(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第3715098号
(P3715098)
(45)発行日　平成17年11月9日(2005.11.9)　(24)登録日　平成17年9月2日(2005.9.2)

(51)Int.Cl.[7]　F I
H04L 12/56　H04L 12/56　200F

請求項の数 6　(全 9 頁)

(21)出願番号	特願平10-27573
(22)出願日	平成10年2月9日(1998.2.9)
(65)公開番号	特開平10-313324
(43)公開日	平成10年11月24日(1998.11.24)
審査請求日	平成11年7月8日(1999.7.8)
審判番号	不服2002-19979(P2002-19979/J1)
審判請求日	平成14年10月11日(2002.10.11)
(31)優先権主張番号	60/037844
(32)優先日	平成9年2月7日(1997.2.7)
(33)優先権主張国	米国(US)
(31)優先権主張番号	08/972424
(32)優先日	平成9年11月18日(1997.11.18)
(33)優先権主張国	米国(US)

(73)特許権者　596092698
ルーセント　テクノロジーズ　インコーポレーテッド
アメリカ合衆国．０７９７４－０６３６　ニュージャーシィ，マレイ　ヒル，マウンテン　アヴェニュー　６００
(74)代理人　100064447
弁理士　岡部　正夫
(74)代理人　100085176
弁理士　加藤　伸晃
(74)代理人　100106703
弁理士　産形　和央
(74)代理人　100081053
弁理士　三俣　弘文

最終頁に続く

(54)【発明の名称】通信ネットワークにおけるパケットの配送装置とその方法

(57)【特許請求の範囲】

【請求項１】

（Ａ）それぞれのソースからパケットを受領し、宛先に配送する前にこの受領したパケットを一時的に記憶する複数の結線毎の待ち行列（２０）と、

（Ｂ）保証されたあらかじめ割り当てられたレートでもって、前記（Ａ）複数の結線毎の各待ち行列（２０）からのパケットをサービスする重み付き公平待ち行列スケジューリング手段（４０）と、

