(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第3556495号
(P3556495)
(45)発行日　平成16年8月18日(2004.8.18)　(24)登録日　平成16年5月21日(2004.5.21)

(51)Int.Cl.[7]	F I
H04L 12/56	H04L 11/20 102Z
	H04L 11/20 102E

請求項の数 20　(全 28 頁)

(21)出願番号	特願平10-356012
(22)出願日	平成10年12月15日(1998.12.15)
(65)公開番号	特開2000-183965(P2000-183965A)
(43)公開日	平成12年6月30日(2000.6.30)
審査請求日	平成14年3月19日(2002.3.19)

(73)特許権者　000003078
株式会社東芝
東京都港区芝浦一丁目１番１号
(74)代理人　100058479
弁理士　鈴江　武彦
(74)代理人　100084618
弁理士　村松　貞男
(74)代理人　100068814
弁理士　坪井　淳
(74)代理人　100092196
弁理士　橋本　良郎
(74)代理人　100091351
弁理士　河野　哲
(74)代理人　100088683
弁理士　中村　誠
最終頁に続く

(54)【発明の名称】パケットスイッチ及びパケット交換方法

(57)【特許請求の範囲】
【請求項１】
入側転送手段からスイッチング手段を経由して所望の出側転送手段へ、パケットを転送するパケットスイッチであって、
所定の転送先毎の輻輳状況を観測するための輻輳状況観測手段と、
前記パケットに、前記輻輳状況観測手段により観測された該パケットの転送先の輻輳状況に基づいて、優先度を付与するための優先度付与手段と、
前記スイッチング手段内部で前記パケットの衝突が発生した場合に、パケットに付与された優先度に基づいて、優先して転送すべきパケットを選択するためのパケット選択手段とを備えたことを特徴とするパケットスイッチ。
【請求項２】
前記優先度付与手段は、パケットの転送先の輻輳の度合いがより大きいほど、より低い優先度を付与することを特徴とする請求項１に記載のパケットスイッチ。
【請求項３】
前記優先度付与手段は、各入側転送手段内に設けられたものであることを特徴とする請求項１または２に記載のパケットスイッチ。
【請求項４】
前記優先度付与手段があるパケットに優先度を付与するために参照する転送先の輻輳状況が未知または無効となっている場合に、該転送先に転送する最初の１つのパケットまたは複数のパケット群に付与する優先度を一時的に高く設定することを特徴とする請求項１な

いし３のいずれか１項に記載のパケットスイッチ。
【請求項５】
前記優先度付与手段は、各入側転送手段毎に１つずつまたは複数の入側転送手段で１つ設けられた、前記輻輳状況観測手段により観測された輻輳状況に基づいて所定の転送先毎に設定される輻輳度を記憶する輻輳度テーブルを参照して、前記パケットに付与すべき優先度を設定することを特徴とする請求項１ないし４のいずれか１項に記載のパケットスイッチ。
【請求項６】
前記輻輳状況観測手段は、各出側転送手段内に設けられたものであることを特徴とする請求項１ないし５のいずれか１項に記載のパケットスイッチ。
【請求項７】
前記輻輳状況観測手段は、前記転送先毎の輻輳状況として、対応する出側転送手段毎、対応する出側転送手段のクラス毎、対応する出側転送手段のポート毎、対応する出側転送手段の各ポートのクラス毎、または対応する出側転送手段の各ポートの各クラスのフロー毎に、前記輻輳状況を観測する手段を含むことを特徴とする請求項６に記載のパケットスイッチ。
【請求項８】
前記輻輳状況観測手段により観測された前記輻輳状況を、前記優先度付与手段に反映させるための輻輳状況通知手段を更に備えたことを特徴とする請求項１ないし７のいずれか１項に記載のパケットスイッチ。
【請求項９】
前記優先度付与手段は、各入側転送手段内に設けられ、前記輻輳状況通知手段は、前記出側転送手段にパケットが到着したことを契機に、該パケットを送出した入側転送手段に対して所定の輻輳状況に関する情報を通知することを特徴とする請求項８に記載のパケットスイッチ。
【請求項１０】
前記入側転送手段毎に、各入側転送手段において複数のパケットが前記スイッチング手段への転送を待っている場合に、該複数のパケットの転送順を制御するためのスケジューリング手段を更に備え、
前記スケジューリング手段は、各パケットの転送先の輻輳状況を考慮して、輻輳していない転送先行きのパケットが優先的に前記スイッチング手段に転送されるように制御することを特徴とする請求項１ないし９のいずれか１項に記載のパケットスイッチ。
【請求項１１】
優先度の付与されたパケットを入側転送手段からスイッチング手段を経由して所望の出側転送手段へ転送し、該スイッチング手段は内部でパケットの衝突が発生した場合に該優先度を考慮して選択したパケットを出側転送手段に対して転送し、その他のパケットを該スイッチング手段内部で廃棄するパケットスイッチであって、
前記入側転送手段は、
前記衝突によるパケット廃棄が検出された場合に、該廃棄されたパケットを再送するための手段と、
前記再送パケットに付与すべき優先度を、もとの廃棄されたパケットに付与した優先度よりも高く設定するための手段とを含むことを特徴とするパケットスイッチ。
【請求項１２】
優先度の付与されたパケットを入側転送手段からスイッチング手段を経由して所望の出側転送手段へ転送し、該スイッチング手段は内部でパケットの衝突が発生した場合に、該優先度を考慮して選択したパケットを出側転送手段に対して優先的に転送するパケットスイッチであって、
前記入側転送手段は、１つのデータが複数のパケットに分割して搭載された際の各パケットを転送する場合に、分割されたデータの後続部分に対応するパケットの優先度を、先頭部分に対応するパケットよりも高く設定するための手段を含むことを特徴とするパケット

スイッチ。
【請求項１３】
前記分割されたデータの先頭部分に対応するパケットの優先度を、分割されたデータを搭載したものではないパケットならば付与するであろう優先度と同等の値に設定することを特徴とする請求項１２に記載のパケットスイッチ。
【請求項１４】
パケットを入側転送手段からスイッチング手段を経由して所望の出側転送手段へ転送するパケットスイッチであって、
前記スイッチング手段を転送されるパケットとともに該パケットを送出した入側転送手段内部の輻輳状況を出側転送手段へ転送するための手段と、前記パケットとともに通知された前記輻輳状況と、該パケットが転送された出側転送手段内部の輻輳状況とを用いて、総合的な輻輳状況を求めるための手段と、
求められた前記総合的な輻輳状況を、通過するパケットフローの量または速度を制御するためにネットワークにおいて行われる輻輳制御に用いるための手段とを備えたことを特徴とするパケットスイッチ。
【請求項１５】
入側転送手段から、パケット衝突の発生しないスイッチング手段を経由して所望の出側転送手段へ、パケットを転送するパケットスイッチであって、
転送先毎の輻輳状況を観測するための輻輳状況観測手段と、
各入側転送手段から送出されるパケットに前記輻輳状況観測手段により観測された該パケットの転送先の輻輳状況に基づいて優先度を付与し、各パケットをパケット衝突の発生するトポロジを有する仮想的なスイッチング網内を転送させたと仮定して、前記パケットの衝突が発生した場合にパケットに付与された優先度に基づき優先して転送すべきパケットを選択するシミュレーションを行い、このシミュレーションにより前記入側転送手段から前記出側転送手段までパケットが到達した結果と同等な結果となるように、前記パケット衝突の発生しないスイッチング手段の接続パターンを決定する接続パターン決定手段とを備えたことを特徴とするパケットスイッチ。
【請求項１６】
前記パケット衝突の発生しないスイッチング手段は、クロスバスイッチであり、前記仮想的なスイッチング網は、単位スイッチにより構成されたスイッチ網であることを特徴とする請求項１５に記載のパケットスイッチ。
【請求項１７】
入側転送手段からスイッチング手段を経由して所望の出側転送手段へ、パケットを転送するパケットスイッチのパケット交換方法であって、
前記入側転送手段は、前記パケットにその転送先の輻輳状況に基づいた優先度を付与してこれを前記スイッチング手段へ送出し、
前記スイッチング手段は、前記入側転送手段から転送された前記パケットをその転送先に従って交換するとともに、その内部でパケットの衝突が発生した場合には各パケットに付与された優先度を考慮して選択したパケットを優先して交換し、
前記パケットの到達した前記出側転送手段は、所定の観測単位についての輻輳状況の観測結果を示す情報を、該パケットを送出した前記入側転送手段に通知することを特徴とするパケット交換方法。
【請求項１８】
優先度の付与されたパケットを入側転送手段からスイッチング手段を経由して所望の出側転送手段へ転送し、該スイッチング手段は内部でパケットの衝突が発生した場合に該優先度を考慮して選択したパケットを出側転送手段に対して転送し、その他のパケットを該スイッチング手段内部で廃棄するパケットスイッチのパケット交換方法であって、
前記入側転送手段は、前記衝突によるパケット廃棄が検出された場合に、該廃棄されたパケットに、もとの廃棄されたパケットに付与した優先度よりも高い優先度を付与して、これを再送することを特徴とするパケット交換方法。

【請求項１９】
優先度の付与されたパケットを入側転送手段からスイッチング手段を経由して所望の出側転送手段へ転送し、該スイッチング手段は内部でパケットの衝突が発生した場合に、該優先度を考慮して選択したパケットを出側転送手段に対して優先的に転送するパケットスイッチのパケット交換方法であって、
前記入側転送手段は、１つのデータが複数のパケットに分割して搭載された際の各パケットを転送する場合に、分割されたデータの先頭部分に対応するパケットに所定の優先度を付与して送出し、該パケットが転送先に到達したならば、該データの後続部分に対応するパケットに該所定の優先度より高い優先度を付与してを送出することを特徴とするパケット交換方法。
【請求項２０】
パケットを入側転送手段からスイッチング手段を経由して所望の出側転送手段へ転送するパケットスイッチのパケット交換方法であって、
前記入側転送手段は、その内部の輻輳状況を示す情報をパケットに付加して送出し、
前記出側転送手段は、その内部の輻輳状況と、前記パケットに付加されて通知された前記輻輳状況とを用いて、総合的な輻輳状況を求め、
求められた前記総合的な輻輳状況を、所定の輻輳制御に用いることを特徴とするパケット交換方法。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】
本発明は、複数のポート間でパケットを交換するパケットスイッチ及びパケット交換方法に関する。
【０００２】
【従来の技術】
近年、目覚しくデータ通信の分野が発展している。伝統的には通信網の代表といえば電話網のことであったが、企業および家庭へのパソコンの普及によりインターネットを代表とするデータ通信網の重要性が急速に高まっている。最近ではデータ通信網の低コスト性や高効率性を生かして、データ通信網上で電話サービスを実現する技術も確立されつつあり、データ通信網に電話網を吸収すべきなどという議論までなされている。
【０００３】
さて、このデータ通信網のほとんどは、送信したいデータに宛先を示す情報を付与したもの、つまりパケット（ＡＴＭにおけるセルの意味も含む）の通信を行なうパケット交換網（ＡＴＭ交換網の意味も含む）である。データ通信のトラフィック量の増大とともに、このパケットを交換するパケットスイッチの大規模化に対する要求が高まっている。
【０００４】
パケットスイッチの構成方法は数多く提案されているが、大きくは２種類に分類できる。一つはやや古いタイプのものでポート数を大きくできる多段スイッチ型、もう一つは比較的新しいタイプのもので、ポート数が少ない単段スイッチ型である。
【０００５】
後者の単段スイッチ型のパケットスイッチは、輻輳している出力ポートがあっても輻輳していない出力ポート行きのパケットの流れがそれに妨げられないようにする輻輳制御機構が備わっていることが多い。これに対して前者の多段スイッチ型のパケットスイッチは、流れ込むパケットの流量が予め決まっていることを前提にしている。したがって、重度の輻輳を想定していないものが多かった。この点は、輻輳の発生には複数の入力ポートと複数の出力ポートとが互いに関係するために、大きいポート数まで拡張することが前提の多段スイッチ型では、輻輳に対処できる機構を組み込むことが困難であった、という事情によるところも大きいと思われる。
【０００６】
データ通信網の代表であるインターネットでは、各ユーザがどのような速度でパケットを

