

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3881663号  
(P3881663)

(45) 発行日 平成19年2月14日(2007.2.14)

(24) 登録日 平成18年11月17日(2006.11.17)

(51) Int. Cl.

H04L 12/56 (2006.01)

F I

H04L 12/56 100Z

請求項の数 17 (全 14 頁)

(21) 出願番号 特願2004-56158 (P2004-56158)  
 (22) 出願日 平成16年3月1日(2004.3.1)  
 (65) 公開番号 特開2004-266837 (P2004-266837A)  
 (43) 公開日 平成16年9月24日(2004.9.24)  
 審査請求日 平成16年3月1日(2004.3.1)  
 (31) 優先権主張番号 2003-012902  
 (32) 優先日 平成15年2月28日(2003.2.28)  
 (33) 優先権主張国 韓国(KR)

(73) 特許権者 390019839  
 三星電子株式会社  
 Samsung Electronics  
 Co., Ltd.  
 大韓民国京畿道水原市靈通区梅灘洞416  
 (73) 特許権者 594006297  
 ポリテクニク ユニヴァーシティー  
 アメリカ合衆国 ニューヨーク州 112  
 01-2990 ブルックリン メトロテ  
 ク センター 6  
 (74) 代理人 100064414  
 弁理士 磯野 道造

最終頁に続く

(54) 【発明の名称】 フィールドレベルツリーを用いたパケット分類装置及び方法

## (57) 【特許請求の範囲】

## 【請求項1】

分類のためのマルチフィールドを有するパケットを階層的構造にてフィールド別に組織化したフィールドレベルツリーを生成し維持する主処理部と、

照会と更新業務を行い、前記フィールドレベルツリーが備える第1の群のフィールドをもとに、IPアドレス探索に代表されるプレフィックス探索を処理する第1の分類部及び前記第1の分類部の結果に基づき、それに属するレンジ探索範囲を処理するために、前記フィールドレベルツリーが備える第2の群のフィールドが有する区間をもとに、当該フィールド別に分類を進める第2の分類部を備える分類エンジンと、

を含むことを特徴とするフィールドレベルツリーを用いたパケット分類装置。

10

## 【請求項2】

前記分類エンジンは、分類プロセッサとメモリを含むことを特徴とする請求項1に記載のフィールドレベルツリーを用いたパケット分類装置。

## 【請求項3】

前記主処理部と前記分類エンジンとは、ブロードキャストバスを介して接続されていることを特徴とする請求項2に記載のフィールドレベルツリーを用いたパケット分類装置。

## 【請求項4】

前記第1の分類部は、プレフィックス形態のフィールドを格納し検索するTCAMを使用することを特徴とする請求項1に記載のフィールドレベルツリーを用いたパケット分類

20

装置。

【請求項 5】

前記第 2 の分類部は、用途及びスペックに応じて適宜の  $k$  値を有する  $k$ -way 検索方法を使用することを特徴とする請求項 1 に記載のフィールドレベルツリーを用いたパケット分類装置。

【請求項 6】

前記  $k$  値は、前記第 2 の分類部のメモリインターフェースの大きさによって決められることを特徴とする請求項 5 に記載のフィールドレベルツリーを用いたパケット分類装置。

【請求項 7】

前記主処理部は、前記ブロードキャストバスを介して前記複数の分類エンジンに更新指示を送り、前記更新指示を受けた分類エンジンは、前記メモリのコンテンツを変更するように指示することを特徴とする請求項 3 に記載のフィールドレベルツリーを用いたパケット分類装置。

10

【請求項 8】

前記フィールドレベルツリーは、第 1 の群のフィールドが上位レベルに示され、第 2 の群のフィールドが下位レベルに示される構造にて組織化されることを特徴とする請求項 1 に記載のフィールドレベルツリーを用いたパケット分類装置。

【請求項 9】

前記第 1 の群のフィールドは、プレフィックス・インプリメンテーションされたフィールドであることを特徴とする請求項 8 に記載のフィールドレベルツリーを用いたパケット分類装置。

20

【請求項 10】

前記第 2 の群のフィールドは、レンジ・インプリメンテーションされたフィールドであることを特徴とする請求項 8 に記載のフィールドレベルツリーを用いたパケット分類装置。

【請求項 11】

前記フィールドレベルツリーは、あるレベルで 2 つのノードが共通の子ノードを有すると、単に 1 つのノードが生成されそれを共有することを特徴とする請求項 1 に記載のフィールドレベルツリーを用いたパケット分類装置。

【請求項 12】

前記フィールドレベルツリーは、プレフィックス探索のためのレベルが互いに結合されたプレフィックスの対を有する単に 1 つのレベルとして存在することを特徴とする請求項 1 に記載のフィールドレベルツリーを用いたパケット分類装置。

30

【請求項 13】

分類のためにマルチフィールドを有するパケットを階層的構造にてフィールド別に組織化したフィールドレベルツリーを生成し維持する主処理部と、

照会と更新業務を行い、前記フィールドレベルツリーが備える第 1 の群のフィールドをもとに、IP アドレス探索に代表されるプレフィックス探索を処理する第 1 の分類部と、