（Ｃ）前記待ち行列内のパケットの存否を検出する手段と、

（Ｄ）前記パケットの不存在を検出すると、その結果発生する余剰のバンド幅を再分配する状態依存性スケジューリング手段（２００）と、

からなり、

前記（Ｄ）状態依存性スケジューリング手段（２００）は、前記待ち行列（２０）の性能特性に応じた状態に従って、これらの待ち行列（２０）にサービスし、

その結果、それぞれの待ち行列（２０）のパケットをルーティングする遅延特性と分離特性が保存される

ことを特徴とする通信ネットワークにおけるパケットの配送装置。

【請求項２】

前記待ち行列（２０）は、最大量のパケットを有する待ち行列を含み、前記（Ｄ）状態依存性スケジューリング手段（２００）は、最長の待ち行列に最初にサービスする

ことを特徴とする請求項１の装置。

【請求項３】
前記待ち行列（２０）は、サービスされる為に最も長時間待機している最大遅延のパケットを有する待ち行列を含み、
前記（Ｄ）状態依存性スケジューリング手段（２００）は、最大遅延の待ち行列に最初にサービスする
ことを特徴とする請求項１の装置。
【請求項４】
前記待ち行列（２０）は、そのバッファメモリがオーバフローする可能性が最も大きい待ち行列を含み、
前記（Ｄ）状態依存性スケジューリング手段（２００）は、バッファがオーバフローする最も可能性の高い待ち行列に最初にサービスする
ことを特徴とする請求項１の装置。
【請求項５】
前記接続に対し、最悪の場合の公平尺度が満足されている
ことを特徴とする請求項１の装置。
【請求項６】
（Ａ）それぞれのソースからパケットを受領し、宛先に配送する前に、この受領したパケットを一時的に記憶する待ち行列（２０）を形成するステップと、
（Ｂ）保証されたあらかじめ割り当てられたレートでもって前記複数の結線毎の待ち行列（２０）からのパケットにサービスするステップと、
（Ｃ）前記待ち行列内のパケットの存否を検出するステップと、
（Ｄ）前記パケットの不存在を検出すると、その結果発生する余剰のバンド幅を再分配するステップと、
からなり、
前記（Ｄ）再分配ステップは、前記待ち行列の性能特性に応じた状態変化に従ってこれらの待ち行列にサービスし、
それぞれの待ち行列のパケットをルーティングする遅延特性と分離特性が保存される
ことを特徴とする通信ネットワークにおけるパケットの配送方法。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】
本発明は、パケット通信システムに関し、特に、このパケット通信システムのルータと交換機内で実行される公平待ち行列システム（fair-queuing system ）に関する。
【０００２】
【従来の技術】
沢山の研究が、理想的な流れ（fluid flow）モデル（即ち複数のソースから送信されるデータパケットは、無限に分解できるものとし、かつ複数のソースは、そのデータを例えば１本の物理的通信リンクに同時に送信できるようなモデル）をできるだけエミュレート（模擬）するパケット通信ネットワーク用の待ち行列システムの開発に費やされている。パケットを無限に分割できることは、実際にはありえないことである。
【０００３】
パケットネットワークにおいては、通常パケットがリンクを介して伝送されると、すべてのパケットを伝送しなければならない、即ちその間に別のパケットの伝送を阻止することはできない。パケットネットワーク内で、サービスの質（Quality of Service（ＱｏＳ））を保証するような要望があるために、データパケット交換器、あるいはルータ内でトラフィックスケジューリング方法の実現が必要とされている。
【０００４】
このスケジューリング方法の機能とは、交換器の各出力用リンクにとってその出力リンクを共有する通信期間に属するパケットの中から次のサイクルに送信すべきパケットを選択することである。このような選択は、個々のトラフィック期間に対し、保証されるサービ

スの質（ＱｏＳ）（例えば、最大遅延の上限）を満足するように行う必要がある。この方法の実行は、ハードウェアあるいはソフトウェアで行われるが、速度のことを考慮するとスケジューリングは通常ＡＴＭ交換器および高速ルータ内のハードウェア内で実行される。

【０００５】

多くの様々な種類のスケジューリング方法が、汎用プロセッサシェアリング（Generalized Processor Sharing（ＧＰＳ））として知られる理論的スケジューリング原理を近似すべく提案されている。このスケジューリング原理は、「流体」モデルを用いて定義される原理である。このようなＧＰＳによりリンク上で通信できる各期間に割り当てられたバンド幅をきっちりと制御できる。しかし、ある期間に伝送されるパケットは、これ以上分割することができないので複数のソースからのデータは、パケットの間（境界）にインターリーブ（挟み込む）しなければならない。このためＧＰＳ原理は、実際のパケット交換ネットワークでは、実行することができない。

【０００６】

単純なＦＩＦＯ技術，ラウンドロビン技術，公平待ち行列技術により、個々の待ち行列をサービスすることは公知である。しかし、この流体システムを最もよく近似する重み付けされた公平待ち行列（Weighted fair-queuing（ＷＦＱ））系が開発されている。特に、A. Demers, S. Keshav, S. Shenker著の "Analysis and Simulation of a Fair Queuing Algorithm" Internetworking: Research and Experience, pp, 3-26, vol. 1, 1990 の文献によれば、流体フローＧＰＳシステムをシミュレートし、このＧＰＳシステム内の出発の順番に対し、パケットのスケジューリング決定にバイアスを掛けることによりＧＰＳをエミュレートする公平待ち行列系が記載されている。

【０００７】

重み付け公平待ち行列においては、スケジューリング方法により制御される出力リンクを共有する各トラフィック期間ｉに、その期間の予約バンド幅に対応するϕ_iが割り当てられている。この値ϕ_iは、リンク上の期間ｉの予約バンド幅が次式で与えられるように計算される。

【数１】

$$\frac{\phi_i}{\sum_{j=1}^{\nu}\phi_j}$$

ここで、分母はリンクを共有する全てのυ期間のϕ_iの値の和を計算している。

【０００８】

図１に従来技術に係る重み付け公平待ち行列ＷＦＱシステム１０を示す。この重み付け公平待ち行列ＷＦＱシステム１０は、複数の結線用（結線毎に用意される）行列２０ａ，...，２０ｉを有し、各待ち行列はデータ端末のようなソースからの入力トラフィックである情報パケットを、共有メモリ２５内の別々の場所にパケットを一時的に記憶する。様々な種類のトラフィック、例えばオーディオ，ビデオ，データ等を収納する様々な種類の待ち行列が存在する。

