(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第5593517号
(P5593517)

(45)発行日　平成26年9月24日(2014.9.24)　(24)登録日　平成26年8月15日(2014.8.15)

(51)Int.Cl.　　F I
H04L 12/28 (2006.01)　　H04L 12/28　2OOD
H04L 12/825 (2013.01)　　H04L 12/825

請求項の数 12　(全 27 頁)

(21)出願番号　特願2011-156939(P2011-156939)
(22)出願日　平成23年7月15日(2011.7.15)
(65)公開番号　特開2013-26680(P2013-26680A)
(43)公開日　平成25年2月4日(2013.2.4)
審査請求日　平成25年11月5日(2013.11.5)

(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　100114236
弁理士　藤井　正弘
(74)代理人　100075513
弁理士　後藤　政喜
(74)代理人　100120260
弁理士　飯田　雅昭
(72)発明者　早川　仁
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　對馬　雄次
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
最終頁に続く

(54)【発明の名称】ネットワーク装置及び送信フレームの制御方法

(57)【特許請求の範囲】
【請求項１】
ノードとのリンクを流れる複数のトラフィックのフレームの送受信を行い、送信フレームの送信を行う送信部と受信フレームの受信を行う受信部とを備えるポートと、
前記ポートを介して受信した受信フレームが一時停止の指令のときには、前記ポートからの送信フレームの送信を所定時間まで一時的に停止し、
前記複数のトラフィックから選択した少なくとも１以上のトラフィックの送信帯域を制限し、
前記所定時間を経過した後、前記一時停止の指令を含む受信フレームを受信する間隔が拡大したときは、前記選択したトラフィックの送信帯域の制限を強化する制御部と、
を備えたことを特徴とするネットワーク装置。
【請求項２】
請求項１に記載のネットワーク装置であって、
前記制御部は、
所定時間を経過した後、前記一時停止の指令を含む受信フレームを受信する間隔が拡大しないときは、前記選択したトラフィックの送信帯域の制限を緩和することを特徴とするネットワーク装置。
【請求項３】
請求項１に記載のネットワーク装置であって、
前記制御部は、

予め設定した順序に従って前記送信帯域を制限する少なくとも１以上のトラフィックを選択することを特徴とするネットワーク装置。

【請求項４】

請求項１に記載のネットワーク装置であって、

前記一時停止の指令は、ＰＡＵＳＥフレームであることを特徴とするネットワーク装置。

【請求項５】

請求項１に記載のネットワーク装置であって、

前記制御部は、

前記一時停止の指令を受信する間隔が拡大した場合には、当該選択したトラフィックを、輻輳を引き起こすトラフィックとして推定することを特徴とするネットワーク装置。

【請求項６】

プロセッサと、メモリと、ノードとのリンクを流れる複数のトラフィックのフレームの送受信を行い、送信フレームの送信を行う送信部と受信フレームの受信を行う受信部とを備えるポートと、を備えたネットワーク装置で送信するフレームの輻輳を抑制する送信フレームの制御方法であって、

前記ポートを介して受信した前記受信フレームが一時停止の指令のときには、前記ポートからの前記送信フレームの送信を所定時間まで一時的に停止する第１のステップと、

前記複数のトラフィックから選択した少なくとも１以上のトラフィックの送信帯域を制限する第２のステップと、

前記所定時間を経過した後、前記一時停止の指令を含む受信フレームを受信する間隔が拡大したときは、前記選択したトラフィックの送信帯域の制限を強化する第３のステップと、

を含むことを特徴とする送信フレームの制御方法。

【請求項７】

請求項６に記載の送信フレームの制御方法であって、

前記第３のステップは、

所定時間を経過した後、前記一時停止の指令を含む受信フレームを受信する間隔が拡大しないときは、前記選択したトラフィックの送信帯域の制限を緩和することを特徴とする送信フレームの制御方法。

【請求項８】

請求項６に記載の送信フレームの制御方法であって、

前記第２のステップは、

予め設定した順序に従って前記送信帯域を制限する少なくとも１以上のトラフィックを選択することを特徴とする送信フレームの制御方法。

【請求項９】

請求項６に記載の送信フレームの制御方法であって、

前記一時停止の指令は、ＰＡＵＳＥフレームであることを特徴とする送信フレームの制御方法。

【請求項１０】

請求項６に記載の送信フレームの制御方法であって、

前記第３のステップは、

前記一時停止の指令を受信する間隔が拡大した場合には、当該選択したトラフィックを、輻輳を引き起こすトラフィックとして推定することを特徴とする送信フレームの制御方法。

【請求項１１】

ノードとのリンクを流れる複数のトラフィックのフレームの送受信を行い、送信フレームの送信を行う送信部と受信フレームの受信を行う受信部とを備えるポートと、

前記ポートを介して受信した受信フレームが一時停止の指令のときには、前記ポートからの送信フレームの送信を所定時間まで一時的に停止し、

前記複数のトラフィックから選択した少なくとも１以上のトラフィックの送信帯域を制限し、
前記所定時間を経過した後、前記一時停止の指令を含む受信フレームを受信する間隔が拡大したときは、前記選択したトラフィック以外のトラフィックの送信を促進する制御部と、
を備えたことを特徴とするネットワーク装置。
【請求項１２】
ノードとのリンクを流れる複数のトラフィックのフレームの送受信を行い、送信フレームの送信を行う送信部と受信フレームの受信を行う受信部とを備えるポートと、
前記ポートを介して受信した受信フレームが一時停止の指令のときには、前記ポートからの送信フレームの送信を所定時間まで一時的に停止し、
前記複数のトラフィックから選択した少なくとも１以上のトラフィックの送信帯域を制限し、
前記所定時間を経過した後、前記一時停止の指令を含む受信フレームを受信する間隔が拡大しないときは、前記選択したトラフィックの送信帯域の制限を緩和する制御部と、
を備えたことを特徴とするネットワーク装置。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、コンピュータネットワークでフレームを転送する装置及び方法に関し、特に、トラフィックの輻輳を制御する技術の改良に関する。
【背景技術】
【０００２】
コンピュータネットワークを実現するイーサネット（登録商標）は、データをフレームという単位に分割して送受信し、中継装置等の通信系路上でのフレーム廃棄を許容する標準規格となっている。フレーム廃棄によるデータ損失を許容しないアプリケーションに対しては、拡張規格としてＩＥＥＥ８０２．３ｘにフロー制御用のＰＡＵＳＥフレームが規定されている。フロー制御は送信側と受信側を結ぶリンクに対して行われる。リンクの受信側で、輻輳等の理由でフレームを正しく受信できないと想定される場合、リンクの受信側から送信側にＰＡＵＳＥフレームを送信する。ＰＡＵＳＥフレームを受信したリンクの送信側ではフレーム送信を一時停止する。これらリンクの送信側及び受信側での処理により、リンクのフロー制御が行われ、リンクの受信側でフレームを受信できないことによるフレーム廃棄を防ぐ。
【０００３】
一方、サーバの高密度化や、Local Area Network（ＬＡＮ）に用いられるイーサネットの高速化を背景として、これまでＬＡＮとは別に構築されていたストレージアクセスのためのStorage Area Network（ＳＡＮ）を、イーサネットを用いて構築してＬＡＮに統合し、ネットワークインタフェースやケーブル、ネットワーク機器を削減して運用の効率化をはかるＬＡＮ／ＳＡＮ統合の動きが加速している。また、サーバ仮想化の進展により、１台の物理サーバ上で複数の仮想マシンを稼動させ、１つのＬＡＮを複数のネットワークで共用する環境が一般化している。これらの環境では物理的に１本のイーサネットケーブル上に性質や要求の異なる複数のトラフィックを流す必要がある。これらのトラフィックを論理的に区別する手法としてＩＥＥＥ８０２．１Ｑに規定されるVirtual LAN （ＶＬＡＮ）が広く普及している。ＶＬＡＮでは各フレームにＶＬＡＮタグを付加し、ＶＬＡＮタグに含まれる優先度およびＶＬＡＮ　ＩＤでトラフィックを区別する。ＶＬＡＮを用いることで１本の物理的なリンクを論理的に分割して用いることが可能となる。
【０００４】
ＬＡＮ／ＳＡＮ統合におけるストレージアクセスのトラフィックはフレーム廃棄を許容しない。しかしながら、ＰＡＵＳＥによるフロー制御は物理的なリンク単位で行うため、ＶＬＡＮにより論理的に分割されたリンクのいずれかで輻輳が発生してＰＡＵＳＥフレー

ムが送信されて送信が一時停止されると、同じ物理的なリンクを共有する他の論理的なリンクを流れるトラフィックのフレーム送信も一時停止してしまう。すなわち、複数のトラフィックで１つのリンクを共有するため、一部トラフィックでの輻輳が他のトラフィックの通信を妨げるという問題が生じる。

【０００５】

これに対する従来技術として、論理的なリンク毎に独立したＰＡＵＳＥフレームを規定する手法が知られている（例えば、特許文献１および非特許文献１）。

【０００６】

特許文献１には「複数の入出力ポートから入力する伝送フレームの輻輳を検出する輻輳検出手段と、前記輻輳検出手段が輻輳を検出した入出力ポートにＰＡＵＳＥフレームを送出するＰＡＵＳＥフレーム送出手段とを備えたネットワークスイッチを有する輻輳制御システムにおいて、前記伝送フレームの代わりに仮想ＬＡＮ情報を付加した拡張伝送フレームと、前記ＰＡＵＳＥフレームの代わりに前記ＰＡＵＳＥフレームに仮想ＬＡＮ情報を付加した拡張ＰＡＵＳＥフレームとを使用し、前記ＰＡＵＳＥフレーム送出手段の代わりに前記拡張ＰＡＵＳＥフレームを前記入出力ポートに送出する拡張ＰＡＵＳＥフレーム送出手段を設けたことを特徴とする。」と記載されている。さらに「特に、前記仮想ＬＡＮ情報をＩＥＥＥ８０２．３ｘ規格のＶＬＡＮタグのＶＬＡＮ－ＩＤ値とすることを特徴とする。或いは、前記仮想ＬＡＮ情報をＩＥＥＥ８０２．３ｘ規格のＶＬＡＮタグのプライオリティ値とすることを特徴とする。」と記載されている。

【０００７】

また、非特許文献１では、ＩＥＥＥ８０２．１ＱｂｂとしてＶＬＡＮタグのプライオリティ毎のＰＡＵＳＥ指示を含む拡張したＰＡＵＳＥフレームおよび当該ＰＡＵＳＥフレームを用いた論理的なリンク毎のフロー制御を規定している。

【先行技術文献】

【特許文献】

【０００８】

【特許文献１】特開２００７－１７４１５２号公報

【非特許文献】

【０００９】

【非特許文献１】IEEE P802.1Qbb/D2.3, "IEEE Draft Standard for Local and Metropolitan Area Networks - Virtual Bridged Local Area Networks - Amendment: Priority-based Flow Control", Publication No.5570060, Issue date Sept. 9 2010

【発明の概要】

【発明が解決しようとする課題】

【００１０】

上記特許文献１および非特許文献１の処理では、リンクの送信側と受信側の双方がそれぞれ対応する必要があるため、既存の設備を全面的に置き換えまたは改造することが必要となる。

【００１１】

また、上記特許文献１では、ＩＥＥＥの標準規格として規定されているＰＡＵＳＥおよびＶＬＡＮを組み合わせたものであるが、これらを組み合わせた処理について、上記標準には規定されていないため、既存の装置にそのまま適用することはできず、また、上記特許文献１の技術を利用した装置も広範に利用可能な状況ではない。

【００１２】

一方、上記非特許文献１は現在策定中の標準規格であり、将来的に対応する装置が増えると考えられるが、非特許文献１の技術で扱うことが可能な論理的なリンクの数は８までという制限がある。これは仮想マシンの増加やトラフィックの優先度制御を考慮すると不十分であり、８を超えた数のトラフィックを流す場合は、１つの論理的なリンク内を複数のトラフィックが流れる状況が生じ、先に挙げた一部トラフィックでの輻輳が他のトラフィックの通信を妨げるという問題は解決できない。