前記プレフィックス探索の処理結果に基づき、それに属するレンジ探索範囲を処理するために、前記フィールドレベルツリーが備える第 2 の群のフィールドが有する区間をもとに、当該フィールド別に分類を進める第 2 の分類部と、

40

を含むことを特徴とするフィールドレベルツリーを用いたパケット分類装置。

【請求項 14】

ルーティングシステムにおけるパケット分類方法において、

マルチフィールドを有するパケットを分類するためにフィールドレベルツリーを形成するステップと、

前記フィールドレベルツリーを用いて、前記フィールドレベルツリーが備える第 1 の群のフィールドをもとに、パケット分類規則に対しプレフィックス探索を行なうステップと、

プレフィックス探索が行われた後、前記フィールドレベルツリーが備える第 2 の群のフ

50

フィールドが有する区間をもとに、レンジ探索を行なうステップと、  
を含むフィールドレベルツリーを用いたパケット分類方法。

【請求項 15】

前記フィールドレベルツリーは、第1の群のフィールドが上位レベルに示され、第2の群のフィールドが下位レベルに示される構造にて組織化されることを特徴とする請求項14に記載のフィールドレベルツリーを用いたパケット分類方法。

【請求項 16】

前記フィールドレベルツリーは、あるレベルで2つのノードが共通の子ノードを有すると、単に1つのノードが生成されそれを共有することを特徴とする請求項14に記載のフィールドレベルツリーを用いたパケット分類方法。

10

【請求項 17】

前記フィールドレベルツリーは、プレフィックス探索のためのレベルが互いに結合されたプレフィックスの対を有する単に1つのレベルとして存在することを特徴とする請求項14に記載のフィールドレベルツリーを用いたパケット分類方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、通信システムにおけるパケット等処理する分野に係り、ルータでパケットを分類するに際し、フィールドレベルツリーを用いてパケットを分類するパケット分類方法及びその方法を実現させるためのパケット分類装置に係る。

20

【背景技術】

【0002】

インターネットのような通信システムにおいてパケットを処理するに際し、通常は、宛先アドレスを計算し、パケット経路上の中間ノードのそれぞれからどの出力ノードまたはリンク上にパケットを送信すべきかを定める。各種の通信システムでは、各パケットのヘッダにある宛先アドレス、送信元アドレスまたはその他のデータによって様々な形態のサービスが提供される。サービス形態の差というものとしては、パケットが処理または送信される優先順位やその伝送に対し支払うべき料金または特定の送信元に対してはパケット処理を拒否すること等が挙げられる。

【0003】

30

今日のシステムでは、膨大な量のパケット（一般には、データエントリ）を処理しなければならないため、パケットと一緒に受信した内容を読み取り、それに応じてどのような形態の処理を選択すべきかを決定することを非常に短時間内に高速で行われなければならない。

【0004】

未来のIPネットワーク上でユーザーに対しサービスレベルアグリーメント（service level agreements）、VPN、QoS等のより発展したサービスを提供するために、ルータ等で入力されるIPパケットを所望する規則に沿って分類可能でなければならない。これをパケット分類といい、パケット分類は、パケット内の複数のフィールドの値を参照しなければならないマルチフィールド探索（multi-field lookup）である。

40

【0005】

言い換えれば、パケット分類は、既存のIP宛先アドレス探索とは異なって、1つのパケット内のソースアドレス、宛先アドレス、プロトコルID、ポート番号等の複数のフィールドを参照してパケットを処理しなければならない。従って、基本的により多くの時間とメモリを必要とすることはもとより、これを解決する方法に対する研究も足りないのが実情である。

【0006】

図1乃至図3は、従来のパケット分類方法の一例を示す図である。

【0007】

50

既存方法を構成するパケット分類ツリーの例を、次の表 1 に表す。

【 0 0 0 8 】

【表 1】

F i l t e r	F 1	F 2
R <sub>1</sub>	0 0 *	1 1 *
R <sub>2</sub>	0 0 *	1 *
R <sub>3</sub>	1 0 *	1 *
R <sub>4</sub>	0 *	0 1 *
R <sub>5</sub>	0 *	1 0 *
R <sub>6</sub>	0 *	1 *
R <sub>7</sub>	*	0 0 *

10

【 0 0 0 9 】

従来のパケット分類方法としては、図 1 に示すようなグリッド・オブ・ツリー ( g i r d - o f - T r e e s ) に代表される伝統的な方法と、図 2 のレンジ探索 ( r a n g e l o o k u p ) に代表される幾何学的な方法、そして、図 3 の帰納的フロー分類 ( r e c u r s i v e f l o w c l a s s i f i c a t i o n ) に代表される経験的な方法等が挙げられる。

【 0 0 1 0 】

このうち、グリッド・オブ・ツリーのデータ構造は、標準階層ツリーとセット・ブルーニングツリーの長所を組み合わせたものであり、N 分類規則が適用される W ビット長のヘッダフィールド d ( すなわちディメンション ) の数に基づく 0 ( d W ) の照会 ( q u e r y ) 時間の複雑度と 0 ( N d W ) の格納複雑度を有する。この特徴は、スイッチングポイントをデータ構造に導入することにより得られる。しかし、スイッチングポイントは、ビットレベルで組み合わせられるため、0 ( N d W ) の格納複雑度を保証し、照会過程はビットずつ行なう必要がある。実行が求められる実際のアプリケーションにおいてビットずつの照会過程 ( または分類 ) は受け入れられない。

