

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4368979号
(P4368979)

(45) 発行日 平成21年11月18日(2009.11.18)

(24) 登録日 平成21年9月4日(2009.9.4)

(51) Int.Cl. F I
HO4L 12/56 (2006.01) HO4L 12/56 F
 HO4L 12/56 200Z

請求項の数 32 (全 28 頁)

(21) 出願番号	特願平11-220626	(73) 特許権者	506355419
(22) 出願日	平成11年8月4日(1999.8.4)		エリクソン アーバー
(65) 公開番号	特開2000-69053(P2000-69053A)		Ericsson AB
(43) 公開日	平成12年3月3日(2000.3.3)		スウェーデン, ストックホルム, トーシャ
審査請求日	平成18年7月5日(2006.7.5)		マンズガーテン 23
(31) 優先権主張番号	09/129284	(74) 代理人	100066728
(32) 優先日	平成10年8月5日(1998.8.5)		弁理士 丸山 敏之
(33) 優先権主張国	米国 (US)	(74) 代理人	100100099
			弁理士 宮野 孝雄
		(74) 代理人	100111017
			弁理士 北住 公一
		(74) 代理人	100119596
			弁理士 長塚 俊也
		(74) 代理人	100141841
			弁理士 久徳 高寛

最終頁に続く

(54) 【発明の名称】 簡易な明示的レート表示アルゴリズムの方法及び装置

(57) 【特許請求の範囲】

【請求項1】

ネットワークのエンティティからのセルをスイッチする装置であって：

ネットワークから全てのセルを受信する入力メカニズムであって、J個(Jは2以上の整数)の入力ポートメカニズムを含み、全ての入力ポートメカニズムは、一緒に、セルをある着信レートで受信する入力メカニズムと；

セルの占有場所を有し、入力メカニズムに接続されたメモリメカニズムであって、全てのセルが装置に格納されるように、入力メカニズムが受信した全てのセルを格納するメモリメカニズムと；

セルを装置からネットワークへ送信する出力メカニズムと；

メモリメカニズム及び出力メカニズムに接続され、あるサービスレートでサービスをセルへ提供するサーバと；

サーバ及びメモリメカニズムに接続され、セルによるメモリメカニズムの占有のファンクションとして、サービスをセルに提供するスケジューラであって、メモリメカニズムの占有を監視する監視メカニズムを含むスケジューラと；

全ての入力ポートメカニズムがセルを受信した時の着信レートを変更する変更メカニズムであって、各エンティティは最小のセルレートを有しており、サービスを要求するエンティティの各々に割り当てられるべき追加の帯域幅の公平なシェアを計算する、変更メカニズムと、

を具備している、セルスイッチング装置。

【請求項 2】

監視メカニズムは、セルのメモリメカニズムへの着信レートと、メモリメカニズムのセルが利用可能なサーバからのサービスレートとの差に基づいて、エンティティが利用可能な追加の帯域幅を決定する、請求項 1 に記載の装置。

【請求項 3】

各々の入力ポートメカニズムは、関連するエンティティに対応するセルのエンティティ着信レートを有し、変更メカニズムは、入力ポートメカニズムを通じてセルを供給する各エンティティのエンティティ着信レートを変更するようにしており、全ての入力ポートメカニズムからのセルの着信レートは、メモリメカニズムのセルが利用可能なサーバからのサービスレートと等しくなり、この着信レートは、全てのエンティティの着信レートの合計である、請求項 2 に記載の装置。

10

【請求項 4】

変更メカニズムは、入力ポートメカニズムを通じてセルを供給する各エンティティのエンティティ着信レートを、等しい量だけ変更する、請求項 3 に記載の装置。

【請求項 5】

メモリメカニズムは、セルスイッチング装置が受信した全てのセルが格納される共有メモリメカニズム、又は別に独立して設けられたバッファのどちらかから構成され、各入力ポートメカニズムは関連バッファを有しており、該関連バッファは、関連入力ポートメカニズム又は各エンティティに対して独立して設けられたバッファだけが受信した全てのセルを格納する、請求項 4 に記載の装置。

20

【請求項 6】

監視メカニズムは、サーバからのサービスを待機するメモリメカニズムのセルのバックログを有するエンティティの個数を監視する、請求項 5 に記載の装置。

【請求項 7】

エンティティは A B R V C を含んでいる、請求項 6 に記載の装置。

【請求項 8】

エンティティは C B R / V B R V C を含んでいる、請求項 7 に記載の装置。

【請求項 9】

監視メカニズムは、サーバからのサービスを待機するセルのバックログを有する A B R V C の個数、即ち、供給される C B R / V B R V C からの C B R / V B R セル (n_{vbr}) の個数を監視する、請求項 8 に記載の装置。

30

【請求項 10】

公平なシェア F S を共有するための追加の帯域幅は、次式で表され、

【数 1】

$$FS = \frac{\max(C_{abr} + U, 0)}{\max(N_a, 1)}$$

ここで、 C_{abr} は、A B R V C が利用可能な残りの帯域幅から、バックログを有する A B R V C の最小セルレート M C R の合計を差し引いたものであって、次式で表され、

40

【数 2】

$$C_{abr} \left(\frac{N - n_{vbr}}{N} L - S_a \right)$$

ここで、L は、出力リンクの帯域幅であり、N は監視メカニズムがパラメータを測定する測定期間のスロットの数であり、 S_a はバックログを有する A B R V C の M C R の合計の指数平均値であり、U は追加の帯域幅パラメータであり、 N_a はサービスを待機するア

50

クティブ A B R V C の指数平均値である、請求項 9 に記載の装置。

【請求項 1 1】

監視メカニズムは、A B R セルの着信レートと A B R V C が利用可能なサービスレートとの差を、A B R メモリメカニズム占有の変化レートに基づいて決定し、その差は $Q = Q_n - Q_p$ で表され、ここで Q_n は、現在の測定期間の開始時における A B R メモリメカニズム占有の総計であり、 Q_p は、現在の測定期間より前の測定期間の開始時における A B R メモリメカニズム占有の総計である、請求項 1 0 に記載の装置。

【請求項 1 2】

変更メカニズムは、着信レートが、現在利用可能なサービスレートに等しくなるまで U を変更する、請求項 1 1 に記載の装置。

【請求項 1 3】

セルをスイッチする方法であって：

スイッチの入力メカニズムにて、ネットワークから J 個(但し、 J は 1 以上の数)のエンティティのセルを受信するステップ；

スイッチのメモリメカニズムにセルを格納するステップ；

サーバにより、メモリメカニズムのセルへサービスを提供するステップ；

メモリメカニズムのセルが、セルによるメモリメカニズムの占有のファンクションとして、スイッチのサーバからサービスを受信するとき、スイッチのスケジューラを用いてスケジューリングするステップ；

スケジューラの監視メカニズムでメモリの占有を監視するステップ；

セルによるメモリメカニズムの占有のファンクションとして、変更メカニズムにより、セルの着信レートを、入力メカニズムの全ての入力ポートメカニズムへ変更するステップであって、スイッチが受信したメモリメカニズムへのセルの着信レートが、メモリメカニズムのセルがサーバから利用可能なサービスレートに等しくなるようにするステップ；

セルのメモリメカニズムへの着信レートと、メモリメカニズムのセルがサーバから利用可能なサービスレートとの差に基づいて、エンティティが利用可能な追加の帯域幅を決定するステップ；

各エンティティは最小セルレートを有し、サービスを要求するエンティティの各々に割り当てられるべき追加の帯域幅の公平なシェアを計算するステップ；

を有しているセルスイッチング方法。

【請求項 1 4】

A T M スwitch の所定リンクにおける A B R トラフィックフローを制御するために、明示的レートを A B R V C へスタンプする A T M スwitch であって、：

V C のトラフィックをスイッチにキャリーする入力ポートメカニズムと；

V C のセルの占有場所を有し、入力ポートメカニズムに接続されたメモリメカニズムであって、全てのセルがスイッチに格納されるように、入力ポートメカニズムが受信した V C の全てのセルを格納するメモリメカニズムと；

