

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第5710418号  
(P5710418)

(45) 発行日 平成27年4月30日 (2015. 4. 30)

(24) 登録日 平成27年3月13日 (2015. 3. 13)

(51) Int. Cl.

F I

H O 4 L 12/827 (2013.01)

H O 4 L 12/827

請求項の数 9 (全 40 頁)

(21) 出願番号	特願2011-172611 (P2011-172611)	(73) 特許権者	504411166
(22) 出願日	平成23年8月8日 (2011. 8. 8)		アラクスネットワークス株式会社
(65) 公開番号	特開2013-38570 (P2013-38570A)		神奈川県川崎市幸区鹿島田一丁目1番2号
(43) 公開日	平成25年2月21日 (2013. 2. 21)	(74) 代理人	110001689
審査請求日	平成26年2月6日 (2014. 2. 6)		青稜特許業務法人
		(74) 代理人	110000350
			ポレール特許業務法人
		(72) 発明者	石川 有一
			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	矢崎 武己
			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
		審査官	鈴木 崇雅
			最終頁に続く

(54) 【発明の名称】 パケット中継装置、及び方法

(57) 【特許請求の範囲】

【請求項 1】

送信端末を含む送信側と受信端末を含む受信側との間で転送されるデータパケット及び  
当該データパケットに対応する応答パケットを中継するパケット中継装置であって、

複数の入力回線と、

複数の出力回線と、

前記入力回線から受信した前記データパケット或いは前記応答パケットの出力先を検索し、

前記送信側から前記受信側に向けて転送される前記データパケットに含まれる前記送信側と前記受信側に関する情報の組み合わせに対応するフロー毎に、前記送信側と前記受信側との間の輻輳状態を監視し、

前記応答パケットの受信に応じて、当該応答パケットに含まれる前記送信側に関する情報と前記受信側に関する情報とに基づいて、前記送信側と前記受信側に関する情報の組み合わせに対応するフローを特定し、特定したフローが、前記監視の結果により輻輳状態にあると判定すると、前記受信側から受信する当該フローに属する前記データパケットに対応する前記応答パケットのヘッダの内、輻輳状態を示すフィールドを輻輳状態にあることを示す値に書き換え、出力するパケット検索部と、を備え、

前記パケット検索部は、

前記フロー毎の輻輳状態を記録するフロー毎の輻輳状態フラグから構成される輻輳状態管理テーブルを備え、

10

20

前記フローが輻輳状態にあると判定すると、当該フローの前記輻輳状態フラグを輻輳状態にあることを示す値に書き換え、

前記フローが輻輳状態から解消したと判定すると、当該フローの前記輻輳状態フラグを非輻輳状態にあることを示す値に書き換え、

前記応答パケットを受信すると前記輻輳状態管理テーブルを参照して、当該応答パケットが属するフローの前記輻輳状態フラグに従って、当該輻輳状態フラグが輻輳状態を示す場合には前記応答パケットのパケットヘッダのうちネットワークの輻輳状態を示すフィールドの値を輻輳状態にあることを示す値に書き換え、当該輻輳状態フラグが非輻輳状態を示す場合には前記応答パケットのパケットヘッダのうちネットワークの輻輳状態を示すフィールドの値を輻輳状態にあることを示す値に書き換えず、

10

前記送信端末から受信したデータパケットのパケットヘッダ情報のうち、前記送信端末が送信帯域を抑制する輻輳回避制御を行ったことを示す情報が、輻輳回避を行ったことを示す値となっている場合に、フローが輻輳状態から解消したと判定する、

ことを特徴とするパケット中継装置。

【請求項 2】

請求項 1 に記載のパケット中継装置であって、

前記パケット検索部は、

フロー毎の帯域監視の結果、帯域違反と判定されると、前記フローが輻輳状態にあると判定する、

ことを特徴とするパケット中継装置。

20

【請求項 3】

請求項 1 に記載のパケット中継装置であって、

前記フローが、送信元 IP アドレス、宛先 IP アドレス、送信元ポート番号、宛先ポート番号の組み合わせが等しい TCP フローまたは DCCP フローまたは SCTP フロー、または送信元 IP アドレス、宛先 IP アドレスの組み合わせが等しい IP フロー、または同一 VLAN に属する VLAN フローである、

ことを特徴とするパケット中継装置。

【請求項 4】

請求項 1 に記載のパケット中継装置であって、

前記輻輳状態管理テーブルは、フロー毎の輻輳状態フラグを運用管理者が指定した特定フローに対してのみ備える、

30

ことを特徴とするパケット中継装置。

【請求項 5】

請求項 1 に記載のパケット中継装置であって、

受信したパケットを蓄積するパケット受信バッファを更に備え、

前記パケット受信バッファに蓄積されたパケット数がある閾値を超過すると、前記パケット受信バッファが輻輳状態にあると判定する、

ことを特徴とするパケット中継装置。

【請求項 6】

請求項 1 に記載のパケット中継装置であって、

前記パケット検索部は、

輻輳通知機能をサポートしない受信端末から送信された、送受信端末間で輻輳通知機能のサポート可否の情報を通知する制御パケットのパケットヘッダのうち、輻輳通知機能のサポート可否を示すフィールドの値を、輻輳通知機能の非サポートを示す値から輻輳通知機能のサポートを示す値に書き換える、

40

ことを特徴とするパケット中継装置。

【請求項 7】

請求項 6 に記載のパケット中継装置であって、

前記受信端末は、ECN による輻輳通知機能をサポートしない受信端末であり、

前記パケット検索部は、

50

前記制御パケットである、ECNの初期化手順の3 way handshake処理におけるECEフラグのセットされていないSYN - ACKパケットを、ECEフラグのセットされたSYN - ACKパケットに書き換える、

ことを特徴とするパケット中継装置。

【請求項8】

請求項1に記載のパケット中継装置であって、

前記パケット検索部は、前記応答パケットが属するフローを検出するフロー検索において、

送信元IPアドレスと宛先アドレス、送信元ポート番号と宛先ポート番号を交換した条件で検索する、

ことを特徴とするパケット中継装置。

【請求項9】

複数の入力回線と複数の出力回線を備え、送信側と受信側との間で転送されるデータパケット及び当該データパケットに対応する応答パケットを中継するパケット中継装置のパケット中継方法であって、

前記パケット中継装置は、

前記入力回線から受信した前記データパケット或いは当該データパケットに対応する前記応答パケットの出力先を検索し、

前記送信側から前記受信側に向けて転送される前記データパケットに含まれる前記送信側と前記受信側に関する情報の組み合わせに対応するフロー毎に、前記送信側と前記受信側との間の輻輳状態を監視し、

前記応答パケットの受信に応じて、当該応答パケットに含まれる前記送信側に関する情報と前記受信側に関する情報とに基づいて、前記送信側と前記受信側に関する情報の組み合わせに対応するフローを特定し、特定したフローが、前記監視の結果により輻輳状態であると判定すると、前記受信側から受信する当該フローに属する前記データパケットに対応する前記応答パケットのヘッダの内、輻輳状態を示すフィールドを輻輳状態にあることを示す値に書き換え出力し、

前記フロー毎の輻輳状態を記録するフロー毎の輻輳状態フラグから構成される輻輳状態管理テーブルを書き換えおよび参照可能であり、

前記フローが輻輳状態であると判定すると、当該フローの前記輻輳状態フラグを輻輳状態にあることを示す値に書き換え、

前記フローが輻輳状態から解消したと判定すると、当該フローの前記輻輳状態フラグを非輻輳状態にあることを示す値に書き換え、

前記応答パケットを受信すると前記輻輳状態管理テーブルを参照して、当該応答パケットが属するフローの前記輻輳状態フラグに従って、当該輻輳状態フラグが輻輳状態を示す場合には前記応答パケットのパケットヘッダのうちネットワークの輻輳状態を示すフィールドの値を輻輳状態を示す値に書き換え、当該輻輳状態フラグが非輻輳状態を示す場合には前記応答パケットのパケットヘッダのうちネットワークの輻輳状態を示すフィールドの値を輻輳状態を示す値に書き換えず、

前記送信側から受信したデータパケットのパケットヘッダ情報のうち、前記送信側が送信帯域を抑制する輻輳回避制御を行ったことを示す情報が、輻輳回避を行ったことを示す値となっている場合に、フローが輻輳状態から解消したと判定する、

ことを特徴とするパケット中継方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、ネットワークにおけるパケット中継装置、特に輻輳通知機能を備える装置の帯域監視技術に関する。

【背景技術】

【0002】

10

20

30

40

50

輻輳通知技術の一つであるECN (Explicit Congestion Notification) が、LinuxやWindows (登録商標) 等のOS (Operating System) にオプション機能として実装され、またIETF (Internet Engineering Task Force) でも盛んに議論されるなど、脚光を浴びている。

【 0 0 0 3 】

ECNとは、ネットワークを構成するルータ/スイッチ等のパケット中継装置において輻輳が発生した場合に、送信端末から送信されたデータパケットに対しルータ/スイッチ自身で輻輳情報を明示的にマーキングして受信端末に送信することで、受信端末に輻輳情報を通知する輻輳通知技術である (特許文献 1 参照)。

【 先行技術文献 】

【 特許文献 】

10

【 0 0 0 4 】

【 特許文献 1 】 特開2010 - 232876号公報

【 特許文献 2 】 特開2009 - 239401号公報

【 特許文献 3 】 特開平11 - 341076号公報

【 非特許文献 】

【 0 0 0 5 】

【 非特許文献 1 】 RFC2474、 “ Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers ”、 ( <http://www.ietf.org/rfc/rfc2474.txt> )

【 非特許文献 2 】 RFC2475、 “ An Architecture for Differentiated Services ” ( <http://www.ietf.org/rfc/rfc2475.txt> )

20

【 非特許文献 3 】 NII Journal No. 3 (2001.11) 特集：情報プラットフォーム解説論文「情報流通プラットフォームにおけるQoS 保証のためのトラフィック制御」Traffic Control for QoS Guarantee in Communication Networks、計宇生、国立情報学研究所、 ( <http://www.nii.jp/journal/pdf/03/03-04.pdf> )

【 発明の概要 】

【 発明が解決しようとする課題 】

【 0 0 0 6 】

従来のTCP (Transmission Control Protocol) が備える輻輳回避機能では、輻輳発生を明示的に送受信端末に通知する手段がなかった。そのため、ネットワークにおいてパケット廃棄が生じたことをタイムアウトや重複ACK受信により送信端末にて検出した場合に、輻輳の発生を送信端末が自律的に判断する仕組みとなっていた。輻輳が発生したと判断された場合には、送信端末が送信帯域を抑制することで輻輳回避を実現する。

30

【 0 0 0 7 】

パケットが廃棄された場合には、廃棄されたパケットのデータを補完しなければ受信端末でのアプリケーション処理を正常に実行できないため、送信端末のTCPは廃棄されたパケットを再送する機能を備えている。廃棄されたパケットは再送されるが、再送パケットが受信端末に到着するタイミングは、パケット廃棄が生じなかった場合に本来パケットが到着したはずのタイミングよりも遅れてしまう。そのため、再送パケットを受信して、パケットとして分割されたデータを再構成してアプリケーション処理できるようになるタイミングが遅れてしまい、通信品質が低下してしまう問題があった。

40

【 0 0 0 8 】

ECNでは、ルータ/スイッチのようなパケット中継装置において輻輳を検出すると、データパケットのIPヘッダのTOSフィールドの下位2ビットにて定義されるECNフィールドを、輻輳検出済の意味をもつCE (Congestion Experienced) の値にルータ/スイッチが書き換えるECNマーキングを実施して、可能な限りパケット廃棄することなく中継する。

【 0 0 0 9 】

CEパケットを受信した受信端末は、それ以降送信端末に返信するACK (Acknowledge) パケットにおいて、輻輳検出したことを送信端末に通知するためECN向けに拡張されたTCP制御フラグのECE (ECN Echo) フラグを ' 1 ' にセットする。送信端末は、ECEフラグがセッ

50

トされたACKパケットを受信すると、ネットワークが輻輳状態にあるものと判断し、従来のECNを備えないTCPがもつ輻輳回避機能に基づいて送信帯域を抑制することで、輻輳回避する。輻輳回避の処理を実行すると、送信端末は送信パケットにTCP制御フラグのCWR (Congestion Window Reduced) フラグをセットする。CWRフラグは、送信端末にて輻輳回避の処理を行った後に、受信端末が返信するACKパケットのECEフラグのセットを停止させるためECN向けに拡張されたTCP制御フラグである。受信端末はCWRフラグがセットされたパケットを受信すると、ACKパケットに対するECEフラグのセットを停止する。

【 0 0 1 0 】

このようにECNを用いることで、ルータ/スイッチにおいて輻輳を検出した際のパケット廃棄を可能な限り抑止し、パケットを中継しつつ輻輳発生を明示的に送受信端末に通知して送信帯域の抑制を促すことが可能となり、パケット廃棄やこれに伴うパケット再送による通信品質の低下を抑制できる。

10

【 0 0 1 1 】

また、従来のパケット廃棄により輻輳の発生を判断する輻輳回避技術では、ビットエラー等の障害によるパケット廃棄を、ネットワークの輻輳によるパケット廃棄と区別することができないため、輻輳とは無関係のビットエラー障害に対しても輻輳回避の処理がなされてしまい、送信帯域の不要な低下を招いてしまう問題があった。しかしECNでは、送受信端末に対して輻輳発生が明示的に通知されるので、ビットエラー障害とネットワークの輻輳を明確に区別してネットワークの輻輳の場合のみ輻輳回避することが可能となり、送信帯域の不要な低下を防止できるようになる。

20

【 0 0 1 2 】

TCPでは、受信端末がデータパケットを受信すると、受信端末はこれに対するACKパケットを送信端末に送信する。送信端末はACKパケットを受信すると送信帯域を増大させるが、ECNに基づいてECEフラグをセットされた輻輳通知用のACKパケットを受信すると、輻輳回避するため送信帯域を減少させる帯域制御を行う。

【 0 0 1 3 】

従来のECNでは、ルータ/スイッチ等のパケット中継装置において輻輳が発生すると、パケット中継装置 受信端末 パケット中継装置 送信端末の経路で輻輳情報を通知する。そのため、パケット中継装置と受信端末間の遅延が増大すると輻輳通知が遅れてしまうこととなる。輻輳通知がなされるまでの間、ネットワークは輻輳状態にあるにも関わらず送信端末は受信したACKパケットの分だけ送信帯域を増大させるので、輻輳状態は更に悪化してしまう。その結果、廃棄パケットが増加し、廃棄パケットを補完するための再送パケットも増加するので、実効帯域の低下、実効遅延の増大を招き、通信品質が低下してしまうという課題がある。

30

【 0 0 1 4 】

本発明の目的は、上記の課題を解決し、パケット中継装置から受信端末の遅延が増大した場合にも、輻輳通知の遅延を防止することが可能なパケット中継装置、及び方法を提供することにある。

【課題を解決するための手段】

【 0 0 1 5 】

40

上記の目的を達成するため、本発明においては、パケットを中継するパケット中継装置であって、複数の入力回線と出力回線と、入力回線から受信したパケットを検索するパケット検索部を備え、パケット検索部は、入力回線から受信したパケットの情報により識別されるパケットの集合が構成されるフローを検出し、検出したフローが輻輳状態にあると判定すると、それ以降に受信する当該フローに対する応答パケットのヘッダの内、輻輳状態を示すフィールドを輻輳状態にあることを示す値に書き換えるパケット中継装置を提供する。

【 0 0 1 6 】

また、上記の目的を達成するため、本発明においては、複数の入力回線と出力回線を備えたパケット中継装置のパケット中継方法であって、入力回線から受信したパケットの情

50

報により識別されるパケットの集合が構成されるフローを検出し、検出したフローが輻輳状態にあると判定すると、それ以降に受信する当該フローに対する応答パケットのヘッダの内、輻輳状態を示すフィールドを輻輳状態にあることを示す値に書き換えるパケット中継方法を提供する。

【発明の効果】

【0017】

本発明によれば、ルータ/スイッチから受信端末の遅延が増大した場合にも、輻輳通知の遅延を防止する。

【0018】

また、本発明を備えていないパケット中継装置で判定された輻輳に対しても、輻輳通知の遅延を防止する。

【0019】

更に、受信端末が輻輳通知機能を備えていない場合にも、輻輳通知システムを実現する。

【図面の簡単な説明】

【0020】

【図1】各種の実施例に係るパケット中継装置による輻輳通知の基本構成を説明するための図である。

【図2】第1の実施例に係る、パケット中継装置の一構成例を示す図である。

【図3】第1の実施例に係る、パケット中継装置のパケット受信回路の一例を示す図である。

【図4A】第1の実施例に係る、パケット受信バッファの一構成例を示す図である。

【図4B】第1の実施例に係る、パケットヘッダ情報の一例を示す図である。

【図5】第1の実施例に係る、TOSフィールドの一例を示す図である。

【図6】第1の実施例に係る、TCPフラグの一例を示す図である。

【図7】第1の実施例に係る、フロー検索部の一構成例を示す図である。

【図8】第1の実施例に係る、フロー検索テーブルの一構成例を示す図である。

【図9】第1の実施例に係る、帯域監視部の一構成例を示す図である。

【図10】第1の実施例に係る、帯域監視テーブルの一構成例を示す図である。

【図11】第1の実施例に係る、リーキーバケツアルゴリズムのフローチャートを示す図である。

【図12】第1の実施例に係る、帯域監視エントリ毎輻輳状態管理テーブルの一構成例を示す図である。

【図13】第1の実施例に係る、輻輳処理を模式的に説明するための図である。

【図14】第1の実施例に係る、フロー検索部の他の構成例を示す図である。

【図15】第1の実施例に係る、フロー検索部のフローアクションテーブルの一例を示す図である。

【図16】第1の実施例に係る、パケット受信回路の他の構成例を示す図である。

【図17】第1の実施例に係る、パケット受信バッファの他の構成例を示す図である。

【図18】第1の実施例に係る、ECNマーキング閾値テーブルの一構成例を示す図である。

【図19】第1の実施例に係る、バッファ蓄積パケット数とバッファ面数、ECNマーキング値の大小関係に応じた輻輳処理を模式的に示す図である。

【図20】第1の実施例に係る、パケット送信回路の一構成例を示す図である。

【図21】第1の実施例に係る、回線毎送信帯域管理テーブルの一構成例を示す図である。

【図22】第1の実施例に係る、パケット送信回路の処理のフローチャートを示す図である。

【図23】第1の実施例に係る、パケット中継装置を用いた通信のフローチャートを示す図である。

10

20

30

40

50

【図 2 4】第 1 の実施例に係る、パケット中継装置を用いた通信のシーケンス図である。

【図 2 5】第 2 の実施例に係る、回線毎送信帯域管理テーブルの一構成例を示す図である。

。

【図 2 6】第 3 の実施例に係る、フロー検索テーブルの一構成例を示す図である。

【図 2 7】第 3 の実施例に係る、フローアクションテーブルの一構成例を示す図である。

【図 2 8】第 3 の実施例に係る、パケット中継装置を用いた通信のフローチャートを示す図である。

【図 2 9】第 3 の実施例に係る、パケット中継装置を用いた通信のシーケンス図である。

【図 3 0】第 4 の実施例に係る、フロー検索テーブルの一構成例を示す図である。

【図 3 1】第 4 の実施例に係る、フローアクションテーブルの一構成例を示す図である。

【図 3 2】第 4 の実施例に係る、パケット中継装置を用いた通信のフローチャートを示す図である。

【図 3 3】第 4 の実施例に係る、パケット中継装置を用いた通信のシーケンス図を示す図である。

【発明を実施するための形態】

【0021】

以下に、本発明を実施するための形態を図面に従い説明する。それに先立ち、本発明の輻輳通知機能を備えるパケット中継装置の基本構成について説明する。本発明のパケット中継装置の基本構成では、複数の入力回線と出力回線を備え、入力回線から受信したパケットの入力物理回線番号、または入力論理回線番号、またはパケットヘッダ情報のうち、少なくとも一つ以上の情報により識別されるパケットの集合から構成されるフローを検出し、フローが輻輳状態にあると判定すると、それ以降に受信した当該フローに対する応答パケットのパケットヘッダのうち、ネットワークの輻輳状態を示すフィールドを輻輳状態にあることを示す値に書き換える。