20

【 0 0 1 1 】

パケットを分類するに際し、分類速度、メモリの大きさ、分類規則数、参照フィールド数、規則更新時間、及び最悪の条件時の性能が主な項目として考慮される。

30

【 0 0 1 2 】

即ち、パケット分類では、使用可能なメモリを使用して最大の性能を確保するために、与えられた容量 ( s t o r a g e l i m i t a t i o n ) で高い性能 ( H i g h P e r f o r m a n c e ) を達成することに関する問題を解決することが重要である。

【発明の開示】

【発明が解決しようとする課題】

【 0 0 1 3 】

そこで、本発明は、前記のような問題点を解決するために、従来から使用されてきたビット単位のツリーではなくフィールド単位のツリーを展開し、T C A M と k - w a y 検索を用いてそれぞれプレフィックス・インプリメンテーションとレンジ・インプリメンテーションを取り扱う、照会性能を改善したフィールドレベルツリー ( F L T ; F i e l d - L e v e l T r e e ) を用いたパケット分類装置及び方法を提供することを目的とする。

40

【課題を解決するための手段】

【 0 0 1 4 】

前記目的を達成するために、本発明のパケット分類装置は、分類のためにマルチフィールドを有するパケットを階層的構造にてフィールド別に組織化したフィールドレベルツリーを生成し維持する主処理部と、照会と更新業務を行い、I P アドレス探索に代表されるプレフィックス探索を処理する第 1 の分類部及び第 1 の分類部の結果に基づき、それに属するレンジ探索範囲を処理するために当該フィールド別に分類を進める第 2 の分類部を備

50

える分類エンジンとを備える。

【0015】

ここで、分類エンジンは、分類プロセッサとメモリを備える。

【0016】

また、主処理部と分類エンジンとは、ブロードキャストバスを介して接続されている。

【0017】

第1の分類部は、プレフィックス形態のフィールドを格納しそれを検索するTCAMを使用することが好ましい。

【0018】

そして、第2の分類部は、用途及びスペックに応じて適宜のk値を有するk-way検索方法を使用することが好ましい。

【0019】

そして、k値は、第2の分類部のメモリインターフェースの大きさによって決められることが好ましい。

【0020】

また、主処理部は、ブロードキャストバスを介して複数の分類エンジンに更新指示を送り、更新指示を受けた分類エンジンは、メモリのコンテンツを変更するように指示する。

【0021】

フィールドレベルツリーは、第1の群のフィールドが上位レベルに示され、第2の群のフィールドが下位レベルに示される構造にて組織化される。

【0022】

第1の群のフィールドは、プレフィックス・インプリメンテーションされたフィールドであり、第2の群のフィールドは、レンジ・インプリメンテーションされたフィールドであることが好ましい。

【0023】

フィールドレベルツリーは、あるレベルで2つのノードが共通の子ノードを有すると、単に1つのノードが生成されそれを共有する。

【0024】

フィールドレベルツリーは、プレフィックス探索のためのレベルが互いに結合されたプレフィックスの対を有する単に1つのレベルとして存在する。

【0025】

一方、本発明に係るフィールドレベルツリーを用いたパケット分類装置は、分類のためにマルチフィールドを有するパケットを階層的構造にてフィールド別に組織化したフィールドレベルツリーを生成し維持する主処理部と、照会と更新業務を行い、IPアドレス探索に代表されるプレフィックス探索を処理する第1の分類部と、プレフィックス探索の処理結果に基づき、それに属するレンジ探索範囲を処理するために当該フィールド別に分類を進める第2の分類部とを含む。

【0026】

前記目的を達成するために、本発明のパケット分類方法は、ルーティングシステムにおけるパケット分類方法において、与えられたマルチフィールドを有するパケットをフィールド別に展開してフィールドレベルツリーを形成する第1のステップと、フィールドレベルツリーを用いてパケット分類規則に対しプレフィックス探索を行なう第2のステップと、プレフィックス探索が行われた後、レンジ探索を行なう第3のステップとを含む。

【発明の効果】

【0027】

本発明によると、フィールド単位のツリーを展開することによって、優れた照会性能が確保された超高速のネットワークのためのパケット分類を実現することができる。また、約50万個の分類ツリーのルールを処理することができる。

10

20

30

40

50

【発明を実施するための最良の形態】

【0028】

以下、図面を参照して本発明をより詳細に説明する。

【0029】

図4は、本発明に係るフィールドレベルツリーを用いたパケット分類装置の構成図である。

【0030】

同図に示すように、本発明に係るフィールドレベルツリーを用いたパケット分類装置は、主処理部10と複数の個別分類エンジン20を備える。

【0031】

主処理部10は、FLT (Filed Level Tree) データ構造を構成及び管理し、該情報を加工してそれぞれの分類エンジン20に提供する。それぞれのインターフェイスに属する分類エンジン20は、入力する個別パケットを実際に分類する。