【００１３】
したがって、上記従来技術では、標準規格に準拠し、技術導入時に必要な機器の変更が最小限で、フレーム廃棄を許容しない複数のトラフィックを扱い、かつ８を超える多数のトラフィックに対応可能な輻輳制御を行うことができないという課題があった。
【課題を解決するための手段】
【００１４】
本発明は、ノードとのリンクを流れる複数のトラフィックのフレームの送受信を行い、送信フレームの送信を行う送信部と受信フレームの受信を行う受信部とを備えるポートと、前記ポートを介して受信した受信フレームが一時停止の指令のときには、前記ポートからの送信フレームの送信を所定時間まで一時的に停止し、前記複数のトラフィックから選択した少なくとも１以上のトラフィックの送信帯域を制限し、前記所定時間を経過した後、前記一時停止の指令を含む受信フレームを受信する間隔が拡大したときは、前記選択したトラフィックの送信帯域の制限を強化する制御部と、を備える。
【発明の効果】
【００１５】
したがって、本発明によれば、既存のネットワーク機器をそのまま利用しながらフレームの廃棄を許容しない複数のトラフィックの輻輳制御が可能となる。
【図面の簡単な説明】
【００１６】
【図１】本発明の実施形態を示し、ネットワークシステムの構成の一例を示すブロック図である。
【図２】本発明の実施形態を示し、フレーム転送装置のネットワークシステムの構成の一例を示すブロック図である。
【図３】本発明の実施形態を示し、転送ポート管理表の一例を示す図である。
【図４】本発明の実施形態を示し、トラフィック設定表の一例を示す図である。
【図５】本発明の実施形態を示し、キュー設定表の一例を示す図である。
【図６】本発明の実施形態を示し、トラフィック履歴表の一例を示す図である。
【図７】本発明の実施形態を示し、帯域割当表の一例を示す図である。
【図８】本発明の実施形態を示し、輻輳トラフィック推定部における制限帯域計算処理の一例を示すフローチャートである。
【図９】本発明の実施形態を示し、輻輳トラフィック推定部における制限帯域補正処理の一例を示すフローチャートである。
【図１０Ａ】本発明の実施形態を示し、送信フレーム選択部における送信帯域計算処理の一例を示すフローチャートで、前半部である。
【図１０Ｂ】本発明の実施形態を示し、送信フレーム選択部における送信帯域計算処理の一例を示すフローチャートで、後半部である。
【発明を実施するための形態】
【００１７】
以下、本発明の一実施形態を添付図面に基づいて説明する。
【００１８】
図１は、本発明の実施形態を示し、ネットワークシステムの構成の一例を示すブロック図である。
【００１９】
＜システムおよびハードウェアの構成＞
図１は本実施形態におけるネットワークシステムの構成を説明するブロック図である。図１において、ネットワークシステムはフレーム転送装置１の入出力ポート１１２－１にケーブル１１５－１を介して接続されたノード１１７－１、および、入出力ポート１１２－ｎにケーブル１１５－ｎを介して接続されたノード１１７－ｍから構成される。なお、ノード１１７－１～１１７－ｍは物理計算機やネットワーク装置あるいはストレージ装置等で構成される。また、フレーム転送装置１は、ｎ個の入出力ポート１１２－１～１１２

－ｎを有する。以下では、入出力ポートの総称を１１２とし、ケーブルの総称を１１５とし、ノードの総称を１１７とする。

【００２０】

本実施形態におけるネットワークシステムはノード１１７－１とノード１１７－ｍの間でケーブル１１５－１およびケーブル１１５－ｎとフレーム転送装置１を介したイーサネット（登録商標）による通信を実現する。ノード１１７－１～ノード１１７－ｍはイーサネットに準拠した入出力ポートを有し、ＰＡＵＳＥフレームによるフロー制御をサポートしたネットワークノードであり、例えばネットワークインタフェースカードを装着した計算機や、ルータ装置、スイッチ装置等のネットワーク機器等である。ケーブル１１５はイーサネットによる通信に適合するケーブルである。フレーム転送装置１はイーサネットフレームを転送するブリッジの機能を持つネットワーク装置である。

【００２１】

イーサネットについてはＩＥＥＥ　８０２．３に規定されており、ブリッジ機能はＩＥＥＥ　８０２．１ＤおよびＩＥＥＥ　８０２．１Ｑに規定されている。本実施形態においては、フレーム転送装置１は、ブリッジ機能に加えて、イーサネットに関して、ＰＡＵＳＥフレームによるフロー制御、特にＩＥＥＥ　８０２．３ｂｄ及びＩＥＥＥ　８０２．１Ｑｂｂに規定されるＰＦＣ（Priority-based Flow Control）を対象として、ノード１１７－１～ノード１１７－ｍの間で交換されるイーサネットフレームに関して、ＭＡＣ（Media Access Control）アドレスやＶＬＡＮ　ＩＤ（Virtual Local Area Network Identifier）、ＴＣＰ（Transmission Control Protocol）ポート番号等で区別される論理的なトラフィックに対して輻輳制御機能を提供する。ただし、本発明はイーサネットに限定されるものではなく、例えばＦｉｂｒｅＣｈａｎｎｅｌにおけるＢｕｆｆｅｒ－ｔｏ－Ｂｕｆｆｅｒクレジット等、受信側の状態について送信側に通知するフロー制御の機能を持ったプロトコルであれば、本実施形態で説明する輻輳制御方法を適用可能である。なお、本実施形態において、ノードは１１７－１～１１７－ｍのｍ台の場合で説明を行うが、ノードの個数は任意であり、例えば、より多くのノードを用意してケーブルによりフレーム転送装置１とそれぞれ接続し、より多くのノード間でフレーム転送装置１を介した通信をできるようにしてもよい。

【００２２】

以下、他の図や図１の他の部分においても構成要素の数が限定されないことを同様の点線によって表現するものとする。また、本実施形態ではフレーム転送装置１およびノード１１７－１および１１７－ｍは物理的にそれぞれ異なる装置として実施した場合を説明するが、本発明の実施形態はこれに限定されるものではない。例えば、仮想マシンを稼動させる仮想化ソフトウェアを用いて物理的に同一装置内で複数の仮想マシンにより実施してもよい。仮想化ソフトウェアによる仮想化の実施ではケーブル１１５－１～ケーブル１１５－ｍは共有メモリやメモリ間コピー等に置き換えられる。仮想化ソフトウェアを用いた仮想マシンによる実施により、柔軟な接続関係の変更や、ハードウェアの利用効率を高めること等が可能となる。

【００２３】

続いて、図１を用いてフレーム転送装置１内部のハードウェア構成を説明する。フレーム転送装置１は、演算処理装置１０１、記憶装置１０２、入出力装置１０３、ネットワークインタフェース１０４－１～ネットワークインタフェース１０４－ｎより構成され、共有バス１０６を通して互いに接続されている。本実施形態において、フレーム転送装置１はネットワークインタフェース１０４－１～１０４－ｎがそれぞれケーブル１１５－１およびケーブル１１５－ｎに接続され、ノード１１７－１～ノード１１７－ｍとイーサネットフレームを交換して、ノード１１７－１とノード１１７－ｍの間の通信を中継する。

【００２４】

演算処理装置１０１はCentral Processing Unit（ＣＰＵ）に代表される演算処理装置であり、記憶装置１０２上に展開されたプログラム（ソフトウェア）を実行する。なお、演算処理装置１０１を物理的または論理的に複数のＣＰＵにより構成してよい。記憶装置

１０２は演算処理装置１０１によって実行されるプログラムならびに当該プログラムが実行されることによって作成されるデータまたは読み込まれるデータおよび設定等を格納する。

【００２５】

記憶装置１０２は例えばメモリ、ハードディスク、もしくは光ディスク等のデータやプログラムを保持可能な記憶媒体を含む装置（遠隔地に設置され、ネットワークインタフェース１０４－１もしくは１０４－ｎ、入出力装置１０３等を通して通信される装置を含む）、またはこれらを組み合わせたもので構成される。記憶装置１０２に格納されるプログラムには、全体制御プログラム１０７（図1においては、プログラムを「ＰＧＭ」と記述、以下同様）、ポート制御プログラム１０８－１～１０８－ｎが含まれる。また、記憶装置に格納されるデータの一例としてはキュー部１１０、転送ポート管理表３００、トラフィック設定表４００、キュー設定表５００、トラフィック履歴表６００、帯域割当表７００が含まれる。これらの詳細については後述するが、以下の説明においてデータの格納場所が明示されていないデータの保存・保持・記憶・記録・格納等の書き込みや参照・取り出し等の読み出しは記憶装置１０２に対して行うものとする。

【００２６】

上記記憶装置１０２のプログラムはフレーム転送装置１を制御するための機能に関して論理的に区別したものであり、例えばすべてを１つのプログラムとして構成して処理させてもよいし、細分化して複数のスレッドやプロセス等に分離して実行したりしてもよい。なお、記憶装置１０２にはフレーム転送装置１を制御するためのプログラムにより作成された他のデータが格納されてもよい。本実施形態では記憶装置１０２にフレーム転送装置１を制御するためのプログラムが展開され、演算処理装置１０１により実行される。

【００２７】

入出力装置１０３は、フレーム転送装置１に対して情報の入出力を行う装置である。入出力装置１０３は、例えば、スイッチ、キーボード、マウス、マイクロホン、ビデオカメラ、ディスプレイ、またはスピーカー等の機器、またはフレーム転送装置１にはこれらの機器に接続可能なインタフェースを備えて、フレーム転送装置１にこれらの機器を接続して機能させる形態をとってもよい。また、入出力装置１０３には信号ケーブルまたは電波及び赤外線等の無線を介して行うシリアル通信などの通信によるものも含まれる。

【００２８】

フレーム転送装置１は、入出力装置１０３によって、フレーム転送装置１のユーザまたは管理者からの指示を受けたり、結果を出力したりすることが可能となる。ネットワークインタフェース１０４－１および１０４－ｎは、イーサネットフレームの送信および受信を行う入出力ポート１１２－１～１１２－ｎを備え、演算処理装置１０１の指示により、ポート制御プログラム１０８－１および１０８－ｎによってそれぞれ制御され、他の機器とイーサネットフレームによる通信を行う。以下、これらの入出力ポート１１２と当該入出力ポート１１２を制御するポート制御プログラム１０８をまとめてポートと呼ぶ。

【００２９】

また、共有バス１０６はフレーム転送装置１の各構成要素間の通信を行うためのものであるが、本発明は共有バスに限定されるものではない。各構成要素間で必要な通信を行えるようになっていれば、共有バスを用いる以外の方法によって接続されてもよい。例えば、各構成要素間を直接接続することのよって、要素間の接続を最適化し、処理に必要な消費電力を減らしたり、処理効率を上げたりすることが可能となる。

【００３０】

＜システムおよびハードウェアの変形例＞

なお、本発明において、フレーム転送装置１の構成要素の個数は図１に限定されるものではなく、例えば入出力装置１０３を２つ用意して冗長化したり、機能分担したりしてもよい。また、本発明は、記憶装置１０２に格納したプログラムに格納したプログラムによって実現される機能のうち、一部または全部の機能をハードウェアとして実装してもよい。ハードウェアとして実装することにより、例えば処理の高速化や低消費電力化が可能と

なる。また、仮想マシンのように、当該ハードウェアが備える機能をプログラムとして実現してもよい。ハードウェアが備える機能をプログラムとして実現することによって、例えば、設置スペースを削減したり、管理を単純化したりできる。また、プログラムによって実現される機能の構成は後述する構成に限るものではなく、複数の機能が統合される構成、または１つの機能が複数に分割される構成でもよい。また、各機能によって実行される処理の順序についても、後述する順序に限るものではなく、処理の依存関係が許すならば、並列で同時に実行したり、順序を入れ替えて実行したりしてもよい。例えば、並列実行することによって処理時間を短縮したり、順序を入れ替えることによって待ち時間を減らしたりすることが可能となる。

【００３１】

＜ソフトウェアの構成＞

図２はフレーム転送装置１におけるソフトウェア構成を説明するブロック図である。なお、図において、四角はフレーム転送装置１で実行されるプログラムによって実現される処理ブロックを表す。また、処理ブロック間の矢印は、矢印の方向に情報または指示・指令が伝えられることを示す。処理ブロック間で伝えられる情報または指示・指令の内容は以下で説明する。