転送するかが予め決められているわけではないため、いつでもどこでも輻輳が発生し得る。そのため、パケットスイッチのアーキテクチャを考える際には、大規模化のための技術と同時に輻輳を扱う技術が重要といえる。輻輳を扱う技術には、パケットスイッチ内で輻輳が発生した場合に、ユーザに対してパケットの転送速度を下げてもらうように通知する技術と、輻輳が発生してもその輻輳に関係のないパケットの流れが悪くならないようにする技術とがある。

【０００７】
以下、このような従来の多段スイッチ型のパケットスイッチの概略について説明する。

【０００８】
多段スイッチ型パケットスイッチは、小さな単位スイッチを多段に組み合わせることで大規模なパケットスイッチを作ることが可能である（ＪｏｓｅｐｈＹ．Ｈｕｉ：“Ｓｗｉｔｃｈｉｎｇ ａｎｄ Ｔｒａｆｆｉｃ Ｔｈｅｏｒｙ ｆｏｒ Ｉｎｔｅｇｒａｔｅｄ Ｂｒｏａｄｂａｎｄ Ｎｅｔｗｏｒｋｓ”、ＩＳＢＮ ０－７９２３－９０６１－Ｘ、Ｋｌｕｗｅｒ Ａｃａｄｅｍｉｃ Ｐｕｂｌｉｓｈｅｒｓ、１９９０）。

【０００９】
図６に、多段スイッチ型パケットスイッチの一構成例を示す。図６は、３入力３出力の単位スイッチ１４５を組合せて、２７入力２７出力のパケットスイッチを構成した例である。

【００１０】
入力ポート１２１から入力されたパケットは、入側転送部１０２からスイッチング部１０４を経由して所望の出側転送部１０６へ転送され、そこから出力ポート１６１へ出力される。この多段スイッチ型パケットスイッチは、単位スイッチの数と段数を増やせば、さらに大きなパケットスイッチを構成することが可能であり、また、例えば単位スイッチをより大きな８入力８出力に置き換えれば、より少ない数の単位スイッチで大規模なパケットスイッチを構築できる。

【００１１】
多段スイッチ型パケットスイッチのよく知られている構成方法の一つが、ランダム網とルーティング網を従属接続する方法である。図６のランダム網１４１とルーティング網１４３は、互いに線対称の関係になっている結合網である。それぞれの網は、どの入力リンクからでも任意の出力リンクへ到達でき、かつ、その経路は一つしかない。ランダム網とルーティング網を従属接続することで、ある入側転送部１０２からある出側転送部１０６への経路が、ランダム網とルーティング網の境界にある単位スイッチの数（図６では９個）だけ存在することになる。

【００１２】
入側転送部１０２から出力されたパケットは、ランダム網１４１内を他のパケットと衝突しないように、ランダムに選ばれた空いているリンクに転送され、ルーティング網１４３の入口へ到達する。このランダムな転送は、ルーティング網１４３の入口（図６では９個）に、パケットを確率的に分散させる目的で行なわれる。パケットは、ルーティング網１４３でその宛先に基づいて転送され、出側転送部１０６に到達する。

【００１３】
ルーティング網１４３の内部においては、１本の単位スイッチの出力リンクに複数のパケットが同時に向かったためにパケットの衝突が発生する可能性がある。同じ出側転送部１０６行きのパケット同士であれば必ず衝突するし、異なる出側転送部１０６行きのパケット同士でも衝突する場合がある。各単位スイッチ１４５が待ち合わせバッファを持たないようなパケットスイッチでは、衝突すると唯一のパケットが転送され、残りのパケットは直ちに廃棄される。このような衝突によってパケットが廃棄された場合、該当する入側転送部１０２は廃棄されたパケットを再送する。

【００１４】
このように、スイッチング部１０４内でパケットが廃棄されても、そのパケットが出側転送部１０６へ無事に転送されるまで何度も入側転送部１０２が再送し続ける。したがって

、スイッチング部１０４の内部でパケットの情報が失われることはない。パケットスイッチの外部から見ればパケットが廃棄されるのは入側転送部１０２および出側転送部１０６の内部のみである。

【００１５】

このような多段スイッチ型パケットスイッチ内におけるパケット転送制御では、３種類の輻輳が発生し得る。これらの輻輳を一つずつ説明する。ここでは、説明を分かり易くするために、スイッチング部１０４として２入力２出力の単位スイッチを３段に組み合わせた４入力４出力のものを用いる（図７～図９）。この小規模のスイッチング部の場合でも、１段目の単位スイッチ１４５から２段目の単位スイッチの入力まではランダム網１４１、２段目の単位スイッチ１４５の出力から３段目の単位スイッチ１４５はルーティング網１４３、という構成は図６と同様であり、ここで行う３種類の輻輳についての議論もしくは説明が図６やそれ以上の大規模なスイッチング部に対しても同様に成立する（適用できる）。

【００１６】

まず、図７を参照しながら、輻輳の第１番目の例について説明する。

【００１７】

図７の例では、入側転送部［Ａ］に出側転送部［Ｃ］行きのパケットが存在し、入側転送部［Ｃ］に出側転送部［Ｄ］に行きのパケットが存在するものとする。この場合、理想的なパケットスイッチであれば、この二つのパケットは宛先が異なるため同時に転送できるはずである。ところが、実際には、ランダム網１４１が二つのパケットを例えばスイッチング部１０４の２段目で（偶然に）下側の単位スイッチ１４５に転送するような状況があり得る。そのような状況が発生した場合、それら二つのパケットは同時に宛先へ到着することはできない。逆にもしそれら二つのパケットが２段目の上側と下側の別々の単位スイッチに分かれて転送されていたなら、この輻輳は回避されていたはずである。このような輻輳を「ルーティング網内衝突」と呼ぶことにする。

【００１８】

このルーティング網内衝突による転送性能の劣化を解決するための方法は古くからいくつかのものが提案されていた。例えば、ルーティング網の前段にランダム網を従属接続しルーティング網内衝突が特定のパケットの集合に偏らないようにパケットを分散させる方法や、この輻輳によってパケットの再送が発生してもパケットスイッチの外部からみたスループットが低下しないようにスイッチング部のパケット転送速度を入力ポートおよび出力ポートの速度よりも高速化する方法が知られている。このようにスイッチアーキテクチャを工夫して継続的な輻輳が発生しないようにすれば、スイッチの外部からこの輻輳の影響を見えないようにすることができる。

【００１９】

次に、図８を参照しながら輻輳の第２番目の例について説明する。

【００２０】

図８の例では、二つの入側転送部［Ａ］，［Ｃ］に同じ出側転送部［Ｃ］行きのパケットが存在する。この二つのパケットはどのような経路でランダム網１４５の内部を転送されても出側転送部［Ｃ］に到着するまでに必ず衝突してしまう。このような輻輳を「出側転送部入口輻輳」と呼ぶことにする。

【００２１】

この輻輳は、基本的には同じ行き先を持つパケットが継続的に集中することで発生する。例えば、図８において、さらに入側転送部［Ｂ］から出側転送部［Ｄ］行きのパケットの流れがあったとすると、出側転送部［Ｃ］行きの過剰なパケット転送が継続する影響で出側転送部［Ｄ］行きのパケットの流れが悪くなってしまう。

【００２２】

従来の多段スイッチ型のようなパケットスイッチは出側転送部の数が多いため、この輻輳に対する適切な対処方法は未解決の課題であった。なぜなら出側転送部の数が多いと、どの出側転送部が混んで、どの出側転送部がすいているかの情報を一度に収集することが困

難だったからである。
【００２３】
次に、図９を参照しながら輻輳の第３番目の例について説明する。
【００２４】
この輻輳は、出側転送部（図９の例ではＣ）の出力ポートの速度が、そのポート行きのパケットの到着速度よりも遅いことが原因で発生する。この輻輳を「出側転送部内輻輳」と呼ぶことにする。図９の例においては、出側転送部［Ｃ］が輻輳している場合、本来、入側転送部［Ａ］は、出側転送部［Ｃ］行きのパケットを必要以上に頻繁に転送する必要はなく、出側転送部［Ｃ］の出力ポートの速度程度で十分である。
【００２５】
従来のパケットスイッチは過剰に多くのパケットを同一の出力ポートへ転送してしまうことにより、先に述べた第１の輻輳（「ルーティング網内衝突」）や第２の輻輳（「出側転送部入口輻輳」）を誘発する原因になっていた。
【００２６】
以上、３種類の輻輳についてそれぞれ説明した。このように、従来の多段スイッチ型パケットスイッチは、単位スイッチを組み合わせることで大規模化することができるという利点を持っていたが、重度の輻輳に対するよい対処方法がなかった。つまり、スイッチング部内に大量の輻輳ポート行きパケットが転送されるおそれがあった。これらはスイッチング部内で廃棄される可能性が高い無駄な転送である。本来は出力ポートの転送速度を越える転送は無意味であり、それを越えて大量に転送されたパケットはスイッチング部内で廃棄され再送を繰り返すという事態が発生し得る。このような事態になると、輻輳していないポート行きのパケットも、輻輳しているポート行きのパケットとスイッチング部内で衝突して廃棄と再送を繰り返し流れが悪くなる。結果的にはパケットスイッチ全体のパケット転送効率が著しく低下するという欠点があった。
【００２７】
【発明が解決しようとする課題】
以上述べたように、従来の多段スイッチ型のパケットスイッチでは、スイッチング部内に大量の輻輳ポート行きパケットが転送されるおそれがあった。これらはスイッチング部内で廃棄される可能性が高い無駄な転送である。本来は出力ポートの転送速度を越える転送は無意味であり、それを越えて大量に転送されたパケットはスイッチング部内で廃棄され再送を繰り返すという事態が発生し得る。このような事態になると、輻輳していないポート行きのパケットも、輻輳しているポート行きのパケットとスイッチング部内で衝突して廃棄と再送を繰り返し流れが悪くなる。結果的にはパケットスイッチ全体のパケット転送効率が著しく低下するという問題点があった。したがって、例えば、トラフィックが急増するインターネット等に適用するためには、ポート数を大きくできる多段スイッチ型のパケットスイッチにおける輻輳制御が要望される。
【００２８】
本発明は、上記事情を考慮してなされたもので、輻輳による影響を回避し転送等の性能をより最大限に引き出すことのできるパケットスイッチ及びパケット交換方法を提供することを目的とする。
【００２９】
本発明は、輻輳しているポート行きのパケットの影響により、輻輳していないポート行きのパケットの流れが妨げられないようにしたパケットスイッチ及びパケット交換方法を提供することを目的とする。
【００３０】
また、本発明は、廃棄されたパケットを再送する場合にそれを廃棄されにくくさせることができるようにしたパケットスイッチ及びパケット交換方法を提供することを目的とする。
【００３１】
また、本発明は、複数のパケットに分割した１つデータを一塊りに転送させることができ