【0022】

図 1 を用いて、本発明のパケット中継装置の基本構成の処理と効果を説明する。送信端末 1 と受信端末 3 間は ECN (Explicit Congestion Notification) を用いて輻輳通知しているものとする。送信端末 1 から送信されたデータパケット 14 をパケット中継装置 2 が受信した際に、データパケット 14 が属するフローが輻輳状態にあると判定した場合、パケット中継装置 2 は、これ以降、受信端末 3 から送信された TCP の応答パケット (ACK) 16 を受信すると、この ACK パケット 16 の ECE (ECN Echo) フラグを 0 から 1 に書き換えて、ACK パケット (ECE) 17 として送信端末 1 に送信する。これにより、パケット中継装置 2 から送信端末 1 への経路で輻輳通知できる。この輻輳通知を受け、送信端末 1 は、ウィンドウサイズを縮小し、その旨を通知する CWR (Congestion Window Reduce) フラグをセットしてデータ (CWR) 18 を送信する。

【0023】

一方、以上説明した本発明の基本構成を用いない場合のパケット中継装置の処理の流れの場合、送信端末 1 から送信されたデータパケット 14 をパケット中継装置 2 が受信した際に、当該データパケット 14 が属するフローが輻輳状態にあると判定すると、パケット中継装置 2 はデータパケット 14 の ECN フィールドを 11 (Congestion Experienced: CE) に書き換えて、データパケット (CE) 15 として受信端末 3 に送信する。受信端末 3 はデータパケット (CE) 15 を受信すると、ECN フィールドが 11 (CE) という輻輳発生を示す値にあるので、これ以降送信する ACK パケットの ECE (ECN Echo) フラグを 1 として、ACK (ECE) を送信する。パケット中継装置 2 は ACK (ECE) を受信するとそのまま ACK (ECE) 17 として送信端末 1 に送信する。従って、パケット中継装置 2 から受信端末 3、受信端末 3 からパケット中継装置 2、パケット中継装置 2 から送信端末 1 の経路で輻輳通知することとなる。

【0024】

以上より明らかなように、本発明のパケット中継装置の基本構成によれば、パケット中継装置 2 から受信端末 3、受信端末 3 からパケット中継装置 2 の分だけ輻輳通知の経路を省いて短縮することができる。その結果、パケット中継装置 2 と受信端末 3 間の遅延が増大して

10

20

30

40

50

も、輻輳通知は遅れないこととなり、通信品質の低下を防止できるという効果を得ることとなる。

【実施例 1】

【0025】

図2～図24を用いて、第1の実施例の packets 中継装置を説明する。以下においては、TCP/IP packets かつ、輻輳通知にはECNを用いるものとして説明するが、DCCP (Data Congestion Control Protocol)、SCTP (Stream Control Transmission Protocol) 等の他のトランスポートプロトコルの packets、他の輻輳通知のプロトコルであっても構わない。

【0026】

まず、図2に本実施例の packets 中継装置の一構成例を示す。packets 中継装置2には、packets 入力用の入力回線20と、packets 出力用の出力回線30が接続されている。また、packets 中継装置2の内部には、packets の受信処理を行う packets 受信回路21と、入力 packets に対する packets 検索部22と、packets 検索部22の経路検索部23で宛先IPアドレスから経路検索を行って判定された出力回線番号に基づき packets をスイッチングする packets 中継処理手段26と、入力 packets に対する packets 検索部27と、packets を読み出して packets の送信処理を行う packets 送信回路28を備える。packets 検索部22内の経路検索部23で宛先IPアドレスから経路検索により出力回線番号を判定する検索方式は本実施例の構成では特定しないが、経路検索方式の一例は、例えば特許文献3に開示されている。

【0027】

なお、packets 検索部22と packets 検索部27は、受信側と送信側が異なる他は、フロー検出部24と帯域監視部25などに関し同様の構成となっている。本明細書において、packets 検索部と称する場合、packets 検索部22、packets 検索部27の何れか一方または両方を意味している。また、packets 受信回路21と packets 送信回路28は、packets ヘッダのECNフィールドとECEフィールドの書き換え機能を備えている。

【0028】

図2の構成において、入力回線20と出力回線30がそれぞれ1回線ずつ記載されているが、packets 中継装置2は、複数の入力回線20、複数の出力回線30を備える。packets 受信回路21とこれに接続された packets 検索部22は、複数の入力回線20を収容することができる。また、packets 受信回路21とこれに接続された packets 検索部22を複数備えて、それぞれが互いに異なる複数の入力回線20を収容する構成としても良い。同様に packets 送信回路28とこれに接続された packets 検索部27は、複数の入力回線を収容することができる。また、packets 送信回路28とこれに接続された packets 検索部27を複数備えて、それぞれが互いに異なる複数の出力回線30を収容する構成としても良い。また、packets 中継装置2には管理端末19が接続されており、レジスタを経由して packets 中継装置2の管理と各種設定を行う。

【0029】

packets 中継装置2は、packets 受信回路21に接続される入力回線20から packets を受信する。送信端末1から送信された packets は入力回線20を経由して packets 受信回路21に入力される。

【0030】

図3に本実施例の packets 受信回路21の一構成例を示す。packets 受信回路21は packets 受信バッファ制御部210と packets 受信バッファ211とバッファ蓄積 packets 数管理部212から構成される。packets 受信バッファ制御部210は、packets 受信バッファ211に対する読み出し制御と書き込み制御、バッファ面に対する上書きを許可する開放制御を行う。バッファ蓄積 packets 数管理部212は、packets 受信 packets 受信バッファ211に蓄積されている packets 数を計数して管理する。

【0031】

図4Aに packets 受信バッファ211の内部構成の一例を示す。packets 受信バッファ211は

10

20

30

40

50



複数のバッファ面2111～2113から構成されている。各バッファ面には、図4Bに示すパケットのパケットヘッダ情報1000として、L2ヘッダ部102として宛先MACアドレス1020、送信元MACアドレス1021、イーサタイプ1022、VLAN (Virtual Local Area Network) ID1023、L3ヘッダ部103としてIPバージョン1030、TOS (Type of Service) 1031、L4プロトコル1032、送信元IPアドレス1033、宛先IPアドレス1034、L4ヘッダ部104として送信元ポート番号1040、宛先ポート番号1041、TCPフラグ1042、そしてパケット受信回路1にて付加される内部ヘッダ101として入力回線番号1010とパケットのバイト長を示すLEN1011が書き込みされる。なお、図4Aに示す、パケットヘッダ情報1000の構成は一例であり、本実施例の構成を限定するものではない。

【 0 0 3 2 】

10

次に、図5に示すように、TOS 1031は、ネットワークにおける転送の優先度を示すDSCP (Differentiated Services Code Point) フィールド10310と、ECNによる輻輳通知に用いられるECNフィールド10311から構成される。なお、DSCPの定義に関しては非特許文献1の第3章に、DSCPが適用されるQOSアーキテクチャのDiffservモデルに関しては非特許文献2の第2章に説明されている。ECNフィールド10311は、“01”または“10”の値のときにECNをサポート (ECN: ECN - Capable - Transport) していること、“11”のときにパケットが輻輳を経験したこと (CE: Congestion Experienced) を示す。ECNをサポートしている送信端末はECNフィールドを“01”または“10”に設定したパケットを送信し、パケット中継装置2にて輻輳が発生していると判断された場合、このフィールドはパケット中継装置2により“11”に書き換えられる。

20

【 0 0 3 3 】

図6に示すように、TCPフラグ1042は、既存のフラグURG10422、ACK10423、PSH10424、RST10425、SYN10426、FIN10427に対し、ECNの拡張によりCWR10420、ECE10421が追加されている。

【 0 0 3 4 】

図2、図3に示した本実施例のパケット受信回路21は、入力回線20からパケットが入力されると、パケットが入力された順序に従って、パケット受信バッファ制御部210により、パケット受信バッファ211のバッファ面2111からバッファ面2113の順に、入力パケットのパケットヘッダ情報1000を書き込む。なお、バッファ蓄積パケット数管理部212にて管理するバッファ蓄積パケット数が全バッファ面数に相当するn個に達し、全てのバッファ面にn個のパケットヘッダ情報1000が書き込みされてしまうと、最初に書き込みされたバッファ面2111が開放されるまでの間、パケット受信回路21はパケットの受信を停止し、停止中に入力回線20から新たにパケットが入力した場合はこれを廃棄する。

30

【 0 0 3 5 】

最初に書き込みされたバッファ面2111に書き込みされたパケットが、パケット検索部22での処理を完了してパケット中継処理手段26に出力され、そのパケットヘッダ情報1000をバッファ面2111に蓄積しておく必要がなくなると、パケット受信バッファ制御部210はバッファ面2111を開放する。そして、これ以降、入力回線20から新たにパケットが入力した場合は、パケット受信バッファ制御部210が、開放されたバッファ面2111に、当該パケットのパケットヘッダ情報1000を書き込みする。

40

【 0 0 3 6 】

そして、パケット検索部22がパケットを受付可能なタイミングになると、パケット受信バッファ211に蓄積したパケットヘッダ情報1000を送信する。また、パケット検索部22での処理が完了してパケットをパケット中継処理手段26に送信するまでの間は、パケットヘッダ情報1000を一時的にパケット受信バッファ211に蓄積したままとしておく。以降、パケット中継処理手段26にパケットを新たに出力すると、その都度バッファ面2112、バッファ面2113の順でバッファ面を開放して、開放されたバッファ面に対し開放された順に入力回線20から新たに入力したパケットを蓄積する。パケット検索部22は、出力回線12を識別するための出力回線番号を判定する経路検索部3と、フローを検索するフロー検索部4を備える。

50

## 【 0 0 3 7 】

図7に、本実施例のフロー検索部24の一構成例を示す。フロー検索部24は、フロー検索テーブル41と、フロー検索テーブル41に対するテーブル書き込み、テーブル読み出し、テーブル検索起動等の各種制御を行うフロー検索テーブル制御部40とから構成される。フロー検索テーブル41は、物理的には例えばCAM (Content Addressable Memory) により構成することができる。CAMの概要については、例えば、特許文献2を参照されたい。

## 【 0 0 3 8 】

図8に、本実施例のフロー検索テーブル41の一構成例を示す。フロー検索テーブル41は、複数のフロー検索エントリ410、411、・・・418より構成される。各フロー検索エントリ410～418は、図8に示す通り、VLANID、送信元IPアドレス、宛先IPアドレス、TOS、プロトコル、送信元ポート番号、宛先ポート番号、TCPフラグ等の各種パケットヘッダの情報から構成される。なお、図8に示すフロー検索エントリを構成するパケットヘッダ情報項目はあくまで一例であり、本実施例の対象範囲を限定するものではない。

## 【 0 0 3 9 】

図8の各フロー検索エントリ410～418は、後述する帯域監視エントリと対応付けられており、各帯域監視エントリにより帯域監視する対象とするフローの条件を、フロー検索エントリで設定する。各フロー検索エントリの設定値は、パケット中継装置2の運用管理者が帯域監視の対象として決定した、フローの対象とするパケットヘッダの条件に応じて定まる。そして、パケット中継装置2の運用管理者がフローの対象とするパケットヘッダの条件を、管理端末19から入力することにより、この入力値がレジスタ9を経由して、フロー検索部24のフロー検索テーブル制御部40に入力され、フロー検索テーブル制御部40がフロー検索テーブル41に対し書き込み制御することにより、フロー検索テーブル41に設定される。

## 【 0 0 4 0 】

以下では、パケットがTCPの応答パケットでない場合、つまり当該パケットのパケットヘッダ情報1000のTCPフラグフィールド1042のACK10423が0である場合の処理を説明する。TCPの応答パケットに対するフロー検索テーブル41の制御は後述する。

## 【 0 0 4 1 】

フロー検索部24のフロー検索テーブル制御部40からフロー検索テーブル41を構成するCAMにパケットが入力して検索起動されると、CAMはフロー検索エントリのアドレスが上位の410 - aにあるフロー検索エントリ410からアドレス下位の418 - aにあるフロー検索エントリ418に向かって順にフロー検索エントリを検索して、パケットヘッダの情報の項目とフロー検索エントリに記載された情報の項目の全てが一致するフロー検索エントリを検索する。この際、「d.c.」と記載されたパケットヘッダ情報の項目に関しては、入力したパケットに記載された値を問わず当該のパケットヘッダ情報に一致と判定する。そして、最初に一致したフロー検索エントリのアドレスをヒットアドレスと判定して、ヒットアドレスをフロー検索テーブル制御部40に出力する。ヒットアドレスから、当該パケットはいずれのフロー検索エントリに一致し、後述するいずれの帯域監視エントリに対応付けられるフローであるか判定できる。

## 【 0 0 4 2 】

図9に、本実施例の帯域監視部25の一構成例を示す。フロー検索部24からのヒットアドレスの情報は、フロー検索テーブル制御部40から帯域監視部25に出力される。図9において、帯域監視部25は、帯域監視テーブル制御部50と、帯域監視テーブル51と、現在水量判定部52と、監視結果判定部53と、帯域監視エントリ毎輻輳状態管理部54から構成される。

## 【 0 0 4 3 】

帯域監視テーブル制御部50は、帯域監視テーブル51に対する読み出し制御と書き込み制御を行う。パケット中継装置2の運用管理者が決定したフロー毎の帯域監視に関する帯域の条件は、パケット中継装置2の運用管理者が管理端末9から入力することにより、この入力値がレジスタ10を経由して帯域監視部25の帯域監視テーブル制御部50に入力され、帯域監視テーブル制御部50が帯域監視テーブル51に対し書き込み制御することにより、帯域監視

視テーブル51に設定される。

【 0 0 4 4 】

図10に、本実施例の帯域監視テーブル51の一構成例を示す。帯域監視テーブル51は、フロー検索テーブル41の各フロー検索エントリに対応付けられた複数の帯域監視エントリ5101～510nから構成される。各帯域監視エントリは、監視帯域R 510と、前回パケット到着時刻511と、パケット水量CNT 512と、閾値THR 513、ECNマーキング閾値THRM 514から構成される。

【 0 0 4 5 】

パケットが入力すると、帯域監視テーブル制御部50において、フロー検索部24にて判定されたヒットアドレスを帯域監視テーブル51の読み出しアドレスとして帯域監視テーブル51を読み出し、ヒットアドレスに対応付けられた帯域監視エントリを参照する。参照された帯域監視エントリに記載のR 510、TLST 511、CNT 512、THR 513、THRM 514は各々、R蓄積手段522、TLST蓄積手段523、CNT蓄積手段524、THR蓄積手段533、THRM蓄積手段534に蓄積され、帯域監視の判定に用いられる。帯域監視のアルゴリズムは、例えば、リーキーバケツアルゴリズムである。

【 0 0 4 6 】

図11にリーキーバケツアルゴリズムの一例を示す。同図において、パケットが順次入力（1100）されているとして、このアルゴリズムに沿って説明する。まず、現在水量判定部52における演算について説明する。現在水量演算回路520では、タイマ521が示す現在時刻から、TLST蓄積手段523に蓄積されたTLSTを減算することで、経過時間Tを計算する。R蓄積手段522に蓄積されたRとTの積 $R \times T$ と、CNT蓄積手段524に蓄積されたCNTを大小比較する（1101）。その結果、 $R \times T$ の方が大きい場合は、CNTをゼロとする（1103）。その他の場合には、CNTから $R \times T$ を減算する（1102）。以上の処理により決定されたCNTは、NOWCNT蓄積手段531に蓄積する。

【 0 0 4 7 】

次に、帯域監視結果が遵守であるか違反であるかを判定する監視結果判定部53における演算について説明する。LEN蓄積手段532には、パケット受信回路1より送信された、パケットの内部ヘッダ101のLEN1011が蓄積される。LEN1011は、パケットのバイト長を示す値である。監視結果判定回路530では、NOWCNT蓄積手段531に蓄積されたCNTとTHR蓄積手段533に蓄積されたTHRを大小比較し（1104）、CNTの方が大きい場合には帯域違反となり、輻輳検出と判定してパケットを廃棄する（1105）。