【0032】

高速ノコアルータにおいて、パケット分類は、一般にそれぞれのラインカードで並列に行われる。従って、それぞれのラインカードは、少なくとも1つの内蔵された分類エンジンを有する。伝送リンクの要求帯域幅によって、これらの分類エンジンは照会 (query) 機能に最適化する。

【0033】

本発明に係るフィールドレベルツリーを用いたパケット分類方式は、TCAM (Ternary Content Addressable Memory) 技術とk-way検索とを組み合わせるプレフィックス・インプリメンテーションとレンジ・インプリメンテーションを取り扱うハイブリッド方式である。

【0034】

しかし、分類は動的であり、データ構造は、刻々と変更する必要がある。分類エンジンにおいてデータの組織は、ルールの挿入または削除のようなアップデートに適しないことがある。

【0035】

分類は、それぞれのラインカードで同一であり、中心のラインカードで実行されるアップデート機能を有することがより好ましいが、必ずしも必要ではない。

【0036】

主処理部10は、通常、強力なプロセッサを備え、ラインカードとは別のボードに位置する。

【0037】

主処理部10は、分類のために使用された全体のデータ構造を維持し、分類のアップデートが必要とされる時に変更する。図6乃至図8は、本発明に係るデータ構造を示し、図6に示したデータ構造は、主処理部10で維持され、図7、図8(a)及び(b)に示したデータ構造は、分類エンジン20で使用される。

【0038】

図示したデータ構造において、それぞれのノードにルールを格納する必要はない。図8(a)及び(b)に示したように、プレフィックスの対、k-way検索値及びポインタが分類エンジン20で維持される。

【0039】

主処理部10は、更新指示をブロードキャスティングバス12を介して全ての分類エンジン20に送る。これらの更新指示は、分類エンジン20がメモリ(及び/またはTCAM)のコンテンツを変更するように指示するのに使用される。

【0040】

図5は、図4中の分類エンジン20の構成図である。

それぞれの分類エンジン20は、IPソース/宛先アドレス探索に代表されるプレフィックス探索を処理する部分21と、1次の分類の結果に基づき、それに属する2次分類を

10

20

30

40

50

処理する部分 2 2 とからなる。2 次分類において、レンジ探索範囲を処理するために当該フィールド別に分類を進めるための下位マイクロエンジンが備えられる。

【 0 0 4 1 】

2 次分類を行うために、用途及びスペックに応じて適当な値を有する k - w a y 検索方法を使用する。

【 0 0 4 2 】

分類エンジン 2 0 は、分類プロセッサ 3 2 と、T C A M 4 1 と、一般のメモリ ( S D R A M 及び / または S S R A M ) 4 2 を備える外部メモリと、からなる。

【 0 0 4 3 】

分類プロセッサ 3 2 は、その内部にマイクロエンジン 3 1 とメモリアンターフェイス 3 0 の主要構成要素を備えている。マイクロエンジン 3 1 は、特定のアプリケーション用または一般用の R I S C プロセッサである。データ構造は、外部メモリに格納されている。内部マイクロエンジン 3 1 は、照会動作に対し制御機能を行い、メモリアンターフェイス 3 0 を介してデータに接続する。

10

【 0 0 4 4 】

取り出された全てのフィールドを有するパケットヘッダは、左側から分類プロセッサ 3 2 に入る。第 1 のマイクロエンジン 3 1 は、プレフィックスフォーマットでインプリメンテーションされた 1 番目の k フィールドを取り、T C A M 4 1 に送る。T C A M 4 1 では検索を実行し、その結果を第 1 マイクロエンジン 3 1 にフィードバックする。その結果を、複数のマイクロエンジンがレンジでインプリメンテーションした次のフィールドの k - w a y 検索に対して存在する第 2 の段に送る。例えば、図 5 において 1 つのマイクロエンジン 3 1 は第 2 の段でそれぞれのフィールドに用いられる。しかし、実際の動作において、1 つのマイクロエンジンは複数のフィールドを収容することができる。マイクロエンジンの数は、それぞれのフィールドで検索のために要する外部メモリの帯域幅、マイクロエンジンの速度及び平均帯域幅によって決められる。

20

【 0 0 4 5 】

照会機能の他にも、分類エンジンは、主処理部から送信された更新指示に基づいてその外部メモリの内容 ( コンテンツ ) を更新する必要がある。これら 2 つの作業におけるコンテンツを解決する 2 つの方法がある。a ) 分類エンジンは、照会または更新を何時でも実行可能なタスクをインタリーブする。時間消費に対して更新費用が少ない時に実現可能である。b ) メモリにおいてデータ構造の 2 つのコピーを作る。一方は照会のために使用され、他方は更新のために使用される。一つの更新動作が終了すると、更新されたコピーが照会用に転換され、他方のコピーは更新のために使用される。この方法は、インタリーブ方法の照会性能が非常に低くなる程更新費用が高い時に適用される。

30

【 実施例 】

【 0 0 4 6 】

次いで、本発明に係るフィールドレベルツリーの分類構造の実施例と、これに係るフィールドレベルツリーの分類方法について、図面を参照して詳細に説明する。

【 0 0 4 7 】

フィールドレベルツリーの分類構造は、マルチフィールドを有する分類ツリーの分類を目標にし、それぞれのフィールドは、個々のプレフィックス・インプリメンテーションまたはレンジ・インプリメンテーションで・インプリメンテーションされる。