るようにしたパケットスイッチ及びパケット交換方法を提供することを目的とする。
【００３２】
【課題を解決するための手段】
本発明（請求項１）は、入側転送手段からスイッチング手段を経由して所望の出側転送手段へ、パケット（例えば、パケットデータ、またはパケットデータの転送に先立つリクエスト）を転送するパケットスイッチであって、所定の転送先毎の輻輳状況を観測するための輻輳状況観測手段と、前記パケットに、前記輻輳状況観測手段により観測された該パケットの転送先の輻輳状況に基づいて、優先度を付与するための優先度付与手段と、前記スイッチング手段内部で前記パケットの衝突が発生した場合に、パケットに付与された優先度に基づいて、優先して転送すべきパケットを選択するためのパケット選択手段とを備えたことを特徴とする。
【００３３】
本発明によれば、パケットにその転送先の輻輳の程度に応じて優先度を付与することができ、例えば輻輳していないポート行きのパケットには相対的に高い優先度を付ける、といった制御が可能となる。そして、スイッチング手段内でパケットの衝突が発生しても、優先度を考慮したパケット選択を行うので、例えば輻輳していないポート行きのパケットの流れは相対的に高い優先度が付与されていれば妨げられることがなく、結果としてパケットスイッチの転送効率の低下を回避することができる。
【００３４】
スイッチング手段内部の衝突で負けたパケットについては、例えば、待ち合わせバッファで待たせるようにしてもよいし、あるいは即座に廃棄し該当する入側転送手段から再送するようにしてもよい。入側転送手段がスイッチング手段内でパケットが廃棄されたことを検出する方法は、明示的に再送要求を受信したことによる方法、暗黙的に到着確認情報が期待する時刻までに戻ってこなかったことによる方法などが考えられる。
【００３５】
なお、１つの出側転送手段には、１つの輻輳状況が存在する場合の他に、複数の輻輳状況が存在する場合がある。例えば、出側転送手段毎でもフロー毎やクラス毎に輻輳状況が違うケース、１つの出側転送手段に複数の低速ポートがぶら下がっているケースなどが考えられる。
【００３６】
好ましくは、前記優先度付与手段は、パケットの転送先の輻輳の度合いがより大きいほど、より低い優先度を付与する（パケットの転送先の輻輳の度合いがより小さいほど、より高い優先度を付与する）ようにしてもよい。
【００３７】
好ましくは、前記優先度付与手段は、各入側転送手段内に設けられたものであるようにしてもよい。
【００３８】
好ましくは、前記優先度付与手段があるパケットに優先度を付与するために参照する転送先の輻輳状況が未知または無効となっている場合に、該転送先に転送する最初の１つのパケットまたは複数のパケット群に付与する優先度を一時的に高く設定するようにしてもよい。
【００３９】
好ましくは、前記優先度付与手段は、各入側転送手段毎に１つずつまたは複数の入側転送手段で１つ設けられた、前記輻輳状況観測手段により観測された輻輳状況に基づいて所定の転送先毎に設定される輻輳度を記憶する輻輳度テーブルを参照して、前記パケットに付与すべき優先度を設定するようにしてもよい。
【００４０】
好ましくは、前記輻輳状況観測手段は、各出側転送手段内に設けられたものであるようにしてもよい。
【００４１】

好ましくは、前記輻輳状況観測手段は、前記転送先毎の輻輳状況として、対応する出側転送手段毎、対応する出側転送手段のクラス毎、対応する出側転送手段のポート毎、対応する出側転送手段の各ポートのクラス毎、または対応する出側転送手段の各ポートの各クラスのフロー毎に、前記輻輳状況を観測する手段を含むようにしてもよい。

【００４２】

好ましくは、前記輻輳状況観測手段により観測された前記輻輳状況を、前記優先度付与手段に反映させるための輻輳状況通知手段を更に備えるようにしてもよい。

【００４３】

好ましくは、前記優先度付与手段は、各入側転送手段内に設けられ、前記輻輳状況通知手段は、前記出側転送手段にパケットが到着したことを契機に、該パケットを送出した入側転送手段に対して所定の輻輳状況に関する情報を通知するようにしてもよい。より具体的には、例えば、ＡＣＫ信号に輻輳状況を載せる方法や、逆向きのパケットを作ってそれに輻輳状況を載せる方法が考えられる。

【００４４】

なお、全出側転送手段の輻輳状況を全入側転送手段に定期的に通知するようにしてもよい。

【００４５】

また、前記優先度付与手段は、前記入側転送手段内に設ける代わりにスイッチング手段内に設けるようにしてもよい。

【００４６】

また、前記輻輳状況観測手段や前記輻輳状況通知手段は、前記出側転送手段内に設ける代わりにスイッチング手段内に設けるようにしてもよいし、前記出側転送手段内とスイッチング手段内の両方に設けるようにしてもよい。

【００４７】

好ましくは、前記入側転送手段毎に、各入側転送手段において複数のパケットが前記スイッチング手段への転送を待っている場合に、該複数のパケットの転送順を制御するためのスケジューリング手段を更に備え、前記スケジューリング手段は、各パケットの転送先の輻輳状況を考慮して、輻輳していない転送先行きのパケットが優先的に前記スイッチング手段に転送されるように制御するようにしてもよい。

【００４８】

また、例えば、パケットにリアルタイムあるいはベストエフォートといったクラスの属性がついている場合に、出力すべきクラスをまず決定し、そのクラスの中で輻輳状況を考慮してパケットを選ぶようにしてもよい。このような場合には、輻輳度の小さな順にパケットが転送されるとは限らないことになる。

【００４９】

なお、このスケジューリング手段は、独立して実施することができる。すなわち、本発明は、入側転送手段からスイッチング手段を経由して所望の出側転送手段へパケットを転送するパケットスイッチであって、入側転送手段毎に、各入側転送手段において複数のパケットが前記スイッチング手段への転送を待っている場合に、該複数のパケットの転送順を制御するためのスケジューリング手段と、所定の転送先毎の輻輳状況を観測するための輻輳状況観測手段とを備え、各スケジューリング手段は、各パケットの転送先の輻輳状況を考慮して、輻輳していない転送先行きのパケットが優先的に前記スイッチング手段に転送されるように制御することを特徴とする。

【００５０】

本発明（請求項１１）は、優先度の付与されたパケット（例えば、パケットデータ、またはパケットデータの転送に先立つリクエスト）を入側転送手段からスイッチング手段を経由して所望の出側転送手段へ転送し、該スイッチング手段は内部でパケットの衝突が発生した場合に該優先度を考慮して選択したパケットを出側転送手段に対して転送し、その他のパケットを該スイッチング手段内部で廃棄するパケットスイッチであって、前記入側転送手段は、前記衝突によるパケット廃棄が検出された場合に、該廃棄されたパケットを再

送するための手段と、前記再送パケットに付与すべき優先度を、もとの廃棄されたパケットに付与した優先度よりも高く設定するための手段とを含むことを特徴とする。
【００５１】
本発明によれば、廃棄されたパケットに付与する優先度をより高くすることにより、該パケットを廃棄されにくくさせることができる。
【００５２】
なお、連続廃棄回数に応じて優先度を順次高くしていくようにしてもよい。
【００５３】
入側転送手段がスイッチング手段内でパケットが廃棄されたことを検出する方法は、明示的に再送要求を受信したことによる方法のほか、暗黙的に到着確認情報が期待する時刻までに戻ってこなかったことによる方法などが考えられる。
【００５４】
この発明は、これまでの各発明と組み合わせて実施することができる。
【００５５】
本発明（請求項１２）は、優先度の付与されたパケット（例えば、パケットデータ、またはパケットデータの転送に先立つリクエスト）を入側転送手段からスイッチング手段を経由して所望の出側転送手段へ転送し、該スイッチング手段は内部でパケットの衝突が発生した場合に、該優先度を考慮して選択したパケットを出側転送手段に対して優先的に転送するパケットスイッチであって、前記入側転送手段は、１つのデータが複数のパケットに分割して搭載された際の各パケットを転送する場合に、分割されたデータの後続部分に対応するパケットの優先度を、先頭部分に対応するパケットよりも高く設定するための手段を含むことを特徴とする。好ましくは、前記分割されたデータの先頭部分に対応するパケットの優先度を、分割されたデータを搭載したものではないパケットならば付与するであろう優先度と同等の値に設定するようにしてもよい。
【００５６】
本発明によれば、一度データグラムの先頭に対応するパケットが宛先の出側転送部に到達すればその後の部分も高優先で続けて宛先に到達させることができ、結果として複数のパケットからなるデータグラムを一塊りに出側転送部へ転送させることができる。
【００５７】
この発明は、これまでの各発明と組み合わせて実施することができる。
【００５８】
本発明（請求項１４）は、パケット（例えば、パケットデータ、またはパケットデータの転送に先立つリクエスト）を入側転送手段からスイッチング手段を経由して所望の出側転送手段へ転送するパケットスイッチであって、前記スイッチング手段を転送されるパケットとともに該パケットを送出した入側転送手段内部の輻輳状況を出側転送手段へ転送するための手段と、前記パケットとともに通知された前記輻輳状況と、該パケットが転送された出側転送手段内部の輻輳状況とを用いて、総合的な輻輳状況を求めるための手段と、求められた前記総合的な輻輳状況を、通過するパケットフローの量または速度を制御するためにネットワークにおいて行われる輻輳制御（輻輳制御とは、ＥＦＣＩ、ＥＣＮ、ｓｏｕｒｃｅ　ｑｕｅｎｃｈ、ＢＥＣＮ、ＲＥＤなど）に用いるための手段とを備えたことを特徴とする。
【００５９】
例えば、ある入側転送部のフローのパケットキューの長さ（例えば２８９とする）をパケットとともに出側転送部へ転送し、これに出側転送部のそのフローのパケットキューの長さ（例えば１２とする）を合計することにより、パケットスイッチ内に存在するそのフローの待ちパケット量（＝２８９＋１２＝３０１）を把握できる。そして、例えば、これをユーザへ通知することで、パケットフローの速度を制御し、これによって輻輳制御を制御することができる。
【００６０】
この発明は、これまでの各発明と組み合わせて実施することができる。