【 0 0 4 8 】

次に、CNTの方が小さい場合には、NOWCNT蓄積手段531に蓄積されたCNTとTHRM蓄積手段534に蓄積されたTHRMを大小比較し（1106）、CNTの方が大きい場合には帯域違反（1109）となり、輻輳検出と判定してECNマーキングを実施する（1107）。CNTの値をCNT2蓄積手段に蓄積し、タイマ521が示す現在時刻をTLST蓄積手段537に蓄積する。そして、帯域監視テーブル51の、パケットが一致したフローの帯域監視エントリのCNT512をCNT2蓄積手段538に蓄積されたCNTで書き換え、ヒットアドレスに対応付けられた帯域監視エントリのTLST511をTLST蓄積手段537に蓄積されたタイマ値で書き換える。

【 0 0 4 9 】

上述のCNT > THRまたはCNT > THRMの条件に一致して帯域違反（1109）と判定され輻輳が検出されると、監視結果判定回路530から輻輳検出を示す情報とヒットアドレスを輻輳状態管理部54に出力する。輻輳状態管理部54は、帯域監視エントリ毎輻輳状態管理テーブル制御部540と帯域監視エントリ毎輻輳状態管理テーブル541から構成される。帯域監視エントリ毎輻輳状態管理テーブル制御部540は、帯域監視エントリ毎輻輳状態管理テーブル541に対する読み出し制御と書き込み制御を行う。

【 0 0 5 0 】

図12に、本実施例の帯域監視エントリ毎輻輳状態管理テーブル541の一構成例を示す。帯域監視エントリ毎輻輳状態管理テーブル541は、フロー検索テーブル41の各フロー検索エントリに対応付けられた複数の帯域監視エントリ毎輻輳状態管理エントリ5411～541n

10

20

30

40

50

から構成される。各輻輳状態管理エントリ5411~541nは、フロー検索エントリにより定義されたフロー毎の輻輳状態を示す輻輳状態フラグ5410から構成される。輻輳状態フラグ5410は、初期状態では輻輳状態に無いことを示す値となっている。輻輳状態管理部54は、監視結果判定回路530から輻輳検出を示す情報とヒットアドレスを入力されると、帯域監視エントリ毎輻輳状態管理テーブル制御部540によりヒットアドレスを書き込みアドレスとして帯域監視エントリ毎輻輳状態管理テーブル541の帯域監視エントリ毎輻輳状態管理エントリ5411~541nの輻輳状態フラグ5410に輻輳検出を示す情報を書き込む。

【0051】

図11において、 $CNT > THR$ または $CNT > THRM$ の条件に一致しない場合には帯域遵守(1110)となり、CNTにLENを加算した値をCNT2蓄積手段538に蓄積し(1108)、タイマ521が示す現在時刻をTLST蓄積手段537に蓄積する。そして、帯域監視テーブル51の、ヒットアドレスに対応付けられた帯域監視エントリのCNT512をCNT2蓄積手段538に蓄積されたCNTで書き換え、ヒットアドレスの帯域監視エントリのTLST511をTLST蓄積手段537に蓄積されたタイマの値で書き換える。また監視結果判定回路530から輻輳状態に無いことを示す情報が出力された場合には、帯域監視エントリ毎輻輳状態管理テーブル541の帯域監視エントリ毎輻輳状態管理エントリ5411~541nの輻輳状態フラグ5410の書き換え処理は実施しない。

【0052】

なお帯域監視のアルゴリズムとして、クレジット方式によるリーキーバケツアルゴリズムの他、ウインドウ方式によるJumping Windowアルゴリズムも知られており(非特許文献3参照)、これを用いて帯域監視しても良い。その場合は、時間ウインドウW毎にパケット受信回路21より送信されたパケットのバイト長の積算値Bと、時間ウインドウWの間に許容されるバイト数 $W \times R$ とを大小比較し、Bの方が大きい場合には帯域違反、それ以外の場合には帯域遵守として、リーキーバケツアルゴリズムの際と同様な制裁の処理を行う。

【0053】

$CNT > THR$ に一致せず、かつ $CNT > THRM$ に一致する場合は、ECNマーキングを実施(1107)することを示す情報をパケット受信回路21へ出力する。

【0054】

図13に、以上に説明した、リーキーバケツアルゴリズムにおけるパケット水量CNT12と、閾値THR513、ECNマーキング閾値THRM514の関係を示す。

パケット受信回路21では、ECNマーキングを実施することを示す情報を入力されると、パケット受信回路21が備えるパケット受信バッファ211に蓄積された当該パケットのパケットヘッダ情報1000における図5のTOSフィールド1031のECNフィールド10311の値を、パケット受信バッファ制御部210により読み出し、読み出されたECNフィールド10311の値が10または01であった場合には11に書き換える。読み出されたECNフィールド10311の値が00または11の場合には書き換えしない。以上が、本実施例のパケット中継装置2に入力したパケットがデータパケットであった場合の、帯域監視に関する処理の一例である。

【0055】

次に、本実施例のパケット中継装置2に入力したパケットがTCPのACKパケットである場合、つまり図6に示した、当該パケットのパケットヘッダ情報1000のTCPフラグフィールド1042のACK10423が1である場合の処理を説明する。送信端末1から受信端末3に送信されるデータパケットに対するTCPのACKパケットでは、データパケットの送信元IPアドレスがACKパケットの宛先IPアドレスに、データパケットの宛先IPアドレスがACKパケットの送信元IPアドレスに、データパケットの送信元ポート番号がACKパケットの宛先ポート番号に、データパケットの宛先ポート番号がACKパケットの送信元ポート番号になる。

【0056】

従って、フロー検索部24のフロー検索テーブル制御部40で入力パケットのACK=1であると判定された場合は、送信元IPアドレスと宛先IPアドレス、送信元ポート番号と宛先ポート番号を交換した条件で、フロー検索テーブル制御部40からフロー検索テーブル41に対する検索起動をかける。すると、ACKパケットに対応するデータパケットが一致するフロー検索エントリのヒットアドレスがフロー検索テーブル41から出力される。このヒットアド

10

20

30

40

50

レスを、フロー検索テーブル制御部40から帯域監視部25の帯域監視エントリ毎輻輳状態管理テーブル制御部540に出力し、このヒットアドレスを読み出しアドレスとして帯域監視エントリ毎輻輳状態管理テーブル541を読み出し制御する。

【0057】

すると、ACKパケットに対応するデータパケットに対する帯域監視エントリに対応する輻輳状態フラグが帯域監視エントリ毎輻輳状態管理テーブル制御部540に読み出され、ACKパケットに対応するデータパケットの帯域監視に関する輻輳状態を判定することができる。この輻輳状態フラグが輻輳状態にあることを示す値である場合には、ACKパケットのECEフラグを ' 1 ' に書き換えるECEマーキングを実施することを示す情報をパケット受信回路21へ出力する。

10

【0058】

パケット受信回路21では、ECEマーキングを実施することを示す情報を入力されると、パケット受信回路21が備えるパケット受信バッファ211に蓄積された当該パケットのパケットヘッダ情報1000におけるTCPフラグフィールド1042のECEフィールド10421の値を1に書き換える。これにより、ACKパケットに対するECEマーキングを、パケット中継装置2における帯域監視の輻輳状態に基づいて、受信端末3ではなくパケット中継装置2で実施することができる。すなわち、パケット中継装置2は、ACKパケットのパケットヘッダのうち、ネットワークの輻輳状態を示すフィールドとして、TCPパケットのECEフラグを用いる。その結果、パケット中継装置2から受信端末3、受信端末3からパケット中継装置2、パケット中継装置2から送信端末1の経路で輻輳通知することができる。輻輳状態フラグが輻輳状態に無いことを示す値である場合には、このECEマーキングは実施しない。

20

【0059】

本実施例のパケット中継装置に入力したパケットがCWR = 1のデータパケットDATA (CWR) であった場合には、帯域監視エントリ毎輻輳状態管理テーブル制御部540は、帯域監視エントリ毎輻輳状態管理テーブル541において当該パケットが属するフローに対応する帯域監視エントリ毎輻輳状態管理エントリ5411 ~ 541nの輻輳状態フラグ5410の値を輻輳状態に無いことを示す値に書き換える。すなわち、本実施例のパケット中継装置は、受信したTCPパケットのCWRフラグが1である場合に、フローが輻輳状態から解消したと判定することを意味する。これにより、それ以降パケット中継装置が受信したACKパケットに対するECEマーキングは実施されないこととなる。

30

【0060】

以上、実施例1の受信側のパケット検索部22が備える帯域監視部25にて、輻輳状態の検出と輻輳通知の指示を行う構成を説明したが、送信側のパケット検索部27でも同様に帯域監視部25を備えることができ、送信側の帯域監視に本実施例の構成を適用することもできる。

【0061】

次に、本実施例のパケット受信回路にて、輻輳状態の検出と輻輳通知の指示を行う場合の変形例について説明する。なお、パケットの優先度を考慮する拡張的なバッファ構成に関する適用例を開示するため、パケット受信回路はパケットの優先度に応じたバッファを複数備え、バッファ毎に輻輳状態の検出と輻輳通知の指示を行う構成とする。また、パケットの優先度はフロー検索部にて決定する構成とする。

40

【0062】

まず、図14に、本変形実施例において、パケットの優先度を決定するフロー検索部の一構成例を示す。フロー検索部140は、先述した図7のフロー検索部24の構成に加え、フローアクションテーブル制御部142とフローアクションテーブル143を備える。

【0063】

図15にフローアクションテーブル143の一構成例を示す。フローアクションテーブル143は、フロー1~n毎の優先度に応じてパケット受信回路21のパケット受信バッファ211にて蓄積するバッファ番号を示すフローアクションエントリ1431 ~ 143nから構成される。フロー検索部24にパケットが入力すると、フロー検索テーブル制御部40によりパケットヘッダ情

50

報1000がフロー検索テーブル41に入力され、当該のパケットヘッダ情報1000に一致するフロー検索エントリのヒットアドレスがフロー検索テーブル41を構成するCAMからフロー検索テーブル制御部40に出力される。このヒットアドレスをフロー検索テーブル制御部40からフローアクションテーブル制御部142へ出力する。フローアクションテーブル制御部142にて、このヒットアドレスをフローアクションテーブル143に対する読み出しアドレスとしてフローアクションテーブル143を読み出し制御する。すると、当該パケットに対応付けられるフローアクションエントリが読み出され、当該パケットを蓄積すべきパケット受信回路21のパケット受信バッファ211のバッファ番号が判定できる。

【0064】

図16に本実施例のパケット受信回路の他の構成例を示す。パケット受信回路160は、図3に示したパケット受信回路21の基本構成、すなわち、パケット受信バッファ制御部1600、パケット受信バッファ1601、バッファ蓄積パケット数管理部1602に加え、ECNマーキング可否判定部1603を備える。パケット受信バッファ1601は、パケット受信バッファ211を拡張したもので、パケットの優先度に応じた複数のバッファを備える。

【0065】

図17にパケット受信バッファ1601の一構成例を示す。パケット受信バッファ1601は、バッファ面0～バッファ面mのm+1個の面をもつバッファ16011～バッファ1601nから構成される。パケットがパケット受信回路21に入力すると、パケット受信バッファ制御部210にてパケットのバッファ番号が示すバッファ1601i (i:1～n) に対し、パケットのヒットアドレスを書き込みアドレスとしてパケットヘッダ情報1000をバッファ面0から順に書き込み制御する。そして、パケット受信バッファ毎に蓄積しているパケット数をバッファ蓄積パケット数管理部1602にて計数する。バッファ蓄積パケット数管理部1602にて計数されるバッファ毎蓄積パケット数は、パケット受信バッファ制御部1600を経由してECNマーキング閾値テーブル制御部1603に出力される。ECNマーキング閾値テーブル制御部1603では、パケットのバッファ番号を読み出しアドレスとしてECNマーキング閾値テーブル1604を読み出し制御する。

【0066】

図18にECNマーキング閾値テーブル1604の一構成例を示す。ECNマーキング閾値テーブル1604は、バッファ番号毎のECNマーキング閾値16041～1604nから構成される。ECNマーキング閾値テーブル制御部1603では、パケットのバッファ番号を読み出しアドレスとしてECNマーキング閾値テーブル1604を読み出すと、当該パケットが蓄積されるバッファ番号に対するECNマーキング閾値が読み出される。このECNマーキング閾値がECNマーキング閾値テーブル制御部1603を経由してECNマーキング可否判定部1605に出力される。ECNマーキング可否判定部1605では、バッファ毎蓄積パケット数とECNマーキング閾値の大小比較の判定がなされ、ECNマーキング閾値<バッファ毎蓄積パケット数を満たす場合にECNマーキングを実行すべきと判定する。ECNマーキングの実行可否の情報をECNマーキング可否判定部1605から出力して、ECNマーキング閾値テーブル制御部1603を経由してパケット受信バッファ制御部1600に出力する。

【0067】

またこのとき、当該パケットが蓄積されているバッファ番号を読み出しアドレスとして、バッファ毎輻輳状態管理テーブル制御部1606からバッファ毎輻輳状態管理テーブル1607を読み出し制御する。バッファ毎輻輳状態管理テーブル1607の構成は、図12に示す帯域監視エントリ毎輻輳状態管理テーブル541と同様であるが、帯域監視エントリ毎輻輳状態管理テーブル541は帯域監視エントリ毎に輻輳状態フラグを備えていたのに対し、バッファ毎輻輳状態管理テーブル1607はバッファ番号毎に輻輳状態フラグを備える点異なる。そして、バッファ毎輻輳状態管理テーブル制御部1606により、当該パケットに対応するデータパケットを蓄積しているバッファ番号に対する輻輳状態フラグを、輻輳を示す値に書き換える制御がなされる。

【0068】

次に、パケット受信バッファ制御部1600では、パケット中継装置2に入力したパケット

10

20

30

40

50

がデータパケットであり、ECNマーキングを実行すべきと判定された場合には、パケット受信回路160が備えるパケット受信バッファ1601に蓄積された当該パケットのパケットヘッダ情報1000におけるTOSフィールド1031のECNフィールド10311の値をパケット受信バッファ制御部1600により読み出し、読み出されたECNフィールドの値が10または01であった場合には11に書き換える。読み出されたECNフィールド10311の値が00または11の場合には書き換えしない。

【0069】

本実施例のパケット中継装置2に入力したパケットがTCPのACKパケットであった場合には、送信元IPアドレスと宛先IPアドレス、送信元ポート番号と宛先ポート番号を交換した条件で、フロー検索部140のフロー検索テーブル制御部40からフロー検索テーブル41に対する検索起動をかける。すると、ACKパケットに対応するデータパケットが一致するフロー検索エントリのヒットアドレスがフロー検索テーブル41からフロー検索テーブル制御部40を経由してフローアクションテーブル制御部142に出力され、このヒットアドレスを読み出しアドレスとしてフローアクションテーブル143を読み出し制御する。

【0070】

すると、このACKパケットに対応するデータパケットを蓄積しているバッファ番号を読み出され、このバッファ番号をフローアクションテーブル制御部142、フロー検索テーブル制御部40を経由して、フロー検索部140からパケット受信回路160に送信し、パケット受信バッファ制御部1600を経由してバッファ毎輻輳状態管理テーブル制御部1606に出力する。ACKパケットに対応するデータパケットを蓄積しているバッファ番号を読み出しアドレスとして、バッファ毎輻輳状態管理テーブル制御部1606からバッファ毎輻輳状態管理テーブル1607を読み出し制御する。

【0071】

そして、ACKパケットに対応するデータパケットを蓄積しているバッファ番号に対する輻輳状態フラグがバッファ毎輻輳状態管理テーブル制御部1606に読み出され、バッファ毎輻輳状態管理テーブル制御部1606からパケット受信バッファ制御部1600に出力されると、この輻輳状態フラグが輻輳状態にあることを示す値である場合には、パケット受信バッファ1601に蓄積された当該パケットのパケットヘッダ情報1000におけるTCPフラグフィールド1042のECEフィールド10421の値を1に書き換える。これにより、ACKパケットに対するECEマーキングを、パケット中継装置におけるパケット受信バッファ1601の輻輳状態に基づいて、受信端末3ではなくパケット中継装置2で実施することができる。従って、パケット中継装置から受信端末、受信端末からパケット中継装置、パケット中継装置から送信端末の経路で輻輳通知することができる。輻輳状態フラグが輻輳状態に無いことを示す値である場合には、ECEマーキングは実施しない。

【0072】

本実施例のパケット中継装置に入力したパケットがCWR=1のデータパケットDATA(CWR)であった場合には、バッファ毎輻輳状態管理テーブル制御部1606は、バッファ毎輻輳状態管理テーブル1607において当該パケットが属するバッファに対応するバッファ毎輻輳状態管理エントリの輻輳状態フラグの値を輻輳状態に無いことを示す値に書き換える。これにより、それ以降パケット中継装置が受信したACKパケットに対するECEマーキングは実施されないこととなる。

【0073】

図19に、以上に説明した本実施例の受信側のパケット受信回路160が備えるバッファ面nのパケット受信バッファ1601のバッファ管理におけるECNマーキング閾値とバッファ蓄積パケット数の関係を模式的に示した。

【0074】

以上、受信側のパケット受信回路160が備えるパケット受信バッファ1601のバッファ管理において、輻輳状態の検出と輻輳通知の指示を行う構成で説明したが、図2の送信側のパケット送信回路28でも同様にパケット送信バッファを備えることができ、送信側のバッファ管理に本実施例の構成を適用することもできる。特に送信側のバッファ管理において

、バッファの送信帯域を制御可能としたものをシェーパと呼ぶ。

【 0 0 7 5 】

図20に、シェーパを構成するパケット送信回路の一構成例を示す。パケット送信回路2000の構成は、図16に示したパケット受信回路160におけるパケット受信バッファ1601、パケット受信バッファ制御部1600をパケット送信バッファ2001、パケット送信バッファ制御部2000と読み替えた同様の構成に加え、タイマ2017、送信時刻計算回路2018、回線毎送信帯域管理テーブル制御部2019、回線毎送信帯域管理テーブル2020を備える。パケット送信バッファ2001は、パケット受信バッファ2011とは異なりパケットの優先度毎には備えず、回線毎に備える。