セルをスイッチからネットワークへ送信する出力ポートメカニズムと；

V C のセルを、入力ポートメカニズムから出力ポートメカニズムへスイッチするスイッチングファブリックと；

明示的レート(E R)とバックワード R M セルにおけるスタンプレートとを計算する A B R - E R メカニズムであって、各々の A B R V C は最小セルレートを有し、サービスを要求する A B R V C の各々に割り当てられるべき追加の帯域幅の公平なシェアを計算する A B R - E R メカニズムと；

メモリメカニズム及び出力ポートメカニズムに接続され、V C のセルに対して、あるサービスレートでサービスを提供するサーバと；

サーバ及びメモリメカニズムに接続され、V C のセルによるメモリメカニズムの占有のファンクションとして、V C のセルに対してサービスを提供するスケジューラと；
を具備している A T M スwitch。

【請求項 1 5】

A B R - E R メカニズムは、メモリメカニズムの占有を監視し、メモリメカニズムの占有のファンクションとして、E R 計算ファンクションを変更するメカニズムを含んでいる、請求項 1 4 に記載のスイッチ。

【請求項 1 6】

A B R - E R メカニズムは、A B R V C のセルがサーバから利用可能なサービスレートに基づいて、A B R V C が利用可能な追加の帯域幅を決定する、請求項 1 5 に記載のスイッチ。

【請求項 1 7】

A B R - E R メカニズムは、全てのエンティティからのV C の全てのセルの着信レートの合計が、メモリメカニズムのセルがサーバから利用可能なサービスレートに等しくなるように、各々のA B R V C ソースにスタンプされるべきレートを計算する、請求項 1 6 に記載のスイッチ。

10

【請求項 1 8】

メモリメカニズムは、セルスイッチング装置が受信した全てのセルが記憶される共有メモリメカニズム、又は別に独立して設けられたバッファのどちらかから構成され、各V C は関連バッファを有しており、該関連バッファはV C の全てのセルを記憶する、請求項 1 7 に記載のスイッチ。

【請求項 1 9】

A B R - E R メカニズムは、サーバからのサービスを待機するメモリメカニズムのセルのバックログを有するA B R V C の個数を監視する、請求項 1 8 に記載のスイッチ。

20

【請求項 2 0】

A B R - E R メカニズムは、アクティブA B R V C の個数 N_a と、C B R / V B R V C に属するセルの個数 n_{vbr} と、必要に応じてアクティブA B R V C のM C R の合計 S_a を、予め設定された時間のN個のセルスロットの測定期間中監視する、請求項 1 9 に記載のスイッチ。

【請求項 2 1】

N_a 及び S_a などの測定量は、急速な変化を少なくするために、部分的に又は全部が指数平均化される、請求項 2 0 に記載のスイッチ。

【請求項 2 2】

A B R - E R メカニズムは、測定結果から、A B R V C とそれらのM C R が利用可能な合計の帯域幅 C_{abr} を計算する際、 $C_{abr} = (N - n_{vbr}) L / N - S_a$ に基づいて行われる、請求項 2 1 に記載のスイッチ。

30

【請求項 2 3】

A B R - E R メカニズムは、追加の帯域幅パラメータUと呼ばれるパラメータを維持し、該パラメータは、

(a) スイッチでバックログされるV C に対して、別の場所でバックログされるV C により未使用の帯域幅を再分配し、

(b) S_a などの量について、測定結果が正しくないとき、又は測定が行われなかったとき、その補正を行なう、請求項 2 2 に記載のスイッチ。

【請求項 2 4】

40

M C R + 均等シェアの公平性基準を用いるとき、E R 計算メカニズムは、利用可能な帯域幅の公平なシェアF S を、

$$F S = \max (C_{abr} + U , 0) / \max (N_a , 1)$$

として計算する、請求項 2 3 に記載のスイッチ。

【請求項 2 5】

M C R の公平性基準の比例関係を用いることにより、E R 計算メカニズムは、i 番目のA B R V C の公平なシェアF S_iを、

$$F S_i = \max (C_{abr} + U , 0) M C R_i / \max (S_a , 1)$$

として計算する、請求項 2 4 に記載のスイッチ。

【請求項 2 6】

50

A B R - E Rメカニズムは、A T MフォーラムT M 4 . 0によって規定された全ての公平性基準を満たすことができる、請求項 2 5 に記載のスイッチ。

【請求項 2 7】

追加の帯域幅パラメータは、A B Rバッファ占有量 Q_n と、A B Rバッファ占有の変化レートであるデルタ Q を用いて計算される、請求項 2 6 に記載のスイッチ。

【請求項 2 8】

U は、

$Q_n < N$ のとき、 $U = + L$ であり、

そうでなく $Q_n > N (B - N)$ のとき、 $U = - L$ であり、

そうでないときは、 $U = U - \text{アルファ} * \text{ガンマ} * \text{デルタ} Q * L / N$ 、

として計算され、

ここで、B はA B Rバッファサイズであり、アルファは指数平均ファクターであり、ガンマはA B Rバッファ占有重み付け関数である、請求項 2 7 に記載のスイッチ。

【請求項 2 9】

A B Rバッファ占有重み付け関数は、厳密なキュー制御を行ない、A B Rバッファ占有のドリフティングを防止する、請求項 2 8 に記載のスイッチ。

【請求項 3 0】

ガンマは、

$\text{ガンマ} = Q_n / B$ であり、

$\text{デルタ} Q < 0$ のとき、 $\text{ガンマ} = \max (1 - \text{ガンマ} , 0)$

として計算される、請求項 2 9 に記載のスイッチ。

【請求項 3 1】

A B R - E Rメカニズムは、V CのF SをそのM C Rに加算することにより、A B R V Cの明示的レートE Rを計算する、請求項 3 0 に記載のスイッチ。

【請求項 3 2】

公平なシェアF Sを共有するための追加の帯域幅は、次式で表され、

【数 3】

$$FS = \frac{\max(C_{abr} + U, 0)}{\max(N_a, 1)}$$

30

ここで、 C_{abr} は、A B R V Cが利用可能な残りの帯域幅から、バックログを有するA B R V Cの最小セルレートM C Rの合計を差し引いたものであって、次式で表され、

【数 4】

$$C_{abr} \left(\frac{N - n_{vbr}}{N} L - S_a \right)$$

40

ここで、L は、出力リンクの帯域幅であり、N は監視メカニズムがパラメータを測定する測定期間のスロットの数であり、 S_a はバックログを有するA B R V CのM C Rの合計の指数平均値であり、U は追加の帯域幅パラメータであり、 N_a はサービスを待機するアクティブA B R V Cの指数平均値である、請求項 1 3 に記載の方法。

【発明の詳細な説明】

【0 0 0 1】

【発明の属する技術分野】

本発明は、メモリメカニズムの占有のファンクション(function)として、スケジューラを用いてセルのスイッチングを行なう方法及び装置に関するものである。より具体的には、本発明は、A T Mスイッチの与えられたリンクにおけるA B Rのトラフィックフローを制

50

御するために、メモリメカニズムの占有のファンクションとして、スケジューラを用いて、A B R V Cへ明示的レートをスタンプするA T Mスイッチに関する。

【 0 0 0 2 】

【従来の技術】

A T Mネットワークのアベラブルビットレートサービス(Available Bit Rate : A B R)は、時変帯域幅(time-varying bandwidth)に適応可能で、セル損失率の低いバーストデータへの適用に限られていた[ATM Forum. Traffic management specification version 4.0, April 1996参照]。A B Rの仮想接続(Virtual Connections : V C)は、利用可能な帯域幅をかなり共有している。ここで、利用可能な帯域幅とは、固定ビットレート(Constant Bit Rate : C B R)と可変ビットレート(Variable Bit Rate : V B R)のトラフィックストリームによる帯域幅の残りのことである。A B R V Cは、利用可能な帯域幅について公平なシェア(fair share)を獲得するだけでなく、最小セルレート(Minimum Cell Rate : M C R)を特定することができ、A B R V Cが一旦許可を与えられると、そのレートは保証される。