【０００９】

さらにまた、重み付け公平待ち行列ＷＦＱシステム１０は、この待ち行列からパケットを割り当てられたレートと全く同一のレートでもって重み付け公平待ち行列スケジューラであるレート比例型サーバー４０に転送するシェーパー３０ａ，...，３０ｉを具備する。この重み付け公平待ち行列スケジューラは、重みはそれぞれ各結線用待ち行列２０ａ，...，２０ｉに関連づけられ、その結果スケジューラによりこれらの待ち行列の１つに提供されるサービスは、待機中のパケットがある場合には、常に重みに比例することになる。

【００１０】

例えば、リンクの容量（バンド幅）をＣ＝１０パケット／秒と仮定する。同時にスケジュ

ーラも３つの待ち行列にサービスするものとし、待ち行列Ｑ１は重みＷＱ１＝２０％、待ち行列Ｑ２は重みＷＱ２＝３０％、待ち行列Ｑ３は重みＷＱ３＝５０％とする。すると、この３個の全ての待ち行列が待機パケットを有している場合には、Ｑ１は２パケット／秒の保証されたバンド幅を、Ｑ２は３パケット／秒の保証されたバンド幅を、Ｑ３は５パケット／秒の保証されたバンド幅を受領する。

【００１１】

しかし、例えばＱ３が待機中のパケットを持っていない場合には、過剰（余分）のバンド幅は５パケット／秒に等しくなる。ＷＦＱシステムにおいては、この過剰バンド幅は、待機中のパケットを有する待ち行列の重みに比例して再分配される。上記の実施例では、Ｑ３は待機パケットを有していないので過剰のバンド幅はＱ１とＱ２に比例して分配され、その結果Ｑ１は４パケット／秒、Ｑ２は６パケット／秒の瞬間的なバンド幅を受領することになる。それぞれのシェーパー３０から出ていった各パケットは、直接、レート比例型サーバー４０に転送され、このレート比例型サーバー４０はこのパケットをさらに出力リンク５０に転送する。このレート比例型サーバー４０は重み付け公平待ち行列の変形例である。

【００１２】

このようなＷＦＱ系においては、端末間の遅延が保証されるような利点がある。例えば、各パケットはストリーム内の各パケットフローに対しあるレートが保証され、ストリーム間の隔離が確保され、例えば誤動作をするソースは他のストリームのフローには影響を及ぼさない。さらにまた、容量の利用が十分でない場合には、例えばフローが特にバースト型であり、送信しない空き時間が存在する場合には、このＷＦＱシステムは作業保存特性（work-conservation property）を保存するために未使用のバンド幅の再分配を容易にする。

【００１３】

現在複数の待ち行列の間で未使用のバンド幅の容量を再分配する特性は、流体フローモデル（fluid-flow model）に特有な方法、例えば特定の待ち行列の重みに従って実行される。かくしてパケットの待ち行列がないアイドル状態の時には、過剰の（余分）のバンド幅は長期の要件に基づくその重みに比例して予備接続（backlogged connections）に再分配される。

【００１４】

全ての公平待ち行列システムがＧＰＳの近似を受け継ぐが、このＧＰＳの欠点は、ＧＰＳが状態依存性バンド幅の共有を厳しく制限してしまうことである。ＧＰＳにおける状態の依存性のみが予備接続の数になる。従ってこれ以上の自由度は存在せず、接続の長期間のニーズに基づいて設定された保証レートによってのみ決定される。

【００１５】

バンド幅の共有に対するこの制限は、公平待ち行列の特性を維持するため、およびリーキーバケット制御のトラフィックソースに対する最悪の遅延限界を保証するためには、必要以上に厳しいものである。したがって、ＧＰＳの余剰バンド幅共有を次善策としてエミュレートするような公平待ち行列システムに対するニーズは存在しない。

【００１６】

【発明が解決しようとする課題】

したがって本発明の目的は、ＧＰＳをエミュレートする重み付け公平待ち行列システムにおいて、残りのトラフィックフローの瞬時のニーズを反映するような状態に依存する方法で未使用のバンド幅の再分配を効率よく達成する方法を提供することである。