【００６１】
本発明（請求項１５）は、入側転送手段から、パケット衝突の発生しないスイッチング手段を経由して所望の出側転送手段へ、パケット（例えば、パケットデータ、またはパケットデータの転送に先立つリクエスト）を転送するパケットスイッチであって、転送先毎の輻輳状況を観測するための輻輳状況観測手段と、各入側転送手段から送出されるパケットに前記輻輳状況観測手段により観測された該パケットの転送先の輻輳状況に基づいて優先度を付与し、各パケットをパケット衝突の発生するトポロジを有する仮想的なスイッチング網内を転送させたと仮定して、前記パケットの衝突が発生した場合にパケットに付与された優先度に基づき優先して転送すべきパケットを選択するシミュレーションを行い、このシミュレーションにより前記入側転送手段から前記出側転送手段までパケットが到達した結果と同等な結果となるように、前記パケット衝突の発生しないスイッチング手段の接続パターンを決定する接続パターン決定手段とを備えたことを特徴とする。
【００６２】
好ましくは、前記パケット衝突の発生しないスイッチング手段は、クロスバスイッチであり、前記仮想的なスイッチング網は、単位スイッチにより構成されたスイッチ網（例えば、ランダム網とルーティング網を接続したもの）であるようにしてもよい。
【００６３】
本発明では、接続パターン決定手段がシミュレーションにより、例えば、各入側転送部のデータ転送のリクエストをトーナメントさせ、勝ち残ったリクエストに対応するデータを交換するようにクロスバスイッチ等の接続パターンを決定する。トーナメントの競合では、輻輳度テーブルの輻輳度を考慮して設定もしくは変更された優先度によって、勝ち残るリクエストを決定する。このようにして出側転送部へ向けて転送を許可するリクエストの集合が求まると、接続パターンが求まる。この接続パターンによって設定されたクロスバスイッチ等によってパケットが交換される。
【００６４】
従来の接続パターン決定に要する計算量は、スイッチのポート数をＮとすると$Ｎ^2$のオーダとなり、ポート数が多くなるとともに計算が困難になる問題点が知られているが、本発明によれば、スイッチのポート数をＮとするとＮ・ｌｏｇＮのオーダで出側転送部へ転送してよいリクエストを求めることができ、比較的高速に計算することができるとともに、出側転送部の輻輳状況を考慮しているので公平性も確保できるという利点がある。
【００６５】
本発明（請求項１７）は、入側転送手段からスイッチング手段を経由して所望の出側転送手段へ、パケットを転送するパケットスイッチのパケット交換方法であって、前記入側転送手段は、前記パケットにその転送先の輻輳状況に基づいた優先度を付与してこれを前記スイッチング手段へ送出し、前記スイッチング手段は、前記入側転送手段から転送された前記パケットをその転送先に従って交換するとともに、その内部でパケットの衝突が発生した場合には各パケットに付与された優先度を考慮して選択したパケットを優先して交換し、前記パケットの到達した前記出側転送手段は、所定の観測単位についての輻輳状況の観測結果を示す情報を、該パケットを送出した前記入側転送手段に通知することを特徴とする。
【００６６】
本発明（請求項１８）は、優先度の付与されたパケットを入側転送手段からスイッチング手段を経由して所望の出側転送手段へ転送し、該スイッチング手段は内部でパケットの衝突が発生した場合に該優先度を考慮して選択したパケットを出側転送手段に対して転送し、その他のパケットを該スイッチング手段内部で廃棄するパケットスイッチのパケット交換方法であって、前記入側転送手段は、前記衝突によるパケット廃棄が検出された場合に、該廃棄されたパケットに、もとの廃棄されたパケットに付与した優先度よりも高い優先度を付与して、これを再送することを特徴とする。
【００６７】
本発明（請求項１９）は、優先度の付与されたパケットを入側転送手段からスイッチング

手段を経由して所望の出側転送手段へ転送し、該スイッチング手段は内部でパケットの衝突が発生した場合に、該優先度を考慮して選択したパケットを出側転送手段に対して優先的に転送するパケットスイッチのパケット交換方法であって、前記入側転送手段は、１つのデータが複数のパケットに分割して搭載された際の各パケットを転送する場合に、分割されたデータの先頭部分に対応するパケットに所定の優先度を付与して送出し、該パケットが転送先に到達したならば、該データの後続部分に対応するパケットに該所定の優先度より高い優先度を付与してを送出することを特徴とする。

【００６８】

本発明（請求項２０）は、パケットを入側転送手段からスイッチング手段を経由して所望の出側転送手段へ転送するパケットスイッチのパケット交換方法であって、前記入側転送手段は、その内部の輻輳状況を示す情報をパケットに付加して送出し、前記出側転送手段は、その内部の輻輳状況と、前記パケットに付加されて通知された前記輻輳状況とを用いて、総合的な輻輳状況を求め、求められた前記総合的な輻輳状況を、所定の輻輳制御に用いることを特徴とする。

【００６９】

なお、以上の本発明は、パケット網全体の輻輳の制御にも適用可能である。

【００７０】

また、装置に係る本発明は方法に係る発明としても成立し、方法に係る本発明は装置に係る発明としても成立する。

【００７１】

また、装置または方法に係る本発明は、コンピュータに当該発明に相当する手順を実行させるための（あるいはコンピュータを当該発明に相当する手段として機能させるための、あるいはコンピュータに当該発明に相当する機能を実現させるための）プログラムを記録したコンピュータ読取り可能な記録媒体としても成立する。

【００７２】

【発明の実施の形態】

以下、図面を参照しながら発明の実施の形態を説明する。

【００７３】

（第１の実施形態）

図１に、本発明の第１の実施形態に係る多段スイッチ型のパケットスイッチの構成例を示す。また、併せて、該パケットスイッチの動作概要を示す。

【００７４】

本パケットスイッチの基本的な全体構成は、図６と同様で、入力ポートから入力されたパケットは、入側転送部２からスイッチング部４を経由して所望の出側転送部６へ転送され、そこから出力ポートへ出力される。本パケットスイッチの入力ポート数、出力ポート数は、任意である。また、１つの入側転送部や出側転送部に複数の入力ポートや出力ポートが対応しても構わない。本パケットスイッチのスイッチング部４の内部の網構成も任意であり、その網を構成する１つの単位スイッチの入力数、出力数も任意である。もちろん、本パケットスイッチが図６と同じ全体構成であってもよい。なお、図１では、４つの入側転送部２と４つの出側転送部６を示している。また、図１では、スイッチング部４の内部構成の詳細は省略している。

【００７５】

ここで、本実施形態における「パケット」とは、本パケットスイッチで交換したいデータそのものであってもよいし、また、データの転送に先立ってスイッチング部の内部を転送される情報であってもよい（前者をデータ・パケットと呼び、後者のような情報を「リクエスト」と呼ぶものとする）。

【００７６】

「パケット」という語句が「リクエスト」を指す場合には、まず、入側転送部２が、データの転送に先立って、宛先情報が書き込まれたリクエストをスイッチング部４へ送出する。このリクエストには優先度が書き込まれる。リクエストが無事に宛先の出側転送部６に

到達した場合もしくは到達したと判断された場合には、そのリクエストに対応するデータ・パケットが入側転送部２から出側転送部６へ転送される。リクエストが経路を確保し、その経路に従ってデータ・パケットを転送すれば、データ・パケットは途中で他のデータ・パケットと衝突することなく安全に出側転送部６へ転送される。データ・パケットを転送している時間に次のデータ・パケットに対応するリクエストを転送するようにすれば、パイプライン的に効率良くデータを入側転送部２から出側転送部６へ交換することができる。

【００７７】
本実施形態では、交換したいデータそのものとデータ・パケットの転送に先立つリクエストとを区別せずに「パケット」の語句を用いて説明するが、いずれの場合であっても本発明は有効に作用する。

【００７８】
また、本発明は、リクエストを用いる（もしくは経路を確保する）パケットスイッチにも、リクエストを用いない（もしくは経路を確保しない）パケットスイッチにも適用可能である。

【００７９】
これらの点は、後述する他の実施形態についても同様である。

【００８０】
さて、本実施形態のパケットスイッチでは、パケットを入側転送部２からスイッチング部４を経由して所望の出側転送部６へ転送する際に、各パケットに優先度を付与し、スイッチング部においてこの優先度に基づいて転送パケット間の優先制御を行う。つまり、この優先度は、パケットスイッチ内で局所的に使われるものである。

【００８１】
優先度は、パケットのパケットスイッチ内における到達すべき転送先、例えば、出側転送部、あるいは出力ポート、あるいは出力ポートをさらに細分化したクラスもしくはパケットの属するフロー（フローは、例えば、そのパケットの宛先アドレス、あるいはこれにポート番号やプロトコル番号等の情報を適宜組み合わせたものによって規定される；ＡＴＭ通信であればフローは仮想コネクションである）についての輻輳状況を少なくとも考慮して設定される。優先度を設定するにあたっては、他の情報、例えば、そのパケットの持つ所定の属性（例えば、パケットの属するクラスなど）や、そのパケットの持つ特別の条件（例えば、再送パケットであることなど）をも考慮するようにしてもよい。

【００８２】
優先度の段階については、特に制限はなく、何段階としても本発明は有効に作用する。

【００８３】
パケットに優先度を付与する場所については、それがどこであっても本発明は有効に作用するが、本実施形態では、各入側転送部２においてパケットに優先度を付与する場合を例にとって説明する。

【００８４】
輻輳状況を観測する場所については、例えばスイッチング部４内でも出側転送部６内でも本発明は有効に作用するが、後者がより適しているので、本実施形態では、各出側転送部６において輻輳状況を観測し、その情報を入側転送部２に通知する場合を例にとって説明する。

【００８５】
図１に示した輻輳状況を考慮して設定した優先度による制御を実現するためのパケットスイッチは、一構成例として、各入側転送部２毎に優先度付与部２１と輻輳度テーブル１２を持ち、出側転送部４毎に輻輳状況観測部６１を持ち、各単位スイッチ４５毎にパケット選択部４５１を持つ構成となっている。

【００８６】
なお、全入側転送部で１つの輻輳度テーブルを共用する構成としてもよい。
この場合には、出側転送部側から輻輳状況に関する情報の通知を受け、これを共用の輻輳

度テーブルに反映させる輻輳度テーブル管理部（図示せず）を持つ構成とすればよい。

【００８７】

また、一部の単位スイッチ４５にはパケット選択部４５１を設けないようにしてもよい。

【００８８】

概略的には、入側転送部２の優先度付与部２１は、出側転送部６からその輻輳状況観測部６１によって観測された輻輳状況に関する情報の通知を受け、これを輻輳度テーブル１２に反映させること（輻輳度テーブル管理機能）と、当該入側転送部２からパケットが送出される際に輻輳度テーブル１２の内容を考慮してそのパケットに付与する優先度を決定すること（優先度付与機能）を行う。

【００８９】

輻輳度は、通知された輻輳状況に関する情報そのものであってもよいし、通知された輻輳状況に関する情報を変換した値であってもよい。輻輳度の設定は、出側転送部単位に行なってもよいし、ポート単位に行なってもよいし、さらに細分化してクラス単位あるいはフロー単位に行なってもよい。

【００９０】

優先度の設定については、基本的な考え方としては、輻輳度テーブル１２からそのパケットに該当する輻輳度を参照し、輻輳度の高いものほど優先度を低く設定し、輻輳度の低いものほど優先度を高く設定するというものである。

【００９１】

ただし、輻輳度と優先度との関係については、多種多様なバリエーションが考えられる。輻輳度と優先度とが線形の関係にあってもよいし非線形の関係にあってもよい。また、輻輳度の段階数と優先度の段階数を同じにしてもよいし、異なるものにしてもよい。また、前述したように、輻輳度と他の情報から優先度を求めるようにしてもよい。

【００９２】

また、輻輳度から優先度を求める手順もしくは関数等を、輻輳度の設定と同じ単位（例えばポート単位）で定義してもよいし、輻輳度の設定より細かい単位（例えば、ポート単位に対してフロー単位）で設けてもよい。

【００９３】

概略的には、単位スイッチ４５のパケット選択部４５１は、パケットが衝突したときにそれらに付与された優先度を考慮して優先して転送すべきパケットを選択するものである。

【００９４】

パケットの選択については、基本的な考え方としては、最も高い優先度を持つパケットを選択するというものであるが、優先度に加えて他の情報（例えば、パケットヘッダ内に記載されている何らかの情報）をも考慮して選択するようにしてもよい。

【００９５】

なお、最も高い優先度を持つパケットが複数存在するなどにより、通常行う判断では１つのパケットを選択できない場合のために何らかの特別の選択基準を設けておいてもよいし、通常行う判断では１つのパケットを選択できない場合には最も優越する複数のパケットのうちからランダムに１つを選択するようにしてもよい。

【００９６】

概略的には、出側転送部４の輻輳状況観測部６１は、当該出側転送部４について輻輳状況を観測するものである。観測された輻輳状況に関する情報は、当該出側転送部４から入側転送部２に通知される。

【００９７】

輻輳状況の観測は、出側転送部単位に行なってもよいし、ポート単位に行なってもよいし、さらに細分化してクラス単位あるいはフロー単位に行なってもよい。

【００９８】

通知すべき輻輳状況に関する情報としては、輻輳状況に関する所定の測定値（例えば、単位時間に転送したパケットの個数もしくはバイト数、パケットキューのキュー長（パケット数もしくはバイト数）、パケットキューから廃棄されたパケット数もしくはバイト数な