10

【 0 0 0 3 】

ネットワークは、A B Rソースに対して、利用可能な帯域幅に関するフィードバック情報を提供する。フィードバック制御方式に関する研究は次の2つのものがある。クレジットベース(credit-based)の方式では、ネットワークのノードは、上流のノードにV Cごとのクレジットを送り、上流ノードでは、V Cに属するセルを送信する際に、そのクレジットを使い切る。レートベース(rate-based)の方式では、ソースはリソースマネジメント(Resource Management : R M)のセルを定期的を送り、このセルは、目的地(destination)に達するとリターンする。V Cルートにあるスイッチは、リターンするR Mセルの輻輳状態(congestion status)を表示する。

20

【 0 0 0 4 】

クレジットベースの方法(例えば、[H. Kung and R. Morris. Credit-based flow control for ATM networks. IEEE Networks, 9(2):40-48, March/April 1995]を参照)は、概念的にはより簡単であり、セル損失が起こらないことを保証しているけれども、A T Mフォーラムは、レートベースの方式(例えば、[F. Bonomi and K. Fendick. The rate-based flow control framework for the available bit rate ATM service. IEEE Networks, 9(2):25-39, March/April 1995]を参照)を選択した。その理由は、レートベースの方式の方が、構造的な自由度(フレキシビリティ)が大きいためである。レートベースの方法の場合、スイッチアルゴリズムが複雑であるので、輻輳ビット(congestion bit)のマーキングは簡単であるが、明示的レート(Explicit Rate : E R)の計算は複雑である。

30

【 0 0 0 5 】

A T Mフォーラムは、エンドシステムの動作(behavior)を規定したが、A B Rスイッチアルゴリズムのデザインの詳細については、A T Mスイッチの設計者に委ねていた。これまでに、幾つかのE R計算スイッチアルゴリズムが提案されているが、それらのアルゴリズムは、公平なシェアを厳密に計算するアルゴリズムと、公平なシェアを近似的に計算するアルゴリズムの2つのカテゴリーに分類される。公平なシェアの厳密計算アルゴリズムとして、例えばE R I C A、E R I C A +がある[R. Jain, S. Kalyanaraman, R. Goyal, Fahmy S., and R. Viswanathan. ERICA switch algorithm: A complete description. ATM Forum Document Number: ATM Forum/96-1172, August 1996]。E D E R A [N. Ghani. Available Bit Rate Service in ATM Networks. PhD thesis, University of Waterloo, 1997]は、スイッチ出力ポートのネット利用可能帯域幅と、ポートを通過するアクティブA B R V Cの数を測定する。次に、利用可能な帯域幅をアクティブA B R V Cの数で割ることにより、公平なシェアの計算が行われる。一方、公平なシェアの近似的計算アルゴリズムとして、例えばP R C A、E P R C Aがある[L. Roberts. Enhanced PRCA (proportional rate-control algorithm), ATM Forum Document: AF-TM 94-0735R1, August 1994. D M R C AとE D M R C A [F. M. Chiussi and Y. T. Wang. An ABR rate-based congestion control algorithm for ATM switches with per-VC queueing. In GLOBECOM '97, pa

40

50

ges 771-778, 1997]は、浮動小数点(floating-point)の除算は行わない。それは、ポート変数を維持し、該ポート変数はスイッチ出力ポートの輻輳状況に応じて調節される。各VCのERは、これらのポート変数から求められる。一般に、公平なシェアの計算アルゴリズムは、厳密計算アルゴリズムの方が近似的計算アルゴリズムよりも、性能的にすぐれている[N. Ghani. Available Bit Rate Service in ATM Networks. PhD thesis, University of Waterloo, 1997参照]。

【0006】

しかし、公平なシェアの厳密計算方法は重大な問題があり、それは、スイッチでボトルネック(bottle-necked)されたABRVCに対して、どこか他の場所でボトルネックされたABRVCから帯域幅を再分配することである。例えば、各々がゼロMCRを有する3個のアクティブABRVCを考えてみることにする。ネット利用可能な帯域幅を30Mビット/秒とし、各々のABRVCの公平なシェアを10Mビット/秒とする。最初の2つのABRVCはスイッチでボトルネックされると仮定し、第3のABRVCはどこか別のスイッチでボトルネックされると仮定する。なお、この別のスイッチは、第3のABRVCが5Mビット/秒を超えないようにしている。これが意味するのは、 $30 - (10 + 10 + 5) = 5$ Mビット/秒の帯域幅は、最初の2つのABRVCの間で再分配されなければならないということである。よって、3つのVCの公平なシェアは、それぞれ、12.5、12.5、5.0Mビット/秒となる。

【0007】

輻輳回避のための明示的レート表示(Explicit Rate Indication for Congestion Avoidance; ERICA)、及びERICA+[R. Jain, S. Kalyanaraman, R.Goyal, Fahmy S., and R. Viswanathan. Determining the number of activeABR sources in switching algorithms. ATM Form Document Number: ATM Forum/98-0154, February 1998; R. Jain, S. Kalyanaraman, R. Goyal, Fahmy S.,and R. Viswanathan. ERICA switch algorithm: A complete description.ATM Form Document Number: ATM Forum/96-1172, August 1996]は、再分配のために複雑なスキームを採用しており、VCごとのサービスレートの測定を必要とする。例えば、ERICA+は、利用可能な帯域幅を、アクティブABRVCの合計数ではなく、アクティブABRVCの有効数で割ることにより、再分配の問題を解決している。アクティブABRVCの有効数は次のように定義される。スイッチでボトルネックされるABRVCは1としてカウントされ、他の場所でボトルネックされるABRVCは、(実際のVCレート/公平なシェア)に等しい部分としてのみカウントされる。このように、アクティブABRVCの有効数を計算するためには、各VCのセルレートは、スイッチポートで測定されなければならない。VCレートの測定には、 $O(n)$ 個の空間コンプレキシティ(space complexity)及び/又は $O(n)$ 個の時間的コンプレキシティ(time complexity)を有するアルゴリズムを必要とする。ここで、 n はポートを通過するVCの数である。このため、このようなスイッチアルゴリズムを高速ハードウェアで実行するには、費用が非常にかかる。

【0008】

分散を向上させた明示的レートアルゴリズム(Enhanced Distributed Explicit Rate Algorithm(EDERA))[N. Ghani. Available Bit Rate Service in ATM Networks. PhD thesis, University of Waterloo, 1997]は、スイッチでボトルネックされたABRVCと、他の場所でボトルネックされたABRVCのトラックを、明示的に保持することにより、再分配の問題を解決する。EDERAでは、ABRソースが、その分割された平均ソースレート(Segmented Average Source Rate: SASR)をRMセルに表示すると仮定している。ここでSASRは、2つのRMセル間のデータセル数の、2つのRMセル間の時間に対する比として規定される。VCがネットワーク内のある地点でボトルネックされると、SASRは、許可セルレート(Allowed Cell Rate: ACR)よりも小さくなることに留意すべきである。ATMフォーラムの標準[ATM Forum. Traffic management specification version 4.0, April 1996]では、RMセルにSASR表示を含まないようにすることを決定したので、EDERAスキームは適宜修正されなければならない。このような修正を

10

20

30

40

50

行うためには、スイッチ出力ポートでVCレートを測定する必要がある。

【0009】

【発明が解決しようとする課題】

本発明は、新規なスイッチアルゴリズム、つまり、簡易な明示的レート表示アルゴリズム (Simple Explicit Rate Indication Algorithm: SERIA)に関するものである。ERICA、ERICA+及びEDERAとは異なり、SERIAは、VCごとのレートを測定する必要がない。実際に、SERIAは、これまでに提案された公平なシェアの厳密計算スイッチアルゴリズムよりもはるかに簡単である。つまり、セル着信(cell arrival)やセル出発(cell departure)の如きイベント(events)の取扱いに必要な算術的及び論理的な演算の数は少なく、VCの数から独立している。SERIAが簡単になると、高速のイン
10 プリメンテーションにとって特に魅力的である[M. Vishnu, D. Basak, and H.S. Kim. Method and apparatus for a simple explicit rate indication algorithm (SERIA) U.S. Patent Pending, 1998]。