40

【 0 0 4 8 】

4 つのフィールドと 7 つの規則を有する分類ツリーの一例が、次の表 2 に表わされている。

【 0 0 4 9 】

【表 2】

R u l e	F 1	F 2	F 3	F 4
R <sub>1</sub>	0 0 *	1 1 0 *	6	[1 0、1 2]
R <sub>2</sub>	0 0 *	1 1 *	[4、8]	1 5
R <sub>3</sub>	1 0 *	1 *	7	9
R <sub>4</sub>	0 *	0 1 *	1 0	[1 0、1 2]
R <sub>5</sub>	0 *	1 0 *	[4、8]	1 5
R <sub>6</sub>	0 *	1 *	1 0	[1 0、1 5]
R <sub>7</sub>	*	0 0 *	7	1 5

10

## 【 0 0 5 0 】

前記表 2 において、最初のフィールド F 1、F 2 はプレフィックスでインプリメンテーションされ、残りの 2 つのフィールド F 3、F 4 はレンジでインプリメンテーションされる。これを F L T で構成したデータ構造図が図 6 に示されている。

## 【 0 0 5 1 】

各個別パケットは、各レベルで該当するフィールドに対し定めた規則に沿って分類される。与えられた全てのレベルのフィールドを経ると分類作業が終り、個別のパケットに定められた規則を与えるようになる。

## 【 0 0 5 2 】

20

次いで、図 6 のフィールドレベルツリーのデータ構造図を詳細に説明する。フィールドレベルツリーのデータ構造は、次の属性を有するように定義される。

## 【 0 0 5 3 】

1 . フィールドレベルツリーは、階層的構造にてフィールドずつ組織化される。ツリーの深さは、フィールドの数 d と同じである。図 6 の一例において F 1 から F 4 まで組織化された 4 つのノードレベルがある。図 6 において、底のノードは別のレベルを形成しない。照会過程が第 4 のレベルで終了する時、どのルールがマッチングするかを示す。

## 【 0 0 5 4 】

2 . ツリーにおいてそれぞれのノードは、親ノードのルールセットのサブセットであるルールセットを含む。ツリーのルートノードは、分類ツリーで全てのルールを含むように

30

## 【 0 0 5 5 】

3 . i 番目のレベルのノード a ( ルートノードは第 1 のレベルにあることと定義される ) は、ノード a に含まれた全てのルールの F i フィールドの値に基づいて i + 1 番目のレベルでその子ノードを生成する。

## 【 0 0 5 6 】

フィールド F i の特定に基づき、子ノードの生成に対する 2 つの他の過程がある。

## 【 0 0 5 7 】

( a ) F i がプレフィックスで特定化されると、a の子ノードの数は、a のルールセットで F i の他の値の数と同一になる。その結果、それぞれの子ノードは、他のプレフィックスと結合される。子ノード b がプレフィックス a と結合されていると仮定すると、b のルールセットに含まれているルール r ( r は、a にも含まれている ) の F i 値が同じであるか、p のプレフィックスになる。

40

## 【 0 0 5 8 】

例えば、図 6 においてルートノードは 7 つのルールの全てを含み、F 1 フィールドには、4 つの他のプレフィックス、0 \*、0 0 \*、1 0 \* があるため、4 つの子ノードが生成される。

## 【 0 0 5 9 】

プレフィックス 0 \* と結合されているノード X は、4 つのルール R 4 ( R 7 を含む。R 4 ( R 6 の F 1 値は 0 \* であり、これはプレフィックスと結合されたものである。R 7 の F 1

50



値は  $0^*$  であり、これは  $0^*$  のプレフィックスである。

【0060】

(b)  $F_i$  がレンジで特定化されると、レンジをナンバーラインに投げだし、1セットの区間を得る。それぞれの区間  $I$ 、子ノード  $b$  が生成される。ルール  $r$  は、 $b$  のルールセットに含まれ、 $r$  の  $F_i$  フィールドにより特定化された範囲は1を含む。例えば、ノード  $y$  は、区間  $[10, 10]$  を有し単一ポイントであるノード  $y$  (、区間  $[6, 6]$  を有するノード  $y$  ( (及び区間  $[4, 5]$  と  $[6, 7]$  を有するノード  $y$  ( (の3つの子ノードを生成する。

【0061】

$i$  番目のレベルでノード  $a$  のルールセットは、同一のレベルにある全てのノードのルールセットの中で特殊になる。

10

【0062】

$i - 1$  番目のレベルで2つのノード  $b$ 、 $c$  が共通の子ノード  $a$  を有すると、単に1つのノード、ノード  $a$  が生成され、それを共有する。ノードがマルチポイントにより指摘される時にノード共有が発生することが図6から分かる。ノード共有は、グリッド・オブ・ツリーでスイッチングポイントのメカニズムと類似しているが、グリッド・オブ・ツリーがビットレベルでノード共有を得るのに対し、本発明で提案された方式は、フィールドレベルでノード共有を得る。フィールドレベル共有は、ルールのコピーが回避できないため照会性能は改善するものの、格納要求は増加する。図6に示したように、ルール  $R_7$  は、第2のレベルで4つのノードに格納される。