【 0 0 7 6 】

図21に、回線毎送信帯域管理テーブル2020の一構成例を示す。回線毎送信帯域管理テーブル2020には、複数の出力回線30毎の送信帯域Rの設定値と遅延タイマ値OTIMEと最後にパケット送信した際のタイマ2017のタイマ値である最終送信タイマ値TLSTとが、回線毎送信帯域管理エントリ20201～2020nに記載される。送信帯域Rの設定値は、出力回線30の回線種別や、パケット中継装置の運用管理者が出力回線30毎に割り当てたい送信帯域の値に応じて決定される。パケット中継装置の運用管理者が、送信帯域Rの設定値を管理端末19から入力し、レジスタ29を経由して回線毎送信帯域管理テーブル制御部2019から回線毎送信帯域管理テーブル2020に書き込み制御することにより設定される。

【 0 0 7 7 】

図22のフローチャートを用いて、本実施例のパケット送信回路の動作について説明する。同図において、バッファ毎のパケット処理が開始されると(2200)、パケット送信回路2000のパケット送信バッファ制御部2000では、タイマ2017が出力するタイマ値に基づいて、現在のタイマ値が基準タイマ値に達したか判断する(2201)。基準タイマ値とは、パケット送信回路2000がパケット送信処理またはパケット受信処理を実行する単位時間である。基準タイマ値は、パケット中継装置0が送受信するパケットの基準バイト数と、送信帯域Rの最大値とに基づいて、以下の式(1)により算出される。

【 0 0 7 8 】

基準タイマ値[s] = 基準バイト数 \* 8[bit] / 送信帯域Rの最大値[bps]・・・(1)

なお、本実施例では、[s]は秒ではなくタイマ値を示す。

【 0 0 7 9 】

現在のタイマ値が基準タイマ値に達したと判断された場合には、受信処理および送信処理を実行する(2202以降)。現在のタイマ値が基準タイマ値に達していないと判断された場合には、現在のタイマ値が基準タイマ値に達したか判断する処理をループする。つまり、本実施例では、上述した基準タイマ値ごとに、受信処理と送信処理とが繰り返し実行されることになる。

【 0 0 8 0 】

受信処理(2202)について説明する。受信処理の実行が開始されると、まず、パケット中継処理手段26からパケットを受信したか否かが判断され(2203)、パケットを受信した場合には、パケット送信バッファ制御部2000によって、そのパケットのバッファリング先が判定される(2204)。パケットを受信しなかった場合には、後述する送信処理(2208)に処理が移行される。

【 0 0 8 1 】

パケットのバッファリング先が判定されると、送信バッファ制御部2000によって、パケット送信バッファ2001が参照され、そのバッファリング先に該当するパケット送信バッファ2001にパケットが蓄積されているかが判断される(2205)。該当バッファにパケットが蓄積されていない場合には、送信時刻計算回路2018により当該バッファからパケットの送信を遅延させる時間(遅延タイマ値OTIME)が計算される。本実施例では、パケット中継処理手段26から送信されたパケットのパケット長LENを、回線毎の送信帯域Rで除算することにより遅延タイマ値OTIMEを計算する(2206)。回線毎の送信帯域Rは、回線毎送信帯域管理テーブル制御部2019によりパケットの出力回線に応じて回線毎送信帯域管理テーブル

10

20

30

40

50



2020を読み出すことにより得られる。この遅延タイム値OTIMEは、以下の式(2)により算出される。この式(2)によれば、遅延タイム値OTIMEはそのパケットを次の機会に送信することができる最も早い時間を示すことになる。

【0082】

$$OTIME = LEN / R$$

・・・(2)

送信時刻計算回路2018により遅延タイム値OTIMEが計算されると、その値が回線毎送信帯域管理テーブル8-20内の該当バッファの回線毎送信帯域管理エントリに記載され、パケット送信バッファ2001の該当バッファの末尾に、受信されたパケットがパケット送信バッファ制御部2000によって蓄積される(2207)。なお、該当キューにパケットが蓄積されている場合には、遅延タイム値OTIMEは計算されずに、受信されたパケットがパケット送信バッファ2001の該当バッファの末尾にそのまま蓄積される。既にパケットが蓄積されている場合には、前回のバッファ毎のパケット処理において、遅延タイム値OTIMEは既に計算済みであり、回線毎送信帯域管理テーブル2020にその遅延タイム値OTIMEが記録されているからである。以上の受信処理が完了すると、引き続いて送信処理が実行され(2208)、該当バッファに蓄積パケットが有るか否かがチェックされる(2209)。

【0083】

蓄積パケットが確認されると、まず、送信時刻計算回路2018によって、タイム2017から出力された現在のタイム値TNOWと、最終送信タイム値TLSTに遅延タイム値OTIMEを加算した送信予定タイム値とが比較される(2210)。

【0084】

現在のタイム値TNOWと送信予定タイム値とが一致するか、もしくは、現在のタイム値が送信予定タイム値を超えるという条件を満たすバッファが存在する場合には、そのバッファは、パケットを送信する対象の候補となる(以下、「送信対象バッファ」という)。送信時刻計算回路2018は上述の条件に基づいて、送信対象バッファが存在するかどうかを判断する。送信対象バッファが存在しない場合には、当該送信処理は終了する。

【0085】

送信対象バッファが存在する場合には、送信時刻計算回路2018は、それらの送信対象バッファのうち、最も送信予定タイム値の早いバッファを特定する。そして、最も送信予定タイム値の早いバッファが特定されると、パケット送信バッファ制御部2000によって、そのバッファから、パケットの送信が行われる。パケットの送信が行われると、回線毎送信帯域管理テーブル制御部2019は、回線毎送信帯域管理テーブル2020の該当回線の最終送信タイム値TLSTを現在のタイム値TNOWに更新する(2211)。

【0086】

最終送信タイム値TLSTbの更新が行われると、続いて、パケット送信バッファ制御部2000によって、上述の最も送信予定タイム値の早いバッファとして特定されたバッファに、次の蓄積パケットがあるかが判断される(2212)。次の蓄積パケットがなければ、当該送信処理は終了する。一方、次の蓄積パケットが存在していれば、送信時刻計算回路2018によって、そのバッファに関する新たな遅延タイム値OTIMEが計算される。この遅延タイム値DTは、蓄積されているパケットのパケット長LENを、その回線の送信帯域Rで割った値となる(2213)。新たな遅延タイム値OTIMEが計算されると、その値は回線毎送信帯域管理テーブル制御部2019により、回線毎送信帯域管理テーブル2020の該当回線のエントリに書き込まれる。

【0087】

以上で説明した一連のシェーパ処理におけるECNマーキングの処理は、パケット受信回路160における処理と同様である。図20のパケット送信回路200のパケット受信処理において、バッファリングがなされると、パケット送信バッファ毎に蓄積しているパケット数をバッファ蓄積パケット数管理部2002にて計数する。バッファ蓄積パケット数管理部2002にて計数されるバッファ毎蓄積パケット数は、パケット送信バッファ制御部2000を經由してECNマーキング閾値テーブル制御部2012に出力される。

【0088】

10

20

30

40

50

ECNマーキング閾値テーブル制御部2012では、ECNマーキング閾値テーブル2013を読み出し制御する。ECNマーキング閾値テーブル2013を読み出すと、当該パケットが蓄積されるバッファに対するECNマーキング閾値が読み出される。このECNマーキング閾値がECNマーキング閾値テーブル制御部2012を経由してECNマーキング可否判定部2014に出力される。ECNマーキング可否判定部2014では、バッファ毎蓄積パケット数とECNマーキング閾値の大小比較の判定がなされ、ECNマーキング閾値<バッファ毎蓄積パケット数を満たす場合にECNマーキングを実行すべきと判定する。ECNマーキングの実行可否の情報をECNマーキング可否判定部2014から出力して、ECNマーキング閾値テーブル制御部2012を経由してパケット送信バッファ制御部2000に出力する。

【0089】

パケット送信バッファ制御部2000では、当該パケットがデータパケットであり、ECNマーキングを実行すべきと判定された場合には、パケット送信回路200が備えるパケット送信バッファ2001に蓄積された当該パケットのパケットヘッダ情報1000におけるTOSフィールド1031のECNフィールド10311の値をパケット送信バッファ制御部2000により読み出し、読み出されたECNフィールドの値が10または01であった場合には11に書き換える。読み出されたECNフィールド10311の値が00または11の場合には書き換えしない。

【0090】

当該パケットがTCPのACKパケットであった場合には、送信元IPアドレスと宛先IPアドレス、送信元ポート番号と宛先ポート番号を交換した条件で、送信側のパケット検索部27におけるフロー検索部24のフロー検索テーブル制御部40からフロー検索テーブル41に対する検索起動をかける。すると、ACKパケットに対応するデータパケットが一致するフロー検索エントリのヒットアドレスがフロー検索テーブル41からフロー検索テーブル制御部40を経由してフローアクションテーブル制御部142に出力され、このヒットアドレスを読み出しアドレスとしてフローアクションテーブル143を読み出し制御する。

【0091】

すると、このACKパケットに対応するデータパケットを蓄積しているバッファ番号が読み出され、このバッファ番号をフローアクションテーブル制御部142、フロー検索テーブル制御部40を経由して、フロー検索部24からパケット送信回路200に送信し、パケット送信バッファ制御部2000を経由してバッファ毎輻輳状態管理テーブル制御部2015に出力する。ACKパケットに対応するデータパケットを蓄積しているバッファ番号を読み出しアドレスとして、バッファ毎輻輳状態管理テーブル制御部2015からバッファ毎輻輳状態管理テーブル2016を読み出し制御する。

【0092】

そして、ACKパケットに対応するデータパケットを蓄積しているバッファに対する輻輳状態フラグがバッファ毎輻輳状態管理テーブル制御部2015に読み出され、バッファ毎輻輳状態管理テーブル制御部2015からパケット送信バッファ制御部2000に出力されると、この輻輳状態フラグが輻輳状態にあることを示す値である場合には、パケット送信バッファ2011に蓄積された当該パケットのパケットヘッダ情報1000におけるTCPフラグフィールド1042のECEフィールド10421の値を1に書き換える。これにより、ACKパケットに対するECEマーキングを、パケット中継装置におけるパケット送信バッファ2001の輻輳状態に基づいて、受信端末ではなくパケット中継装置で実施することができる。従って、パケット中継装置から受信端末を経由しパケット中継装置、そして、パケット中継装置から送信端末の経路で輻輳通知することができる。輻輳状態フラグが輻輳状態に無いことを示す値である場合には、ECEマーキングは実施しない。

【0093】

当該パケットがCWR=1のデータパケットDATA(CWR)であった場合には、バッファ毎輻輳状態管理テーブル制御部2015は、バッファ毎輻輳状態管理テーブル2016において当該パケットが属するバッファに対応するバッファ毎輻輳状態管理エントリの輻輳状態フラグの値を輻輳状態に無いことを示す値に書き換える。これにより、それ以降パケット中継装置が受信したACKパケットに対するECEマーキングは実施されないこととなる。

10

20

30

40

50

## 【 0 0 9 4 】

以上詳述した本実施例のパケット中継装置におけるECNの処理と、送信端末、受信端末とパケット中継装置間でなされるECNの通信の概要をシーケンス図にまとめて説明する。

## 【 0 0 9 5 】

図23に、本実施例のパケット中継装置による通信のフローチャート図を示す。図23で明らかのように、入力パケットがACKパケットか否かの判定（2301）、入力パケットがACKパケットである場合には、当該フローに対する輻輳状態管理テーブルの輻輳状態管理エントリを参照し（2302）、輻輳状態であるならば（2303）、当該ACKパケットのECEフィールドを1に更新する処理（2304）、入力パケットがDATA（CWR）である場合には（2305）、当該フローの非輻輳状態を輻輳状態管理テーブルに記録する処理（2306）、ECNフィールドを1に更新する場合には（2311）、当該フローの輻輳状態を輻輳状態管理テーブルに記録する処理（2310）が追加となる。

10

## 【 0 0 9 6 】

図24に、本実施例のパケット中継装置を用いた通信のシーケンス図を示す。送信端末1から、例えばルータ（Router）等で構成されるパケット中継装置2を経由して受信端末3の間でパケットの送受信がなされるものとする。ECNの初期化に関わる通信として、初めにTCP制御フラグのECEフラグとCWRフラグを立てたECN - setup SYNパケットが送信端末1から受信端末3に送信される。これは、送信端末1がECN通信を開始しようとすることを受信端末3に示すパケットである。受信端末3がECN - setup SYNパケットを受信すると、受信端末3は、ECNをサポートしている場合にはECEフラグを立てたECN - setup SYN - ACKパケットを送信端末1に返信する。受信端末3がECN通信をサポートしていない場合には、ECEフラグを立てない通常のTCPの3 way handshakeにおけるSYN - ACKパケットを送信端末1に返信する。従って、SYN - ACKパケットのECEフラグの有無により、送信端末1は受信端末3のECNサポート可否を判定でき、その後ECN通信を継続できるか否か判定する。そして、送信端末1通常のTCPの3 way handshakeにおけるACKパケットを受信端末3に返信する。その後、ECNのデータ通信が開始される。受信端末3からのACKパケットを送信端末1が受信した際の送信端末1によるCongestion Windowを増加させる制御、つまりパケット送信帯域を増加させる制御は、通常のTCPと同様である。DATA（ECT）は、ECN - Capable - Transport、つまりECNサポートするデータパケットであり、ECNフィールドが01または10のパケットである。

20

## 【 0 0 9 7 】

先に詳述したように、本実施例のパケット中継装置2では輻輳（Congestion）をパケット中継装置2が検出すると、その後にパケット中継装置2が受信したACK（SEQ = 4001）をパケット中継装置2にてACK（ECE）に書き換える処理が追加となる。従来構成では、送信端末1はSEQ = 5001のACK（ECE）により輻輳回避制御を開始したのに対し、図24に示すように、SEQ = 4001のACK（ECE）により輻輳回避制御を開始できるので、この分だけ早期に輻輳回避制御がなされることとなる。

30

## 【 0 0 9 8 】

この結果、従来のECNのシーケンスでは、輻輳通知が間に合わず、SEQ = 4001のACKによってSEQ = 7001 - 8000のDATA（ECT）が送信されてしまう。そして、SEQ = 7001 - 8000のDATA（ECT）は、パケット中継装置2における輻輳中に送信されているので、輻輳悪化を招き、パケット中継装置2において廃棄される。廃棄されたSEQ = 7001 - 8000のDATA（ECT）に対する重複ACKまたはタイムアウトが送信端末1にて検出されると、送信端末1はSEQ = 7001 - 8000のDATA（ECT）を再送するので、送信端末1とパケット中継装置間の帯域を余分に消費することとなり、実質的なスループットが低下してしまう。

40

## 【 0 0 9 9 】

これに対し、本実施例のパケット中継装置2では、図24に示すように、SEQ = 4001のACK（ECE）によって早期に輻輳通知できるので、SEQ = 7001 - 8000のDATA（ECT）の送信を遅延させることができる、この結果、送信端末1は、輻輳中の過剰なパケット送信を抑止することができて、再送パケットによるスループット低下を防止できる。パケット中継装置2と受信端末3間の遅延が大きくなるほどこの効果は高くなり、従来のECNよりもより早期

50

に輻輳通知して輻輳回避制御できるので、より多くのパケット再送を抑止して、スループット低下を効果的に防止できる。そして、SEQ = 4001のACK (ECE) を送信端末1が受信して、Congestion Windowを減少させて送信帯域を抑制する輻輳回避制御を行った後、SEQ = 7001 - 8000のDATA (CWR) を受信端末3に送信する。DATA (CWR) はCongestion Window Reduced、つまりCongestion Windowを減少させて輻輳回避制御が送信端末1においてなされたことを示すデータパケットである。中継装置がDATA (CWR) を受信すると、中継装置内の輻輳状態管理テーブルにおいて、当該パケットが一致するフローに対する輻輳状態フラグを、輻輳状態でないことを示す値に書き換える。そして、受信端末3がDATA (CWR) を受信すると、受信端末3はACK (ECE) の送信を停止して、これ以降はECEフラグを立てないACKパケットを送信する。

10

**【0100】**

ACK (ECE) を送信端末1が受信すると、送信端末1はCongestion Windowを減少させて送信帯域を抑制する輻輳回避制御を行い、その後DATA (CWR) を受信端末3に送信する。DATA (CWR) はCongestion Window Reduced、つまりCongestion Windowを減少させて輻輳回避制御が送信端末1においてなされたことを示すデータパケットである。受信端末3がDATA (CWR) を受信すると、受信端末3はACK (ECE) の送信を停止して、ECEフラグを立てないACKパケットを送信する。

**【0101】**

なお、輻輳検出処理のタイミングは、パケット受信時であっても良いし、定期的に検出しても良い。

20

**【0102】**

以上詳述したように、実施例1のパケット中継装置によれば、早期に輻輳通知できるので、送信端末は、輻輳中の過剰なパケット送信を抑止することができて、再送パケットによるスループット低下を防止できる。パケット中継装置と受信端末間の遅延が大きくなるほどこの効果は高くなり、従来のECNよりもより早期に輻輳通知して輻輳回避制御できるので、より多くのパケット再送を抑止して、スループット低下を効果的に防止できる。

**【実施例2】****【0103】**

続いて、図25を用いて、第2の実施例のパケット中継装置について説明する。実施例2の説明では、実施例1との差分のみを説明する。実施例2として、パケット中継装置に入力するパケットのパケットヘッダ情報を、パケット中継装置にて自律的に収集し、入力パケットのパケットヘッダ情報に基づいて、フローの条件をパケット中継装置にて自律的にフロー検索テーブル41に登録する構成の実施例を説明する。これにより、パケット中継装置の運用管理者が予めフロー条件を設定する必要なく、入力パケットに対し自律的な中継が可能となる。更に、パケット中継装置にて自律的に登録するフローの対象をパケット中継装置の運用管理者が指定できる構成を説明する。これにより、フロー検索テーブル41や輻輳状態管理テーブルの物理的資源の制約の下で、適用対象とするフローをパケット中継装置の運用管理者が効率的に指定することができる。以下、輻輳検出は帯域監視によるものとして説明するが、実施例1と同様、バッファやシェーパによる輻輳検出に対しても、同様に本実施例の構成を適用できる。すなわち、シェーパにおけるキュー長、またはポリサにおけるパケット水量等を、バッファにおける蓄積パケット数が所定の閾値を超過するか否かの判断の代わりにして、フローが輻輳状態にあると判定することができる。