【００３２】

フレーム転送装置１におけるソフトウェアは全体制御プログラム１０７、ポート制御プログラム１０８－１～１０８－ｎ（以下、制御プログラム１０８－１～１０８－ｎのうちのいずれかを表す場合は「ポート制御プログラム１０８」と記述する。複数ある他のコンポーネントについても同様）、キュー１１０、入出力ポート１１２－１～１１２－ｎの送受信機能から構成される。なお、入出力ポート１１２－１～１１２－ｎの送受信機能はハードウェアで構成してもよい。フレーム転送装置１全体は、入出力ポート１１２－１にてイーサネットフレームを受信して、キュー１１０に格納し、その内容をポート制御プログラム１０８－１で解析し、解析結果から全体制御プログラム１０７にて当該イーサネットフレームを転送すべきポートを決定し、キュー１１０に格納している当該イーサネットフレームを転送すべきと決定されたポートの入出力ポート１１２－ｎから出力することによりイーサネットブリッジとして機能する。

【００３３】

全体制御プログラム１０７は設定受付部２０７、輻輳検出部２０８、ＰＡＵＳＥ送信ポート・時間決定部２０９、トラフィック設定部２１０および格納キュー選択部２１１、トラフィック履歴管理部２１２より構成され、複数のポートにまたがる可能性のある処理およびフレーム転送装置１全体の制御を行う。

【００３４】

ポート制御プログラム１０８－１～１０８－ｎは、割当帯域管理部２１３、輻輳トラフィック推定部（輻輳推定部）２１４、送信フレーム選択部２１５、送信停止タイマー（送信停止部）２１６、ＰＡＵＳＥフレーム作成部２１７、キュー制御部２１８、入出力制御部２１９およびフレーム解析部２２０より構成され、当該ポートに関連する入出力ポート１１２－１の制御およびキュー１１０の制御を行う。キュー１１０は論理的に分割されたキューである論理キュー２２１－１～２２１－ｉにより構成され、入出力ポート１１２－１～１１２－ｎから入出力されるイーサネットフレームを一時的に保持する。

【００３５】

なお、論理キュー２２１－１～２２１－ｉは、１つのイーサネットフレームについて、イーサネットフレームの内容や入出力されるポート等に基づいて、後述する格納キュー選択部２１１が分類したトラフィックに対応する。もちろん、論理キュー２１１の数は図２に示すような３に限定されるものではなく、例えば論理キューの数を１としてすべてのトラフィックのイーサネットフレームを格納してもよいし、論理キューの数を３より大きくしてより細かくトラフィックの種類毎などに分類して格納してもよい。また、論理キュー２２１－１～２２１－ｉの総称を論理キュー２２１とする。

【００３６】

また、格納キュー選択部２１１が行うトラフィックの分類は、１つのイーサネットフレームが複数のトラフィックに同時に分類される処理でもよく、その場合、１つのイーサネットフレームが複数の論理キュー２２１に格納されることとなる。この場合、記憶装置１０２上の表現としては、複数の論理キュー２２１に同一のイーサネットフレームをコピーして記憶装置１０２の制御を簡素化してもよい。あるいは、複数の論理キュー２２１で同一の記憶装置１０２を参照するようにしてメモリを節約してもよい。

【００３７】

１つのイーサネットフレームが複数のトラフィックに同時に分類されることで、例えばポートに接続されたノードで複数の仮想マシンが稼動する場合や、フレーム転送装置１のポートにブリッジが接続され、当該ブリッジに複数のノードが接続された場合にマルチキャストフレームの転送を簡素化できる。

【００３８】

入出力ポート１１２－１～１１２－ｎはフレーム送信部２２４およびフレーム受信部２２５から構成され、当該ポートに接続されたノード１１７とイーサネットフレームを送受信する。以下の説明では、当該ポートにおけるノードとの接続をリンクと呼び、リンクを通してイーサネットフレームが当該ノードと送受信できる状態にあるものとして説明する。

【００３９】

また、本実施形態では、トラフィックはＶＬＡＮ　ＩＤや優先度情報、送信元や宛先等の通信経路等により区別される、ネットワークを流れる情報を示す。そして、イーサネットフレームまたはフレームは、トラフィックに流すデータの単位を表す。

【００４０】

＜ソフトウェアの構成：事前準備＞

続いて、図２の各構成要素間で送受信される支持・指令および情報について、フレーム転送装置１によるイーサネットフレームの転送を例に説明する。まず、イーサネットフレームの転送に先立って、設定受付部２０７は入出力装置１０３またはネットワークインタフェース１０４－１～１０４－ｎを介してフレーム転送装置１のユーザまたは管理者から入力された情報にしたがって、記憶装置１０２に各種の設定を格納する。

【００４１】

具体的には、設定受付部２０７へ入力された情報にしたがって、トラフィック設定部２１０からトラフィック設定表４００と、キュー設定表５００の内容を設定し、設定受付部２０７へ入力された情報とトラフィック設定部２１０のキュー設定表５００の内容にしたがって、割当帯域管理部２１３を通して帯域割当管理表７００の内容を格納する。

【００４２】

これらはフレーム転送装置１の製造時に予め記憶装置１０２に格納しておいてもよい。なお、トラフィック設定表４００、キュー設定表５００、帯域割当管理表７００を含め、以下の説明において用いる表の具体的な内容については後ほど説明する。キュー制御部２１８はキュー設定表５００に従ってキュー１１０に論理キュー２２１の作成・削除等の初期設定を指示し、論理キュー２２１を使用できるように準備する。

【００４３】

＜ソフトウェアの構成：受信時の処理＞

以下、まず、フレーム転送装置１で行われるイーサネットフレーム受信時の処理を説明する。イーサネットフレームはフレーム受信部２２５で受信されると、イーサネットフレームの情報をフレーム解析部２２０が解析し、フレーム解析部２２０がＰＡＵＳＥフレームでない場合は転送すべきフレームとして格納キュー選択部２１１に解析結果を通知する。

【００４４】

フレーム解析部２２０が格納キュー選択部２１１に通知する解析結果は、送信先ＭＡＣアドレス、ＶＬＡＮ　ＩＤ、Ｐｒｉｏｒｉｔｙ等の、格納キュー選択部２１１がトラフィックとして分類するために必要な情報である。これらの情報は、送信するイーサネットフ

レームに関連づけられた情報である。

【００４５】

格納キュー選択部２１１では、フレーム解析部２２０より通知された解析結果とトラフィック設定部２１０のトラフィック設定表４００を用いて当該イーサネットフレームのトラフィックＩＤ４０５（図４参照）を決定し、当該トラフィックＩＤ４０５と当該イーサネットフレームが入力されたポートを表すポートＩＤ３０３（図３参照）およびトラフィック設定部２１０のキュー設定表５００（図５参照）から当該イーサネットフレームを格納する論理キューのＩＤ５０３を決定する。そして、格納キュー選択部２１１がフレーム解析部２２０より通知された解析結果と転送ポート管理表３００を用いて当該イーサネットフレームを転送するポートＩＤ３０３を決定する。

【００４６】

格納キュー選択部２１１は決定されたポートのキュー制御部２１８と入出力制御部２１９に対してフレーム受信部２２５の当該イーサネットフレームを当該論理キューのＩＤ５０３に格納するよう指示する。当該指示に従って、入出力制御部２１９は当該イーサネットフレームを当該論理キューに伝え、キュー制御部２１８は当該イーサネットフレームを当該論理キュー２２１に格納する。

【００４７】

また、格納キュー選択部２１１はトラフィック履歴管理部２１２に、当該イーサネットフレームのトラフィックＩＤとポートＩＤとフレームサイズを通知する。トラフィック履歴管理部２１２では格納キュー選択部２１１から通知された情報を現時刻とともにトラフィック履歴表６００（図６参照）に保存する。トラフィック履歴表６００への保存時に古い時刻のエントリを削除する等によりトラフィック履歴表６００が使用するメモリサイズを節約してもよい。

【００４８】

以上により、リンクで受信されたイーサネットフレームが転送すべきフレームの場合、当該イーサネットフレームが適切な論理キューに格納される。このとき、格納キュー選択部２１１は、送信先ＭＡＣアドレス、ＶＬＡＮ　ＩＤ、Ｐｒｉｏｒｉｔｙ等のフレームの種類に応じてトラフィックを分類し、分類結果に応じた論理キュー２２１へ当該フレームを格納する。

【００４９】

＜ソフトウェアの構成：送信時の処理＞

続いて、各ポートの論理キュー２２１－１～２２１－ｉからイーサネットフレームを送信する処理について述べる。割当帯域管理部２１３は帯域割当管理表７００（図７参照）の内容を輻輳トラフィック推定部２１４および送信フレーム選択部２１５に通知する。

【００５０】

送信フレーム選択部２１５は、割当帯域管理部２１３から通知された帯域割当管理表７００の内容および輻輳トラフィック推定部２１４が通知する制限帯域と、キュー制御部２１８から通知された各論理キュー２２１の使用量と、送信停止タイマー２１６のタイマー値にしたがって、論理キュー（例えば２２１－１）のイーサネットフレームを送信する。送信フレーム選択部２１５による送信すべきイーサネットフレームの選択方法については送信帯域計算処理として図１０Ａ、図１０Ｂで後述する。

【００５１】

送信フレーム選択部２１５による論理キュー２２１－１等からのイーサネットフレームの送信は、具体的には、格納キュー選択部２１１がキュー制御部２１８と入出力制御部２１９に対して論理キュー２２１－１のイーサネットフレームを読み出してリンクに送信するよう指示する。当該指示に従って、キュー制御部２１８は当該イーサネットフレームを当該論理キューから取り出してフレーム送信部２２４に送るよう論理キュー２２１－１に指示し、入出力制御部２１９はフレーム送信部２２４に論理キュー２２１－１から送られた当該イーサネットフレームを入出力ポート１１２－１に接続されたリンクに対して送信するように指示する。

【００５２】
また、入出力制御部２１９は送信したイーサネットフレームの情報を、具体的には属する論理キュー２２１－１～２２１－ｉのキューＩＤと当該イーサネットフレームのフレームサイズを輻輳トラフィック推定部２１４に通知する。輻輳トラフィック推定部２１４は、フレーム解析部２２０から通知されたＰｒｉｏｒｉｔｙ毎のＰＡＵＳＥフレーム受信状況、割当帯域管理部２１３から通知された帯域割当管理表の内容、キュー制御部２１８から通知された論理キュー２２１－１～２２１－ｉの使用量および入出力制御部２１９から通知された送信するイーサネットフレームの情報と、送信停止タイマー２１６を参照して得たＰｒｉｏｒｉｔｙ毎の変数であるＰＡＵＳＥ状態に基づいて、各論理キュー２２１からイーサネットフレームの制限帯域を計算して送信フレーム選択部２１５に通知する。
【００５３】
輻輳トラフィック推定部２１４における制限帯域の計算方法については制限帯域計算処理として図８で後述する。以上により、論理キューからイーサネットフレームがリンクに送信される。
【００５４】
なお、イーサネットフレームの転送（ブリッジ）については、ＩＥＥＥ　８０２．１ＤおよびＩＥＥＥ　８０２．１Ｑとして標準化されており、これらの技術を適宜用いることができる。
【００５５】
＜ソフトウェアの構成：ＰＡＵＳＥ受信＞
続いて、図２におけるＰＡＵＳＥフレームの扱いについて述べる。ＰＡＵＳＥフレームを受信した場合および送信する場合のフレーム転送装置１における処理について説明する。まず、ＰＡＵＳＥフレームを受信した時のフレーム転送装置１が実行する処理について説明する。イーサネットフレーム受信時は、フレーム解析部２２０の解析により、当該イーサネットフレームがＰＡＵＳＥフレームであることが検出される。
【００５６】
ＰＡＵＳＥフレームは転送すべきフレームではないのでフレーム解析部２２０は格納キュー選択部２１１に通知せず、ＰＡＵＳＥフレームの内容にしたがって各Ｐｒｉｏｒｉｔｙで一時停止すべき時間を計算し、送信停止タイマー２１６に指示する。そして、フレーム解析部２２０は輻輳トラフィック推定部２１４に対してＰｒｉｏｒｉｔｙ毎にＰＡＵＳＥフレームの受信状況を通知する。
【００５７】
具体的には、フレーム解析部２２０は輻輳トラフィック推定部２１４にＰＡＵＳＥフレームにより指示された時間を通知する。送信停止タイマー２１６は格納キュー選択部２１１より指示されたＰｒｉｏｒｉｔｙ毎の時間にタイマー値をセットする。送信停止タイマー２１６は一定時間毎に設定された各タイマー値を０になるまで減じる。送信停止タイマー２１６は、ＰＡＵＳＥフレームにより指示された時間がＰＡＵＳＥ状態の解除を意味する０の場合は０をセットする。送信フレーム選択部２１５は送信停止タイマー２１６を参照し、特定のＰｒｉｏｒｉｔｙの送信タイマーが０より大きい場合は当該Ｐｒｉｏｒｉｔｙに属するイーサネットフレームを含む論理キュー２２１からの送信を停止する。
【００５８】
以上により、ＰＡＵＳＥフレームに指定された時間だけ、当該リンクに対するフレーム転送装置１からのフレーム送信が入出力ポート１１２毎に一時停止され、ＰＡＵＳＥフレームによる輻輳制御が実現される。
【００５９】
＜ソフトウェアの構成：ＰＡＵＳＥ送信＞
次に、フレーム転送装置１がＰＡＵＳＥフレームをリンクへ送信する場合の処理について説明する。ＰＡＵＳＥフレームの送信は当該リンクが輻輳状態となり、ノード１１７から送信されたフレームをフレーム転送装置１で正しく受信できない恐れがある場合に行う。本実施形態においては、フレーム転送装置１のポート制御プログラム１０８が受信する