【0010】

広範なシミュレーションを行なった結果、SERIAは、最大-最小フェアネス(max-min fairness)、つまり最大-最小公平性を提供するもので、定常状態に迅速にコンバージ(converge)し、リンク帯域幅を100%利用し、バッファ制御をしっかりと維持することを示している。

【0011】

【課題を解決するための手段】

本発明は、ネットワークのエンティティからセルをスイッチングする装置に関する。セルスイッチング装置は、ネットワークからの全てのセルを受信する入力メカニズムを具えている。装置は、入力メカニズムが受信した全てのセルを格納するメモリメカニズムを具えており、全てのセルは装置内に格納される。メモリメカニズムは、入力ポートメカニズムに接続されている。メモリメカニズムは、セルが占有する占有場所(occupancy)を有している。セルスイッチング装置は、該装置からネットワークへセルを送信する出力メカニズムを具えている。装置は、あるサービスレートでセルにサービスを提供するサーバを具えている。サーバは、メモリメカニズムと出力メカニズムに接続されている。装置は、メモリメカニズムのセルに対してサービスを提供するスケジューラを具えている。スケジューラは、サーバとメモリメカニズムに接続されている。
20

【0012】

本発明は、セルをスイッチングする方法に関する。この方法は、J個(但し、Jは1以上)のエンティティのセルを、スイッチの入力メカニズムでネットワークから受信するステップを含んでいる。次に、セルをスイッチのメモリメカニズムに格納するステップがある。次に、サーバにより、メモリメカニズム内のセルへサービスを提供するステップがある。次に、メモリメカニズムにセルが占有された結果、メモリメカニズムのセルが、スイッチのサーバからのサービスを受信するとき、スイッチのスケジューラを用いてスケジューリングを行なうステップがある。
30

【0013】

本発明は、ATMスイッチの与えられたリンクにおけるABRのトラフィックフローを制御するために、ABRVCへ明示的レートをスタンプするためのATMスイッチに関する。本発明の装置は、VCへのトラフィックをスイッチへキャリーする入力ポートメカニズムを具えている。装置は、入力ポートメカニズムが受信したVCの全てのセルを格納するためのメモリメカニズムを具えており、セルはスイッチの中に格納される。メモリメカニズムは、入力ポートメカニズムに接続されている。メモリメカニズムは、VCのセルを占有する占有場所を有している。装置は、セルをスイッチからネットワークへ送信する出力ポートメカニズムを具えている。装置は、VCのセルを、入力ポートメカニズムから出力ポートメカニズムへスイッチするためのスイッチングファブリックを具えている。装置は、バックワード(backward)RMセルにおける明示的レート(ER)とスタンプレートを計算するABR-ERメカニズムを具えている。装置は、あるサービスレートでVCのセルへ
40
50

サービスを提供するサーバを具えている。サーバは、メモリメカニズム及び出力ポートメカニズムに接続されている。装置は、V Cのセルによるメモリメカニズムの占有のファンクションとして、V Cのセルへサービスを提供するスケジューラを具えている。スケジューラは、サーバとメモリメカニズムに接続されている。

【0014】

【発明の実施の形態】

図面を参照すると、同一又は類似の要素については、幾つかの図を通して、同じ引用符号を付している。特に図41を参照すると、ネットワーク(12)のエンティティからのセルをスイッチングする装置(10)が示されている。装置(10)は、ネットワーク(12)から全てのセルを受信する入力メカニズム(14)を具えている。装置(10)は、入力メカニズム(14)が受信した全てのセルを格納するメモリメカニズム(16)を具えており、全てのセルは、装置(10)の中に記憶される。メモリメカニズム(16)は、入力ポートメカニズムに接続されている。メモリメカニズム(16)は、セルの占有場所(occupancy)を有している。装置(10)は、セルを装置(10)からネットワーク(12)へ送信する出力メカニズム(18)を具えている。装置(10)は、あるサービスレートでセルへサービスを提供するサーバ(20)を具えている。サーバ(20)は、メモリメカニズム(16)及び出力メカニズム(18)に接続されている。装置(10)は、セルによるメモリメカニズムの占有のファンクションとして、セルへサービスを提供するスケジューラ(22)を具えている。スケジューラ(22)は、サーバ(20)及びメモリメカニズム(16)に接続されている。

【0015】

望ましくは、入力メカニズム(14)はJ個の入力ポートメカニズム(24)を含んでおり、Jは2以上の整数である。全ての入力ポートメカニズム(24)は、一緒に、ある着信レートでセルを受信することが望ましく、ここでスケジューラ(22)は、メモリメカニズム(16)の占有を監視する監視メカニズム(26)と、セルが全ての入力ポートメカニズム(24)により受信された着信レートを変更させる変更メカニズム(28)を含んでいる。

【0016】

望ましくは、監視メカニズム(26)は、セルのメモリメカニズム(16)への着信レートと、メモリメカニズム(16)のセルがサーバ(20)から利用可能な提供レートとの差に基づいて、エンティティが利用可能な追加の帯域幅を決定する。各エンティティは最小セルレートを有することが望ましく、変更メカニズム(28)は、サービスを要求するエンティティの各々に割り当てられべき追加の帯域幅の公平なシェアを計算する。

【0017】

望ましくは、入力ポートメカニズム(24)の各々は、エンティティに対応するセルのエンティティ着信レートを有しており、変更メカニズム(28)は、入力ポートメカニズムを通じてセルを供給する各エンティティのエンティティ着信レートを変更するので、全ての入力ポートメカニズム(24)からのセルの着信レートは、メモリメカニズム(16)のセルがサーバ(20)から利用可能なサービスレートに等しく、前記着信レートは全てのエンティティ着信レートの合計となる。変更メカニズム(28)は、入力ポートメカニズムを通じてセルを供給する各エンティティのエンティティ着信レートを、等量だけ変更することが望ましい。

【0018】

望ましくは、メモリメカニズム(16)は、装置(10)が受信した全てのセルが格納される共有メモリメカニズム(shared memory mechanism)(30)か、又は別に設けたバッファ(32)のどちらか一方から構成され、各入力ポートメカニズムは、関連入力ポートメカニズム、又は各エンティティに対して独立して設けられたバッファ(32)だけが受信した全てのセルを格納する関連バッファ(associated buffer)を有している。監視メカニズム(26)は、メモリメカニズム(16)の中でサーバ(20)からのサービスを待機するセルのバックログ(backlog)を有するエンティティの個数を監視することが望ましい。

【0019】

エンティティはA B R V Cを含むことが望ましい。エンティティは、C B R / V B R V Cを含むことが望ましい。監視メカニズム(26)は、サーバ(20)からのサービスを待機するセ

10

20

30

40

50

ルのバックログを有する A B R V C の数、即ち、提供された C B R / V B R V C から C B R / V B R セル (n_{vbr}) の数を監視することが望ましい。

【 0 0 2 0 】

望ましい実施例の実施において、スイッチ (34) の出力ポートメカニズム (40) を考える。N は、ポートを通過するアクティブ A B R V C の集合 (set) を表すものとする。A B R V C は、そのソースがゼロ以外のレートでセルを送信する場合、アクティブである。集合 N は、次式で示されるように、 N_{bs} と N_{be} に分けられた 2 組の集合の和である。

【 数 3 】

$$N_{\alpha} = N_{bs} \cup N_{be}$$

10

ここで N_{bs} は、スイッチ (34) のポートでボトルネックされるアクティブ A B R V C の集合であり、 N_{be} は、別の場所でボトルネックされるアクティブ A B R V C の集合である。N、 N_{bs} 及び N_{be} は、それぞれ、N、 N_{bs} 及び N_{be} の集合における A B R V C の個数とする。 C_{abr} は、利用可能な帯域幅の量を表すものとする。