20

【0063】

次いで、フィールドレベルツリーのノード構造と照会過程について説明する。

【0064】

それぞれのフィールドは、通常、プレフィックスまたはレンジフォームでインプリメンテーションされ、それぞれのインプリメンテーションは、それ自身の好むデータ構造と検索アルゴリズムを有する。従って、分類ツリーは、2群のフィールドを有する。第1の群は、プレフィックス・インプリメンテーションであり、第2の群は、レンジ・インプリメンテーションである。

【0065】

前記表2に示した分類ツリーにおいて、 $F_1$  と  $F_2$  は第1の群であり、 $F_3$  と  $F_4$  は第2の群である。フィールドレベルツリーでは、第1の群のフィールドが上位レベルに示され、第2の群のフィールドが下位レベルに示される構造にて組織化される。

30

【0066】

単にプレフィックス形態のフィールドを含む第1の群に対し、TCAMは、プレフィックスを格納しそれらを検索するために使用される。更に、TCAMは、同時にマルチフィールドを収容することができるため、フィールドの第1の群の照会は、一回だけのメモリ接続により実行可能である。

【0067】

図7は、図6のプレフィックス探索のためのデータ構造(即ち、レベル1、2)を図5の分類エンジン内にTCAM(Ternary Content Addressable Memory)を使用して短縮した構成図である。図7において、 $F_1$ 、 $F_2$  フィールドに対し1つのレベルだけが存在する。ルートノードは、図6中の第3のレベルにおいて最初にある7つの子ノードを有する。それぞれの第2のレベルノードは、互いに結合された  $F_1 / F_2$  プレフィックスの対を有する。かかるそれぞれのプレフィックスの対は、TCAMにおけるエントリーのコンテンツである。プレフィックスの対は、図6中のツリー構造から誘導される。図7の第2のレベルにおけるノードに対応して、図6の第3のレベルでそれぞれのノード  $x$  に対し、ルートノードからノード  $x$  までにおいて最も小さいプレフィックス長の和で経路を発見する。この経路によるプレフィックスは、図7でノードのプレフィックスの対を形成する。全てのプレフィックスの対は、TCAMにおいてプレフィックス長が減少する順に配列される。同一の長さを有するプレフィックスの対に対し、

40

50

それらの相関順序は任意である。図 7 に示すツリー構造の一例に対する T C A M のコンテンツを、次の表 3 に表わしている。第 2 のレベルにおいて適当なノードは、全体の照会過程を継続すると決められる。

【 0 0 6 8 】

F L T と T C A M の特徴を組み合わせ、工程を縮小し、前記表 2 のプレフィックス探索部分を T C A M 内に実現した例が、次の表 3 である。

【 0 0 6 9 】

【表 3】

エントリー	プレフィックスの対	ノード名	長さの和
1	0 0 * / 1 1 0 *	f	5
2	0 0 * / 1 1 *	e	4
3	1 0 * / 1 *	g	3
4	0 * / 0 1 *	d	3
5	0 * / 1 0 *	c	3
6	0 * / 1 *	b	2
7	* / 0 0 *	a	2

10

【 0 0 7 0 】

プレフィックスの対をその長さが減少する順に T C A M で配列することによって、T C A M における精度が高い検索結果が保証され、第 2 のレベルの正確なノードが照会工程全体を継続すると決められる。

20

【 0 0 7 1 】

図 8 ( a ) 及び図 8 ( b ) は、図 5 の分類エンジン内のレンジ探索用マイクロエンジンで行われる k - w a y 検索を説明するための原理図である。2 次分類の最も大きな性能を左右する k - w a y 検索の k 値をメモリインターフェイスの大きさによって決める。与えられた環境下において k 値は大きい程よい。

【 0 0 7 2 】

分類すべきパケットヘッダが与えられると、第 1 の群に属するフィールドが取り出され、検索のために T C A M に送られる。T C A M からの出力は、次に接続すべき第 2 のレベルでノードを表示する。T C A M は、全てのプレフィックス形態のフィールドを収容するため、照会プロセスの残りは、レンジ・インプリメンテーションフィールドに頼る。第 2 または下位レベルに存在するノードに対し、それぞれのノードに対してバイナリ検索ツリー（または、外部メモリの帯域幅によって k - w a y 検索ツリー）を使用する。例えば、ツリーの i ( i ( 1 ) 番目のレベルでノード a がある。a のルールセットでルールの i 番目のフィールドをナンバーラインに投射した後、図 8 a に示すように 8 つの終点 E 1 ( E 8 を有する 7 つのインターバル I 1 ( I 7 が得られる。これらのインターバルを組織化するために 3 - w a y 検索ツリーを用いると、図 8 b に示すようなツリー構造が得られる。これは、4 つのブロックを有する 2 レイヤーツリーである。それぞれのブロックは、k - ポインタと k - 1 終点を含む。内部ブロックにおいてポインタは、k - w a y 検索ツリーで他のブロックを指示するのに対し、リーフブロックにおいてポインタは、フィールドレベルツリーで次段のレベルノードを指示する。k - w a y 検索ツリーにおいて検索過程を例を挙げて説明する。