フレームを格納する論理キュー２２１の使用量が停止閾値（例えば、確保したキュー容量の８割）を超えた場合に、ノード１１７から送信されたフレームを受信できない恐れがあると判定して、当該論理キュー２２１に格納されるフレームを送信しているリンクに送信を停止させるＰＡＵＳＥフレームを送信する。

【００６０】

フレーム転送装置１は、当該論理キュー２２１の使用量が再開閾値（例えば、確保したキュー容量の５割）を下回った場合に受信できない恐れが消滅したと判定して、当該リンクへの送信を再開させるＰＡＵＳＥ解除フレームを接続されているリンクに送信する。

【００６１】

図２において、キュー制御部２１８は論理キュー（例えば、２２１－１）の使用量を輻輳検出部２０８に通知する。輻輳検出部２０８はキュー制御部２１８から通知された論理キューの使用量が停止閾値を越えたもしくは再開閾値を下回ったことを検知すると、ＰＡＵＳＥ送信ポート・時間決定部２０９に通知する。

【００６２】

ＰＡＵＳＥ送信ポート・時間決定部２０９は、トラフィック履歴管理部２１２のトラフィック履歴表６００とトラフィック設定部２１０のトラフィック設定表４００、キュー設定表５００を参照して、当該論理キュー２２１に格納されるイーサネットフレームが入力されている（１または複数の）ポートを検索する。ＰＡＵＳＥ送信ポート・時間決定部２０９は検索したポートのＰＡＵＳＥフレーム作成部２１７に、当該論理キュー２２１のＰｒｉｏｒｉｔｙと、当該Ｐｒｉｏｒｉｔｙでの送信を一時停止させる時間を通知する。なお、当該論理キュー２２１に複数のＰｒｉｏｒｉｔｙに属するフレームを格納するなど、当該論理キューのＰｒｉｏｒｉｔｙが一意に定まらない場合は、複数のＰｒｉｏｒｉｔｙに関して同様の検索をＰＡＵＳＥ送信ポート・時間決定部２０９が行う。また、前記検索を行う代わりにすべてのポートや当該論理キューの所属するＶＬＡＮに参加しているポートを対象として、検索に必要な時間を節約してもよい。

【００６３】

送信を一時停止させる時間については、停止閾値を越えた場合は最大値である６５５３５（単位は５１２ビットを送信する時間）とし、再開閾値を下回った場合は０とする。本発明では、停止閾値を越えた場合の送信を一時停止させる時間が前記の６５５３５に限るものではなく、例えば論理キュー２２１の空き容量や当該論理キューからの送出量、演算処理装置１０１の使用量等に基づいてさらに細かく時間設定をして送信再開ＰＡＵＳＥフレーム（またはＰＡＵＳＥ解除フレーム）の送出を抑制するなどしてもよい。

【００６４】

ＰＡＵＳＥフレーム作成部２１７は、ＰＡＵＳＥ送信ポート・時間決定部２０９からの指示を、当該Ｐｒｉｏｒｉｔｙの時間として埋め込んだＰＡＵＳＥフレームを作成し、フレーム送信部２２４に通知する。

【００６５】

ＰＡＵＳＥフレーム作成部２１７では、ＰＡＵＳＥ送信ポート・時間決定部２０９から指示された時間をＰｒｉｏｒｉｔｙ毎に記憶し、送信停止タイマー２１６と同様に０になるまで一定時間毎に前記指示され時間から１を減じ、各Ｐｒｉｏｒｉｔｙでの送信を一時停止させる時間の現在の値を参照可能にする。また、ＰＡＵＳＥフレーム作成部２１７では、ＰＡＵＳＥ送信ポート・時間決定部２０９から指示されたＰｒｉｏｒｉｔｙ以外のＰｒｉｏｒｉｔｙに関しても一時停止させる時間を埋め込んだＰＡＵＳＥフレームを作成できるようにする。

【００６６】

入出力ポート１１２のフレーム送信部２２４では、ポート制御プログラム１０８のＰＡＵＳＥフレーム作成部２１７から通知されたＰＡＵＳＥフレームを、接続されたリンクに対して送出する。なお、ＰＡＵＳＥフレームについて直接フレーム送信部２２４による送出は行わず、いったんＰＡＵＳＥフレーム用に用意した論理キュー２２１に格納し、送信フレーム選択部２１５で他の論理キュー２２１より優先してフレームを送出するという方

法にし、送出されるフレームの管理を一元化してもよい。

【００６７】

以上により、当該リンクでのフレーム転送装置１によるフレームの受信が一時停止され、ＰＡＵＳＥフレームによる輻輳制御が実現される。

【００６８】

なお、ＰＡＵＳＥフレームのフォーマットや一時停止時間等については、ＩＥＥＥ　８０２．３ｘおよびＩＥＥＥ　８０２．３ｂｄ、ＩＥＥＥ　８０２．１Ｑｂｂとして標準化されており、本実施形態による説明があれば当業者によるＰＡＵＳＥ機能の実施は困難ではない。

【００６９】

＜ソフトウェアの構成のバリエーション＞

図２においては、受信フレームをそのまま論理キュー２２１に格納する形での実施を説明したが、本発明はこれに限るものではなく、例えば入力バッファ、出力バッファのようにバッファを記憶装置１０２に設置してイーサネットフレームを格納し、格納キュー選択部２１１による論理キューの選択や送信フレーム選択部２１５による送信フレーム選択の処理を平準化してもよい。また、フレーム解析部２２０においてＰＡＵＳＥフレームでない場合は転送すべきフレームとしたが、例えばＰＡＵＳＥフレーム以外のフレームであっても転送する必要のないＬＬＤＰフレームの取り扱いや、ＶＬＡＮタグの付加または除去等、一般にイーサネットスイッチやＬ２スイッチと呼ばれるネットワーク機器で行われる拡張を施してもよい。また、本実施形態ではトラフィックの分類に、送信元ＭＡＣアドレス、送信先ＭＡＣアドレス、ＶＬＡＮ　ＩＤ、Ｐｒｉｏｒｉｔｙを用いたが、本発明におけるトラフィックの分類はこれに限定されるものではなく、例えば、入力されたポートや、イーサネットフレームに含まれる情報であるＥｔｈｅｒＴｙｐｅ、ＩＰ（Ｉｎｔｅｒｎｅｔ　Ｐｒｏｔｏｃｏｌ）アドレス、ＴＣＰのポート番号等を加え、各情報単独または複数の情報を組み合わせてトラフィックを決定する形で用いて行ってもよい。たとえばＴＣＰのポート番号やＩＰアドレスの情報を用いることで、特定アプリケーションのトラフィックを区別して輻輳制御を行うことができるようになる。

【００７０】

＜表の説明＞

続いて、図２の説明に用いた表である、転送ポート管理表３００、トラフィック設定表４００、キュー設定表５００、トラフィック履歴表６００、帯域割当管理表７００の具体的な内容について例を用いて説明する。なお、これらの表は記憶装置１０２に保存される。

【００７１】

図３に転送ポート管理表の例を示す。転送ポート管理表３００は、ＭＡＣアドレス３０１、ＶＬＡＮＩＤ３０２、ポートＩＤ３０３を含むエントリで構成される。ＭＡＣアドレス３０１はイーサネットにおけるエンティティを識別するためのアドレスである。ＶＬＡＮ　ＩＤ３０２は、ＩＥＥＥ　８０２．１Ｑにて規定される、イーサネットフレームにおけるＶＬＡＮの識別子で、異なるＶＬＡＮ　ＩＤを持つフレームを別のＶＬＡＮとして扱うために用いられるものである。ポートＩＤ３０３は複数あるポート（入出力ポートおよび対応するポート制御プログラムを表す）のうちの１つを識別するための識別子である。

【００７２】

格納キュー選択部２１１では転送ポート管理表３００を用いて、フレーム解析部２２０から通知された解析結果のうちの送信先ＭＡＣアドレスがＭＡＣアドレス３０１で、同じくＶＬＡＮ　ＩＤがＶＬＡＮＩＤ３０２と両方ともに一致するエントリを検索し、該当するエントリのポートＩＤ３０３を対象として決定する。格納キュー選択部２１１では転送ポート管理表３００に前記一致するエントリが複数あった場合は複数のポートを対象として決定する。一方、前記一致するエントリが見つからなかった場合はＶＬＡＮ　ＩＤが一致する全ポートを対象とする。

【００７３】

転送ポート管理表３００はＩＥＥＥ　８０２．１ＱにおけるＦｉｌｔｅｒｉｎｇ　Ｄａｔａｂａｓｅに対応する。転送ポート管理表３００はフレームの情報と照合して当該フレームを転送すべきポートを決定する機能を持てば、必ずしも図３に挙げた情報に基づかなくてもかまわない。例えば、入力ポートを識別するポートＩＤと転送先のポートＩＤ３０３の対応表といった構成も考えられ、ＶＬＡＮ　ＩＤを省いた構成（ＩＥＥＥ　８０２．１Ｄに規定されるブリッジ機能）としてもよい。また、転送ポート管理表３００は入力されたフレームの送信元ＭＡＣアドレスとＶＬＡＮ　ＩＤおよびポートＩＤの組み合わせをエントリとして追加し、一定時間使用しなかったエントリを削除するＭＡＣ学習により保守する。なお、本発明においてはこれに限るものではなく、例えば、固定的なエントリを作成し、ＭＡＣ学習の処理を省いてもよい。

【００７４】

図４にトラフィック設定表の例を示す。トラフィック設定表４００は、送信元アドレス４０１、送信先アドレス４０２、ＶＬＡＮＩＤ４０３、Ｐｒｉｏｒｉｔｙ４０４、トラフィックＩＤ４０５を含むエントリで構成される。送信元アドレス４０１は送信元のＭＡＣアドレス、送信先アドレス４０２は送信先のＭＡＣアドレスを格納する。ＶＬＡＮＩＤ４０３およびＰｒｉｏｒｉｔｙ４０４はＩＥＥＥ　８０２．１Ｑ　で規定されるＶＬＡＮ　ＩＤおよびＰＣＰの値をそれぞれ表す。トラフィックＩＤ４０５は論理的に区別して扱うトラフィックの識別子を格納する。