30

40

**【0104】**

実施例2では、図8に示したフロー検索テーブル41の最下位アドレスに位置するフロー検索エントリ418 - aに、登録対象とするフローの条件を設定する。同図に示すフロー検索エントリ418 - aのように、登録対象とするフロー検索エントリの全てのフローの条件を "d. c." に設定すると、任意のフローを登録対象として帯域監視して、任意のフローの輻輳状態を管理することができる。

**【0105】**

また、フロー検索エントリ417 - aのように、登録対象とするフローの条件を特定すると

50

、特定されたフローだけを登録対象として帯域監視して、特定のフローだけの輻輳状態を管理することができる。輻輳状態を管理する特定のフローは、予めパケット中継装置の運用管理者が指定する。フロー検索テーブル41や輻輳状態管理テーブルの容量は有限の物理的資源として制約されるので、この制約に応じた数のフローをパケット中継装置の運用管理者が指定する。登録用フローの条件が複数ある場合には、複数のフロー検索エントリに登録対象とするフローの条件を設定しても構わない。

【0106】

実施例2におけるフロー検索部140のフロー検索テーブル41には、登録対象フローを指定するフロー検索エントリである登録用フロー検索エントリと、実施例1にて説明した帯域監視エントリを参照するためのフロー検索エントリである帯域監視用フロー検索エントリとが設定される。この二種類のフロー検索エントリを区別するため、実施例2におけるフロー検索部も実施例1におけるフロー検索部140と同様にフローアクションテーブル制御部とフローアクションテーブルとを備える。ただし、フローアクションテーブル144の構成は、実施例1における図15のフローアクションテーブル143の構成とは異なり、図25のフローアクションテーブル144の構成を有する。

10

【0107】

図25に示すように、本実施例のフローアクションテーブル144は、図15のようなバッファ番号の情報ではなく、フロー検索エントリが登録フロー検索エントリであるか、帯域監視フロー検索エントリであるかを示す情報から構成される。同図において、「帯域監視」と記載されたフローアクションエントリ1441、1442は、帯域監視用フロー検索エントリに対応して設定され、「登録」と記載されたフローアクションエントリ144nは、登録用フロー検索エントリに対応して設定される。

20

【0108】

パケット中継装置2の運用管理者が指定した本実施例の登録対象とするフローの条件の情報を管理端末19から入力すると、この入力値がレジスタ29を経由してフロー検索部140のフロー検索テーブル制御部40に入力され、フロー検索テーブル制御部40がフロー検索テーブル41に対し書き込み制御することにより、フロー検索テーブル41にフロー登録用フロー検索エントリが設定される。

【0109】

一方、本実施例のパケット中継装置2にパケットが入力して、当該パケットのパケットヘッダの情報がフロー検索テーブル41のフロー検索エントリに記載されたフローの条件に一致し、当該フロー検索エントリのアドレスを読み出しアドレスとしてフローアクションテーブル制御部142から読み出したフローアクションテーブル144のフローアクションエントリに「登録」を指示する情報が記載されていた場合は、入力パケットは登録用フロー検索エントリに一致したとフローアクションテーブル制御部142にて判定され、登録用フロー検索エントリ一致を示す情報と入力パケットのパケットヘッダ情報はフロー検索テーブル制御部40を経由してレジスタ29に送信され、管理端末19に送信される。

30

【0110】

管理端末19では、送信されたパケットヘッダ情報1000を解析して、このパケットヘッダ情報1000に基づいたフローの条件をフロー検索テーブル41に帯域監視用フロー検索エントリとして設定することを指示する情報をレジスタ29経由でフロー検索テーブル制御部40に送信する。フロー検索テーブル制御部40は、この情報に従ってパケットヘッダ情報をフロー検索テーブル41にフロー検索エントリとして書き込み制御して設定する。また、当該のフロー検索エントリのアドレスと同一のアドレスを書き込みアドレスとして、フローアクションテーブル制御部142はフローアクションテーブル144のフローアクションエントリに「帯域監視」を指示する情報を書き込み制御する。更に、予めパケット中継装置の運用管理者が指定した帯域監視に関わる図10のパラメータR510、THR513、THRM514の情報を、レジスタ29経由で帯域監視テーブル制御部50に送信する。帯域監視テーブル制御部50は、当該のフロー検索エントリのアドレスと同一のアドレスを書き込みアドレスとして、R510、THR513、THRM514の情報を帯域監視テーブル51の帯域監視エントリに書き込み制御する。

40

50

## 【 0 1 1 1 】

以上で説明した実施例2の処理により、パケット中継装置に入力するパケットのパケットヘッダ情報をパケット中継装置にて自律的に収集し、入力パケットのパケットヘッダ情報に基づいて、フローの条件をパケット中継装置にて自律的にフロー検索テーブルに登録する処理が実現される。

## 【実施例 3】

## 【 0 1 1 2 】

続いて、第3の実施例として、上流に位置し本発明に係る構成を有しないパケット中継装置により、ECNマーキングされたパケットの受信を契機に、当該パケットのACKパケットにECEマーキングを実施する構成の実施例を、図26～図29を用いて、説明する。本実施例3の説明では、実施例2との差分のみを説明する。

10

## 【 0 1 1 3 】

本実施例では、上流に位置する本発明の構成を備えないパケット中継装置によりECNマーキングされたパケットを検出して、フロー検索テーブルに登録するための上流装置連携登録用フロー検索エントリを、実施例2における登録用フロー検索エントリの直上のアドレスに、パケット中継装置の運用管理者の指定に基づいて登録する。また、上流装置連携登録用フロー検索エントリに基づいて登録される上流装置連携ECEマーキング用フロー検索エントリと上流装置連携解除用フロー検索エントリとを、上流装置連携登録用フロー検索エントリよりも上位アドレスに、パケット中継装置にて自律的に登録する。

20

## 【 0 1 1 4 】

図26に、帯域監視用フロー検索エントリと登録用フロー検索エントリを省略した、本実施例におけるフロー検索テーブル145の構成を示す。フロー検索テーブル145は、TOSフィールドをDSCPフィールドとECNフィールドに分割して設定する点がフロー検索テーブル41と異なる。1451、1453、1455、1457は上流装置連携ECEマーキング用フロー検索エントリ、1452、1454、1456、1458は上流装置連携解除用フロー検索エントリ、1459は上流装置連携登録用フロー検索エントリである。1451と1452、1453と1454、1455と1456、1457と1458は対になるフロー検索エントリとして本実施例のパケット中継装置2により自律的に登録され、上流装置連携ECEマーキング用フロー検索エントリではACK = 1、上流装置連携解除用フロー検索エントリではACK = 0かつCWR = 1とする他は、全てのフローの条件に関し互いに一致している。

30

## 【 0 1 1 5 】

実施例2で説明した登録用フロー検索エントリではECNフィールド = d.c.として上流パケット中継装置による輻輳検出有無によらずにパケット検出するのに対し、上流装置連携登録用フロー検索エントリ1459はECNフィールド = CEと設定して上流パケット中継装置による輻輳検出済パケットのみを検出する点が異なる。また、実施例2で説明した帯域監視用フロー検索エントリではTOSフィールドの下位2ビットとしてECNフィールドも入力パケットのパケットヘッダの値を登録したのに対し、上流装置連携ECNマーキング用フロー検索エントリ1451、1453、1455、1457と上流装置連携解除用フロー検索エントリ1452、1454、1456、1458ではECNフィールド = d.c.として、登録時の入力パケットのECNフィールド = CEであっても、上流装置連携ECNマーキング用フロー検索エントリではECN = CEのパケットも検出する点が異なる。

40

## 【 0 1 1 6 】

これは、ECNの規定では、受信端末がECN = CEのパケットを受信すると、その後CWR = 1のパケットを受信するまでは、ECN = CEのパケットを受信した場合であってもECE = 1のACKパケットを送信し続ける処理に対応している。本実施例のパケット中継装置がECN = CEのパケットを受信すると、その後CWR = 1のパケットを受信するまでは、ECN = CEのパケットを受信した場合であっても、受信したACKパケットのECEを ' 1 ' に書き換えて送信し続ける。

## 【 0 1 1 7 】

次に、図27に、実施例2における帯域監視用フロー検索エントリに対応するフローアク

50

ションエントリと登録用フロー検索エントリに対応するフローアクションエントリを省略した、本実施例におけるフローアクションテーブル146の構成を示す。フローアクションテーブル146では、上流装置連携ECEマーキング用フロー検索エントリに対応するフローアクションエントリ1461には「上流装置連携ECEマーキング」を指示する情報を記載し、上流装置連係解除用フロー検索エントリに対応するフローアクションエントリ1462には「上流装置連係解除」を指示する情報を記載し、上流装置連携登録用フロー検索エントリに対応するフローアクションエントリ146nには「上流装置連携登録」を指示する情報を記載する。

#### 【0118】

本実施例のパケット中継装置に、上流に位置する本発明の構成を備えないパケット中継装置によりECNマーキングされたECN=CEのパケットが入力すると、当該パケットのパケットヘッダ情報はフロー検索テーブル145において、予めパケット中継装置の運用管理者の指定に基づいて登録された上流装置連携登録用フロー検索エントリ1459に記載されたフローの条件に一致する。すると、当該上流装置連携登録用フロー検索エントリ1459のアドレスを読み出しアドレスとしてフローアクションテーブル制御部142から読み出したフローアクションテーブル146のフローアクションエントリに「上流装置連携登録」を指示する情報が記載されていた場合は、入力パケットは上流装置連携登録用フロー検索エントリに一致したものと、フローアクションテーブル制御部142にて判定される。

#### 【0119】

上流装置連携登録用フロー検索エントリ一致を示す情報と入力パケットのパケットヘッダ情報はフロー検索テーブル制御部40を経由してレジスタ29に送信され、管理端末19に送信される。管理端末19では、送信されたパケットヘッダ情報を解析して、このパケットヘッダ情報に基づいたフローの条件をフロー検索テーブル41に上流装置連携ECEマーキング用フロー検索エントリと上流装置連係解除用フロー検索エントリとして設定することを指示する情報をレジスタ29経由でフロー検索テーブル制御部40に送信する。この際、上流装置連携ECEマーキング用フロー検索エントリと上流装置連係解除用フロー検索エントリのECNフィールドはd.c.として登録する。フロー検索テーブル制御部40は、この情報に従ってパケットヘッダ情報をフロー検索テーブル41に上流装置連携ECEマーキング用フロー検索エントリと上流装置連係解除用フロー検索エントリを書き込み制御して設定する。また、当該の上流装置連携ECEマーキング用フロー検索エントリのアドレスと同一のアドレスを書き込みアドレスとして、フローアクションテーブル制御部142はフローアクションテーブル146のフローアクションエントリに「上流装置連携ECEマーキング」を指示する情報を書き込み制御する。また、当該の上流装置連携解除用フロー検索エントリのアドレスと同一のアドレスを書き込みアドレスとして、フローアクションテーブル制御部142はフローアクションテーブル146のフローアクションエントリに「上流装置連携解除」を指示する情報を書き込み制御する。

#### 【0120】

次に、実施例3のパケット中継装置に、TCPのACKパケットが入力した場合の処理について説明する。TCPのACKパケットが本実施例のパケット中継装置に入力すると、パケットヘッダ情報の送信元IPアドレスと宛先IPアドレス、送信元ポート番号と宛先ポート番号を交換するパケットヘッダ条件交換を行ってから、フロー検索テーブル制御部40からフロー検索テーブル41に対し検索起動をかける。検索の結果、上流装置連携ECEマーキング用フロー検索エントリとして登録したフローの条件に対し、パケットヘッダ条件交換を行った後に一致するパケットヘッダ情報をもつパケットが本実施例のパケット中継装置に入力すると、フロー検索テーブル145において一致する上流装置連携ECEマーキング用フロー検索エントリのアドレスを読み出しアドレスとしてフローアクションテーブル制御部42から読み出したフローアクションテーブル146のフローアクションエントリに「上流装置連携ECEマーキング」を指示する情報が記載されていた場合は、入力パケットは上流装置連携ECEマーキング用フロー検索エントリに一致したとフローアクションテーブル制御部142にて判定され、上流装置連携ECEマーキング用フロー検索エントリ一致を示す情報はフロー検索

10

20

30

40

50

テーブル制御部40を経由してパケット受信回路21に送信される。すると、このACKパケットのパケットヘッダ情報1000を、パケット受信バッファ制御部210がパケット受信バッファ211から読み出し制御して、このパケットヘッダ情報1000におけるTCPフラグフィールド1042のECEフィールド10421の値を1に書き換える。これにより、ACKパケットに対するECEマーキングを、パケット中継装置におけるパケット受信バッファ211の輻輳状態に基づいて、受信端末ではなくパケット中継装置で実施することができる。従って、パケット中継装置から受信端末、受信端末からパケット中継装置、パケット中継装置から送信端末の経路で輻輳通知することができる。当該のACKパケットが上流装置連携ECEマーキング用フロー検索エントリに一致しない場合は、ECEマーキング処理はしない。ACKパケットではACK = 1なので、ACK = 0かつCWR = 1の上流装置連携解除用フロー検索エントリに一致することは無い。

10

#### 【0121】

一方、CWR = 1のデータパケットが実施例3のパケット中継装置に入力した場合には、上述のパケットヘッダ条件交換を行わずにフロー検索テーブル制御部40からフロー検索テーブル41に対し検索起動をかける。検索の結果、当該データパケットのパケットヘッダ情報1000が上流装置連携解除用フロー検索エントリのフローの条件に一致した場合には、フロー検索制御部40から一致した上流装置連携解除用フロー検索エントリと、当該エントリと対として登録された上流装置連携ECEマーキング用フロー検索エントリを削除する制御を行う。次に、一致した上流装置連携解除用フロー検索エントリのアドレスと同一アドレスにあるフローアクションテーブル143のフローアクションエントリの「上流装置連携解除」を指示する情報を削除する。また、当該エントリと対として登録された上流装置連携ECEマーキング用フロー検索エントリのアドレスと同一アドレスにあるフローアクションテーブル143のフローアクションエントリの「上流装置連携ECEマーキング」を指示する情報を削除する。

20

#### 【0122】

以上に述べた実施例3のパケット中継装置における処理をフローチャートにまとめ、送信端末、受信端末とパケット中継装置間でなされる本実施例の通信の概要をシーケンス図にまとめて説明する。

#### 【0123】

図28に、実施例3のパケット中継装置における処理のフローチャートを示す。本実施例では、図23に示した実施例1のフローチャートの処理に加え、入力パケットのECNフィールドが"11"、つまりCEであるか否かの判定が追加となる。入力パケットのECNフィールドが"11"でない場合は、実施例1と同様にパケット中継装置内での輻輳検出を判定する処理に移行する。入力パケットのECNフィールドが、"11"である場合は実施例3に特徴的な処理として、当該パケットのパケットヘッダ情報1000のパケットヘッダ条件交換を行い、当該パケットの属するフローに対し受信端末から送信されるACKパケットに対する上流装置連携ECEマーキング用フロー検索エントリと、当該パケットの属するフローでCWR = 1のデータパケットに対する上流装置連携解除用フロー検索エントリをフロー検索テーブルに追加となる。

30

#### 【0124】

また本実施例では、入力パケットがACKパケットである場合には、当該ACKパケットが上流装置連携ECEマーキング用フロー検索エントリに一致するか否かの判定を行い、一致する場合には、ECEマーキングを行う処理が追加となる。一致しない場合には実施例1と同様に、当該フローに対する輻輳状態管理テーブルを参照し、当該フローが輻輳状態にある場合にはECEマーキングを行う。

40

#### 【0125】

また本実施例では、入力パケットがACKパケットではない場合、つまりデータパケットで、かつCWR = 1である場合には、当該フローのパケットのパケットヘッダ情報が一致する上流装置連携ECEマーキング用フロー検索エントリと、当該フローのパケットのパケットヘッダ情報が一致する上流装置連携解除用フロー検索エントリを削除する処理と、当該フ

50



ローの非輻輳状態を輻輳状態管理テーブルに記録する処理が追加となる。

【 0 1 2 6 】

次に、図29に、実施例3の packets 中継装置を用いた通信のシーケンス図を示す。本実施例のシーケンス図では、実施例1の通信のシーケンス図24における通信構成要素である送信端末1、packets 中継装置2、受信端末3に加え、上流側中継装置（upstream Router）が追加となる。上流側中継装置は、packets 中継装置2よりも上流に位置する中継装置であり、本発明の構成を備えない中継装置であるものとする。packets 中継装置2は、上流側中継装置にてECNマーキングを実施されたSEQ = 4001 - 5000の packets であるDATA（ECE）を受信することにより、上流側中継装置における輻輳を検出し、これを契機としてそれ以降受信した当該 packets に対するACK packets、つまりSEQ = 4001以降のACK packets に対しECEマーキングを実施する。これにより、上流に位置する本発明の機能構成を備えない packets 中継装置によりECNマーキングされた packets 受信を契機として、packets 中継装置2は、当該 packets に対するACK packets にECEマーキングを実現して、送信端末1に送信することができる。

10

【実施例4】

【 0 1 2 7 】

次に、第4の実施例の packets 中継装置を説明する。以下の説明では、実施例4の特徴的な処理のみを説明する。実施例4として、TCPの3 way handshakeによる初期化処理を含め、受信端末におけるECNに関する処理の全てをpackets 中継装置にて代行する構成について説明する。この構成により、受信端末がECNをサポートしていない場合にも、送信端末とpackets 中継装置間でECNをサポートした通信が可能となる。なお本実施例の packets 中継装置では、ECNサポートのデータ packets であるDATA（ECT）や、ECNにおける輻輳回避制御が実施済であることを示すデータ packets DATA（CWR）を、ECN非サポートのデータ packets DATA（not - ECT）に書き換える隠蔽処理を行うことにより、受信端末に対するECN通信はなされないようにする。