ど）をそのまま用いてもよいし、所定の測定値を１つの基準値によって輻輳有りまたは無しに分類した結果を用いてもよいし、所定の測定値を複数の基準値によって分類したレベル値を用いてもよい。

【００９９】
なお、輻輳状況の観測の単位（例えばポート単位）と輻輳度テーブルのエントリの単位（例えばポート単位）を一致させてもよいし、輻輳状況の観測の単位（例えばフロー単位）に比べて輻輳度テーブルのエントリの単位（例えばポート単位）をより荒くするようにしてもよい。

【０１００】
以下では、本実施形態についてより詳しく説明する。

【０１０１】
最初に、パケットの衝突の発生した場合について説明する。

【０１０２】
スイッチング部４の内部でパケットの衝突（例えばルーティング網内衝突）が生じた場合に、その箇所にあたる単位スイッチ４５では、パケット選択部４５１により、衝突した各パケットに付与されている優先度を考慮して、優先して転送すべきパケットを選択し、その選択されたパケットのみを転送する。

【０１０３】
衝突で負けた他のパケットについては、例えば、即座に廃棄し、該当する入側転送部２が後で再送する（この場合、入側転送部２は、パケットの再送を考慮して、送出したパケットを一定時間保持しておく必要がある）。再送の契機を与える方法としては、例えば、パケットを廃棄した単位スイッチが入側転送部２に再送メッセージを出す方法や、あるいはパケットが到達した出側転送部４が転送元の入側転送部２にパケット到達メッセージを出すものとして入側転送部２が到達先となる出側転送部４からパケット到達メッセージを一定時間経過しても受信しない場合にパケットが廃棄されたとみなす方法などがある。

【０１０４】
あるいは、全部または一部の単位スイッチを、パケットバッファを備える構成とし、上記のように衝突で負けたパケットを即座に廃棄する代わりに、該パケットを該単位スイッチ内のパケットバッファに蓄積して一時的に待ち合わせるようにしてもよい。

【０１０５】
続いて、輻輳状況の観測、通知について説明する。

【０１０６】
本実施形態では、その特徴の一つとして、パケットが転送される経路の輻輳状況を観測し、その輻輳状況に関する情報を、優先度をパケットに付与する部分（本実施形態では入側転送部２）へ伝えるようにしている。前述したように、輻輳状況を観測する場所は、本実施形態では、より適している出側転送部６としている。この点について説明する。

【０１０７】
出側転送部２にスイッチング部４からパケットが流れ込む速度よりも、出側転送部６からパケットを出力ポートへ出力する速度が低くなっていれば、スイッチング部４内の輻輳状況が出側転送部６の内部の輻輳（例えば、出力用バッファの蓄積パケット数もしくはバイト数の増大）に反映される。つまり、出側転送部入口輻輳が発生すれば必ず出側転送部の内部も輻輳になる。

【０１０８】
また、スイッチング部４に輻輳が発生しなくても、出側転送部６の内部のみに輻輳（前述の出側転送部内輻輳）が生じることもある。これは、一つの出側転送部６が複数の出力ポートを備えている構成で発生しやすい。

【０１０９】
以上から、出側転送部入口輻輳も出側転送部内輻輳も、出側転送部６だけを観測すれば検出できることになる。

【０１１０】

なお、輻輳状況は、出側転送部毎、出力ポート毎、クラス毎、パケットのフロー毎（例えば、仮想コネクション毎）、またはこれらの組合せ、で観測する。図１では、輻輳状況をフロー毎に観測し、輻輳テーブル１２をフロー毎に設定する例を示している。

【０１１１】

さて、このように観測された輻輳状況は、本実施形態では入側転送部２へ伝えられるが、輻輳状況を入側転送部２へ伝える方法としては、いくつかの方法が考えられる。

【０１１２】

例えば、出側転送部６から入側転送部２へ転送されるＡＣＫ／ＮＡＣＫ信号とともに輻輳状況を送る方法が考えられる。従来のパケットスイッチにおいて、出側転送部から入側転送部へ転送される信号として、パケットが無事に出側転送部に到着したことを通知するＡＣＫ（Ａｃｋｎｏｗｌｃｄｇｅｍｅｎｔ：受信応答）信号、または、出側転送部にパケットが到着しても出側転送部がそのパケットを何らかの理由で拒否するＮＡＣＫ（Ｎｏｎ－ａｃｋｎｏｗｌｅｄｇｅｍｅｎｔ）信号がある。一般的によく知られている方法によれば、これらの信号は、出側転送部からそのパケットを送出した入側転送部へ、パケットの転送経路を逆向きにたどって返送される。すなわち、このＡＣＫ／ＮＡＣＫ信号とともに輻輳状況に関する情報を送るというのが、一つの方法である。

【０１１３】

また、他の方法としては、各出側転送部が、その観測単位（例えばポート）についてそれに一定個数（一定個数を１としてもよい）のパケットが到着する毎に、または一定の時間が経過する毎に、そのときにパケットを送出した入側転送部２（あるいは前回の通知からの間にパケットを送出した入側転送部２）に向かって、輻輳状況に関する情報を通知するパケットを生成して返送するようにしてもよい。

【０１１４】

以上のようにして輻輳状況を転送すれば、ポート数が大きなパケットスイッチであっても、容易に出側転送部６の輻輳状況を入側転送部２へ伝えることが可能となる。

【０１１５】

なお、上記のいずれの場合においても、観測単位を出側転送部毎より細分化された単位（例えばポート毎）としている場合に、通知すべき輻輳状況に関する情報を、当該到着したパケットと同じ転送先（例えばポート）についての情報とする代わりに、出側転送部を同じくする全ての転送先についての情報としてもよいし、全ての出側転送部についての全ての情報としてもよい。

【０１１６】

また、一定周期等の所定のタイミングで、全ての出側転送部についての全ての輻輳状況に関する情報を、全ての入側転送手段に通知するようにしてもよい。

【０１１７】

なお、上記のように出側転送手段で輻輳状況を観測し通知する代わりに、その直前または複数段前の単位スイッチで観測し通知するようにしてもよい。また、出側転送手段と単位スイッチの双方で輻輳状況を観測し通知するようにしてもよい。

【０１１８】

続いて、優先度によるスイッチ内パケット転送制御のいくつかの例について説明する。

【０１１９】

まず、最も代表的な制御、すなわち輻輳状態に応じて行う優先度制御について説明する。

【０１２０】

パケットの優先度はもともと例えばリアルタイム情報あるいはベストエフォート情報などといったそのパケットやそのパケットのフローが属するクラスに基づいて決定することが一般的であるが、本実施形態では、その優先度を輻輳状況に応じて設定する（あるいは、上記のようにしてもともと付与されていた優先度を変更する）。

【０１２１】

例えば、輻輳している到達先（輻輳ポートとする）行きのパケットは優先度を低くする。優先度の高いパケットはスイッチング部４の内部で衝突しても優先されるため、優先度の

高いパケットにとって輻輳しているポート行きの優先度の低いパケットは存在しないのと同じことになる。

【０１２２】

したがって、本実施形態により、輻輳していないポート行きのパケットの流れが輻輳ポート行きのパケットの流れによって乱されることが非常に少なくなるという効果がある。輻輳ポート行きのパケットにとっても、優先度が変化させられるだけで、その転送速度は抑制されていないため、優先度の高いパケットの転送のすき間を利用して出側転送部へ転送する試みを継続することができる。そのため、輻輳ポートへもスイッチング部４の転送能力の限りパケットを転送し続けることができる。

【０１２３】

本実施形態とは異なった方法として優先度を用いるのではなく輻輳ポート行きパケットの転送速度を輻輳度に応じて厳密に抑制するという複雑な制御が考えられるが、本実施形態によれば、そのような転送速度抑制の制御は不要であり、入側転送部の構造も簡単にできるという利点がある。

【０１２４】

なお、従来からパケットに優先度を付与して転送する方法は知られているが、その従来の方法に比べ、本実施形態は輻輳度に応じて動的に優先度を設定もしくは変更することが本質的に異なるものである。

【０１２５】

次に、再送パケットに対する優先度制御について説明する。

【０１２６】

スイッチング部４の内部の衝突で廃棄されたパケットを入側転送部２が再送する方式のパケットスイッチでは、同じパケットが連続して廃棄されてしまうと、そのパケットの後で転送を待っているパケット全体の流れが悪くなってしまうおそれがある。

【０１２７】

これを解決する方法として、スイッチング部４の内部でパケットが廃棄された場合に、その再送パケットの優先度をもとのパケットの優先度よりも高くするようにしてもよい。このような制御によって、再送パケットが再び廃棄されることが少なくなる利点がある。もし再送パケットが再び廃棄された場合には、それに対する再送パケットの優先度を更に高くするようにしてもよい。これによって、何度も再送され続けている不運のパケットを優先して転送することが可能となる。

【０１２８】

次に、プローブパケットに対する優先度制御について説明する。

【０１２９】

上記の再送パケットのように優先度を一時的に変更するとよい状況は他にも種々考えられる。

【０１３０】

例えば、出側転送部６にパケットが到達したことを契機として該出側転送部６から該パケットを送出した入側転送部２に輻輳状況に関する情報を通知する構成においては、入側転送部２が新しい宛先に初めて（または輻輳状況が変化している程度に久しぶりに）パケットを転送する場合、その宛先の輻輳状況はわからない（もしくはその宛先の実際の輻輳状況はわからない）。このように宛先の輻輳状況が未知のとき（もしくは未知に等しいようなとき）に送るパケットは、宛先の輻輳状況を知るための探査信号の意味合いが強い。

【０１３１】

この探査を行うパケット（プローブパケット）の優先度を高くすれば、スイッチング部４で廃棄されにくくなるため、早期に宛先の輻輳状況が判明し、その後で転送するパケットの優先度を素早く適切な値に設定することができるという利点がある。優先度を高く設定するパケットは最初の１パケットだけで十分であるが、実装上の都合などで複数のパケットの優先度が高くなったとしても効果がある。

【０１３２】

次に、分割されたデータグラムに対する優先度制御について説明する。

【０１３３】

優先度を一時的に変更するとよい他の例としては、複数に分割されたデータグラムをスイッチング部４が転送する場合に、そのデータグラムの最初の部分に対応するパケットの優先度を他の部分に対応するパケットよりも低く設定することである。他の部分に対応するパケットの優先度を最初の部分に対応するパケットよりも高くしてもよい。例えば、先頭のパケットの優先度を、本実施形態の方法で設定し、２番目以降のパケットの優先度を非常に高めの値にしてもよい。

【０１３４】

また、先頭のパケットの最初の送出では宛先の出側転送部６に到達しなかった場合には、優先度をさらに高く設定して繰り返し転送を試みるようにしてもよい。

【０１３５】

このようにすれば、一度データグラムの先頭に対応するパケットが宛先の出側転送部６に到達すればその後の部分も高優先で続けて宛先に到達させることができ、結果として複数のパケットからなるデータグラムを一塊りに出側転送部６へ転送することができる。出側転送部６が複数に分割されたデータグラムを再び一つに再構成して外部へ出力するような場合などでは、スイッチング部４でこのような転送を行うことにより、データグラムを再構成するために必要なバッファ量を少なくすることができる利点がある。

【０１３６】

以上のように、パケットの優先度は、輻輳度のみによって決定する方法だけでなく、例えば、パケットまたはフローの属するクラスによる優先度に、輻輳状況による変更と再送回数に応じた変更などを行なって、最終的に決定する（もしくは、輻輳度と、クラスおよび／またはその他のファクターとをパラメータとして求める）方法がある。