【００７５】

格納キュー選択部２１１ではトラフィック設定表４００を用いて、フレーム解析部２２０から通知された解析結果のうちの対象フレームの送信先ＭＡＣアドレスと送信先アドレス４０２を、送信元ＭＡＣアドレスと送信元アドレス４０１を、ＶＬＡＮ　ＩＤとＶＬＡＮＩＤ４０３を、ＰｒｉｏｒｉｔｙとＰｒｉｏｒｉｔｙ４０４をそれぞれ照合し、すべて一致したエントリのトラフィックＩＤ４０５を取得し、転送ポート管理表３００およびキュー設定表５００と併せて対象フレームの格納すべき論理キューを決定する。

【００７６】

したがって、トラフィック設定表４００ではトラフィックの分類に用いる情報と、一致した場合のトラフィックＩＤを対応付ける役割を担う。そのため、先に例示したようにトラフィックの分類に用いる情報は例えばＥｔｈｅｒＴｙｐｅなどでもよく、本発明ではトラフィック設定表４００に含まれる情報は図４に限定されるものではない。一致するエントリが見つからず対象トラフィックが確定しない場合は、あらかじめ定められたトラフィックＩＤ（例えば０）を用いる。

【００７７】

図５にキュー設定表５００の例を示す。キュー設定表５００はトラフィックＩＤ５０１、ポートＩＤ５０２、キューＩＤ５０３を含むエントリで構成される。トラフィックＩＤ５０１は図４に示したトラフィック設定表４００のトラフィックＩＤ４０５と同様、論理的に区別して扱うトラフィックの識別子を格納する。ポートＩＤ５０２は図３の転送ポート管理表３００のポートＩＤ３０３と同様にポートを区別するための識別子を格納する。キューＩＤ５０３はキュー１１０における論理キュー２２１－１～２２１－ｉを区別するための識別子を表す。格納キュー選択部２１１では転送ポート管理表３００およびトラフィック設定表４００の結果をキュー設定表５００と照合し、トラフィックＩＤ５０１とトラフィックＩＤ４０５、および、ポートＩＤ５０２とポートＩＤ３０３の双方が方一致するエントリを検索し、対象フレームを格納する論理キューを検索されたエントリのキューＩＤ５０３から決定する。

【００７８】

格納キュー選択部２１１ではフレーム解析部２２０から通知された解析結果と、転送ポート管理表３００、トラフィック設定表４００およびキュー設定表５００を用いて、対象フレームを格納する論理キュー２２１を決定する。そのため、これら３種の表は必ずしも別に保存される必要はなく、フレーム解析部２２０から通知された解析結果から対象フレームを格納する論理キュー２２１を決定できればよい。例えば、転送ポート管理表３００

、トラフィック設定表４００およびキュー設定表５００を１つの表としてまとめることでデータ構造を簡素化することができる。

【００７９】

図６にトラフィック履歴表の例を示す。トラフィック履歴表６００は、トラフィックＩＤ６０１、ポートＩＤ６０２、時刻６０３、フレームサイズ６０４を含むエントリで構成される。トラフィックＩＤ６０１は図４、図５のトラフィックＩＤ４０５、トラフィックＩＤ５０１と同様、論理的に区別して扱うトラフィックの識別子を格納する。ポートＩＤ６０２は図３、図５のポートＩＤ３０３およびポートＩＤ５０２と同様、ポートを区別するための識別子を格納する。時刻６０３は当該エントリが作成された時刻を表す値で、例えばある時点からの経過秒数を用いることができるがこれに限定されるものではない。例えば、一定周期でカウントアップされるカウンタを用いることもでき、この場合、ハードウェア実装の簡素化が期待できる。フレームサイズ６０４は対象とするイーサネットフレームのフレームサイズを表す。フレームサイズ６０４として格納する情報はフレームサイズに限定するものではなく、大まかなフレーム送信量を把握できればよい。例えば、フレームサイズ６０４として１を格納して処理を単純化してもよい。

【００８０】

ＰＡＵＳＥ送信ポート・時間決定部２０９では、トラフィック履歴表６００を用いて対象トラフィックが入力されているポートを決定する。そのため、トラフィック履歴表６００は対象トラフィックの入力ポートを調査可能であればよく、図６に示す形式に限定されるものではなく、例えば、各トラフィックＩＤと入力のあったポート番号といった形式でもよい。形式を変更することでトラフィック履歴表６００に必要なメモリ量を削減したりアクセス時間を削減したりできる。またエントリは追加のみでなく、例えば時刻６０３を現時刻と比較して一定時間経過したもの、フレームサイズ６０４が一定値より小さいもの等を削除したり参照時に除外したりしてもよい。このような対象エントリの限定により、例えば論理キュー２２１の輻輳に与える影響の大きなポートからのフレーム入力を抑制して他の通信に関しては制限せずに通すといった柔軟な輻輳制御がイーサネットフレーム受信時に可能となる。

【００８１】

図７に帯域割当管理表の例を示す。帯域割当管理表７００はキューＩＤ７０１と割当帯域７０２を含むエントリで構成される。キューＩＤ７０１は図５のキューＩＤ５０３と同様、キュー１１０における論理キュー２２１を区別するための識別子を格納する。割当帯域７０２は当該論理キュー２２１からリンクに対して送信する保証帯域を格納する。割当帯域７０２は絶対的な送信帯域のほか、送信帯域の比のような形で指定してもよい。帯域割当管理表７００を用いることで各論理キュー２２１からリンクに対して送信する際に保証すべき帯域を参照することが可能となる。

【００８２】

以上に説明した表のデータ構造は必ずしも表である必要はなく、例えばハッシュや２分木等でもよい。ハッシュを用いることで一致するエントリの検索時間短縮が期待できる。また、それぞれの照合は完全一致に限るものではなく、例えば部分一致等、与えられた情報に基づいて対象とするエントリを決定できればよい。例えば部分一致を用いることにより、表に必要なメモリ容量を削減することができる。

【００８３】

続いて、輻輳トラフィック推定部２１４における制限帯域計算処理および送信フレーム選択部２１５における送信帯域計算処理についてフローチャートを用いて説明する。なお以下の説明において、変数に格納された値は当該変数名により表すこととする。

【００８４】

＜輻輳トラフィック推定部の処理＞

図８は輻輳トラフィック推定部２１４で行う制限帯域計算処理の一例を説明するフローチャートである。輻輳トラフィック推定部２１４は、論理キュー２２１から読み出されてリンクへ送信されるイーサネットフレームについて、キューＩＤによりトラフィックを区

別したうえで各トラフィックの送信帯域を計算する。そして、輻輳を引き起こすトラフィックを推定し、当該トラフィックが輻輳の要因と推定される場合には当該トラフィックの送信帯域を減らす。

【００８５】

このため、輻輳トラフィック推定部２１４は、フレーム解析部２２０から通知されたＰＡＵＳＥフレーム受信状況と、割当帯域管理部２１３から通知された帯域割当管理表の内容と、キュー制御部２１８から通知された論理キューの使用量と、入出力制御部２１９から通知された送信したイーサネットフレームの情報と、送信停止タイマー２１６を参照して得たＰｒｉｏｒｉｔｙ毎のＰＡＵＳＥ状態（タイマー値が０なら非ＰＡＵＳＥ状態、それ以外ならＰＡＵＳＥ状態）を用いて、各トラフィックで送信可能な制限帯域を制限帯域計算処理により計算して送信フレーム選択部２１５に通知する。

【００８６】

制限帯域計算処理は一定時間毎（例えば０．０１秒毎）に繰り返して行う。なお、本発明では制限帯域計算処理実行の契機は一定周期に限るものではなく、例えば、ＰＡＵＳＥフレームの受信を制限帯域計算処理実行の契機としてもよいし、時間間隔をトラフィックの流量等に応じて変更してもよい。なお、以下の説明は、ＰＦＣの１つのＰｒｉｏｒｉｔｙに対する処理を対象としており、当該ＰｒｉｏｒｉｔｙのＰＡＵＳＥフレームにより当該Ｐｒｉｏｒｉｔｙに属する全トラフィックの送信が停止されるものとして説明する。これはＰＦＣではない一般のＰＡＵＳＥで当該リンクを流れる全トラフィックの送信が停止されるのと同様である。なお、ＰＦＣにおける複数のＰｒｉｏｒｉｔｙに対応するには、制限帯域計算処理をＰｒｉｏｒｉｔｙ毎に当該Ｐｒｉｏｒｉｔｙに属するトラフィックを対象として行う。以下の説明中で「トラフィック」は対象Ｐｒｉｏｒｉｔｙに属するトラフィックを、「キュー」は当該トラフィックのイーサネットフレームが格納されている論理キュー２２１をそれぞれ表す。また「保証帯域」は帯域割当管理表の内容である各トラフィックの割当帯域７０２を表す変数である。「累積送信量」は入出力制御部２１９から通知された送信済みのイーサネットフレームの情報を用いてトラフィック毎にこれまで転送したフレームサイズの合計値を保存する変数である。すなわち、各トラフィックの累積送信量は入出力制御部２１９から通知されるたびに加算していく。また変数「制限帯域」、「前回の制限帯域」の初期値は変数「保証帯域」の値とする。

【００８７】

輻輳トラフィック推定部２１４は、制限帯域計算処理が開始状態Ｓ８０１から開始されると、ステップＳ８０２に進む。

【００８８】

ステップＳ８０２では輻輳トラフィック推定部２１４が送信停止タイマー２１６を参照して取得したＰＡＵＳＥ状態であるか否かを判定し、真（ＰＡＵＳＥ中）であればステップＳ８０３に進み、偽ならステップＳ８０４に進む。ステップＳ８０３では輻輳トラフィック推定部２１４が変数「ＰＡＵＳＥ状態」を参照し、前回ＰＡＵＳＥ状態ではなく今回ＰＡＵＳＥ状態となった、すなわちＰＡＵＳＥが開始されたか否かを判定し、真ならステップＳ８０５に進み、偽であればステップＳ８１８に進む。

【００８９】

ステップＳ８０５では、輻輳トラフィック推定部２１４が変数「ＰＡＵＳＥ間隔」を変数「前回のＰＡＵＳＥ間隔」に保存し、変数「ＰＡＵＳＥ開始時刻」を変数「前回のＰＡＵＳＥ時刻」に保存し、変数「ＰＡＵＳＥ開始時刻」に現在の時刻を代入し、変数「ＰＡＵＳＥ間隔」にＰＡＵＳＥ開始時刻から前回のＰＡＵＳＥ時刻を減じた時間を代入し、ステップＳ８０６に進む。

【００９０】

ステップＳ８０６では、輻輳トラフィック推定部２１４が各トラフィックについて変数「累積転送量」の値を変数「前回の累積転送量」に代入し、変数「累積送信量」の値を変数「累積転送量」に代入し、ステップＳ８０７に進む。ステップＳ８０７では、輻輳トラフィック推定部２１４が「前回のＰＡＵＳＥ間隔」および「ＰＡＵＳＥ間隔」が有意であ

ることを判定し、真ならステップＳ８０８に進み、偽ならステップＳ８１７に進む。なお、ステップＳ８０７における「有意」は、値が、変数が未定義値でなく、かつ、一定値（例えば０．０５秒）より大きいこと表す。これは、ＰＡＵＳＥ間隔が初期状態であったり、誤差といえるほど小さかったりした場合に後述する制限帯域補正処理を行わないようにするためである。

【００９１】

ステップＳ８０８では、輻輳トラフィック推定部２１４が変数「帯域制限トラフィック数」と「帯域制限強化トラフィック数」にそれぞれ０を代入し、ステップＳ８０９に進む。ステップＳ８０９では、各トラフィックについて、ステップＳ８１０からステップＳ８１１までを繰り返す。

【００９２】

ステップＳ８１０では、輻輳トラフィック推定部２１４が後述する図９の制限帯域補正処理を実行して、ステップＳ８１１に進む。ステップＳ８１１では、輻輳トラフィック推定部２１４がステップＳ８０９からの繰り返し処理を各トラフィックについて終了したら、ステップＳ８１２に進む。

【００９３】

ステップＳ８１２では輻輳トラフィック推定部２１４が変数「帯域制限トラフィック数」が正であるか否かを判定し、真ならステップＳ８１３に進み、偽ならステップＳ８１７に進む。ステップＳ８１３では輻輳トラフィック推定部２１４が変数「帯域制限強化トラフィック数」が０であるか否かを判定し、真ならステップＳ８１４に進み、偽ならステップＳ８１７に進む。