【 0 0 2 1 】

$N_{be} =$ (N_{be} が空集合) のとき、公平なシェア (Fair Share : F S) は、式 (1) に等しい。

20

【 数 4 】

$$FS = \frac{C_{abr} - \sum_{i \in N_a} MCR_i}{N_a} \quad (1)$$

なお、 MCR_i は、 i 番目の A B R V C の最小セルレート (minimum cell rate : M C R) である。また、 i 番目の V C の明示的レート (E R) は、 $FS + MCR_i$ となる。

【 0 0 2 2 】

30

N_{be} (N_{be} が空集合でない) のとき、 N_{be} 中の A B R V C は、割り当てられた F S を利用することができないため、結果として、リンク帯域幅の利用不足 (under-utilization) となる。リンク帯域幅の 100% 利用を達成するには、 N_{be} 中の A B R V C が未使用の帯域幅は、 N_{bs} 中の A B R V C に再分配されなければならない。未使用の帯域幅の再分配は、公平なシェアの厳密な計算 A B R アルゴリズムにおいて、重大な要素の 1 つである。

【 0 0 2 3 】

B_i は、 i 番目の A B R V C が利用できる帯域幅の最大量を表すものとし、L はリンク帯域幅を表すものとする。 i 番目の A B R V C が N_{bs} にある場合には、 $B_i = L$ と設定する。帯域幅の再分配により、F S の値は次の式 (2)、即ち式 (3)、を満足させるようにする。

【 数 5 】

40

$$\sum_{i \in N_a} \min(B_i, MCR_i + FS) = C_{abr} \quad (2)$$

【 数 6 】

$$\sum_{i \in N_{ba}} (MCR_i + FS) + \sum_{j \in N_{ba}} B_j = C_{abr} \quad (3)$$

これは、次の式(4)を満足することを意味する。

【数7】

10

$$FS = \frac{C_{abr} - \sum_{i \in N_a} MCR_i - \sum_{j \in N_{ba}} B_j}{N_{bs}} \quad (4)$$

【0024】

FSを計算するために、公平なシェアの厳密な計算スイッチ(34)のアルゴリズムは、次式のトラックを保持(keep)する必要があり、これは些細なタスクではない。

20

【数8】

$$A = -\sum_{i \in N_{ba}} MCR_i - \sum_{j \in N_{ba}} B_j$$

30

【0025】

Aのトラックをキープする代わりに、それを推定(estimation)することができる。Aがアンダーエスティメート(又はオーバーエスティメート)される場合、つまり、Aが低く推定(又は高く推定)される場合、ABRVCセルのネット着信レートは、ABRVCが利用可能なネットサービスレートよりも少なく(又は多く)なるであろう。SERIAは、Aの推定にこの事実を利用する。SERIAは、Uで表される追加の帯域幅パラメータと呼ばれる変数Uを維持しており、この変数は、式(5)で示されるように、公平なシェアを計算するときに C_{abr} へ加算される。

【数9】

40

$$FS = \frac{\max(C_{abr} + U, 0)}{\max(N_{bs}, 1)} \quad (5)$$

式(6)中、分母の $\max(\)$ は、ゼロによる割算を回避するためのものである。SERIAによるUの調節は、ABRセルのネット着信レートがABRVCの利用可能なネットサービスレートに等しくなるまで行われる。

50

【 0 0 2 6 】

本明細書中で用いられる記号を表 1 にリストアップしている。S E R I A の疑似コードの完全な記載は、図 4 4 及び図 4 5 に示している。

【 0 0 2 7 】

【表 1】

N	測定期間のスロットの数	
N_a	アクティブ ABR VC の集合 (セット)	10
N_{bs}	スイッチポートでボトルネックされたアクティブ ABR VC の集合	
N_{be}	別の場所でボトルネックされたアクティブ ABR VC の集合	
N_s	アクティブ ABR VC の数	
N_{bs}	スイッチポートでボトルネックされたアクティブ ABR VC の数	
N_{bs}^n	N_{bs} の測定値	
L	出力リンク帯域幅	
B_i	i 番目の ABR VC が使用できる帯域幅の最大量	
N_g	現在の測定期間に供給された CBR/VBR セルの数	
T	VC ごとキューの長さの閾値	20
q_i	i 番目の VC の VC ごとキューの長さ	
Q_n	現在の ABR バッファ占有	
Q_p	以前の ABR バッファ占有	
ΔQ	差 $\Delta Q = Q_n - Q_p$	
α_u	U に対する指数平均ファクター	
α_n	N_{bs} に対する指数平均ファクター	
C_{abr}	VBR/CBR VC による帯域幅の残り	
U	追加の帯域幅パラメータ	
γ	バッファの重み付け関数	30
FS	公平なシェア	
B	ABR バッファの全容量	
β	バッファのフラクシオン	
L	1 秒当たりのセルの出力リンク帯域幅	
R	比率 $R = L / N$	

【 0 0 2 8 】

〔 C_{abr} の計算 〕

N 個のスロットの各測定期間に、S E R I A は、帯域幅が保証された VC (即ち CBR VC と、VBR VC) から提供されたセル n_g のトラックを保持する。なお、CBR VC と VBR VC による帯域幅の残り (bandwidth leftover) は次の式 (6) に等しい。

【数 1 0】

40

$$C_{abr} = \frac{N-n}{N} g L \quad (6)$$

【 0 0 2 9 】

〔 U の計算 〕

前述したように、S E R I A は、A B R セルのネット着信レートが A B R V C の利用可能なネットサービスレートに等しくなるまで、追加の加帯域幅パラメータ U を調節する。S E R I A は、A B R V C のネット着信レートが、A B R V C の利用可能なネットサービスレートよりも小さい(大きい)とき、U を増加(減少)させる。着信レートとサービスレートとの差は、A B R バッファの占有が変化するレートにより決定される。

10

【 0 0 3 0 】

B は、A B R バッファのサイズとし、 Q_n は、現在の測定間隔の A B R バッファ占有の総計(aggregate)を表し、 Q_p は、以前の測定間隔の A B R バッファ占有の総計を表すものとする。そして、 $Q = Q_n - Q_p$ とする。各測定間隔において、 Q_n が小さい ($< N$) ならば、U はラインレート + L に設定される。 Q_n が非常に大きい ($> (B - N)$) ならば、U は - L に設定される。それ以外の場合には、U は次の式(7)のように更新される。

20

【 数 1 1 】

$$U = U - \alpha_u Y \Delta Q \frac{L}{N} \quad (7)$$

ここで α_u は、指数平均化ファクター(exponential averaging factor)であり、Y は後記する重み付け関数(weighting function)である。A B R バッファの占有が増加すれば(即ち、 $Q > 0$ のとき)、U は減少する。同じ様に、A B R バッファの占有が減少すれば、U は増加する。指数平均化ファクター α_u は、U のオシレーションを小さくするために用いられる。

30

【 0 0 3 1 】

リンク中のネット負荷(net load)が 1 0 0 % よりも少ないとき、式(7)は、リンク帯域幅を十分に利用しようとして、U の値を非常に高い値まで上昇させる。U の値がそこまで高くなると、ネット負荷が後に増加したとき、U の値は急速に低下することができない。それゆえ、アルゴリズムは、その高慣性(high inertia)により、速やかに反応することができず、ネットワーク(12)の負荷状態に迅速に変化することができない。U がそのような高い値まで上昇しないように、U の上限を定めなければならない。これに対して、A B R セルのネット着信レートが、A B R V C の利用可能なネットサービスレートよりも大きい場合、A B R ソースをしてそれらのセルレートを低下させるために、式(7)は U の値を低下させる。ラウンドトリップ遅延(round trip delay)が大きい場合、ソースの反応は遅れる。この期間に、U は非常に大きな負の値(negative value)まで低下する。U の値がそこまで低くなると、後になって帯域幅が利用可能になったときに、正の値(positive value)へ速やかに変化することができない。従って、U の下限も同じ様に設定するのが有利である。 C_{abr} の範囲は $[0, L]$ であり、F S の必要範囲もまた $[0, L]$ であるので、U の必要十分な範囲は $[-L, +L]$ である。