20

【 0 1 2 8 】

図30に、本実施例におけるフロー検索テーブルの構成を示す。フロー検索テーブル147は、SYNフィールドとECEフィールドを設けた点がフロー検索テーブル145と異なる。1471はSYN - ACK packets 検出用フロー検索エントリ、1472はACK packets 検出用フロー検索エントリ、1473はDATA（ECT） packets 検出用フロー検索エントリ、1474はDATA（CWR） packets 検出用フロー検索エントリ、1475はFIN packets 検出用フロー検索エントリ、1476はECN - setup SYN packets 検出用フロー検索エントリである。1471～1475は一組のフロー検索エントリとして本実施例の packets 中継装置により自律的に登録され、ACK、SYN、ECE、CWR、FINフィールドから構成されるTCPフラグフィールドとECNフィールドの他は、全てのフローの条件に関し互いに一致している。ECN - setup SYN packets 検出用フロー検索エントリ1476は、任意のACK = 0かつSYN = 1かつECE = 1かつCWR = 1のECN - setup SYN packets を検出する。

30

【 0 1 2 9 】

次に、図31に、実施例4における上述のフロー検索エントリに対応するフローアクションテーブル148の構成を示す。フローアクションテーブル148では、SYN - ACK packets 検出用フロー検索エントリ1471に対応するフローアクションエントリ1481には「ECEマーキング」を指示する情報が記載される。これは、ECNサポートしていない受信端末が送信したECN非サポートを意味するSYN - ACK packets、つまりECE = 0のSYN - ACK packets を、ECNサポートを意味するSYN - ACK packets、つまりECE = 1のSYN - ACK packets に、本実施例4の packets 中継装置にて書き換えることを指示する。

40

【 0 1 3 0 】

ACK packets 検出用フロー検索エントリ1472に対応するフローアクションエントリ1482には「輻輳時ECEマーキング」を指示する情報が記載される。これは、packets 中継装置にて輻輳を検出した場合、それ以降CWR = 1のデータ packets を検出するまでの間は、ACK packets 検出用フロー検索エントリ1472にて検出されたACK packets に対し、本実施例の

50

パケット中継装置にてECEマーキングを行うことを指示する。

【 0 1 3 1 】

DATA (ECT) パケット検出用フロー検索エントリ1473に対応するフローアクションエントリ1483には「not - ECTマーキング」を指示する情報が記載される。これは、DATA (ECT) パケット検出用フロー検索エントリ1473にて検査検出されたDATA (ECT) パケット、つまりECNフィールド = " 01 " または " 10 " または " 11 " のデータパケットを、DATA (not - ECT) パケット、つまりECNフィールド = " 00 " のデータパケットに、本実施例のパケット中継装置にて書き換えることを指示する。

【 0 1 3 2 】

DATA (CWR) パケット検出用フロー検索エントリ1474に対応するフローアクションエントリ1484には「ECEマーキング停止&not - ECTマーキング」を指示する情報が記載される。これは、DATA (CWR) パケット検出用フロー検索エントリ1474にて検出されたDATA (CWR) パケット、つまりCWR = 1のデータパケットを、DATA (not - ECT) パケット、つまりECNフィールド = " 00 " かつCWR = 0のデータパケットに、本実施例のパケット中継装置にて書き換え、かつ輻輳状態管理テーブルの輻輳状態管理エントリを輻輳状態で無いことを示す値に書き換えることにより、それ以降のACKパケットのECEマーキングを停止することを指示する。

【 0 1 3 3 】

FINパケット検出用フロー検索エントリ1475に対応するフローアクションエントリ1485には「ECNプロキシ削除」を指示する情報が記載される。これは、ECN - setup SYNパケット検出用フロー検索エントリ1476で検出されたフローに属するSYN - ACKパケットを検出するSYN - ACKパケット検出用フロー検索エントリ1471、データパケットを検出するDATA (ECT) パケット検出用フロー検索エントリ1473、DATA (CWR) パケット検出用フロー検索エントリ1474、ACKパケットを検出するACKパケット検出用フロー検索エントリ1472、FINパケットを検出する本FINパケット検出用フロー検索エントリ1475をフロー検索テーブル147から削除し、これらのフロー検索エントリに対応するフローアクションエントリ1481 ~ 1485をフローアクションテーブル148から削除して、それ以降のTCPの3 way handshakeによる初期化処理を含めたECNプロキシ処理を停止することを指示する。

【 0 1 3 4 】

ECN - setup SYNパケット検出用フロー検索エントリ1476に対応するフローアクションエントリ1486には「ECNプロキシ登録」を指示する情報が記載される。これは、ECN - setup SYNパケット検出用フロー検索エントリ1476で検出されたフローに属するSYN - ACKパケットを検出するSYN - ACKパケット検出用フロー検索エントリ1471、データパケットを検出するDATA (ECT) パケット検出用フロー検索エントリ1473、DATA (CWR) パケット検出用フロー検索エントリ1474、ACKパケットを検出するACKパケット検出用フロー検索エントリ1472、FINパケットを検出するFINパケット検出用フロー検索エントリ1475をフロー検索テーブル147に登録し、これらのフロー検索エントリに対応するフローアクションエントリ1481 ~ 1485をフローアクションテーブル148に登録することを指示する。

【 0 1 3 5 】

次に、各処理の詳細について説明する。

ACK = 0かつSYN = 1かつECE = 1かつCWR = 1かつFIN = 0のECN - setup SYNパケットが本実施例のパケット中継装置に入力すると、フロー検索部140のフロー検索テーブル制御部40からフロー検索テーブル147にパケットヘッダ情報1000を入力して検索起動制御がなされ、検索の結果、当該パケットのパケットヘッダ情報1000に一致するECN - setup SYNパケット検出用フロー検索エントリ1476のアドレス1476 - aがフロー検索テーブル147から出力される。すると、このアドレス1476 - aがフローアクションテーブル制御部42に入力され、アドレス1476 - aを読み出しアドレスとしてフローアクションテーブル143を読み出し制御する。すると、「ECNプロキシ登録」を指示する情報が予め記載されたフローアクションエントリ1486が読み出される。すると、この「ECNプロキシ登録」を示す情報とパケットヘッダ情報1000がレジスタ29を経由して管理端末19に送信される。管理端末19では、送信さ

10

20

30

40

50

れたパケットヘッダ情報1000を解析して、このパケットヘッダ情報1000に基づいたフローの条件をフロー検索テーブル147に帯域監視用フロー検索エントリとして設定することを指示する情報をレジスタ10経由でフロー検索テーブル制御部40に送信する。管理端末19では、送信されたパケットヘッダ情報1000を解析して、このパケットヘッダ情報1000に基づいたフローの条件をフロー検索テーブル147に、SYN - ACKパケット検出用フロー検索エントリ1471、ACKパケット検出用フロー検索エントリ1472、DATA (ECT) パケット検出用フロー検索エントリ1473、DATA (CWR) パケット検出用フロー検索エントリ1474として設定することを指示する情報をレジスタ29経由でフロー検索テーブル制御部40に送信する。

【0136】

そして、フロー検索テーブル制御部40からフロー検索テーブル147に対し、SYN - ACKパケット検出用フロー検索エントリ1471、ACKパケット検出用フロー検索エントリ1472、DATA (ECT) パケット検出用フロー検索エントリ1473、DATA (CWR) パケット検出用フロー検索エントリ1474、FINパケット検出用フロー検索エントリ1475の書き込み制御を行う。なお実際には、DATA (ECT)、DATA (CWR) は入力パケットのECNフィールド値によらずECN = 01または10または11の3エントリを設定する。DATA (ECT) に対応するフローアクションエントリ1483も、ECN = 01または10または11の3つのDATA (ECT) パケット検出用フロー検索エントリ1473に対応した3つのエントリから構成される。DATA (CWR) に対応するフローアクションエントリ1484も、ECN = 01または10または11の3つのDATA (CWR) パケット検出用フロー検索エントリ1474に対応した3つのエントリから構成される。

【0137】

図30に示す通り、SYN - ACKパケット検出用フロー検索エントリ1471では、パケットヘッダ情報1000に基づいたフローの条件と、パケットヘッダ情報1000によらずACK = 1、SYN = 1、ECE = 0、CWR = 0かつFIN = 0を設定する。ACKパケット検出用フロー検索エントリ1472では、パケットヘッダ情報1000に基づいたフローの条件と、パケットヘッダ情報1000によらずACK = 1、SYN = 0、ECE = 0、CWR = 0かつFIN = 0を設定する。DATA (ECT) パケット検出用フロー検索エントリ1473では、パケットヘッダ情報1000に基づいたフローの条件と、パケットヘッダ情報1000によらずACK = 0、SYN = 0、ECE = 0、CWR = 0かつFIN = 0と、パケットヘッダ情報1000によらずECN = 01のエントリと、同じくECN = 10のエントリと、同じくECN = 11のエントリの3エントリを設定する。DATA (CWR) パケット検出用フロー検索エントリ1474では、パケットヘッダ情報1000に基づいたフローの条件と、パケットヘッダ情報1000によらずACK = 0、SYN = 0、ECE = 0、CWR = 1かつFIN = 0と、パケットヘッダ情報1000によらずECN = 01のエントリと、同じくECN = 10のエントリと、同じくECN = 11のエントリの3エントリを設定する。

【0138】

また、管理端末19では、フローアクションテーブル148のフローアクションエントリ1481に「ECEマーキング」を指示する情報を設定し、フローアクションエントリ1482に「輻輳時ECEマーキング」を指示する情報を設定し、フローアクションエントリ1483に「not - ECTマーキング」を指示する情報を設定し、フローアクションエントリ1484には「ECEマーキング停止&not - ECTマーキング」を指示する情報を設定することを指示する情報をレジスタ10経由でフローアクションテーブル制御部42に送信する。そして、フローアクションテーブル制御部42からフローアクションテーブル148に対し、フローアクションエントリ1481、フローアクションエントリ1482、フローアクションエントリ1483、フローアクションエントリ1484の書き込み制御を行う。

【0139】

次に、ACK = 1かつSYN = 1かつECE = 0かつCWR = 0かつFIN = 0の、ECN非サポートのSYN - ACKパケットが本実施例4のパケット中継装置に入力すると、フロー検索部140のフロー検索テーブル制御部40からフロー検索テーブル147にパケットヘッダ情報1000を入力して検索起動制御がなされ、検索の結果、当該パケットのパケットヘッダ情報1000に一致するSYN - ACKパケット検出用フロー検索エントリ1471のアドレス1471 - aがフロー検索テーブル147から出力される。すると、このアドレス1471 - aがフローアクションテーブル制御部142

に入力され、アドレス1471 - aを読み出しアドレスとしてフローアクションテーブル143を読み出し制御する。すると、「ECEマーキング」を指示する情報がフローアクションエントリ1481から読み出される。すると、この「ECEマーキング」を示す情報がパケット受信回路160のパケット受信バッファ制御部1600に送信され、当該パケットのパケットヘッダ情報1000をパケット受信バッファ1601から読み出し制御して、当該パケットのECE = 1に書き換える。すると、本実施例のパケット中継装置にてECN非サポートのSYN - ACKパケットがECNサポートのECN - setup SYN - ACKパケット、つまりACK = 1かつSYN = 1かつECE = 1かつCWR = 0かつFIN = 0のSYN - ACKに書き換えられて送信されることとなる。このECN - setup SYN - ACKパケットを送信端末が受信して、送信端末がACKパケットを受信端末に送信すると、ECNサポート通信に関するTCP 3 way handshake処理が完了する。

10

#### 【 0 1 4 0 】

次に、ECNサポート通信が開始されると、ECN = 01または10または11のDATA ( ECT ) パケットが本実施例のパケット中継装置に入力する。ACK = 0かつSYN = 0かつECE = 0かつCWR = 0かつFIN = 0のDATA ( ECT ) が本実施例パケット中継装置に入力すると、フロー検索部140のフロー検索テーブル制御部40からフロー検索テーブル147にパケットヘッダ情報1000を入力して検索起動制御がなされ、検索の結果、当該パケットのパケットヘッダ情報1000に一致するDATA ( ECT ) パケット検出用フロー検索エントリ1473のアドレス1473 - aがフロー検索テーブル147から出力される。すると、このアドレス1473 - aがフローアクションテーブル制御部142に入力され、アドレス1473 - aを読み出しアドレスとしてフローアクションテーブル143を読み出し制御する。すると、「not - ECTマーキング」を指示する情報がフローアクションエントリ1483から読み出される。すると、この「not - ECTマーキング」を示す情報がパケット受信回路160のパケット受信バッファ制御部1600に送信され、当該パケットのパケットヘッダ情報1000をパケット受信バッファ1601から読み出し制御して、当該パケットのECN = 00に書き換える。すると、本実施例のパケット中継装置にてECNサポートのDATA ( ECT ) パケットがECN非サポートのDATA ( not - ECT ) パケットに書き換えられて送信されることとなる。これにより、送信端末から送信されたECNサポートのデータパケットを、本実施例のパケット中継装置にて非ECNサポートのデータパケットに書き換えて、受信端末に対するECN通信を非ECN通信として隠蔽することができる。

20

#### 【 0 1 4 1 】

次に、本実施例のパケット中継装置にて書き換えられた非ECNサポートのデータパケットに対するACK = 1かつSYN = 0かつECE = 0かつCWR = 0かつFIN = 0のACKパケットが本実施例のパケット中継装置に入力すると、フロー検索部140のフロー検索テーブル制御部40からフロー検索テーブル147にパケットヘッダ情報1000を入力して検索起動制御がなされ、検索の結果、当該パケットのパケットヘッダ情報1000に一致するACKパケット検出用フロー検索エントリ1472のアドレス1472 - aがフロー検索テーブル147から出力される。すると、このアドレス1472 - aがフローアクションテーブル制御部142に入力され、アドレス1472 - aを読み出しアドレスとしてフローアクションテーブル143を読み出し制御する。すると、「輻輳時ECEマーキング」を指示する情報がフローアクションエントリ1482から読み出される。すると、この「輻輳時ECEマーキング」を示す情報がパケット受信回路160のバッファ毎輻輳状態管理テーブル制御部1606に送信され、バッファ毎輻輳状態管理テーブル制御部1606によりバッファ毎輻輳状態管理テーブル1607の読み出し制御を行い、当該ACKパケットに対応するデータパケットが蓄積されているバッファの輻輳状態フラグを読み出す。この輻輳状態フラグが輻輳状態にあることを示す値であった場合には、パケット受信バッファ制御部1600によりパケット受信バッファ1601の当該ACKパケットのパケットヘッダ情報1000の読み出し制御を行い、当該ACKパケットのECEフラグを1に書き換える「輻輳時ECEマーキング」の処理を行う。

30

40

#### 【 0 1 4 2 】

次に、ECEマーキングされたACKパケットにより輻輳を検出して、送信端末1にて輻輳回避制御がなされたことを示すACK = 0かつSYN = 0かつECE = 0かつCWR = 1かつFIN = 0のDATA ( CWR ) パケットが本実施例のパケット中継装置に入力すると、フロー検索部140のフロー検

50

フロー検索テーブル制御部40からフロー検索テーブル147にパケットヘッダ情報1000を入力して検索起動制御がなされ、検索の結果、当該パケットのパケットヘッダ情報1000に一致するDATA (CWR) パケット検出用フロー検索エントリ1474のアドレス1474 - aがフロー検索テーブル147から出力される。すると、このアドレス1474 - aがフローアクションテーブル制御部142に入力され、アドレス1474 - aを読み出しアドレスとしてフローアクションテーブル143を読み出し制御する。すると、「ECEマーキング停止&not - ECTマーキング」を指示する情報がフローアクションエントリ1484から読み出される。すると、この「ECEマーキング停止」を示す情報がバッファ毎輻輳状態管理テーブル制御部1 - 15に送信され、バッファ毎輻輳状態管理テーブル制御部1606からバッファ毎輻輳状態管理テーブル1607の当該パケットが蓄積されるバッファに対する輻輳状態管理エントリの輻輳状態フラグの値を、輻輳状態に無いことを示す値に書き換える。これにより、これ以降本実施例のパケット中継装置に入力するACKパケットに対するECEマーキングの処理は停止する。また、「not - ECTマーキング」を示す情報がパケット受信回路160のパケット受信バッファ制御部1600に送信され、当該パケットのパケットヘッダ情報1000をパケット受信バッファ1601から読み出し制御して、当該パケットのECN = 00に、CWR = 0に書き換える。すると、本実施例のパケット中継装置にてECNサポートのDATA (CWR) パケットがECN非サポートのDATA (not - ECT) パケットに書き換えられて送信されることとなる。これにより、送信端末から送信されたECNサポートのDATA (CWR) パケットを、本実施例のパケット中継装置にて非ECNサポートのDATA (not - ECT) パケットに書き換えて、受信端末に対するECN通信を非ECN通信として隠蔽することができる。

#### 【 0 1 4 3 】

次に、通信終了を示すACK = 0かつSYN = 0かつECE = 0かつCWR = 0かつFIN = 1のFINパケットが本実施例のパケット中継装置に入力すると、フロー検索部140のフロー検索テーブル制御部40からフロー検索テーブル147にパケットヘッダ情報1000を入力して検索起動制御がなされ、検索の結果、当該パケットのパケットヘッダ情報1000に一致するFINパケット検出用フロー検索エントリ1475のアドレス1475 - aがフロー検索テーブル147から出力される。すると、このアドレス1475 - aがフローアクションテーブル制御部142に入力され、アドレス1475 - aを読み出しアドレスとしてフローアクションテーブル143を読み出し制御する。すると、「ECNプロキシ削除」を指示する情報がフローアクションエントリ1485から読み出される。すると、この「ECNプロキシ削除」を示す情報がフローアクションエントリ1482から読み出される。すると、フロー検索テーブル制御部40からフロー検索テーブル147において本FINパケット検出用フロー検索エントリ1475と、本エントリより上位にある4つのエントリとしてSYN - ACKパケット検出用フロー検索エントリ1471、ACKパケット検出用フロー検索エントリ1472、DATA (ECT) パケット検出用フロー検索エントリ1473、DATA (CWR) パケット検出用フロー検索エントリ1474を削除する。また、フローアクションテーブル制御部142からフローアクションエントリ1481 ~ 1485を削除する。これにより、ECN通信の完了後は当該フローに属するECNプロキシ関連フロー検索エントリとECNプロキシ関連フローアクションエントリを全て削除する。