【００９４】

ステップＳ８１４では、輻輳トラフィック推定部２１４がトラフィックを選択し、帯域制限を実施し、ステップＳ８１７に進む。このステップＳ８１４におけるトラフィック選択の例として、輻輳トラフィック推定部２１４が、変数「トラフィック番号」（初期値は０）を参照し、対応するトラフィックの変数「前回の送信量」が正なら当該トラフィックを選択し、そうでなければトラフィック番号に１を加えて次のトラフィックを調べる。そして、輻輳トラフィック推定部２１４は、トラフィック番号がトラフィックの数以上となった場合は０に戻す。本発明においてステップＳ８１４におけるトラフィック選択は前述の方法に限定されるものではなく、例えばランダムにトラフィックを選択するようにして処理を単純化してもよい。また、複数のトラフィックを順番に選択してもよい。さらに、例えば送信履歴に基づいて、直近にイーサネットフレームを多く送信しているトラフィックから優先して選択するようにしてもよい。また、ステップＳ８１４における帯域制限は、選択したトラフィックの変数「新しい制限帯域」に、例えば、

（２×保証帯域＋実効帯域）／３

を代入する。各トラフィックの実効帯域は後述の制限帯域補正処理で計算される。なお、ステップＳ８１４における「新しい制限帯域」の計算式は、制限帯域を保証帯域から実効帯域に近づけるためのものであり、本発明はこれに限るものではない。例えば、より保証帯域に近づけるようにして、帯域制限を緩やかにしてもよい。

【００９５】

次に、ステップＳ８０２の判定でＰＡＵＳＥ中でない場合のステップＳ８０４では、輻輳トラフィック推定部２１４が、現在の時刻と変数「ＰＡＵＳＥ開始時刻」の差が一定時間（例えば変数「ＰＡＵＳＥ間隔」の２倍）より大きいか否かを判定し、真ならステップＳ８１５に進み、偽ならステップＳ８１８に進む。

【００９６】

ステップＳ８１５では、輻輳トラフィック推定部２１４が帯域制限を緩和して、ステップＳ８１６に進む。ここで、ステップＳ８１５における帯域制限の緩和は、各トラフィックについて、「新しい制限帯域」として、例えば、

（前回の制限帯域＋保証帯域）／２

を代入する。なお、ステップＳ８１５における「新しい制限帯域」の計算式は、制限帯域

を保証帯域に近づけるものであり、本発明はこれに限定されるものではない。例えば、より保証帯域に近づけるようにして、帯域制限の緩和幅を大きくしてもよい。

【００９７】

次に、ステップＳ８１６では、輻輳トラフィック推定部２１４が変数「ＰＡＵＳＥ間隔」を大きくしてステップＳ８１７に進む。ここで、ステップＳ８１６はステップＳ８１５が実行される頻度を減らすために行うもので、帯域制限の緩和を緩やかにする効果がある。変数「ＰＡＵＳＥ間隔」を大きくする例としては、ＰＡＵＳＥ間隔に

（ＰＡＵＳＥ間隔×２）

を代入するなどが考えられるが、本発明は必ずしもこれに限定されるものではなく、例えば定数値を足すなどでもよい。また、ステップＳ８１６を省略して処理を簡素化してもよい。

【００９８】

次に、ステップＳ８１７では、輻輳トラフィック推定部２１４が各トラフィックについて、変数「前回の制限帯域」に変数「制限帯域」の値を代入し、変数「制限帯域」に「新しい制限帯域」の値を代入し、ステップＳ８１８に進む。ステップＳ８１８では輻輳トラフィック推定部２１４が送信停止タイマー２１６を参照して得たＰＡＵＳＥ状態を変数「ＰＡＵＳＥ状態」に保存し、終了状態８１９に進む。終了状態８１９では制限帯域計算処理を終了する。

【００９９】

以上の処理によって、輻輳トラフィック推定部２１４は、論理キュー２２１から読み出されてリンクへ送信されるイーサネットフレームについて、キューＩＤに基づいてトラフィックを区別したうえで各トラフィックの送信帯域を計算する。そして、輻輳を引き起こすトラフィックを推定し、当該推定されたトラフィックの送信帯域を減らして制限することが可能となる。そして、ＰＡＵＳＥを解除する際には、帯域制限を徐々に緩和することで保証帯域に戻すことができる。

【０１００】

図９は輻輳トラフィック推定部２１４で行う制限帯域計算処理のステップＳ８１０で実行される制限帯域補正処理の一例を示すフローチャートである。

【０１０１】

制限帯域補正処理では輻輳トラフィック推定部２１４が、１つのトラフィックに関して制限帯域の値を補正する。したがって、制限帯域補正処理における変数はトラフィック毎に値を保持する。

【０１０２】

輻輳トラフィック推定部２１４は制限帯域補正処理が開始状態Ｓ９０１から開始されると、ステップＳ９０２に進む。ステップＳ９０２では変数「新しい制限帯域」に変数「保証帯域」の値を代入してステップＳ９０３に進む。ステップＳ９０３では、輻輳トラフィック推定部２１４が、変数「前回の送信量」に変数「送信量」の値を代入し、変数「送信量」に変数「累積転送量」から変数「前回の累積転送量」を減じたものを代入し、ステップＳ９０４に進む。

【０１０３】

ステップＳ９０４では、輻輳トラフィック推定部２１４は、変数「実効帯域」に

（送信量／ＰＡＵＳＥ間隔）

を代入し、ステップＳ９０５に進む。ステップＳ９０５では、輻輳トラフィック推定部２１４が、「保証帯域」より「制限帯域」が小さい、すなわち帯域制限されているか否かを判定し、真ならステップＳ９０６に進み、偽なら終了状態９１３に進む。

【０１０４】

ステップＳ９０６では、輻輳トラフィック推定部２１４が、変数「前回の制限帯域」より「制限帯域」が小さい、すなわち帯域制限を前回強化したか否かを判定し、真ならステップＳ９０７に進み、偽ならステップＳ９１１に進む。

【０１０５】

ステップＳ９０７では、輻輳トラフィック推定部２１４が、「前回のＰＡＵＳＥ間隔」より「ＰＡＵＳＥ間隔」が大きい、すなわちＰＡＵＳＥ間隔が拡大しているか否かを判定し、真ならステップＳ９０８に進み、偽ならステップＳ９０９に進む。ステップＳ９０８では、輻輳トラフィック推定部２１４が帯域制限を強化してステップＳ９１０に進む。ステップＳ９０８における帯域制限の強化は、例えば、「新しい制限帯域」に、

（前回の制限帯域＋２×実効帯域）／３

を代入することで行う。ただし、この計算式は、制限帯域を小さくするためのものであり、本発明ではこれに限らない。例えば、「新しい制限帯域」に実効帯域を代入することとして、帯域制限の強化の幅を大きくしてもよい。

【０１０６】

ステップＳ９１０では輻輳トラフィック推定部２１４が、変数「帯域制限強化トラフィック数」を１増やし、ステップＳ９１２に進む。一方、ステップＳ９０９では、輻輳トラフィック推定部２１４は帯域制限を緩和してステップＳ９１１に進む。ステップＳ９０９における帯域制限の緩和は、例えば「新しい制限帯域」に、

（２×前回の制限帯域－実効帯域）

を代入することで行う。ただし、この計算式は、制限帯域を大きくするためのものであり、本発明ではこれに限らない。例えば、「新しい制限帯域」に保証帯域を代入することとしてもよい。

【０１０７】

ステップＳ９１１では輻輳トラフィック推定部２１４が、帯域制限を緩和してステップＳ９１２に進む。ここで、ステップＳ９１１における帯域制限の緩和は、例えば「新しい制限帯域」に、

（制限帯域×１．０１）

を代入することで行う。ただし、ただし、この計算式は、制限帯域を大きくするためのものであり、本発明ではこれに限らない。例えば、「新しい制限帯域」に保証帯域を代入することとしてもよい。また、ステップ９０９を省略し、ステップ９１１にて制限帯域を大きくする処理を行ってステップ数を減らしてもよい。ステップＳ９１２では輻輳トラフィック推定部２１４が変数「帯域制限トラフィック数」を１増やし、終了状態９１３に進む。終了状態９１３では制限帯域補正処理を終了する。

【０１０８】

なお、図８に示した制限帯域計算処理において、各トラフィックの制限帯域が保証帯域より大きい場合は制限帯域に保証帯域を代入し、また制限帯域が実効帯域より小さい場合は制限帯域に実効帯域を代入することで、制限帯域の範囲を限定する。

【０１０９】

なお、本発明では、制限帯域の下限は実効帯域に限定されるものではなく、例えば制限帯域を０として対象トラフィックの送信を停止してもよい。また、帯域制限緩和の計算時のみ保証帯域を大きくしたり、帯域制限の計算時のみ実効帯域を小さくしたりして制限帯域の変化を大きくしてもよい。

【０１１０】

以上に説明した制限帯域計算処理により、輻輳トラフィック推定部２１４はトラフィック毎の制限帯域を計算して送信フレーム選択部２１５に通知し、フレーム転送装置１はトラフィック毎の送信帯域を制限する。

【０１１１】

制限帯域計算処理の要点は以下のようなものである。ＰＡＵＳＥ開始時は輻輳が発生したことを意味するので、図８に示した制限帯域計算処理で帯域制限が強化されたトラフィックがなければ（ステップＳ８１３が真）、適当なトラフィックを選択して帯域制限を行う（ステップＳ８１４）。そして、図９に示した制限帯域補正処理では前記帯域制限の結果、ＰＡＵＳＥ間隔が拡大していれば（ステップＳ９０７が真）、当該帯域制限が輻輳緩和に効果があったとみなして当該帯域制限を強化し（ステップＳ９０８）、拡大していなければ当該帯域制限の効果がないとみなして当該帯域制限を緩和する（ステップＳ９０９

）。
【０１１２】
なお、ＰＡＵＳＥにより一時停止している時間も含めた実効的な送信帯域（ステップＳ９０４の実効帯域）は、当該トラフィックで使用できた帯域の実績とみなせるので、実効帯域を下限として帯域制限を行ってもよい。また、ＰＡＵＳＥがしばらく来ない場合（ステップＳ８０４が真）や帯域制限の強さが変化していない場合（ステップＳ９０６が偽）、帯域制限をしていたトラフィックで利用可能な帯域が変化した可能性も考慮して輻輳状態が緩和されたとみなして帯域制限を緩和してもよい（ステップＳ８１５およびステップＳ９１１）。
【０１１３】
すなわち、本発明では、ＰＡＵＳＥでフロー制御されるリンクについて、ＰＡＵＳＥフレーム受信側（すなわち制御対象のフローの送信側）において、ＰＡＵＳＥの間隔の変化から輻輳状態の変化（間隔拡大は輻輳緩和、間隔縮小は輻輳激化）を検出する。そして、当該リンクに送信するトラフィックのうち、当該トラフィックの送信帯域制限（送信の抑制）により輻輳緩和（ＰＡＵＳＥ間隔の拡大）が検出される場合に、当該トラフィックが輻輳の原因であると推定する。また、ＰＡＵＳＥフレーム受信時に、適当なトラフィックを選択して送信を抑制してみて輻輳が緩和されるか否かを試し、輻輳が緩和されれば送信抑制を強化する。逆に輻輳が緩和されなければ送信抑制を緩和して他のトラフィックの送信を抑制してみるという試行錯誤により輻輳の原因と推定されるトラフィックを探索する。そして輻輳の原因であると推定されたトラフィックの送信を抑制することで、当該リンクのＰＡＵＳＥ発生を抑制して通信速度を改善する。本発明では制限帯域計算処理が輻輳制御の対象となるトラフィック数の制限要因を含まないため、必要に応じて対象トラフィック数を増減することができる。また、輻輳の原因と推定されるトラフィックの探索では、２分木探索にならって複数のトラフィックを選択して送信抑制して輻輳の状態変化を確認し、輻輳緩和が確認された複数のトラフィックの中でさらに一部トラフィックを選択して送信抑制を行うような、段階的な探索を行ってもよい。本発明の実施は、２分木探索のほか、リスト探索、木探索、グラフ探索など、一般に探索アルゴリズム、検索アルゴリズムと呼ばれる考え方を用いても行うことができる。なお、本発明における前述した送信の抑止は、一部トラフィック以外のトラフィックの送信促進により、相対的に当該一部トラフィックの送信を抑制してもよい。また本発明は送信の抑制よる輻輳の抑制に限るものではなく、一部トラフィックの帯域変更による輻輳状態の変化に基づいて、目的とする輻輳状態、例えば、あらかじめ定めたＰＡＵＳＥフレームの受信間隔となるように各トラフィックの送信帯域を変更してもよい。
【０１１４】
＜送信フレーム選択部の処理＞
図１０Ａ、図１０Ｂは送信フレーム選択部２１５における送信帯域計算処理を説明するフローチャートである。送信フレーム選択部２１５では、論理キュー２２１から読み出されてリンクへ送信されるイーサネットフレームについて、キューＩＤによりトラフィックを区別したうえで各トラフィックを送信する量を計算する。このため、送信フレーム選択部２１５は、割当帯域管理部２１３から通知された帯域割当管理表７００の内容および輻輳トラフィック推定部２１４から通知された各トラフィックの制限帯域、キュー制御部２１８から通知された各論理キューの使用量、送信停止タイマー２１６を参照して得たＰｒｉｏｒｉｔｙ毎のＰＡＵＳＥ状態（タイマー値が０なら非ＰＡＵＳＥ状態、それ以外ならＰＡＵＳＥ状態）を用いて、各トラフィックの送信帯域を送信帯域計算処理により計算する。そして、送信フレーム選択部２１５は、計算された前記送信帯域内でフレームが送信されるようにキュー制御部２１８および入出力制御部２１９を制御する。
【０１１５】
前記送信帯域計算処理は一定時間毎（例えば０．０１秒毎）に繰り返して行う。なお、本発明では送信帯域計算処理実行の契機は一定周期に限るものではなく、例えば、ＰＡＵＳＥフレーム受信を送信帯域計算処理実行の契機としてもよいし、時間間隔をトラフィッ