40

【 0 0 3 2 】

〔 N_{bs} の計算 〕

50

N_{bs} 中のアクティブなABRVCは、スイッチ(34)のポートに、VCごとのキューで認識可能な長さのものを一切作らない。従って、 N_{bs} の良好な推定は、VCごとのキュー長さ(per-VC queue length)が幾つかの低閾値Tを越えるABRVCの集合(set)である。Tの値の適当な範囲は、3乃至10セルである。 N_{bs}^n は測定値を表し、 N_{bs} は指数平均値を表すものとする。 N_{bs}^n は、次の要領で測定される。ポートは、初期状態では、 N_{bs}^n と N_{bs} は両方共、ゼロに設定される。セルがABRVCに到着し、そのVCごとのキュー長さをT-1からTまで増加させる時はいつでも、 N_{bs}^n は増分される(incremented)。セルが発し、VCごとのキュー長さをTからT-1まで減少させる時はいつでも、 N_{bs}^n は減分される(decremented)。これは、次の式(8)で表される。

【数12】

10

$$N_{bs}^n = (1 - \alpha) N_{bs}^n + \alpha N_{bs}^n \quad (8)$$

なお、 n は指数平均化パラメータである。

【0033】

〔Q重み付け関数〕

20

ABRスイッチ(34)アルゴリズムが、 Q_n の値を考慮せずに、Qのみに基づいてFSの調節を行なう場合、 Q_n はゆっくりと高い値ヘドリフト、その結果、セルの損失は多くなる。従来のABRスイッチ(34)のアルゴリズムでは、閾値を用いてABRバッファをいくつかの領域に分割し、 Q_n がこれらの閾値を越えた時に、FSの計算値を漸進的(progressively)に低下させることにより、このようなドリフトを防止していた。

【0034】

SERIAでは、より簡易な手法を用いており、Qは、 Q_n (及びQ)の関数であるYによって重み付けされる。適当な重み付け関数として、次の例を挙げることができる。

$Y = Q_n / B$ であって、

もし、($Q < 0$)のとき、

30

$Y = \max(1 - Y, 0)$ であり、

Bは、ABRVCが利用可能なバッファ空間の合計であり、 B は、パラメータであり、現在0.8に設定されている。

【0035】

ABRバッファの占有が増加している(即ち、 $Q > 0$)と仮定すると、Uは減少するだろう。重み付けの結果、ABRバッファが略完全に占有される場合の方が、ABRバッファが殆ど占有されていない場合よりも、Uの減少は確実に多くなる。同じ様に、ABRバッファの占有が減少している(即ち、 $Q < 0$)場合、Uは増加する。重み付けの結果、ABRバッファが殆ど占有されていない場合の方が、ABRバッファが略完全に占有される場合よりも、Uの増加は確実に多くなる。

40

【0036】

Qの重み付けにより、 Q_n は約 $B/2$ となる。この結果、フリースロットを満たすために、ABRVCに属する十分なセルがABRバッファ内で確実に利用可能となる。さらに、シミュレーションの結果では、Qを重み付けすることにより、従来のスイッチ(34)のアルゴリズムで採用された方法よりも、 Q_n の安定性がより高められることを示している。

【0037】

図1は、ATMスイッチ(34)のポートの構成を簡略化して示している。出口ユニット(egress unit)は、FSの計算、セルバッファの管理、スケジューリング、ポリシーリング(policying)その他の機能を行なう。FSの計算値は入口ユニット(ingress unit)へ送られ、該

50

ユニットは、ER値をバックワードRMセルにスタンプする。セルはセルメモリに記憶され、VCごとの情報はテーブルメモリに記憶される。SERIAは、これまでの公平なシェアの厳密計算アルゴリズムよりも、高速ATMスイッチに組み込むのが遙かに簡単である。その理由は次の通りである。

【0038】

SERIAは、FSを計算するのにどんなメモリアクセスを実行する必要がない。SERIAで用いられる幾つかの変数は、出口ユニットのレジスタに保存することができる。これに対して、VCごとのフラグ又はVCごとの変数を用いるスイッチ(34)のアルゴリズムでは、スロットごとに $O(n)$ 回のテーブルメモリアクセスを実行する必要がある。

【0039】

第2に、バックワードRMセルが到着するとき、VCI/VPIのルックアップを実行する必要がない。計算されたFSは全てのABRVCに共通しており、RMセルはそのMCRを含んでいるので、VCの特定の情報にアクセスするためにVCI/VPIをルックアップする必要はない。

【0040】

第3に、SERIAは、バッファの管理やサービスのスケジューリングに関して、特定の要求を一切行わない。従って、SERIAは様々なスケジューリングとバッファ管理のスキームと共に作業することができる。

【0041】

〔公平性(フェアネス)〕

利用可能な帯域幅は、数多くの公平性(フェアネス)の基準(criteria)を用いて分配されることができる。MCR+均等シェアの公平性(MCR plus equal share fairness)の基準では、利用可能な帯域幅は、アクティブABRVCの間で等しく分けられる。MCR比例(proportional to MCR)の基準では、利用可能な帯域幅は、MCRに比例して分けられる。MCR比例基準の主な欠点は、MCRを有しないABRVCが追加の帯域幅を全く獲得できないことである。このため、多くのABRアルゴリズムでは、MCR+均等シェアの公平性基準となるように設計されている。図44及び図45に示された疑似コードは、MCR+均等シェアの公平性がもたらされるSERIAを示している。それは、ATMフォーラム[ATM Forum. Traffic management specification version 4.0, April 1996]に掲げられたその他の公平性基準に修正することは容易である。

【0042】

最大-最小フェアネスの定義[D. Bertsekas and R. Gallager. Data Networks, Second Edition. Printice Hall, Inc., 1992]は、ATMフォーラムドキュメント[N. Yin. Fairness definition in ABR service model. ATM Forum Document: AF-TM 94-0928R2, Number 1994]のMCRも含まれるように拡張された。拡張された最大-最小フェアネス基準の定義は、次の通りである。FSを式(2)の唯一の解とすると、次式の関係性を有するとき、 i 番目のABRVCが $\min(B_i, MCR_i + FS)$ の帯域幅を割り当てられる場合にも、拡張された最大-最小フェアネス基準は充足される。

【数13】

$$i \in N_a$$

SERIAは、拡張された最大-最小フェアネス基準を満たす。

【0043】

〔バーストVBRVCが存在するときのロバストパフォーマンス〕

ABRスイッチ(34)アルゴリズムの実テストは、非常にバーストなVBRストリームの存在下でのパフォーマンスである。即ち、ABRVCは、割り当てられていない帯域幅を利

10

20

30

40

50

用しなければならないだけでなく、C B R及びV B R V Cが一時的に使用していない帯域幅を使用することができなければならない。多くのA B Rアルゴリズムは、このようなシナリオではうまく実行できない[F.M. Chiussi and Y.T. Wang. An ABR rate-based congestion control algorithm for ATM switches with per-VC queueing. In GLOBECOM '97, pages 771-778, 1997参照]。S E R I Aの強さの1つは、非常にバーストなV B Rストリームのときにも、非常にうまく実行できることである。

【0044】

〔多数のA B R V Cが存在するときのロバストパフォーマンス〕

公平なシェアの近似計算アルゴリズムは、 N_a または N_{bs} のトラックを保持しないので、A B R V Cの数が多いと、大きなオシレーションを受ける。このアルゴリズムが自由帯域幅を検出すると、各A B R V CのF Sを増加させて、着信レートを大幅に増加させる。これによって、アルゴリズムはそのF Sを急に低下させるので、各A B R V Cのセルレートは減少する。これは、オシレーションを生じ、過渡期間(transient period)は長くなる。この問題は、公平なシェアの厳密計算アルゴリズムでは起こらない。シミュレーション結果では、多数のA B R V Cが存在するときにも、S E R I Aのロバストパフォーマンスを示している。