【００１７】

【課題を解決するための手段】

本発明は、適応型再分配系を実行する重み付け公平待ち行列への改良型アプローチである。このような再分配系においては、接続毎の各フローは、適応しながら再分配される余剰のバンド幅とリンクバンド幅の特定の共有が保証される。この再分配系により、最悪の場合の端末間の遅延限界を与えるような公平な待ち行列機能を保存でき、そしてこの再分配

系は、余剰のバンド幅が存在しない場合でも公平待ち行列のように働く。この余剰のバンド幅は個々の条件に従って再分配される。

【００１８】

状態依存性の条件の例は、１）現在最も長い遅延を有するフローにサービスする最長遅延優先条件（Longest delay first（ＬＤＦ））と、 ２）許容された最大遅延と現在の遅延との差が最小のフローにサービスする対オーバフロー最短時間条件（Least time to overflow（ＬＴＯ））と、３）最悪のケースの到着が発生したときに、バッファが最初にオーバフローを引き起こすようなフローにサービスするリーキーバケットのオーバフローへの最短時間条件（Least time to overflow with leaky buckets（ＬＴＯ－ＬＢ））である。

【００１９】

この適応型バンド幅再分配系は、常に各接続に必要とされる最小限の保証を与え、最悪の場合の公平さを満足させている。

【００２０】

このＬＤＦ方式は、余剰のバンド幅を用いて遅延分配の変動を低減し、これにより音声ソース，ビデオソースに対しプレイアウト（使い切った）バッファサイズを縮小できるメリットがある。ビデオトレースとボイストラフィックのシミュレーションによれば、この方式は最悪の場合の保証を犠牲にすることなくＧＰＳよりは、性能がよい。

【００２１】

許容最大遅延からの変動を考慮にいれてないので、遅延制限が小さいフロー（例、音声）は、遅延制限が大きいフローの存在の元では、余剰のバンド幅をほとんど得ることができない。割当重みの不正確さは、大きな遅延制限のフローよりは、これらの小さな遅延制限のフローの方がはるかに大きな損失を受けることになる。

【００２２】

ＬＴＯ方式は、最も早くオーバフローする可能性のあるフローが、最も瞬時のバンド幅を必要とするという仮定の元で、余剰バンド幅を割り当てることによりパケットの喪失を最小にしようとしてる。このようにする際に、許容最大遅延からの各フローの現行の変動を考慮にいれている。ＣＢＲと音声ソースとビデオソースを異なる遅延限界で組み合わせてシミュレーションすると、この方式は全てのクラスに対しパケットの喪失を低減できかつ各クラスの遅延の変動を低減することができる。

【００２３】

【発明の実施の形態】

図２は適応型余剰バンド幅再分配を用いた公平待ち行列方法を表すブロック図。本発明のシステムは、図１に示した従来のシステムに加えて状態依存性スケジューラ２００を有し、この状態依存性スケジューラ２００へは各結線用待ち行列２０ａ，...，２０ｉからパケットが入力され、状態依存性スケジューラ２００の出力はマルチプレクサ３５０で出力リンク５０と多重化される。パケットは一時的に結線用待ち行列２０ａ，...，２０ｉ内に記憶される。重みがこれらの各待ち行列に割り当てられ、この重みは、各待ち行列に到着したパケットに割り当てられなければならない出力バンド幅の一部を表す。

【００２４】

シェーパー３０ａ，...，３０ｉは、結線用待ち行列２０ａ，...，２０ｉからのパケットを割り当てられたレートと等しいレートでレート比例型サーバー４０に転送する。このような構成により、余剰のバンド幅は、再分配用に利用できることになる。本発明によれば、他のソース即ち他の待ち行列からのトラフィックは、適応型バンド幅再分配メカニズムである状態依存性スケジューラ２００にも入力される。結線毎の待ち行列，シェーパー，スケジューラ等のハードウェア構成は、当業者には公知である。

【００２５】

上述したようにシェーパー３０ａ，...，３０ｉは、パケットを割り当てられたレートに等しいレートで、パケットを状態依存性スケジューラであるレート比例型サーバー４０に送る。シェーパー３０ａ，...，３０ｉよりそれぞれの待ち行列の接続「ｉ」に時間間隔

の間提供されるサービスは、S_i（τ，ｔ） で表す。パケットはシェーパー３０から無限の容量を有するサーバー４０に転送されるものとする。レート比例型サーバー４０により提供されるサービスは、R_i（τ，ｔ） で表す。