30

40

【 0 0 7 3 】

インターバル I 3 内にポイント P が存在すると仮定すると、検索過程は、ルートブロック X から始まる。P を X に格納されている 2 つの終点 E 3、E 6 と比較することによって、それらの間の順序が E 3 ( P ( E 6 であることが分かる。インターバル I 3 に結合された第 1 のポインタが、フィールドレベルツリーの次段のレベルでのノードの後にこななければならない。

【 0 0 7 4 】

50

k - w a y 検索は、レンジ探索問題のための効率よいアルゴリズムである。k - w a y 検索ツリーのレイヤーの数は、 $\log_k M$ により決めることができ、Mは、インターバルの数である。k - w a y 検索ツリーにおいてそれぞれのブロックは、メモリに格納された基本単位であり、1回の読み取り/書き込み動作に対し1回のメモリへの接続を必要とする。その結果、検索過程の間、メモリへの接続の数は、k - w a y 検索ツリーのレイヤーの数と同一である。ここで、kの数は、ブロックの大きさによって制限され、ブロックの大きさは、メモリ帯域幅によって決められる。

#### 【0075】

フィールドレベルツリーの照会過程は、全てのフィールドに対しプレフィックス・インプリメンテーションを有するTCAMから始まる。レンジ・インプリメンテーションを有するフィールドに達した後、照会過程は、1回に1つのレベルに対して進め、k - w a y 検索は、接続する次の子ノードを探し出すために実行される。照会過程は、リーフノードが到達した時に終了され、整合したルールが結果として復帰する。

#### 【産業上の利用可能性】

#### 【0076】

本発明は、通信システムにおけるパケット等の項目を処理する分野で利用でき、ルータでパケットを分類するに際し、フィールドレベルツリーを用いてパケットを分類するパケット分類方法及びその方法を実現させるためのパケット分類装置に適用できる。

#### 【図面の簡単な説明】

#### 【0077】

【図1】従来のパケット分類方法におけるグリッド・オブ・ツリーのデータ構造図を示す図である。

【図2】従来のパケット分類方法におけるレンジ探索の原理を示す図である。

【図3】従来のパケット分類方法における帰納的フロー ( r e c u r s i v e f l o w ) 分類の原理を示す図である。

【図4】本発明に係るフィールドレベルツリーを用いたパケット分類装置を示す構成図である。

【図5】図4中の分類エンジンを示す構成図である。

【図6】表2の分類ツリーをフィールドレベルツリーで構成したデータ構造を示す図である。

【図7】図6中のプレフィックス探索のためのデータ構造を図5の分類エンジン内にTCAM ( T e r n a r y C o n t e n t A d d r e s s a b l e M e m o r y ) を使用して短縮した構成を示す図である。

【図8】( a ) は図5中の分類エンジン内のレンジ探索のためにマイクロエンジンで行われるk - w a y 検索を説明するための原理を示す図であり、( b ) は図5中の分類エンジン内のレンジ探索のためにマイクロエンジンで行われるk - w a y 検索を説明するための原理を示す図である。

#### 【符号の説明】

#### 【0078】

- 10 主処理部
- 12 ブロードキャスティングバス
- 20 分類エンジン
- 30 メモリインターフェイス
- 31 マイクロエンジン
- 32 分類プロセッサ
- 41 TCAM
- 42 メモリ