10

【0045】

〔パーキングロットのようなトポロジにおけるロバストパフォーマンス〕

重要なことは、スイッチ(34)のアルゴリズムは、ホップの数が少ないA B R V Cが有利となるように、ホップの数が多いたA B R V Cを識別しないことである。従来のスイッチ(34)のアルゴリズムはこの問題を抱えており、文献の中で、「ビートダウン問題(beat-down problem)」と称されている[F. Bonomi and K. Fendick. The rate-based flow control framework for the available bit rate ATM service. IEEE Networks, 9(2):25-39, March/April 1995]。シミュレーションの結果でも、S E R I Aは、数多くのホップを有するA B R V Cを区別しないことが示されている。

20

【0046】

〔シミュレーション結果〕

全てのリンクはO C - 3リンク(155.52Mビット/秒)であり、スイッチは、8KセルのA B Rバッファサイズを有するS E R I Aを用いている。スイッチは、V Cごとのキューイングと、簡易に重み付けされたラウンドロビン(Weighted Round Robin: W R R)スケジューラ(22)を用いている。W R Rスケジューラ(22)の重み付けは、A B R V Cの場合にはM C Rに比例して設定され、V B R V Cの場合にはP C Rに比例して設定される。測定間隔Nは64に設定される。V Cごとキューの閾値Tは5セルに設定される。指数平均化パラメータ α_n 及び α_u は共に1/8に設定される。

30

【0047】

広帯域端末装置(Broadband Terminal Equipments: B T E)は、A T Mフォーラムの文献の中で定義されたエンドシステムアルゴリズムを用いている[ATM Forum. Traffic management specification version 4.0, April 1996]。B T Eアルゴリズムのパラメータ値を表2に示す。これらの値は、A T Mフォーラムの文献の中で推奨されるデフォルト値と同一である[ATM Forum. Traffic management specification version 4.0, April 1996]。これらの文献はその引用を以て本願への記載加入とする。

40

【0048】

V B Rソースは、50msのO NとO F Fの周期で、O N/O F Fのストリームを発生させる。O N周期では、ソースは指示されたピークセルレート(Peak Cell Rate: P C R)でセルを送信し、O F F周期では、ソースはアイドル状態である。

【0049】

シミュレーション実験で用いられる多くのA B R V Cは持続性である(persistent)。即ち、それらは、ネットワークから指示された許可セルレート(Allowed Cell Rate: A C R)と等しいレートで送信するのに十分なセルを常に有している。

【0050】

50

【表 2】

エンドシステムアルゴリズムで使用されるパラメータ

R I F	1 / 1 6
N r m	3 2セル
M r m	2
R D F	1 / 1 6
C R M	3 2セル
A D T F	5 0 0秒
T r m	1 0 0秒
C D F	1 / 1 6
T C R	1 0セル/秒

10

20

【 0 0 5 1 】

〔実験 1〕

〔セットアップ〕

図 2 は、使用したトポロジを示している。セットアップは、10 個の A B R ソースからなる。A B R 1 の M C R は 5 0 M ビット / 秒、A B R 2 の M C R は 1 0 M ビット / 秒である。一方、A B R 3 - A B R 1 0 の M C R は 1 M ビット / 秒である。A B R 2 以外の全ての A B R V C は持続性であるが、A B R 2 はソースでボトルネックされて、5 M ビット / 秒となる。リンク L 1 - L 1 0 と L 1 s - L 1 0 s の長さは 1 0 0 km であるのに対し、リンク L a の長さは 1 0 0 0 km である。

【 0 0 5 2 】

〔考察〕

実験 1 の結果を図 3 乃至図 1 2 に示す。リターン R M セルのソースに対してネットワーク (12) により指示された A C R 値のプロット、実際の A B R ソースレート、並びにボトルネックスイッチ s w 1 のパラメータのプロット及びソースレートが示されている。式 (2) に代入して F S の解を求めた結果を式 (9) に示している。

【数 1 4】

$$(-50 + FS) + 5 + 8(1 + FS) = 155.52 \rightarrow FS = 10.28 \text{ Mbits/sec}$$

(9)

40

【 0 0 5 3 】

従って、A B R 1 は、6 0 . 2 8 M ビット / 秒の帯域幅を得ることになる。A B R 2 は 5 M ビット / 秒だけ使うことになる。A B R 3 - A B R 1 0 は、各々が 1 1 . 2 8 M ビット / 秒を得る。シミュレーションの結果では、これらの値と一致している。A C R 中の初期オーバーシュート (initial overshoots) は、A B R バッファを安定点まで満たすためであることに留意すべきである。

【 0 0 5 4 】

50

〔実験2〕

〔簡易トポロジにおける10個のABRVCと1個のバーストVBRVC〕

〔セットアップ〕

使用したトポロジを図14乃至図21に示す。セットアップは、10個のABRVCと1個のバーストVBRソースからなる。ABRVCは全て持続性である。ABR1のMCRは50Mビット/秒であり、その他のABRVCのMCRは1Mビット/秒である。VBR1のPCRは50Mビット/秒である。リンクL1-L11及びL1s-L11sの長さは10kmであるのに対し、リンクLa及びLbの長さは各々1000kmである。

【0055】

〔観察〕

実験2の結果を図14乃至図21に示す。ACRは、VBRソースがOFFのときに起こるパルスでパルス整形される(pulse shaped)。VBRがONのとき、 $C_{abr} = L - VBR$ レート} - MCR = 155.52 - 50 - (50 + 9 × 1) = 46.52 Mビット/秒である。従って、各々のABRVCは、そのMCRに加えて46.52 / 10 = 4.65 Mビット/秒を得る。VBRがOFFのとき、 $C_{abr} = 155.52 - 0 - (50 + 9 × 10) = 96.52$ Mビット/秒である。従って、各々のABRVCは、それらのMCRに加えて96.52 / 10 = 9.65 Mビット/秒を得る。

【0056】

〔実験3〕

〔簡易トポロジにおける100個のABRVCと1個のバーストVBRVC〕

〔セットアップ〕

使用したトポロジは図2のものと同様であるが、100個の全く同じABRソース及び1個のバーストVBRソースを用いている。全てのABRVCは持続性であり、それらのMCRは1.0Mビット/秒である。VBR1のPCRは55Mビット/秒である。リンクL1-L101及びL1s-L101sの長さは10kmであり、一方、リンクLaの長さは1000kmである。

【0057】

実験3の結果を図22乃至図29に示す。ACRは、VBRソースがOFFのときに起こるパルスでパルス整形される。VBR1がONのとき、 $C_{abr} = 155 - 55 - (100 × 1) = 0$ Mビット/秒である。従って、各々のABRVCは、そのMCR = 1.0 Mビット/秒だけを得る。VBR1がOFFのとき、 $C_{abr} = 155 - 0 - (100 × 1) = 55$ Mビット/秒である。従って、各々のABRVCは、そのMCRに加えて55 / 100 = 0.55 Mビット/秒を得る。即ち、その各々は1.55 Mビット/秒を得る。

【0058】

〔実験4〕

〔パーキングロットトポロジにおける5個のABRVCと1個のバーストVBRVC〕

パーキングロットトポロジを図30に示す。全てのABRVCは持続性である。ABR2以外のABRVCのMCRは全て10Mビット/秒であるが、ABR2のMCRは20Mビット/秒である。VBR1のPCRは90Mビット/秒である。リンクL1-L5及びL1aの長さは各々が10kmであるのに対し、中央リンクLa-Leの長さは各々が1000kmである。

【0059】

実験4の結果を図31乃至図40に示す。ACRは、VBRソースがOFFのときに起こるパルスでパルス整形される。VBRがONのとき、 $C_{abr} = 155 - 90 - (20 + 4 × 10) = 5$ Mビット/秒である。従って、各々のABRVCは、そのMCRに加えて5 / 5 = 1 Mビット/秒を得る。VBRがOFFのとき、 $C_{abr} = 155 - 0 - (20 + 4 × 10) = 95$ Mビット/秒である。従って、各々のABRVCは、そのMCRに加えて95 / 5 = 19 Mビット/秒を得る。