【００２６】

サービスを受ける資格のなくなったパケットは、シェーパー内の対応する接続待ち行列内に保存され、一方全ての資格のあるパケットは、レート比例型サーバー４０内でサービスを受けるために待機する。サービスはパケットがそこにある限りレート比例型サーバー４０から与えられる。状態依存性スケジューラ２００の待ち行列の全てが空の時には、状態依存性スケジューラ２００は結線用待ち行列２０ａ，...，２０ｉからパケットを選択するよう起動され、シェーパー３０の状態に影響を及ぼさない接続に提供されるサービスでもって伝送する。図２に示すように、シェーパー３０によりタイムインターバルの間待ち行列接続「ｉ」へ提供されるサービスは、D_i（τ，ｔ） で示し、状態依存性スケジューラ２００により提供されるサービスは、Z_i（τ，ｔ） で示す。

【００２７】

シェーパー３０は当業者に公知のカレンダー（暦）待ち行列を用いて実現できる。シェーピング用のカレンダー待ち行列のメカニズムは、D. Stiliadis と A. Verma 著の A General Methodology for Designing Scheduling and Shaping Algorithms, in Proceedings of IEEE INFOCOM '97 を参照のこと。

【００２８】

重み付き公平待ち行列サーバーであるレート比例型サーバー４０は、公知の公平待ち行列メカニズムを用いて実現できる、これに関しては、D. Stiliadis と A. Varma の Traffic Scheduling System and Method for Packet-Switched Networks, 米国特許出願第０８／６３４，９０４号（１９９６年４月１５日提出）を参照のこと。

【００２９】

ある時点でＲＰＳスケジューラ２００内にパケットが存在しないと決定された場合には、これはフリーの（だれでもが使用可能な）バンド幅があることをを意味する。このためパケットは、状態依存の方法に基づいて状態依存性スケジューラ２００によりサービスされる。この状態依存の方法により、決定はシステムに関連するある変数の現在の状態に基づいて行われる。状態依存の決定の２つの例は、最長遅延優先（Longest Delay First） とオーバフローへの最短時間（Least Time to Overflow）である。

【００３０】

最長遅延優先方式においては、状態依存性スケジューラ２００は結線用待ち行列２０ａ，...，２０ｉの中からこの保証されたのと等しいレートでもってこれらの待ち行列がサービスされた場合には、遅延が最長となるパケットを有する待ち行列を選択する。これを実行するために、スケジューラは各待ち行列に保証レートにより除算された待ち行列のサイズに等しい遅延値を割り当てる。この値は、この待ち行列の最後のパケットが受ける遅延を表す。状態依存性スケジューラ２００は、一組の数の中から最大値を選択するために何らかのメカニズムを用いて、最大値を有する待ち行列を選択する。この選択メカニズムは公知である。

【００３１】

オーバフローへの最短メカニズムにおいては、状態依存性スケジューラ２００は結線用待ち行列２０ａ，...，２０ｉの中から最短時間でオーバフローする可能性のある待ち行列を選択する。このメカニズムは、最大サイズは待ち行列に関係していると見なしている。パケットが到着したり、あるいは待ち行列によりサービスされる毎にオーバフローするパケット（packet-to-overflow）の変数は、最大容量を超えることなく待ち行列に付加されるパケットの数をカウントする。

【００３２】

第２の変数（オーバフローする時間（time-to-overflow）と称する）は、オーバフローする予測時間を表し、関連レートでオーバフローするパケット変数（packets-to-overflow） を除算することにより計算される。状態依存性スケジューラ２００は、最小のオーバ

フローする時間変数（time-to-overflow）を有する待ち行列を伝送用に選択する。この最小値は、公知のメカニズムにより決定できる。

【００３３】

上記の変数は、状態依存性の変数を計算する２種類の方法について説明したが、他の類似の方法も用いることができる。状態依存性の変数に基づいて、状態依存性スケジューラ２００は、選択された待ち行列からパケットをリンク７５を介して、マルチプレクサ３５０により多重化されて次の宛先に転送される。

【００３４】

最悪の場合の公平性の特徴は、本発明のシステムで満足されることが分かる。最悪の場合の公平性においては、異なる割当レートで異なる長さの待ち行列をサービスすることは、第１の待ち行列からのパケットが時間t_1でサービスされ、第２の待ち行列のパケットが時間t_2でサービスされるように、インタリーブされる（間に入れられる）。