以上に述べた本実施例のパケット中継装置における処理をフローチャートにまとめ、送信端末、受信端末とパケット中継装置間でなされる本実施例の通信の概要をシーケンス図にまとめて説明する。

#### 【 0 1 4 4 】

図32に本実施例のパケット中継装置における処理フローチャートを示す。

図32に示すフローチャートでは、受信端末がECN非サポートであることを前提とする。本実施例では、実施例1の図23の処理フローに対し、ECN - setup SYNパケットが本実施例のパケット中継装置に入力すると、ECNプロキシ関連エントリ、つまりSYN - ACKパケット検出用フロー検索エントリ1471、ACKパケット検出用フロー検索エントリ1472、DATA (ECT) パケット検出用フロー検索エントリ1473、DATA (CWR) パケット検出用フロー検索エントリ1474、FINパケット検出用フロー検索エントリ1475を登録する処理が追加となる。また、FINパケットが本実施例のパケット中継装置に入力すると、ECNプロキシ関連エント

り、つまりSYN - ACKパケット検出用フロー検索エントリ1471、ACKパケット検出用フロー検索エントリ1472、DATA (ECT) パケット検出用フロー検索エントリ1473、DATA (CWR) パケット検出用フロー検索エントリ1474、FINパケット検出用フロー検索エントリ1475を削除する処理が追加となる。また、SYN - ACKパケットが本実施例のパケット中継装置に入力すると、当該パケットをECE = 1のECN - setup SYN - ACKパケットに書き換えて送信する処理が追加となる。また、DATA (CWR) パケットが本実施例のパケット中継装置に入力すると、当該パケットをECN = 00のECN非サポートパケットに書き換えて送信する処理が追加となる。また、その他のデータパケットが本実施例のパケット中継装置に入力すると、当該パケットをECN = 00のECN非サポートパケットに書き換えて送信する処理が追加となる。

【0145】

最後に、図33に、本実施例のパケット中継装置を用いた通信のシーケンス図を示す。本実施例のシーケンス図では、実施例1のシーケンス図24に対し、3 way - handshakeによるInitializeを含め、受信端末3におけるECNの手続きをパケット中継装置2で代行するECNプロキシの処理と、パケット中継装置2で、DATA (ECT) やDATA (CWR) をDATA (not - ECT) に書き換えて、受信端末3に対するECN通信を隠蔽する処理が追加となる。

【0146】

以上、本発明の種々の実施例の説明を行ったが、本発明は上記した実施例に限定されるものではなく、様々な変形例が含まれる。例えば、上記した実施例は本発明を分かりやすく説明するために詳細に説明したのであり、必ずしも説明の全ての構成を備えるものに限定されるものではない。また、ある実施例の構成の一部を他の実施例の構成に置き換えることが可能であり、また、ある実施例の構成に他の実施例の構成を加えることが可能である。また、各実施例の構成の一部について、他の構成の追加・削除・置換をすることが可能である。

【0147】

また、上記の各構成、機能、処理部、処理手段等は、それらの一部又は全部を、例えば集積回路で設計する等によりハードウェアで実現してもよい。また、上記の各構成、機能等は、それぞれの機能を実現するプログラムを実行することで実現したが、これらの各機能を実現するプログラム、テーブル、ファイル等の情報は、メモリや、ハードディスク、SSD (Solid State Drive) 等の記録装置、または、ICカード、SDカード、DVD等の記録媒体におくことができるし、必要に応じてネットワーク等を介してダウンロード、インストールすることも可能である。

【0148】

以上詳述したように、本明細書においては、特許請求の範囲に記載した発明以外に、種々の発明が開示されている。その一部を下記に適用例として列記する。

【0149】

[適用例1]

複数の入力回線と出力回線を備え、

入力回線から受信したパケットの入力物理回線番号、または入力論理回線番号、またはパケットヘッダ情報のうち少なくとも一つ以上の情報により識別されるパケットの集合から構成されるフローを検出し、

フローが輻輳状態にあるとパケット中継装置が判定すると、

それ以降にパケット中継装置が受信した当該フローに対する応答パケットのパケットヘッダのうちネットワークの輻輳状態を示すフィールドを輻輳状態にあることを示す値に書き換えることを特徴とするパケット中継装置。

[適用例2]

適用例1のパケット中継装置において、

フロー毎の輻輳状態を記録するフロー毎の輻輳状態フラグから構成される輻輳状態管理テーブルを備え、

フローが輻輳状態にあるとパケット中継装置が判定すると、当該フローの輻輳状態フラグを輻輳状態にあることを示す値に書き換え、

10

20

30

40

50

フローが輻輳状態から解消したとパケット中継装置が判定すると、当該フローの輻輳状態フラグを非輻輳状態にあることを示す値に書き換え、  
パケット中継装置が応答パケットを受信すると輻輳状態管理テーブルを参照して、応答パケットが属するフローの輻輳状態フラグに従って、  
輻輳状態フラグが輻輳を示す場合には応答パケットのパケットヘッダのうちネットワークの輻輳状態を示すフィールドの値を輻輳状態にあることを示す値に書き換え、  
輻輳状態フラグが非輻輳状態を示す場合には応答パケットのパケットヘッダのうちネットワークの輻輳状態を示すフィールドの値を非輻輳状態にあることを示す値に書き換えることを特徴とするパケット中継装置。

【 0 1 5 0 】

10

[ 適用例 3 ]

適用例 2 のパケット中継装置において、  
受信したパケットのパケットヘッダ情報のうち、送信端末が輻輳回避制御を行ったことを示す情報が、輻輳回避を行ったことを示す値となっている場合に、フローが輻輳状態から解消したとパケット中継装置が判定することを特徴とするパケット中継装置。

【 0 1 5 1 】

[ 適用例 4 ]

適用例 3 のパケット中継装置において、  
受信したTCPパケットのCWRフラグが1である場合に、フローが輻輳状態から解消したとパケット中継装置が判定することを特徴とするパケット中継装置。

20

【 0 1 5 2 】

[ 適用例 5 ]

適用例 2 から適用例 4 いずれか記載のパケット中継装置において、  
フローが、送信元IPアドレス、宛先IPアドレス、送信元ポート番号、宛先ポート番号の組み合わせが等しいTCPフローまたはDCCPフローまたはSCTPフロー、または送信元IPアドレス、宛先IPアドレスの組み合わせが等しいIPフロー、または同一VLANに属するVLANフロー、またはパケット中継装置の運用管理者が指定した任意のパケットヘッダ情報項目の組み合わせで定義されたフローであることを特徴とするパケット中継装置。

【 0 1 5 3 】

[ 適用例 6 ]

30

適用例 2 から適用例 5 いずれかに記載のパケット中継装置において、  
フロー毎の輻輳状態フラグをパケット中継装置の運用管理者が指定した特定フローに対してのみ備えることを特徴とするパケット中継装置。

【 0 1 5 4 】

[ 適用例 7 ]

適用例 1 から適用例 6 いずれかに記載のパケット中継装置において、  
応答パケットのパケットヘッダのうちネットワークの輻輳状態を示すフィールドが、TCPパケットのECEフラグであることを特徴とするパケット中継装置。

【 0 1 5 5 】

[ 適用例 8 ]

40

適用例 1 から適用例 7 いずれかに記載のパケット中継装置において、  
シェーパにおけるキュー長、またはポリサにおけるパケット水量、またはバッファにおける蓄積パケット数が所定の閾値を超過すると、フローが輻輳状態にあると判定することを特徴とするパケット中継装置。

【 0 1 5 6 】

[ 適用例 9 ]

適用例 1 から適用例 7 いずれかに記載のパケット中継装置において、  
受信パケットのパケットヘッダのうちネットワークの輻輳状態を示すフィールドの値に基づいて、フローが輻輳状態にあると判定することを特徴とするパケット中継装置。

【 0 1 5 7 】

50

## [ 適用例 10 ]

適用例 9 のパケット中継装置において、  
受信パケットのパケットヘッダのうちネットワークの輻輳状態を示すフィールドが、IPパケットのECNフィールドであることを特徴とするパケット中継装置。

【 0 1 5 8 】

## [ 適用例 11 ]

適用例 1 から適用例 10 のいずれかに記載のパケット中継装置において、  
輻輳通知機能をサポートしない受信端末から送信された、送受信端末間で輻輳通知機能のサポート可否の情報を通知する制御パケットのパケットヘッダのうち、輻輳通知機能のサポート可否を示すフィールドの値を、輻輳通知機能の非サポートを示す値から輻輳通知機能のサポートを示す値に書き換えることを特徴とするパケット中継装置。

10

【 0 1 5 9 】

## [ 適用例 12 ]

適用例 11 のパケット中継装置において、  
ECNによる輻輳通知機能をサポートしない受信端末から送信された、ECNの初期化手順の 3 way handshake処理におけるECEフラグのセットされていないSYN - ACKパケットを、ECEフラグのセットされたSYN - ACKパケットに書き換えることを特徴とするパケット中継装置。

【 符号の説明 】

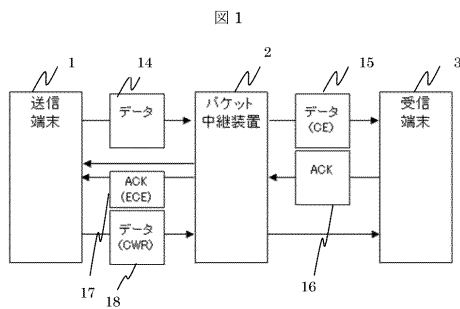
【 0 1 6 0 】

- 1... 送信端末 20
- 2... パケット中継装置
- 3... 受信端末
- 14... データ
- 15... データ ( CE )
- 16... ACK
- 17... ACK ( ECE )
- 18... データ ( CWR )
- 19... 管理端末
- 21、160... パケット受信回路
- 22... パケット検索部 30
- 23... 経路検索部
- 24、140... フロー検索部
- 25... 帯域監視部
- 26... パケット中継処理手段
- 27... パケット検索部
- 28... パケット送信回路
- 29... レジスタ
- 30... 出力回線
- 40... フロー検索テーブル制御部
- 41、145、147... フロー検索テーブル 40
- 50... 帯域監視テーブル制御部
- 51... 帯域監視テーブル
- 52... 現在水量判定部
- 53... 監視結果判定部
- 54... 輻輳状態管理部
- 101... 内部ヘッダ部
- 102... L2ヘッダ部
- 103... L3ヘッダ部
- 104... L4ヘッダ部
- 142... フローアクションテーブル制御部 50

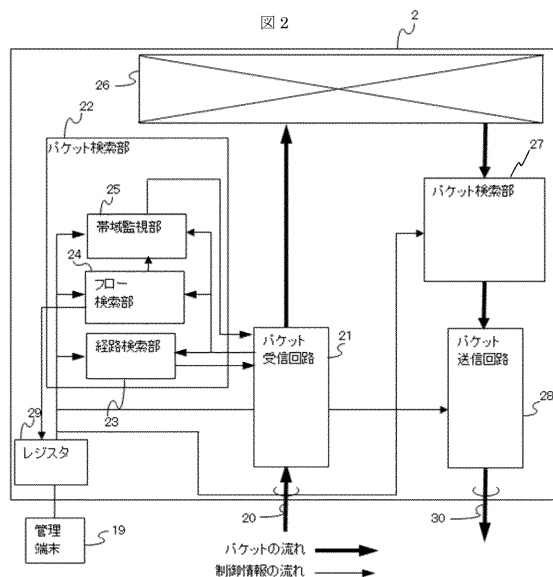


143、144、146、148... フローアクションテーブル  
210、1600... パケット受信バッファ制御部  
211、1601... パケット受信バッファ  
212、1602... バッファ蓄積パケット数管理部  
1000... パケットヘッダ情報