【０１３７】

また、さらにパケットの優先度を高度に制御することにより、スイッチング部４内のパケット転送速度を保証することも可能である。例えば、入側転送部２でフローの転送速度をモニタし、設定した速度を下回りそうになった場合に、優先度を一時的に高くする制御を行なうことで、転送速度の最小値を保証するサービスを提供することができる。このサービスは、スイッチがすいているときには、保証速度を上回る転送が可能である。また、重みをつけてその比で帯域を分割する、といったことも可能である。

【０１３８】

なお、パケットの優先度を設定もしくは変更する優先度付与部の実現方法として、入側転送部２の内部にプロセッサを配置しソフトウェアで優先度を決定するようにすれば、パケットスイッチのポート構成や扱うクラス数などが運用中に変化しても柔軟に対応することができる。

【０１３９】

続いて、輻輳度テーブルと優先度付与部のバリエーションについて説明する。

【０１４０】

本実施形態では、その大きなの特徴として、出側転送部６から通知された輻輳状況に応じて、対応する宛先へと向かうパケットの優先度を変更するようにしている。

【０１４１】

その実現例としては、パケットを転送することにより通知された輻輳状況に関する情報から得られる輻輳度を、そのパケットを転送した入側転送部２の輻輳度テーブル１２で記憶し、そこから転送する同じ宛先行きのパケットの優先度を設定もしくは変更する方法があるが、この方法としては、その他にも、種々のバリエーションが考えられる。

【０１４２】

例えば、輻輳状況に関する情報を複数の場所、例えば複数の入側転送部２で共通に記憶してもよい。輻輳度テーブル１２を共用することで、各入側転送部２は自らが得た輻輳状況に加えて他の入側転送部２が得た輻輳状況をも使用でき、より多くの輻輳状況がわかるため、輻輳に即座に対応できる確率が高くなるという利点がある。

【０１４３】
また、他の方法として、図２に示すように、入側転送部２で輻輳状況を考慮した優先度をパケットに付与するのではなく、（入側転送部２では輻輳状況を考慮しない優先度をパケットに付与し）、スイッチング部４の内部に配置したパケット優先度変更部４２で、パケットに付与されている優先度を、輻輳状況を考慮したものに変更するようにしてもよい。
【０１４４】
この方法の場合、出側転送部６から転送された輻輳状況は、スイッチング部４の内部のパケット優先度変更部４２へ通知される。パケット優先度変更部４２は、その通知された輻輳状況をもとに輻輳度を求めて輻輳度テーブル５２に記憶し、通過するパケットの宛先に応じてそのパケットの優先度を変更する。このパケット優先度変更部５２は必ずしも一つでパケットスイッチの全ての出側転送部６を受け持つ必要はない。一つのスイッチング部４に複数のパケット優先度変更部５２を用意し、それぞれが近くの出側転送部６を担当すれば処理が分散され実装が容易にできる。
【０１４５】
この方法の場合、宛先の輻輳状況に応じた優先度をその宛先に近い場所でパケットに付与できるため、効果的にパケットに優先度を付与することができるという利点がある。
【０１４６】
（第２の実施形態）
第１の実施形態では、輻輳度に応じて優先度を制御したが、第２の実施形態では、スケジューリング（複数のパケットがスイッチング部への転送を待っている場合に、どのパケットを転送するかを選択する処理）を輻輳度を考慮して行うようにしたものである。
【０１４７】
本発明の第２の実施形態に係る多段スイッチ型のパケットスイッチの基本的な全体構成は、第１の実施形態のように、図６と同様で、入力ポートから入力されたパケットは、入側転送部からスイッチング部を経由して所望の出側転送部へ転送され、そこから出力ポートへ出力される。また、パケットスイッチの入力ポート数、出力ポート数が任意である点、１つの入側転送部や出側転送部に複数の入力ポートや出力ポートが対応しても構わない点、スイッチング部の内部の網構成も任意である点、その網を構成する１つの単位スイッチの入力数、出力数も任意である点は、第１の実施形態と同様である。
【０１４８】
また、本実施形態では、入側転送部の輻輳度テーブル、出側転送部の輻輳状況観測部については、第１の実施形態と同様のものを備えるものとする。また、第１の実施形態の優先度付与部のうち輻輳度テーブルの輻輳度を設定する部分を入側転送部に備えるものとする。
【０１４９】
図３に、本実施形態の多段スイッチ型パケットスイッチの各入側転送部のスケジューリングに関する構成の一例を示す。
【０１５０】
図３においては、スケジューリングするグループ（例えば、フロー単位）毎に設けられたパケットキュー２２、クラス間スケジューリング設定部２４、転送セル選択部２６、輻輳度テーブル３２を示している。
【０１５１】
なお、図３の例では、スケジューリングに輻輳度とクラスを用いるものとする。また、輻輳度テーブルに輻輳度と併せてクラスのフィールドを設けた構成例を示している。
【０１５２】
以下、本実施形態についてより詳しく説明する。
【０１５３】
入側転送部内で転送を待っているパケットは、同じ輻輳状況を共有するグループに分類してパケットキュー２２を作ることにより、効率よく転送できる。例えば、このパケットキュー２２は、パケットのフロー（または仮想コネクション）毎に作る。他の例としては、

パケットスイッチの各出側転送部の各出力ポートのクラス毎に作ることも有効である。さらに、より簡単には各出側転送部のクラス毎に作ってもよい。ただし、クラスの概念を持たないパケットスイッチの場合には、クラス毎の分類は不要である。

【０１５４】

ここでは、フロー毎にパケットキュー２２を作って管理するものとして説明する。

【０１５５】

輻輳状況が出側転送部から転送されてくると、入側転送部はそれをもとに各フローの輻輳度をテーブル３２に記憶・更新する。各フローにクラスの属性がついていれば、その情報も同じテーブル３２に記憶すると便利である。

【０１５６】

図３の転送セル選択部２６は、転送すべきパケットを選択する場合に、この輻輳度とクラス情報のテーブル３２、およびクラス間スケジューリング設定部２４を参照する。クラス間スケジューリング設定部２４は、例えば、リアルタイムクラス（図３のＲＴ）の転送速度を、入側転送部からのパケット送出最大速度の８０パーセントを限度として、ベストエフォートクラス（図３のＢＥ）よりも優先して出力する、などといった、クラス間のパケット選択ポリシーを設定するためのものである。

【０１５７】

本実施形態では、転送セル選択部２６は、まず、（ｉ）このクラス間スケジューリング設定部２４に記憶された設定内容に基づいて、クラスを選択し、その後に、（ｉｉ）その選択さえれたクラスのフローの中から、なるべく輻輳していない宛先行きのフローを選択するようにしている。

【０１５８】

その際に、理想的には、最も輻輳していない宛先行きのフローを選択すべきであるが、実装上の都合により必ずしも厳密に輻輳度の順にフローを選ぶことが簡単にできるとは限らない。しかし、輻輳している宛先行きのフローよりも輻輳していない宛先行きのフローを選ぶことが多いようにすれば、それだけでも十分な効果が得られる。なるべく簡単に実装する方法として、フローを輻輳度により例えば３段階に大きく分類し、最も輻輳していない段階のフローから選択することが考えられる。このようにすると、同じ段階の中では必ずしも輻輳度の順に厳密にフローが選択されるわけではないが、全体的にはおよそ輻輳度の低い順にフローが選択されることになる。

【０１５９】

フローが選択されると、（ｉｉｉ　）そのフローのパケットキュー２２の先頭からパケットを取り出してスイッチング部へ転送する。

【０１６０】

以降、上記の（ｉ）～（ｉｉｉ　）の手順を繰り返し実行する。

【０１６１】

なお、本実施形態を実装する場合には、輻輳度の有効期間を定めると好ましい。例えば、入側転送部において、あるフローの輻輳度が決定されてから一定時間新たな輻輳状況が通知されなかった場合には、その古い輻輳度を無効にする。

【０１６２】

このように本実施形態によれば、入側転送部が輻輳していない宛先へのパケットを優先して転送するので、輻輳して流れが悪くなっているフローの影響で、輻輳していない宛先行きのパケットの流れも悪くなるという問題を容易に解決することができる。

【０１６３】

また、本実施形態では、パケットを転送すると、そのパケットの転送によってそのフローの最新の輻輳状況が通知される。したがって、この輻輳状況によってそのフローの選択優先度が動的に最適な値に変化する利点がある。

【０１６４】

ところで、この第２の実施形態（スケジューリングにおけるパケット選択）は、第１の実施形態（衝突時におけるパケット選択）と組み合わせて実施することが可能である。

【０１６５】
この場合には、さらに次のような効果を得ることができる。すなわち、入側転送部の内部に、輻輳している宛先行きパケットしかない場合には輻輳している宛先へ連続してパケットを出力してしまうが、これらのパケットは廃棄され再送される確率が高い。しかし、第２の実施形態に第１の実施形態を組み合わせた場合、パケットに輻輳度に応じた優先度が付与されるため、出力してもかまわない。このような構成は、輻輳している宛先へパケットを転送しないように抑制する制御よりも、簡単に実現できるという利点がある。転送を抑制する制御では、過剰な抑制のためにスイッチング部の利用効率が低下することがないように厳密な制御を行なう必要があるが、それと比較して本実施形態は簡単な制御でよいわけである。

【０１６６】
なお、第２の実施形態に第１の実施形態を組み合わせる場合、輻輳度テーブル（および通知された輻輳状況から輻輳度を求める部分）は、パケットに優先度を付与するための部分と、スケジューリングのための部分とで、互いに独立した構成にしてもよいし、１つの輻輳度テーブルに共通化してもよい。

【０１６７】
（第３の実施形態）
第１の実施形態、第２の実施形態では、輻輳していない出力ポート等行きのパケットの流れが、それとの関係を持たない他の部分の輻輳の影響を受けないようにする点を中心に説明してきた。

【０１６８】
第３の実施形態では、輻輳の原因となるパケットを送信しているユーザに対してパケット送信速度を下げてもらうように通知するための機構について説明する。

【０１６９】
第３の実施形態は、第１の実施形態、第２の実施形態、第１の実施形態と第２の実施形態を組み合わせたもののいずれにも適用可能であり、また独立して実施することも可能であるが（スイッチング部の構成がどのようなものであっても適用可能である）、第３の実施形態と第１の実施形態および／または第２の実施形態とを組み合わせて実施することで、総合的に輻輳対策に優れたパケットスイッチを構築することができる。

【０１７０】
図４に、本実施形態に係る多段スイッチ型のパケットスイッチの構成例を示す。また、併せて、該パケットスイッチの動作概要を示す。

【０１７１】
ユーザに対して輻輳を通知する機能は、従来、単段スイッチ型パケットスイッチにおいては提案または実用化されている。しかし、この機能を、ポート数の多い多段スイッチ型などのパケットスイッチに効果的に適用することは難しかった。まず、ユーザに対して輻輳を通知する方法として、従来の単段スイッチ型パケットスイッチで用いられている主な３つの方法について説明する。

【０１７２】
輻輳を通知する第１番目の方法は、パケットの廃棄による方法である。例えば、インターネットで広く使用されているプロトコルＴＣＰでは、パケットの廃棄を検出すると、輻輳を緩和するために転送速度（正確にはウィンドウサイズ）を小さくする制御が働く。ＴＣＰから転送されているパケットを扱うパケットスイッチの廃棄制御では、輻輳がひどくなるに従ってパケットの廃棄確率を増やすＲＥＤ（Ｒａｎｄｏｍ　Ｅａｒｌｙ　Ｄｅｔｅｃｔｉｏｎ）と呼ばれている方法が優れていると言われている。パケットの廃棄によって輻輳を通知する場合、パケットキューの末尾のパケットを廃棄するよりも先頭のパケットを廃棄した方が、ユーザに輻輳を早く通知でき、輻輳を重くなる前に緩和できる可能性が高いことが知られている。