【００３５】

本発明においては、最悪の場合の公平性は満足され、その結果、最悪の場合の時間間隔$t_2 - t_1$は、接続の数の関数ではない値により制限されるか、あるいはそれよりも少なく、しかし時間ｔにおいてサービスされる待ち行列のような最長の待ち行列のパケットサイズと、最長の待ち行列の割当レートの関数である。

【００３６】

最悪の場合の公平性の特徴は、本発明の重み付け公平待ち行列システム内の状態依存性スケジューラ２００により満足される。

【図面の簡単な説明】

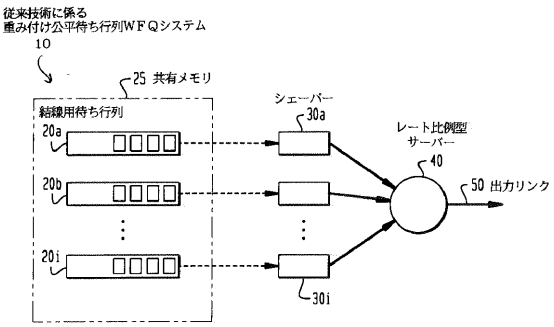
【図１】重み付け公平待ち行列とレート比例型サーバーの特徴を表すデータフロー図

【図２】本発明による重み付け公平待ち行列スケジューリングシステムを表す図

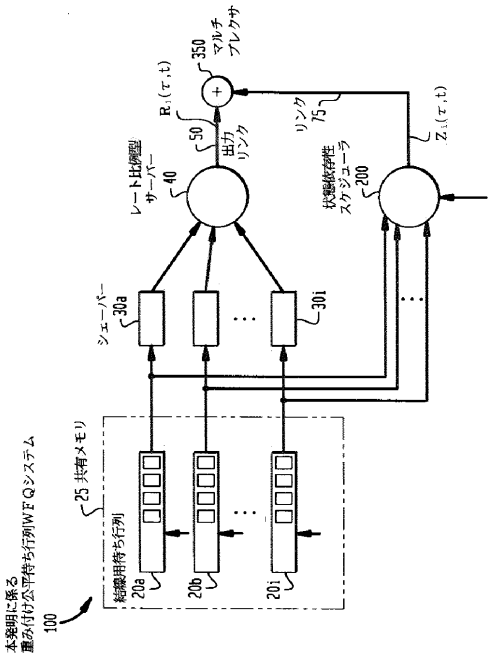
【符号の説明】

１０　従来技術に係る重み付け公平待ち行列ＷＦＱシステム
２０　結線用待ち行列
２５　共有メモリ
３０　シェーパー
４０　レート比例型サーバー
５０　出力リンク
７５　リンク
１００　本発明に係る重み付け公平待ち行列ＷＦＱシステム
２００　状態依存性スケジューラ
３５０　マルチプレクサ

【図１】


従来技術に係る
重み付け公平待ち行列ＷＦＱシステム
10
25 共有メモリ
結線用待ち行列
20a
20b
20i
シェーバー
30a
30i
レート比例型
サーバー
40
50 出力リンク

【図２】


本発明に係る
重み付け公平待ち行列ＷＦＱシステム
100
25 共有メモリ
結線用待ち行列
20a
20b
20i
シェーバー
30a
30i
レート比例型
サーバー
40
出力
リンク
50
Ri(τ,t)
350
マルチ
プレクサ
リンク
75
状態依存性
スケジューラ
200
Zi(τ,t)

フロントページの続き

(72)発明者　ニコラス　ジー．ダフォールド
アメリカ合衆国、０７０３０　ニュージャージー、ホボーケン、＃５０、パークアベニュー　９３３
(72)発明者　ティ、ブイ、ラクシマン
アメリカ合衆国、０７７２４　ニュージャージー、イートンタウン、ヴィクトリアドライブ　１８８
(72)発明者　ディミトリオス　ステリアリス
アメリカ合衆国、０７０３０　ニュージャージー、ホボーケン、パークアベニュー　１０６

合議体
審判長　山本　春樹
審判官　野元　久道
審判官　衣鳩　文彦

(56)参考文献　特開平１０－５１４７２（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.[7]，ＤＢ名)
H04L12/56