であり、設計の複雑性が増加することの犠牲により平均的なレイテンシーを減少させる。

【0032】

本発明はパイプライン型マルチビットトライネットワーク検索エンジンにおけるルックアップ要求のレイテンシー即ち待ち時間を減少させる。検索エンジンパイプラインにおける長いレイテンシーを完全に利用するために十分な数のスレッドを発生することが不可能なNPUは、しばしば、現在のスレッドが関連するルックアップ要求を完了するのを待機している間に、ルーターラインカードに関するそれらの入力キューがバックアップされていることを見出す。減少させた最悪又は平均的な場合のレイテンシーの場合に、ルックアップはより速く完了し且つこのような輻輳を減少させるか又は開始する。

【0033】

以上、本発明の具体的実施の態様について詳細に説明したが、本発明は、これら具体例にのみ制限されるべきものではなく、本発明の技術的範囲を逸脱することなしに種々の変

10

20

30

40

50

形が可能であることは勿論である。

【図面の簡単な説明】

【0034】

【図1】本発明の1実施例に基づく形態特定可能多段パイプライン型マルチビットトライネットワーク検索エンジンを使用した処理システムを示した概略図。

【図2】本発明の1実施例に基づく形態特定可能多段パイプライン型マルチビットトライネットワーク検索エンジンに対するマイクロアーキテクチャを示した概略図。

【図3】本発明の1実施例に基づく形態特定可能多段パイプライン型マルチビットトライネットワーク検索エンジン内のマルチビットトライ検索のハードウェアパイプライン化を例示した概略図。

【図4A】本発明の1実施例に基づく形態特定可能多段パイプライン型マルチビットトライネットワーク検索エンジン内の処理レベルに対するパイプラインユニットの詳細な構造を示した概略図。

【図4B】本発明の1実施例に基づく形態特定可能多段パイプライン型マルチビットトライネットワーク検索エンジン内の処理レベルに対するパイプラインユニットの詳細な構造を示した概略図。

【符号の説明】

【0035】

100 処理システム

101 ネットワーク処理ユニット (NPU)

102 ネットワークパケット検索エンジン (NPSE)

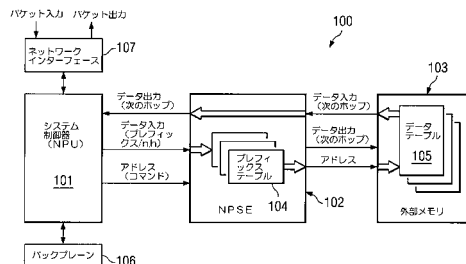
103 外部メモリ

104 プレフィックステーブル

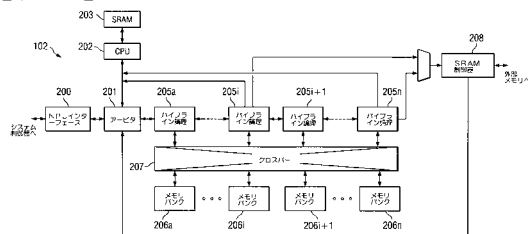
105 データテーブル

107 ネットワークインターフェース

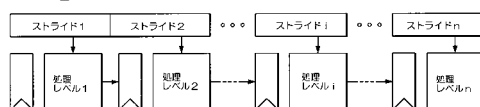
【図1】



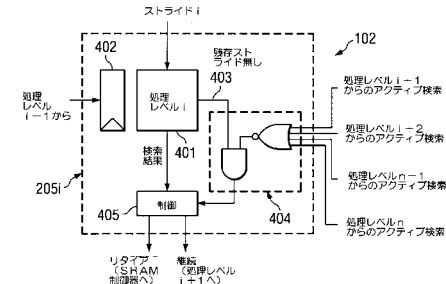
【図2】



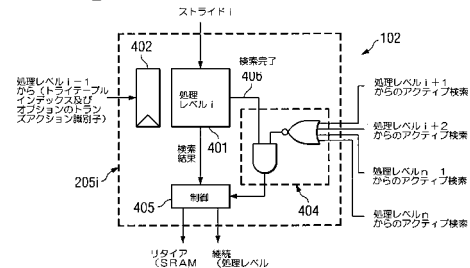
【図3】



【図4A】



【図4B】



10

20

フロントページの続き

(72)発明者 ルン ビン フアン

アメリカ合衆国, カリフォルニア 9 2 1 3 0, サン ディエゴ, ウィロウメアー レーン
5 6 2 9

(72)発明者 ニコラス リチャードソン

アメリカ合衆国, カリフォルニア 9 2 1 0 9, サン ディエゴ, アーチャー ストリート
7 4 0

Fターム(参考) 5K030 GA01 HD03 KA05 LB05