【０１７３】
輻輳を通知する第２番目の方法は、輻輳経験通知による方法である。パケットヘッダの一

部に輻輳経験フラグ領域を設置し、輻輳を経験したパケットは、その領域がマーキングされる。これによりユーザに輻輳を通知する。この方法はＡＴＭ通信の場合にはＥＦＣＩ（Ｅｘｐｌｉｃｉｔ　Ｆｏｒｗａｒｄ　Ｃｏｎｇｅｓｔｉｏｎ　Ｉｎｄｉｃａｔｉｏｎ）と呼ばれている。また、インターネットではＥＣＮ（Ｅｘｐｌｉｃｉｔ　Ｃｏｎｇｅｓｔｉｏｎ　Ｎｏｔｉｆｉｃａｔｉｏｎ）と呼ばれ、現在標準化の議論中である。輻輳経験通知の場合にもパケットを廃棄する方法と同様に、パケットキューの末尾のパケットにマーキングするよりはパケットキューの先頭のパケットにマーキングした方が、ユーザに輻輳を早く通知することができることが知られている。

【０１７４】

輻輳を通知する第３番目の方法は、ＢＥＣＮ（Ｂａｃｋｗａｒｄ　Ｅｘｐｌｉｃｉｔ　Ｃｏｎｇｅｓｔｉｏｎ　Ｎｏｔｉｆｉｃａｔｉｏｎ）による方法である。ＡＴＭ通信では、輻輳が発生するとスイッチはそれを通知するためにＢＥＣＮセルと呼ばれる特殊なセル（ＡＴＭ通信ではパケットのことをセルと呼ぶ）を生成し、送信側ユーザに向けて転送することが考えられている。ＡＴＭのＡＢＲサービスカテゴリの資源管理セルへ情報を載せる処理も送信側ユーザに向けて直接情報を転送する意味で同じである。インターネットにもＳｏｕｒｃｅ　Ｑｕｅｎｃｈと呼ばれる同様の制御メッセージが存在する。ＢＥＣＮの場合には、輻輳を下流側のユーザに対してではなく上流側のユーザに直接通知するため、特にパケットキューの先頭にあるパケットに対して何かを行なう必要はない。しかし、どのユーザに対してＢＥＣＮを送出するかを決定する場合、輻輳している地点つまりパケットスイッチの出側で判定することが、公平な輻輳通知を実現する上では好ましい。というのは、輻輳通知は、輻輳地点毎に、最も輻輳の原因となっているユーザに対して、行なうべきだからである。

【０１７５】

以上、主な３つの輻輳通知方法について述べた。これらに共通することは、なるべくキューの出口で輻輳制御を行なった方がよいということである。これを、これまで説明してきたような各実施形態のパケットスイッチに当てはめて考えると、入側転送部よりも出側転送部においてこれらの輻輳制御を行なうことが望ましいと考えられる。この点について、従来の多段スイッチ型パケットスイッチでは、出側転送部が、入側転送部の内部の輻輳状況を正確に把握することはできず、パケットスイッチ全体の輻輳状況を把握することは難しかった。単段スイッチ型パケットスイッチのようなポート数の少ないパケットスイッチの場合であれば、各出側転送部に全ての入側転送部の輻輳状況をモニタするような結線を行なうことは可能であるが、多段スイッチ型パケットスイッチのようにポート数が多い場合には困難である。

【０１７６】

これを解決する方法として、図４に例示するように、パケットを入側転送部２からスイッチング部４を経由して所望の出側転送部６へ転送するパケットスイッチにおいて、パケットとともに、入側転送部２の内部の輻輳状況を出側転送部６へ通知し、その入側転送部２の内部の輻輳状況から出側転送部６は総合的な輻輳状況を判定して、輻輳制御を行なう方法が考えられる。

【０１７７】

図４の例では、入側転送部［Ｄ］の入側輻輳状況として、フロー毎のパケットキューの長さ（例えば２８９とする）をパケットとともに出側転送部［Ｂ］へ転送する。出側転送部［Ｂ］の内部の輻輳状況をそのフローのパケットキュー長（例えば１２とする）で観測するものとすれば、出側転送部［Ｂ］は、入側のパケットキュー長と出側のパケットキュー長とを合計することにより、パケットスイッチ内に存在するそのフローの待ちパケット量（＝２８９＋１２＝３０１）を把握できる。

【０１７８】

ユーザへの輻輳通知方法がＲＥＤであれば、その待ちパケット量をもとにパケットキューの先頭のパケットを廃棄する確率を決定すればよい。ＥＣＮであれば、その待ちパケット量をしきい値と比較してマーキングするかどうかを決定すればよい。Ｓｏｕｒｃｅ　Ｑｕ ｅｎｃｈであれば、その待ちパケット量をしきい値と比較してＳｏｕｒｃｅ　Ｑｕｅｎｃｈメッセージを転送するかどうかを決定すればよい。

【０１７９】

本実施形態は、輻輳状況としてフロー毎のキュー長だけでなく、クラス毎、転送部毎の輻輳状況なども考慮し、その結果得られる情報を利用しても、有効に作用する。

【０１８０】

従来、ポート数が多いパケットスイッチでは、すべての入側転送部の輻輳状況を出側で観測することは不可能であったが、本実施形態によれば、出側転送部６において入側転送部２の輻輳状況を知ることができ、出側転送部６の内部の輻輳状況と合わせて、パケットが転送される経路の輻輳状況を総合的に把握することができる。ユーザへ輻輳を通知するための制御はパケットキューの出口にて行なうことが好ましく、本実施形態はその要求を満たしており、またポート数が大きな多段スイッチ型パケットスイッチに容易に適用できる利点がある。

【０１８１】

（第４の実施形態）

第１～第３の実施形態では、本発明を、パケットを入側転送部から出側転送部へ転送する際にスイッチング部内で衝突が発生するようなパケットスイッチに適用した場合を中心に説明してきたが（第１の実施形態では、衝突が発生した際に、その宛先の輻輳状況に応じた優先度により、優先して転送すべきパケットを決定する）、本発明は、内部で衝突が発生しないパケットスイッチ、例えば、クロスバ型のスイッチング部を持つパケットスイッチにも適用可能である。言い換えると、第１の実施形態（または第１の実施形態に第２および／または第３の実施形態を組み合わせた実施形態）に相当するような制御の仕組みを利用した、クロスバ型等のパケットスイッチを実現することができる。

【０１８２】

一般に、パケットスイッチは、高速で入側転送部と出側転送部とを接続するパターンを計算しなければならない。例えば、図６のような多段スイッチではパケットが自律分散的にルーティングされ衝突に勝ち残ったパケットの経路を得ることによって入側転送部と出側転送部とを接続するパターンを計算していることになる。一方、クロスバ型スイッチなどでは、このように自律分散的なアプローチではなく、集中して接続パターンを計算するエンジンがスイッチに１箇所存在する。一般に、このエンジン計算量は、スイッチのポート数をＮとするとN^2のオーダとなり、ポート数が多くなるとともに計算が困難になる問題点が知られている。本発明は、この接続パターン計算エンジンのアルゴリズムに適用することが可能である。

【０１８３】

図５に、本実施形態に係るクロスバスイッチ型のパケットスイッチの構成例を示す。また、併せて、該パケットスイッチの動作概要を示す。図５では、第１の実施形態に対応するものを示してある。

【０１８４】

なお、本実施形態のパケットスイッチの基本的な全体構成は、第１の実施形態のように、図６と同様で、入力ポートから入力されたパケットは、入側転送部２からクロスバスイッチ３を経由して所望の出側転送部６へ転送され、そこから出力ポートへ出力される。また、パケットスイッチの入力ポート数、出力ポート数が任意である点、１つの入側転送部や出側転送部に複数の入力ポートや出力ポートが対応しても構わない点、スイッチング部の内部の網構成も任意である点、その網を構成する１つの単位スイッチの入力数、出力数も任意である点は、第１の実施形態と同様である。

【０１８５】

なお、優先度付与部、輻輳状況観測部、輻輳度テーブル、パケット選択部などの機能は第１の実施形態と同様であるが、これらは、接続パターン計算エンジン７のアルゴリズムの中に作り込むものとする。ただし、例えば輻輳状況観測部で輻輳状況の観測のためにカウンタを用いるとする場合などのように、個別のハードウェアを用いる場合には、当該部分

については、これまでの実施形態で示した箇所に実装されることになる。なお、スイッチング部内で行う制御・処理のように接続パターン計算エンジン７のアルゴリズムの中に作り込むことを必須とするもの以外は、これまでの実施形態で示したように実装するようにしてもよい。

【０１８６】

さて、入側転送部２のデータ転送のリクエストを受けて、接続パターン計算エンジン７がクロスバスイッチ３の接続パターンを計算する。また、図５に例示するように、出側転送部６から輻輳状況に関する情報が接続パターン計算エンジン７へ送られる。

【０１８７】

この接続パターン計算エンジン７は、その内部でソフトウェア的に各入側転送部２のデータ転送のリクエストをトーナメントさせ（すなわち、シミュレーションを行い）、勝ち残ったリクエストに対応するデータを交換するようにクロスバスイッチ３へ接続パターンを伝える。このトーナメント表は、予め定めたトポロジ（例えば図６で示されるようなトポロジ）を用いればよい。

【０１８８】

図６で示されるようなトポロジを用いるものとした場合、ランダム網のトポロジを用いてリクエストをランダムな順序に並べ替え、ルーティング網のトポロジを用いてトーナメントを行う。

【０１８９】

トーナメントの競合では、輻輳度テーブル９２の輻輳度を考慮して設定もしくは変更された優先度によって、勝ち残るリクエストを決定する。

【０１９０】

このようにして出側転送部６へ向けて転送を許可するリクエストの集合が求まると、接続パターンが計算され、その接続パターンによって設定されたクロスバスイッチ３によってデータが交換される。

【０１９１】

この方式では、スイッチのポート数をＮとするとＮ・ｌｏｇＮのオーダで出側転送部６へ転送してよいリクエストを求めることができ、比較的高速に計算することができるとともに、出側転送部６の輻輳状況を考慮しているので公平性も確保できるという利点がある。

【０１９２】

なお、第１の実施形態と第２の実施形態および／または第３の実施形態を組み合わせた構成も、同様にして実現可能である。

【０１９３】

なお、以上の各実施形態では、本発明を多段スイッチ型パケットスイッチに適用する場合を中心に説明したが、本発明は単段スイッチ型パケットスイッチにも適用可能である。また、本発明は、パケット網全体の輻輳に対処する方法としても利用可能である。

【０１９４】

なお、各実施形態の構成のうちハードウェアを必須としない各部分は、ソフトウェアとしても実現可能である。

【０１９５】

また、本実施形態の制御機能は、コンピュータに所定の手段を実行させるための（あるいはコンピュータを所定の手段として機能させるための、あるいはコンピュータに所定の機能を実現させるための）プログラムを記録したコンピュータ読取り可能な記録媒体としても実施することもできる。

【０１９６】

本発明は、上述した実施の形態に限定されるものではなく、その技術的範囲において種々変形して実施することができる。

【０１９７】

【発明の効果】

本発明によれば、転送先の輻輳状況に応じた優先度をパケットに付与し、この優先度を考

慮してパケット衝突時の処理を行うようにしたので、各パケットの転送先の輻輳状況に応じたパケット転送制御を行うことができる。例えば、輻輳していないポート行きのパケットに相対的に高い優先度を付与するようにすれば、輻輳しているポート行きのパケットの影響により、輻輳していないポート行きのパケットの流れが妨げられないようにすることができる。