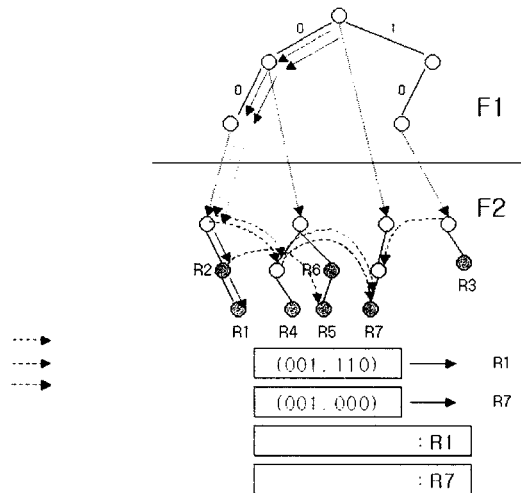
10

20

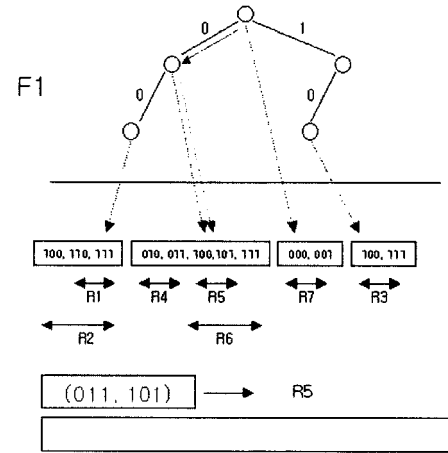
30

40

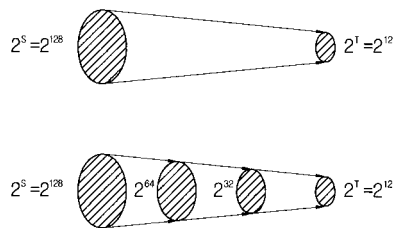
【図 1】



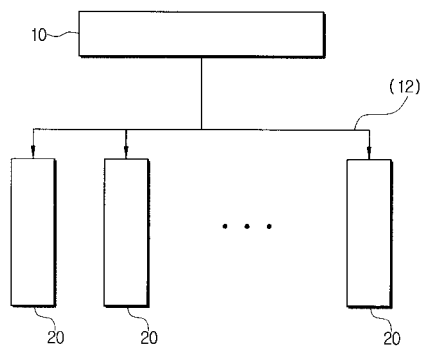
【図 2】



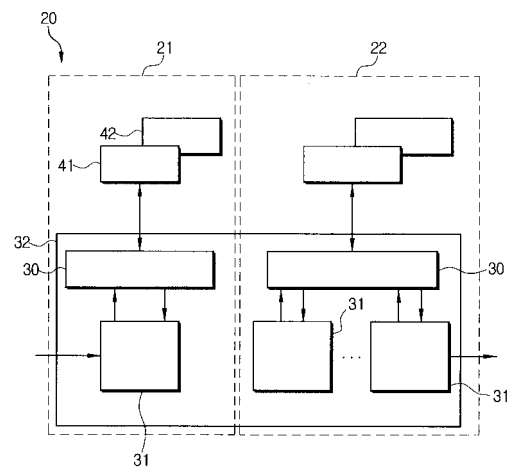
【図 3】



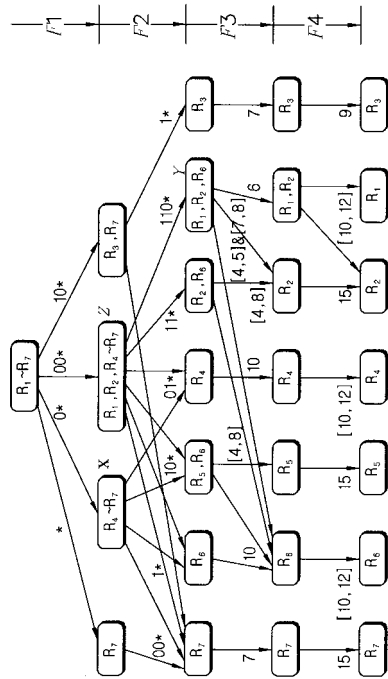
【図 4】



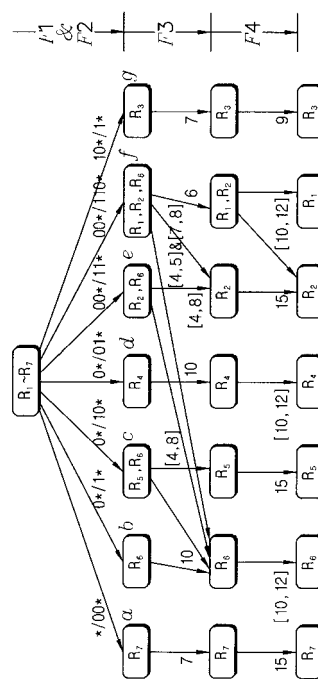
【図 5】



【図 6】



【図 7】

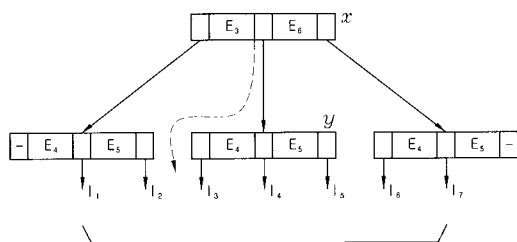


【図 8】

(a)



(b)



---

フロントページの続き

- (72)発明者 鄭 鎮 宇  
大韓民国 京畿道 龍仁市 器興邑 新葛里 道顯マウル 現代アパート 201-501
- (72)発明者 朴 寓 鍾  
大韓民国 ソウル特別市 銅雀區 舍堂2洞 極東アパート 112-1513
- (72)発明者 張 冠 松  
アメリカ合衆国 ニューヨーク州 ブルックリン  
クニック ユニヴァーシティ電気コンピュータ工学部内 メトロテックセンター6 ポリテ
- (72)発明者 ジョナサン チャオ  
アメリカ合衆国 ニューヨーク州 ブルックリン  
クニック ユニヴァーシティ電気コンピュータ工学部内 メトロテックセンター6 ポリテ

審査官 衣鳩 文彦

- (56)参考文献 特開2001-223750(JP,A)  
特開2002-305539(JP,A)  
特開平11-317783(JP,A)  
特開2004-015561(JP,A)  
特開2000-083055(JP,A)  
V.Srinivasan et.al, Fast and Scalable Layer Four Switching , ACM SIGCOMM Computer Communication Review , Proceedings of the ACM SIGCOMM '98 conference on Applications, technologies, architectures, and protocols for computer communication SIGCOMM '98 , 1998年, Volume 28 Issue 4 , p.191-202  
黒田健児 他, データベース:専用LSIでソート処理を30倍に高速化 - レコードの比較方式を工夫, 日経エレクトロニクス, 1998年11月30日, 第731号, p.153-160

- (58)調査した分野(Int.Cl., DB名)  
H04L 12/56