クの流量等に応じて変更してもよい。以下、送信帯域計算処理を説明するが、処理の中で帯域とサイズ（量）を相互変換する必要がある場合、帯域は送信帯域計算処理を実行する時間間隔でサイズを割ることで求められ、サイズは帯域に送信帯域計算処理をかけることで求められる。以下の説明では、これらの変換方法への言及は明示的には行わないが、必要に応じて行うものとする。

【０１１６】

なお、サイズの単位の例としてはバイトまたはビット、帯域の単位の例としてはｂｐｓ（ビット毎秒）とし、１バイトは８ビットとするが、本発明で用いる単位はこれに限らず、例えばオクテット、バイト毎秒等を用いてもよい。また、以下の説明中で「トラフィック」は対象Ｐｒｉｏｒｉｔｙに属するトラフィックを、「キュー」は当該トラフィックのイーサネットフレームが格納されている論理キュー２２１をそれぞれ表す。なお、送信帯域計算処理において、変数「制限帯域」は輻輳トラフィック推定部２１４から通知された各トラフィックの制限帯域を保持するものとし、変数「割当帯域」は割当帯域管理部２１３から通知された帯域割当管理表に基づいて各トラフィックに保証する帯域を保持するものとし、変数「キューの使用量」はキュー制御部２１８から通知された各論理キュー２２１の使用量を各トラフィックについて保持するものとし、各トラフィックについての変数はそれぞれのトラフィックで異なる値を保持できるものとする。

【０１１７】

送信フレーム選択部２１５は送信帯域計算処理が開始状態Ｓ１００１から開始されると、ステップＳ１００２に進む。ステップＳ１００２では、送信フレーム選択部２１５が各トラフィックの変数「送信帯域」に変数「制限帯域」をそれぞれ代入し、ステップＳ１００３に進む。ステップＳ１００３では、変数「合計割当帯域」に全トラフィックの「割当帯域」の合計値を代入し、ステップＳ１００４に進む。ステップＳ１００４では、ステップＳ１００５からステップＳ１０３０までの処理を繰り返す。

【０１１８】

ステップＳ１００５では、各トラフィックについて、ステップＳ１００６からステップＳ１０１２までの処理を繰り返す。ステップＳ１００６では、送信フレーム選択部２１５が、変数「キューの使用量」を、前回送信帯域計算処理を実行してからの経過時間で割った値を変数「最大送信帯域」に代入し、ステップＳ１００７に進む。

【０１１９】

ステップＳ１００７では、送信フレーム選択部２１５が、変数「要求帯域」に変数「最大送信帯域」の値を代入し、ステップＳ１００８に進む。ステップＳ１００８では、送信フレーム選択部２１５が、当該トラフィックがＰＡＵＳＥ状態であるか否かを判定し、真の場合はステップＳ１００９に進み、偽の場合はステップＳ１０１０に進む。ステップＳ１００９では、送信フレーム選択部２１５が、変数「要求帯域」に０を代入しステップＳ１０１０に進む。

【０１２０】

ステップＳ１０１０では、送信フレーム選択部２１５は、変数「制限帯域」が変数「保証帯域」より小さく（すなわち帯域制限されており）変数「送信帯域」が変数「最大送信帯域」より小さいか否かを判定し、真ならステップＳ１０１１に進み、偽ならステップＳ１０１２に進む。ステップＳ１０１１では、変数「要求帯域」に変数「送信帯域」を設定し、ステップＳ１０１２に進む。

【０１２１】

ステップＳ１０１２では、送信フレーム選択部２１５がステップＳ１００５からの繰り返し処理が終了したら、ステップＳ１０１３に進む。

【０１２２】

ステップＳ１０１３では、送信フレーム選択部２１５が変数「最大送信帯域」が０より大きいトラフィックが１トラフィックのみで、かつ当該トラフィックの変数「制限帯域」が変数「保証帯域」より小さい（すなわち帯域制限されている）か否かを判定し、真ならステップＳ１０１４に進み、偽ならステップＳ１０１５に進む。ステップＳ１０１４では

、送信フレーム選択部２１５が各トラフィックについて、変数「要求帯域」に変数「最大送信帯域」を代入し、ステップＳ１０１５に進む。ステップＳ１０１５では、変数「合計余剰帯域」、変数「合計不足帯域」、変数「合計割当済帯域」にそれぞれ０を代入し、ステップＳ１０１６に進む。

【０１２３】

ステップＳ１０１６では、送信フレーム選択部２１５が各トラフィックについて、ステップＳ１０１７からステップＳ１０２２までの処理を繰り返す。ステップＳ１０１７では、変数「分割帯域」に変数「送信帯域」の値を設定し、ステップＳ１０１８に進む。ステップＳ１０１８では、送信フレーム選択部２１５が変数「要求帯域」が変数「分割帯域」より小さいか否かを判定し、真ならステップＳ１０１９に進み、偽ならステップＳ１０２０に進む。ステップＳ１０１９では、変数「合計余剰帯域」に（分割帯域－要求帯域）を加え、ステップＳ１０２１に進む。

【０１２４】

ステップＳ１０２０では、変数「合計不足帯域」に（要求帯域－分割帯域）を加え、ステップＳ１０２１に進む。ステップＳ１０２１では、変数「合計割当済帯域」に分割帯域の値を加え、ステップＳ１０２２に進む。ステップＳ１０２２では、送信フレーム選択部２１５がステップＳ１０１６からの繰り返し処理が終了したら、図１０Ｂのステップ１０２３に進む。図１０Ｂのステップ１０２３では、変数「合計余剰帯域」に

（合計割当帯域　－　合計割当済帯域）

を加え、ステップＳ１０２４に進む。ステップＳ１０２４では、送信フレーム選択部２１５が、変数「合計不足帯域」および変数「合計余剰帯域」が十分小さいか否かを判定し、真なら終了状態１０３１に進み、偽ならステップＳ１０２５に進む。なお、ステップＳ１０２４における「十分小さい」とは例えば１ｂｐｓ以下等、分配するに足らない帯域や制御可能な最小帯域等を表す。本発明では、これに限ることなく、例えばステップＳ１０２４における判定を若干緩め（大き目）にして、繰り返し回数を削減してもよい。

【０１２５】

ステップＳ１０２５では、送信フレーム選択部２１５が、各トラフィックについて、ステップＳ１０２６からステップＳ１０２９までを繰り返す。ステップＳ１０２６では、送信フレーム選択部２１５が、要求帯域が分割帯域より小さい否かを判定し、真ならステップＳ１０２７に進み、偽ならステップＳ１０２８に進む。ステップＳ１０２７では、変数「送信帯域」に要求帯域を代入し、ステップＳ１０２９に進む。

【０１２６】

ステップＳ１０２８では送信フレーム選択部２１５が、変数「送信帯域」に

（合計余剰帯域×（要求帯域－分割帯域）／合計不足帯域）

を代入し、ステップＳ１０２９に進む。なお、ステップＳ１０２８における計算式は、合計余剰帯域の分配例を表しており、本発明はこれに限るものではない。例えば、１つのトラフィックに合計余剰帯域をすべて与え、他のトラフィックへの分配を０とし、処理を簡素化してもよい。

【０１２７】

ステップＳ１０２９では、送信フレーム選択部２１５が、ステップＳ１０２５からの繰り返し処理が終了したら、ステップＳ１０３０に進む。ステップＳ１０３０では、ステップＳ１００４からの繰り返し処理が終了したら、終了状態１０３１に進む。終了状態１０３１では送信帯域計算処理を終了する。

【０１２８】

送信帯域計算処理の要点は以下の通りである。送信フレーム選択部２１５は、各トラフィックについて、送信帯域に制限帯域を代入して（ステップＳ１００２）、以下を繰り返すことで送信帯域を調整する。送信フレーム選択部２１５は、要求帯域をＰＡＵＳＥ状態であれば０とし（ステップＳ１００９）、そうでなければ論理キュー２２１のフレーム送信に必要な帯域（最大送信帯域）とする（ステップＳ１００７）。ただし、帯域が制限されており、要求帯域が送信帯域より大きい場合は、要求帯域は送信帯域とする（ステップ

Ｓ１０１１）。ただし、送信しようとしている（すなわちキューにフレームのある）トラフィックが、帯域制限された１トラフィックのみの場合は（ステップＳ１０１３）、要求帯域はキューのフレーム送信に必要な帯域（最大送信帯域）とする（ステップＳ１０１４）。そして、送信帯域と要求帯域を比べて、不足分及び余剰分を計算し（ステップＳ１０１６）、不足のトラフィックに余剰分を再分配する（１０２５）。不足分及び余剰分が予め設定した一定値以下となって十分小さくなったら、余剰分の再分配を終了する（１０２４）。以上を繰り返す（ステップＳ１０３０）。

【０１２９】

すなわち、本発明では、他のトラフィックがあり、かつ帯域制限されている場合には制限帯域を超えないように送信帯域を設定し、それ以外の場合は最大送信帯域を上限帯域にして、保証帯域を各トラフィックに保証しつつ、各トラフィックで使用していない余剰帯域を帯域が不足しているトラフィックに再配分して、送信帯域を決定する。本発明では送信帯域計算処理が送信対象となるトラフィック数の制限要因を含まないため、必要に応じて対象トラフィック数を増減することができる。

【０１３０】

＜変形例＞

なお、図２の説明でＰＦＣにおける処理を前提として説明したが、本発明はＰＦＣを前提としない実施形態も可能である。その場合も図２の構成や処理は同一で、ＰＦＣにおいて１つのｐｒｉｏｒｉｔｙを用いると限定し、フレームの形式はＩＥＥＥ　８０２．３ｘに規定されるＰＡＵＳＥフレームとする。この場合、本発明はＩＥＥＥ　８０２．３ｘに規定されるＰＡＵＳＥをサポートしたイーサネットに適用可能となる。