【０１９８】

また、本発明によれば、廃棄されたパケットにより高い優先度を付与するようにすれば、廃棄されたパケットを再送する場合にそれを廃棄されにくくすることができる。

【０１９９】

また、本発明によれば、１つのデータの先頭部分に対応するパケットを後続のパケットより低い優先度で転送先に到達させることにより、複数のパケットに分割した１つデータを一塊りに転送させることができるようにすることができる。

【図面の簡単な説明】

【図１】本発明の第１の実施形態に係るパケットスイッチの構成例を示す図

【図２】同実施形態に係るパケットスイッチの他の構成例を示す図

【図３】本発明の第２の実施形態に係るパケットスイッチの要部構成の一例を示す図

【図４】本発明の第３の実施形態に係るパケットスイッチの構成例を示す図

【図５】本発明の第４の実施形態に係るパケットスイッチの構成例を示す図

【図６】多段スイッチ型パケットスイッチの構成例を示す図

【図７】ルーティング網内衝突の一例を示す図

【図８】出側転送部入口輻輳の一例を示す図

【図９】出側転送部内輻輳の一例を示す図

【符号の説明】

２…入側転送部

３…クロスバスイッチ

４…スイッチング部

６…出側転送部

７…接続パターン計算エンジン

１２，５２，９２…輻輳度テーブル

２１…優先度付与部

２２…パケットキュー

２４…クラス間スケジューリング設定部

２６…転送セル選択部

３２…輻輳度テーブル

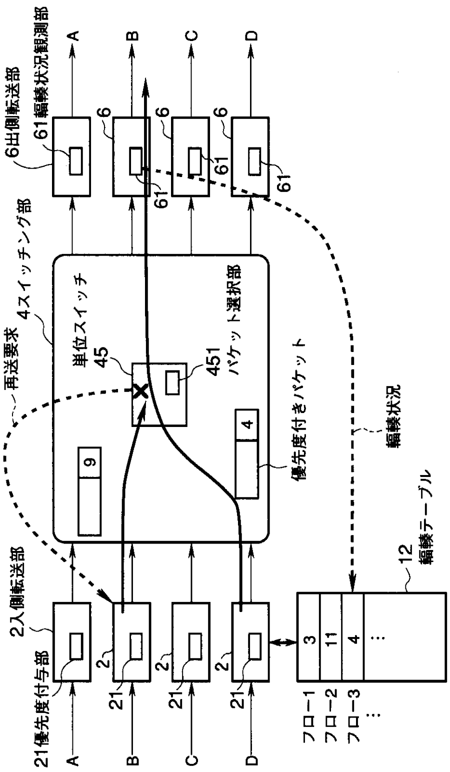
４２…パケット優先度変更部

４５…単位スイッチ

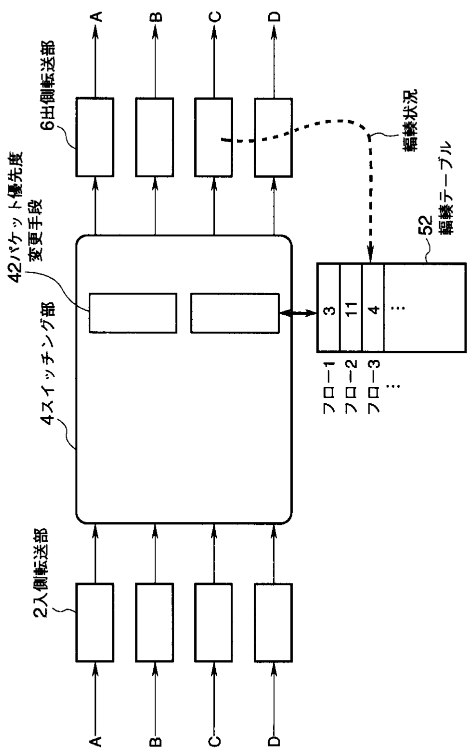
６１…輻輳状況観測部

４５１…パケット選択部

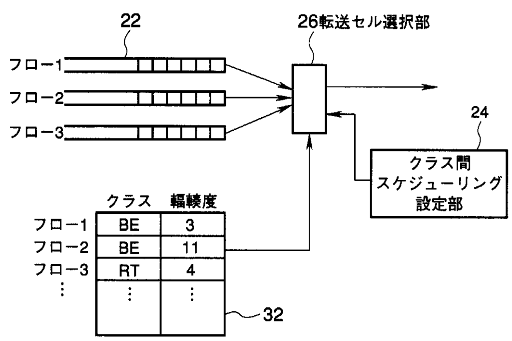
【図１】


6出側転送部
61輻輳状況観測部
4スイッチング部
再送要求
単位スイッチ
45
451
パケット選択部
9
4
優先度付きパケット
輻輳状況
12
輻輳テーブル
2入側転送部
21優先度付与部
フロー−1
フロー−2
フロー−3
3
11
4
A
B
C
D

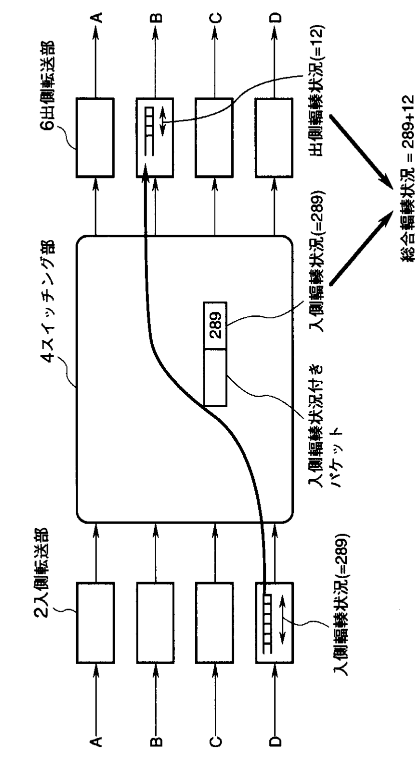
【図２】


6出側転送部
42パケット優先度変更手段
4スイッチング部
2入側転送部
輻輳状況
52
輻輳テーブル
フロー−1
フロー−2
フロー−3
3
11
4
A
B
C
D

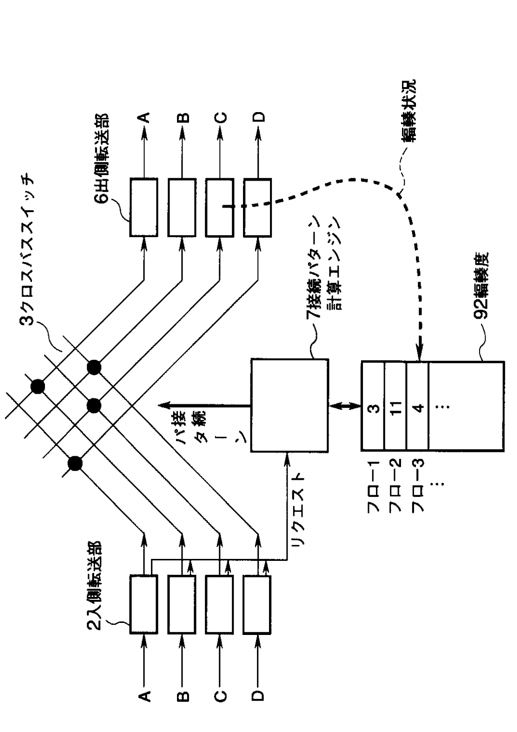
【図３】


22
フロー−1
フロー−2
フロー−3
26転送セル選択部
24
クラス間
スケジューリング
設定部
クラス
輻輳度
フロー−1 BE 3
フロー−2 BE 11
フロー−3 RT 4
32

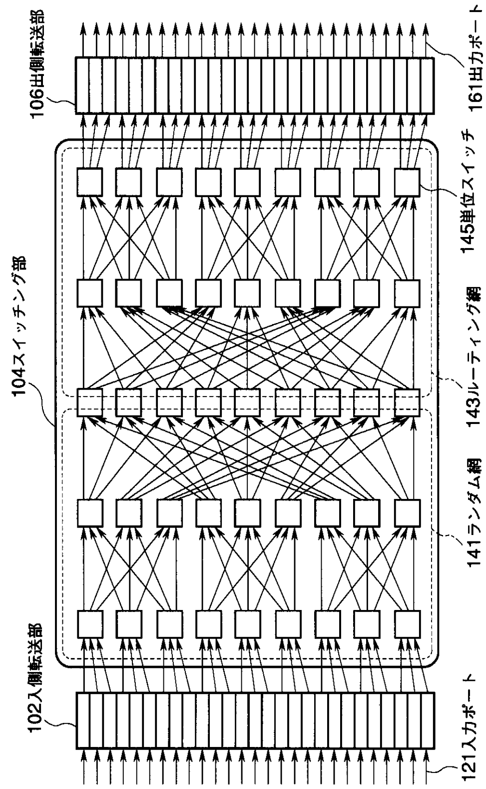
【図４】


6出側転送部
出側輻輳状況(=12)
総合輻輳状況＝289+12
入側輻輳状況(=289)
4スイッチング部
289
入側輻輳状況付き
パケット
2入側転送部
入側輻輳状況(=289)
A
B
C
D

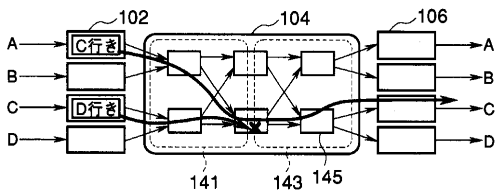
【図５】


3クロスバスイッチ
6出側転送部
2入側転送部
7接続パターン計算エンジン
92輻輳度
輻輳状況
リクエスト
接続パターン
フロー−1
フロー−2
フロー−3
3
11
4
A
B
C
D

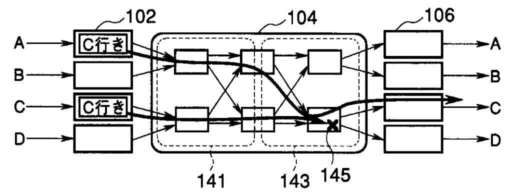
【図６】


106出側転送部
161出力ポート
145単位スイッチ
104スイッチング部
143ルーティング網
141ランダム網
102入側転送部
121入力ポート

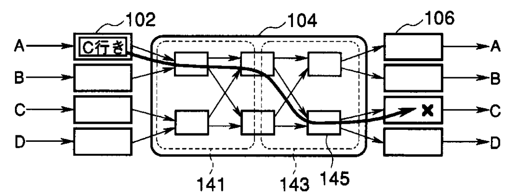
【図７】


102
104
106
A
B
C
D
C行き
D行き
141
143
145

【図８】


102
104
106
A
B
C
D
C行き
C行き
141
143
145

【図９】


102
104
106
A
B
C
D
C行き
141
143
145

フロントページの続き

(74)代理人　100070437
弁理士　河井　将次
(72)発明者　下條　義満
神奈川県川崎市幸区小向東芝町１番地　株式会社東芝研究開発センター内
(72)発明者　中北　英明
神奈川県川崎市幸区小向東芝町１番地　株式会社東芝研究開発センター内

審査官　石井　研一

(56)参考文献　特開平０９－２０５４４１（ＪＰ，Ａ）
特開平０８－２８８９６５（ＪＰ，Ａ）
特開平０７－１７７１７９（ＪＰ，Ａ）
特開平０７－１０７０９５（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.^{7}，ＤＢ名)
H04L 12/56