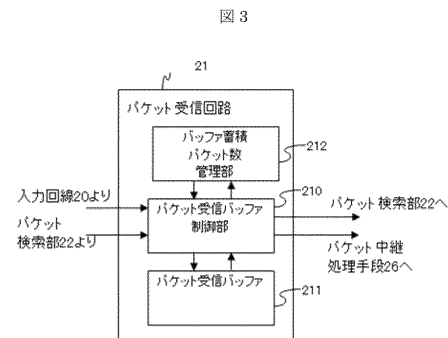
【 図 1 】



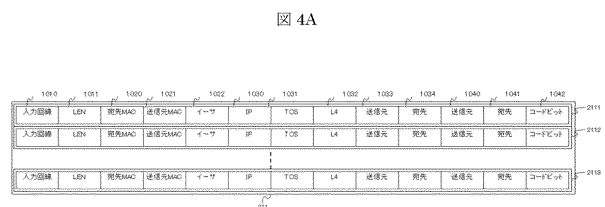
【圖 2】



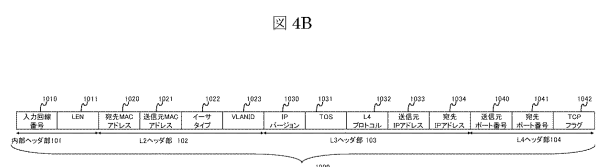
【圖 3】



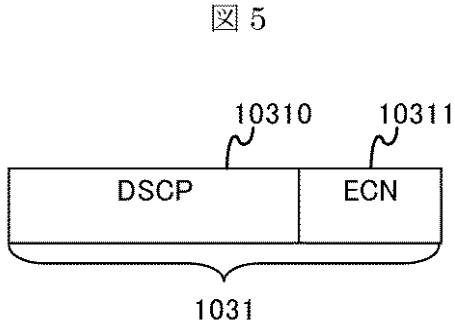
【 図 4 A 】



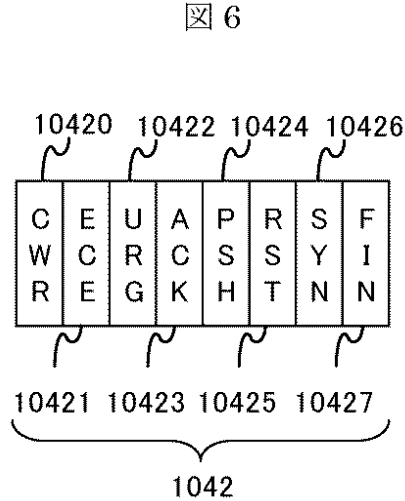
【 図 4 B 】



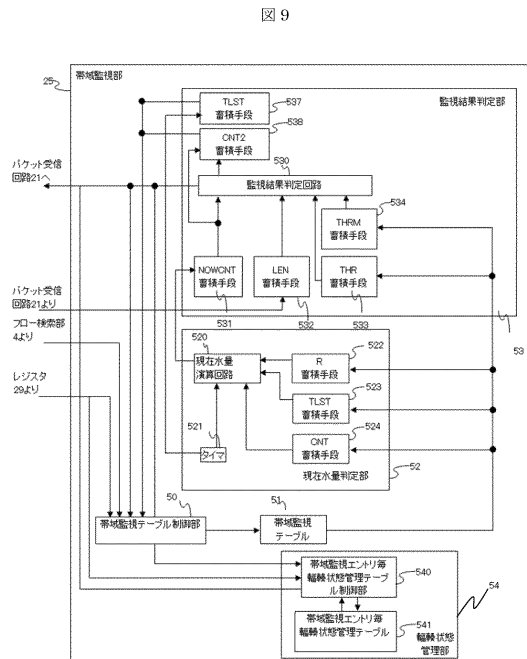
【 図 5 】



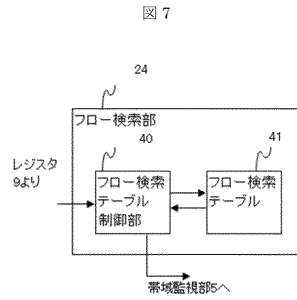
【 図 6 】



【 図 9 】



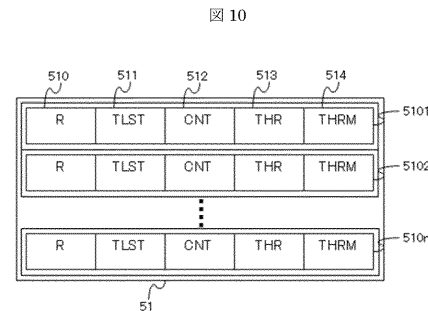
【 図 7 】



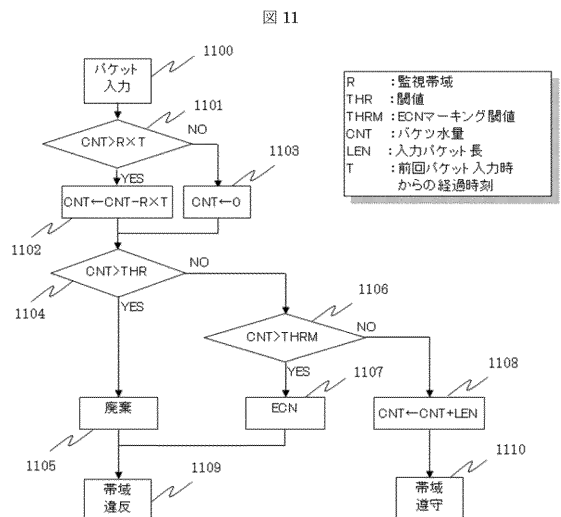
【圖 8】



【 図 1 0 】

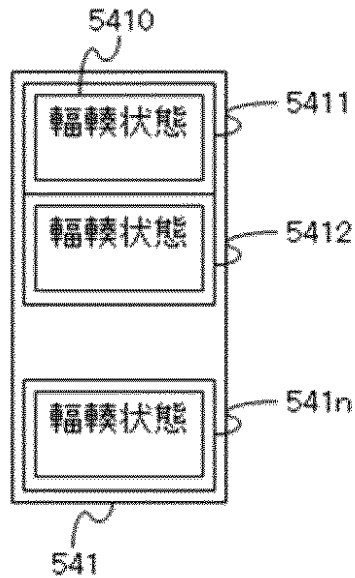


【 図 1 1 】

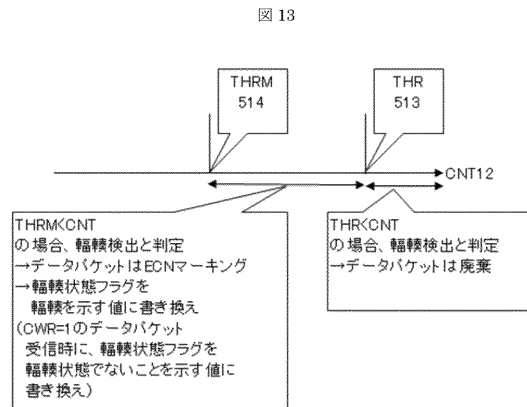


【 図 1 2 】

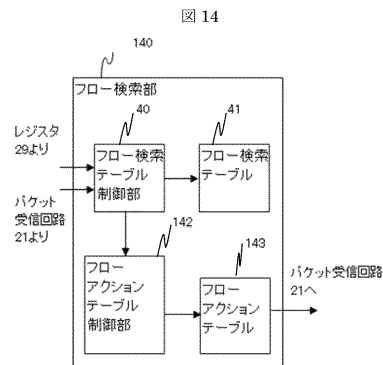
图 12



【 図 1 3 】

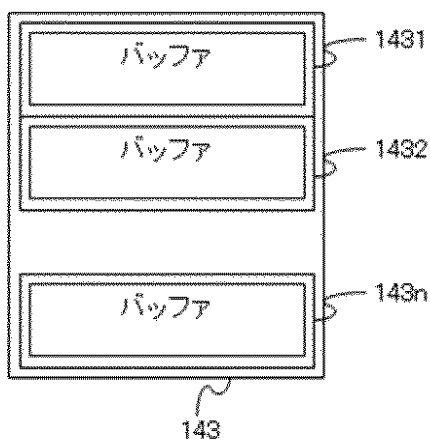


【 図 1 4 】

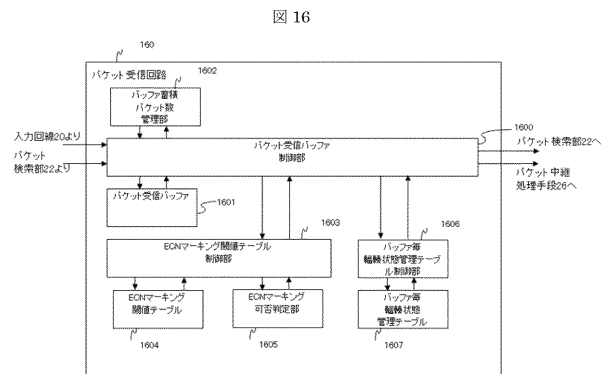


【 図 1 5 】

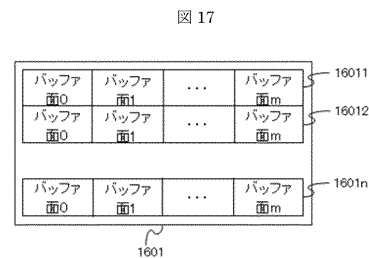
图 15



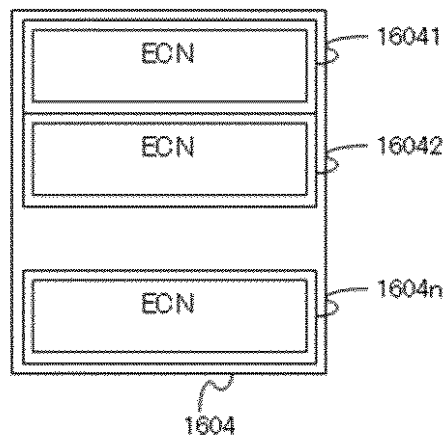
【 図 1 6 】



【圖 17】

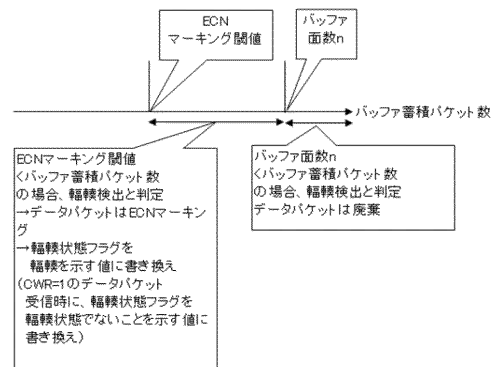


【 図 1 8 】

 18

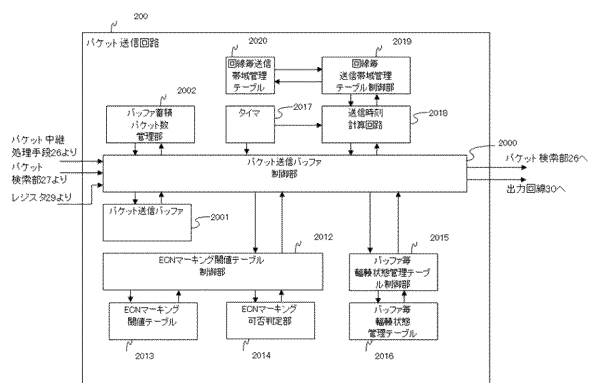
【 図 1 9 】

图 19



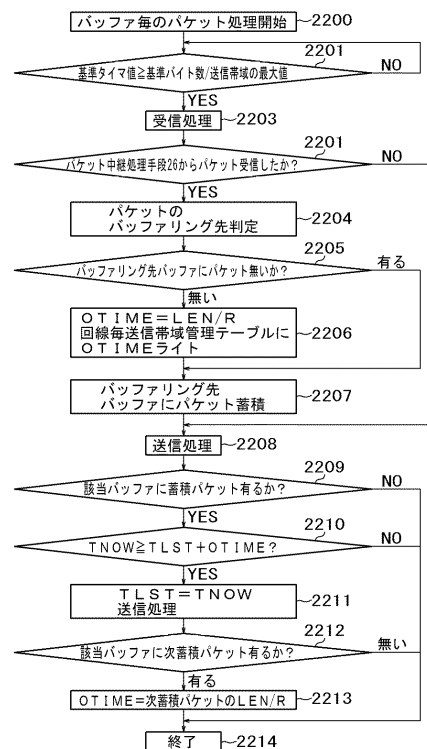
【 図 2 0 】

图 20



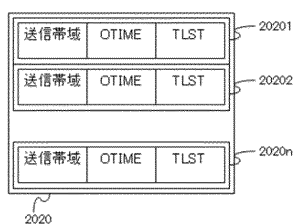
【 図 2 2 】

图 2 2

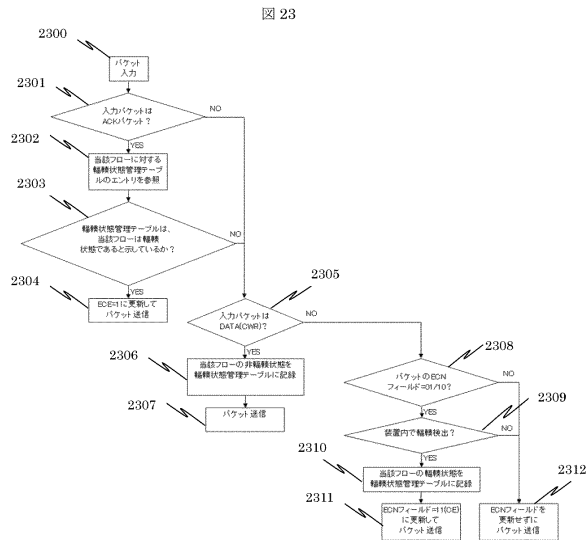


【 図 2 1 】

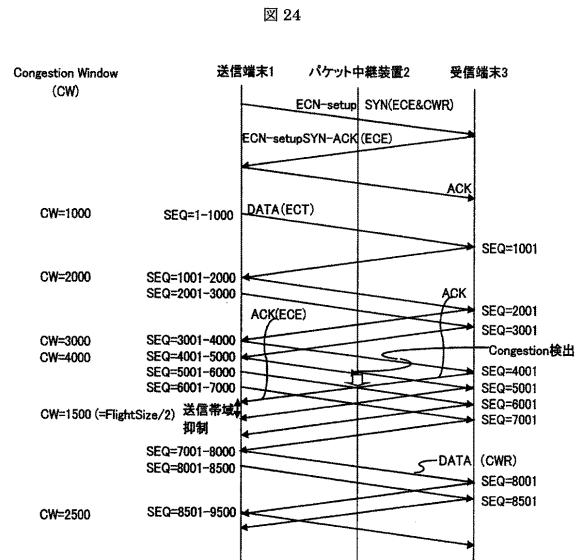
图 21



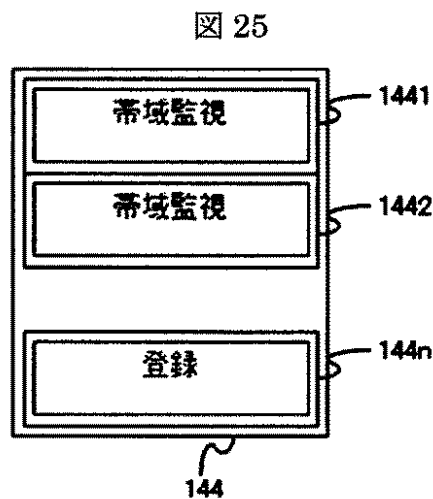
【 図 2 3 】



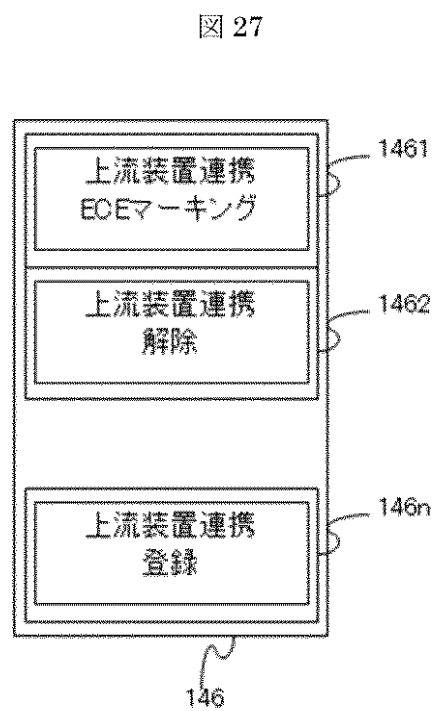
【 図 2 4 】



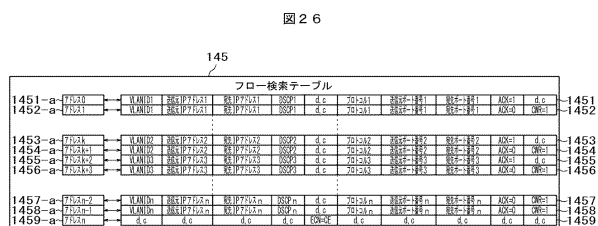
【 図 2 5 】



【 図 2 7 】

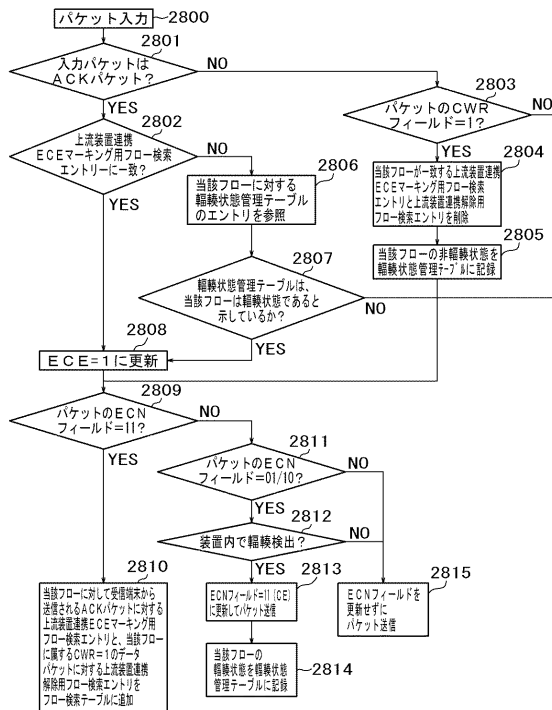


【 図 2 6 】



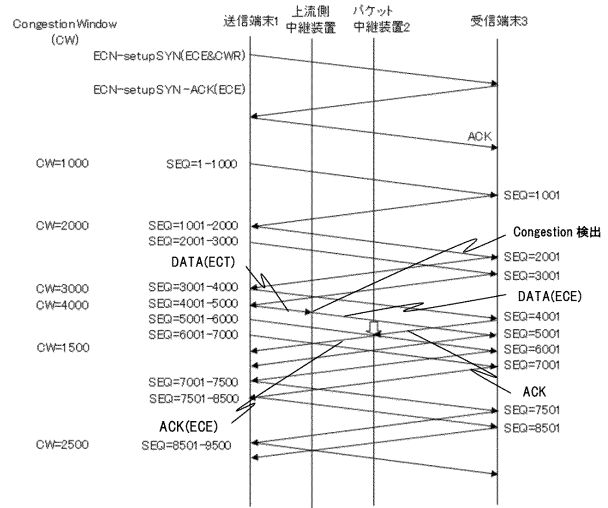
【図 28】

図 28



【図 29】

図 29



【図 30】

図 30

147

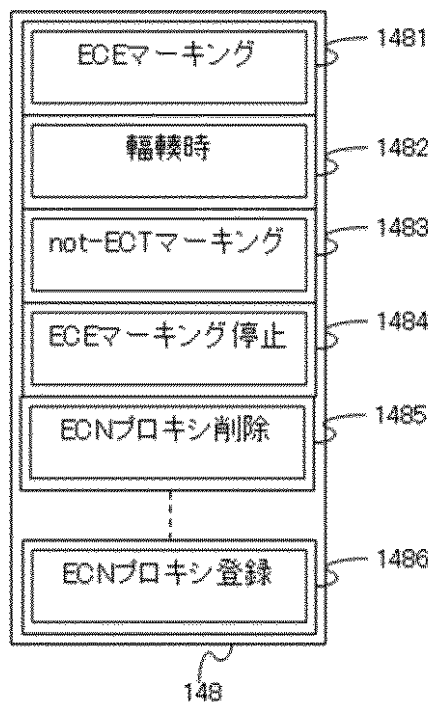
フロー検索テーブル

1471-a	7510	1471	1472	1473	1474	1475	1476	1477	1478	1479	1480	1481	1482	1483	1484	1485	1486	1487	1488	1489	1490	1491	1492	1493	1494	1495	1496	1497	1498	1499	1500
1471-a	7510	1471	1472	1473	1474	1475	1476	1477	1478	1479	1480	1481	1482	1483	1484	1485	1486	1487	1488	1489	1490	1491	1492	1493	1494	1495	1496	1497	1498	1499	1500
1473-a	7512	1473	1474	1475	1476	1477	1478	1479	1480	1481	1482	1483	1484	1485	1486	1487	1488	1489	1490	1491	1492	1493	1494	1495	1496	1497	1498	1499	1500		
1474-a	7513	1474	1475	1476	1477	1478	1479	1480	1481	1482	1483	1484	1485	1486	1487	1488	1489	1490	1491	1492	1493	1494	1495	1496	1497	1498	1499	1500			
1475-a	7514	1475	1476	1477	1478	1479	1480	1481	1482	1483	1484	1485	1486	1487	1488	1489	1490	1491	1492	1493	1494	1495	1496	1497	1498	1499	1500				

1476

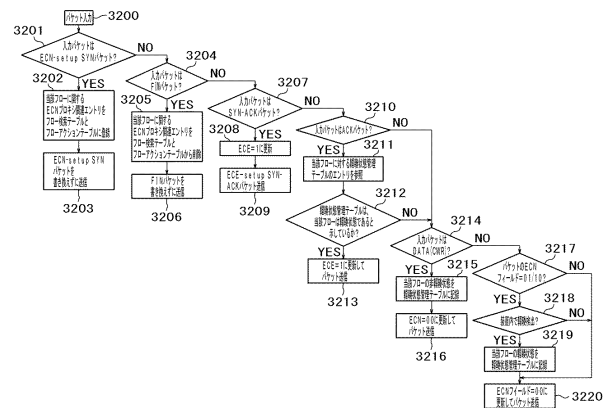
【図 31】

図 31



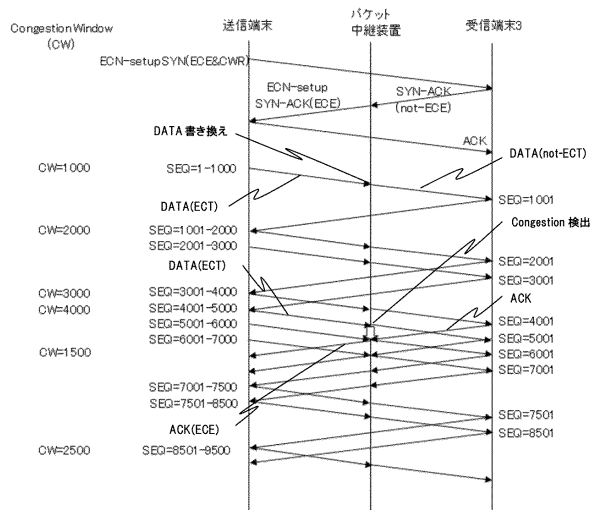
【図 32】

図 32



## 【図 33】

図 33



---

フロントページの続き

- (56)参考文献 特開平09-149077(JP,A)  
特開2000-183965(JP,A)  
特開2010-034708(JP,A)  
特開2011-049658(JP,A)  
国際公開第2008/038390(WO,A1)  
特開2010-232877(JP,A)  
特開2008-104027(JP,A)  
永田 晃、他、「TCP輻輳制御におけるECNを用いたレート制御」,電子情報通信学会技術研究報告,日本,社団法人電子情報通信学会,2001年 7月11日,第101巻 第194号,第39-44頁  
奥山 徹、他、「アプリケーションプログラマが知らなければならないプロトコル TCP講座」,OPEN DESIGN,日本,CQ出版株式会社,2003年 2月 1日,第10巻 第2号,第130-139頁

(58)調査した分野(Int.Cl.,DB名)

H04L 12/00-955