【０１３１】

また、本発明はＰＡＵＳＥの頻度（間隔）に基づいて送信する際のフレーム選択を変更することを含むため、図２で説明したようなフレーム転送装置でなくとも、１つ以上のネットワークインタフェースを持っていれば、実施可能である。このような実施形態を用いることで、計算機用のＮＩＣ（ネットワークインタフェースカード）や仮想マシンの仮想ネットワークインタフェースに対して本発明を実施することが可能となる。また、フレーム転送装置においてもポート単位で本発明を実施することも可能である。

【０１３２】

＜まとめ＞

以上のように、本発明のフレーム転送装置１によれば、リンクの送信先となるノード１１７からＰＡＵＳＥフレームを受信した場合、ＰＡＵＳＥフレームで指定された時間までフレームの送信を一時停止し、送信先のノード１１７で輻輳の原因となっている論理トラフィックを推定する。そして、フレーム転送装置１は、当該ノード１１７への送信再開時には、輻輳の原因として推定された論理トラフィックの送信を抑制し、他の論理トラフィックを優先して送信する。これにより、送信先のノード１１７がＰＡＵＳＥフレームを再度送信するのを防ぐことで当該ノード１１７の論理トラフィックが他の論理トラフィックに与える影響を抑制することができる。これらはすべてＰＡＵＳＥによるフロー制御に基づいているため、確実な通信を保証できる。

【０１３３】

そして、本発明のフレーム転送装置１では、前記従来例のようにノード１１７は既存の構成のままでよく、新たな機能を付加する必要がないため導入コストを抑制できる。さらに、本発明のフレーム転送装置１では、前記従来例の非特許文献１のように、扱うことが可能な論理トラフィックの数が８以下に制限されることはなく、任意の数の論理トラフィックを制御することが可能となる。これにより、マルチコアプロセッサ等を利用した仮想化システムなどでは、仮想マシンの数がネットワークインタフェースの数に制約を受けることがなくなって、可用性を向上させることが可能となる。

【産業上の利用可能性】

【０１３４】

以上のように、本発明は論理トラフィックを転送する装置や計算機及びネットワークシ

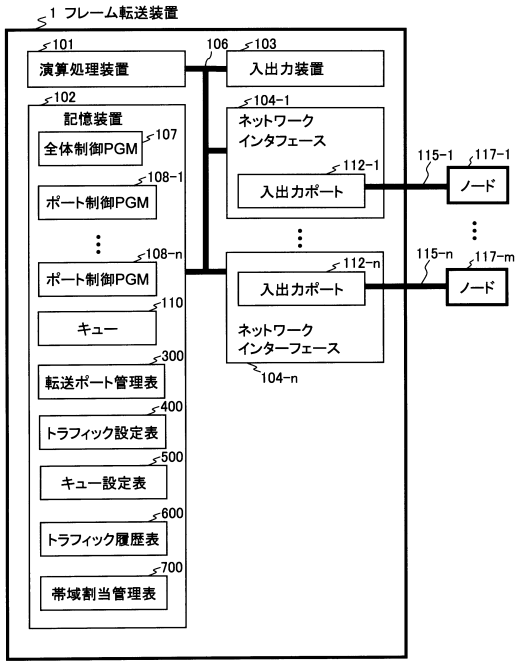
ステムに適用することができる。

【符号の説明】

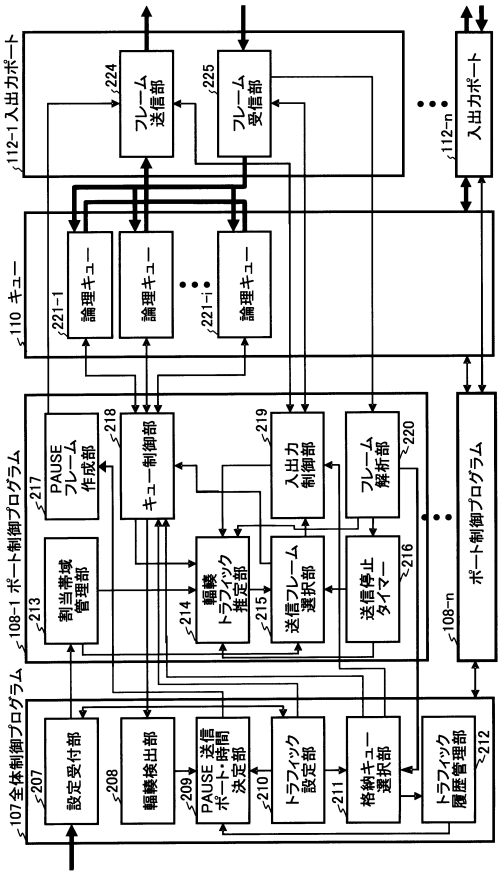
【０１３５】

１　フレーム転送装置
１０１　演算処理装置
１０２　記憶装置
１０３　入出力装置
１０４－１～１０４－ｎ　ネットワークインタフェース
１０７　全体制御プログラム
１０８－１～１０８－ｎ　ポート制御プログラム
１１０　キュー
１１２－１～１１２－ｎ　入出力ポート
１１７－１～１１７－ｍ　ノード
２０７　設定受付部
２０８　輻輳検出部
２０９　ＰＡＵＳＥ送信ポート・時間決定部
２１０　トラフィック設定部
２１１　格納キュー選択部
２１２　トラフィック履歴管理部
２１３　割当帯域管理部
２１４　輻輳トラフィック推定部
２１５　送信フレーム選択部
２１６　送信停止タイマー
２１７　ＰＡＵＳＥフレーム作成部
２１８　キュー制御部
２１９　入出力制御部
２２０　フレーム解析部
２２１－１～２２１－ｉ　論理キュー
２２４　フレーム送信部
２２５　フレーム受信部

【図１】



【図２】



【図３】

300 転送ポート管理表

301 MACアドレス	302 VLAN ID	303 ポートID
MAC1	100	2
⋮	⋮	⋮

【図４】

400 トラフィック設定表

401 送信元アドレス	402 送信先アドレス	403 VLAN ID	404 Priority	405 トラフィックID
MAC 1	MAC 2	100	5	1
⋮	⋮	⋮	⋮	⋮

【図５】

500 キュー設定表

501 トラフィックID	502 ポートID	503 キューID
1	2	3
⋮	⋮	⋮

【図６】

600 トラフィック履歴表

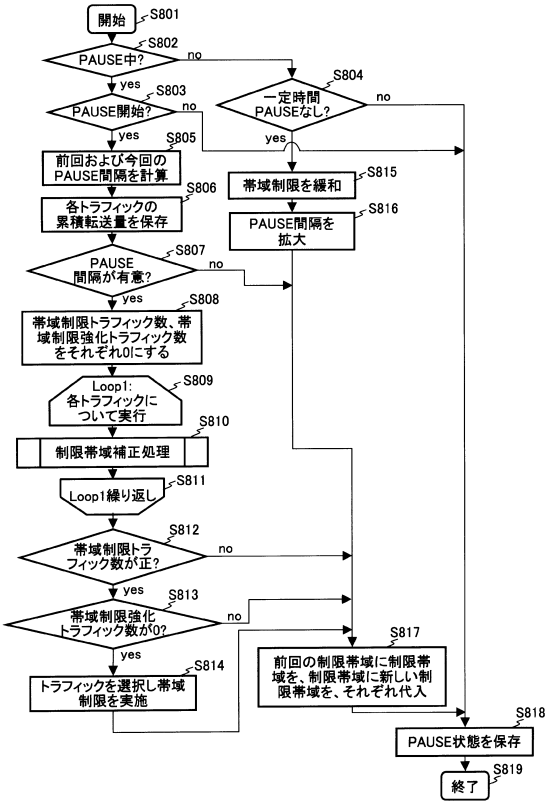
601 トラフィックID	602 ポートID	603 時刻	604 フレームサイズ
1	MAC 2	12:34:56	1200
⋮	⋮	⋮	⋮

【図７】

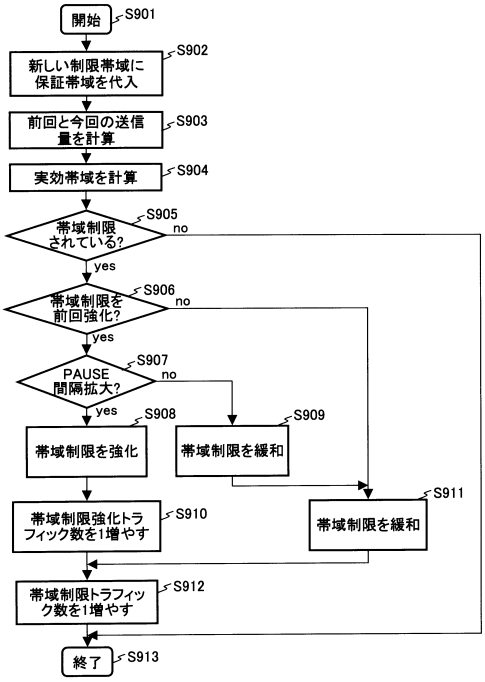
700 帯域割当管理表

701 キューID	702 割当帯域
1	100Mbps
⋮	⋮

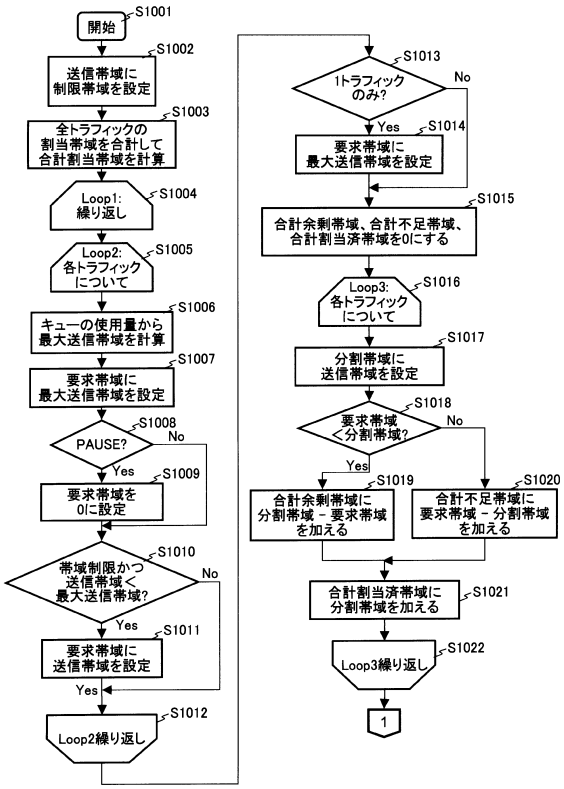
【図８】



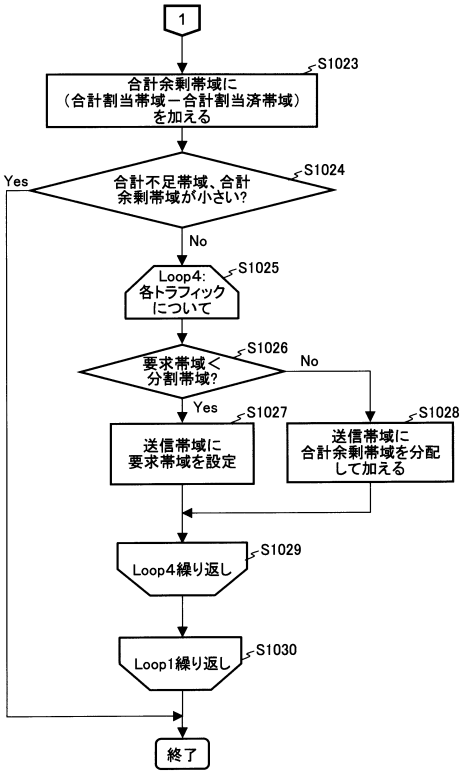
【図９】



【図１０Ａ】



【図１０Ｂ】



フロントページの続き

審査官　安藤　一道

(56)参考文献　特開２００６－０５０３６１（ＪＰ，Ａ）
特開２００８－２３６７３３（ＪＰ，Ａ）
特開２０１１－０１９０４０（ＪＰ，Ａ）
特開２００４－１０４４２７（ＪＰ，Ａ）
特開２００４－１５９２０３（ＪＰ，Ａ）
特開２００６－０３３７１３（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｈ０４Ｌ　１２／２８
Ｈ０４Ｌ　１２／８２５