【0060】

シミュレーションの結果では、SERIAは常に100%のリンク帯域幅を獲得し、AB

10

20

30

40

50

Rバッファ占有の平均値は $B / 2 = 3276$ セルに近いことを示している。SERIAは、ABRバッファの占有制御をきっちりと行なう。

【0061】

実験1は、最大 - 最小フェアネスを得るために、SERIAが、未使用の帯域幅をABR2からスイッチ(34)ポートでボトルネックされた他のABRへ再分配する方法を示している。実験2、実験3及び実験4の結果に示されるように、SERIAは、従来の多くのアルゴリズムとは異なり、非常にバーストなVBRストリームがあるときでさえ、うまく実行する。実験3の結果は、多くの持続性ABRVCを取り扱う際、SERIAがロバストであることを示している。実験4の結果では、SERIAは、長いパスを有するABRVCを区別しない(即ち、ビートダウン問題がない)ことを示している。

10

【0062】

ここに開示したSERIAは、公平なシェアを厳密に計算する新規なABRスイッチ(34)アルゴリズムである。それは、簡易で正確な方法により、スイッチ(34)でボトルネックされたVCに対して、その他の場所でボトルネックされたVCから帯域幅を分配する。SERIAのその他新しい特徴には、これまでの方法よりも、より簡易でより安定したバッファ占有がもたらされるバッファ占有制御メカニズムがある。

【0063】

シミュレーションの結果では、SERIAは、最大 - 最小フェアネス、定常状態への迅速なコンバージェンス、リンクの100%利用、及び安定したABRバッファ占有がもたらされることを示している。また、シミュレーションの結果では、SERIAは、非常にバーストなVBRストリームの存在下でも、ロバストであり、うまく作業が行われることを示している。

20

【0064】

発明を例示するために、実施例を参照して発明を詳細に説明したが、これら説明は、単なる例示であって、当該分野の専門家であれば、特許請求の範囲に記載された発明の精神及び範囲から逸脱することなく、発明に種々の変形をなすことはできると理解されるべきである。

【図面の簡単な説明】

【図1】 ATMスイッチポートの構成の概要説明図である。

【図2】 実験1のトポロジを示す図である。

30

【図3】 実験1の結果を示す図である。

【図4】 実験1の結果を示す図である。

【図5】 実験1の結果を示す図である。

【図6】 実験1の結果を示す図である。

【図7】 実験1の結果を示す図である。

【図8】 実験1の結果を示す図である。

【図9】 実験1の結果を示す図である。

【図10】 実験1の結果を示す図である。

【図11】 実験1の結果を示す図である。

【図12】 実験1の結果を示す図である。

40

【図13】 実験2のトポロジを示す図である。

【図14】 実験2の結果を示す図である。

【図15】 実験2の結果を示す図である。

【図16】 実験2の結果を示す図である。

【図17】 実験2の結果を示す図である。

【図18】 実験2の結果を示す図である。

【図19】 実験2の結果を示す図である。

【図20】 実験2の結果を示す図である。

【図21】 実験2の結果を示す図である。

【図22】 実験3の結果を示す図である。

50

- 【図 2 3】実験 3 の結果を示す図である。
- 【図 2 4】実験 3 の結果を示す図である。
- 【図 2 5】実験 3 の結果を示す図である。
- 【図 2 6】実験 3 の結果を示す図である。
- 【図 2 7】実験 3 の結果を示す図である。
- 【図 2 8】実験 3 の結果を示す図である。
- 【図 2 9】実験 3 の結果を示す図である。
- 【図 3 0】実験 4 のトポロジを示す図である。
- 【図 3 1】実験 4 の結果を示す図である。
- 【図 3 2】実験 4 の結果を示す図である。
- 【図 3 3】実験 4 の結果を示す図である。
- 【図 3 4】実験 4 の結果を示す図である。
- 【図 3 5】実験 4 の結果を示す図である。
- 【図 3 6】実験 4 の結果を示す図である。
- 【図 3 7】実験 4 の結果を示す図である。
- 【図 3 8】実験 4 の結果を示す図である。
- 【図 3 9】実験 4 の結果を示す図である。
- 【図 4 0】実験 4 の結果を示す図である。

10

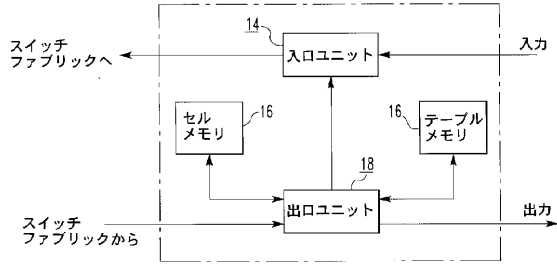
- 【図 4 1】本発明の装置の概略説明図である。
- 【図 4 2】本発明の方法のフローチャート図である。
- 【図 4 3】本発明のスイッチの概略説明図である。
- 【図 4 4】S E R I A の疑似コードを示す図である。
- 【図 4 5】S E R I A の疑似コードを示す図である。

20

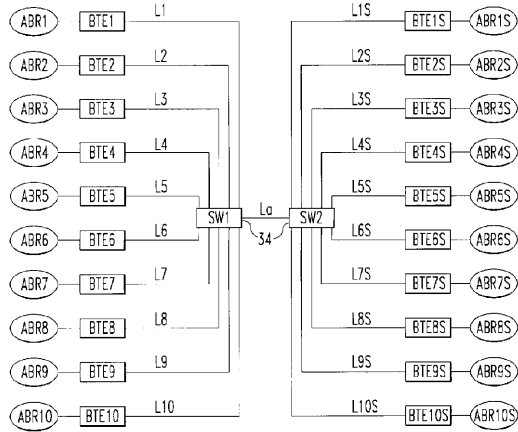
【符号の説明】

- (10) セルスイッチング装置
- (12) ネットワーク
- (16) メモリメカニズム
- (20) サーバ
- (22) スケジューラ

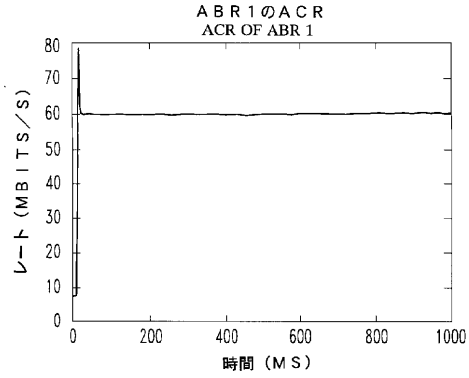
【図1】



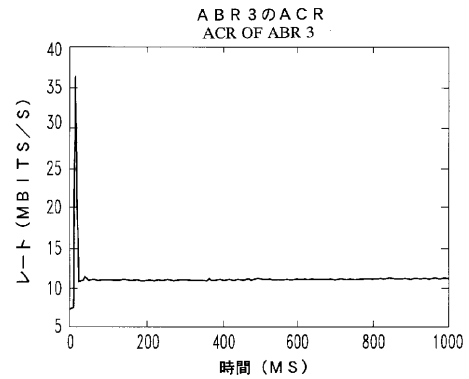
【図2】



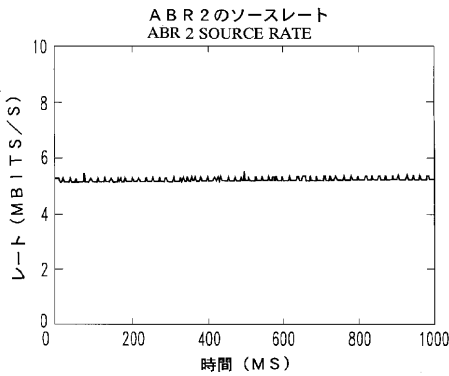
【図3】



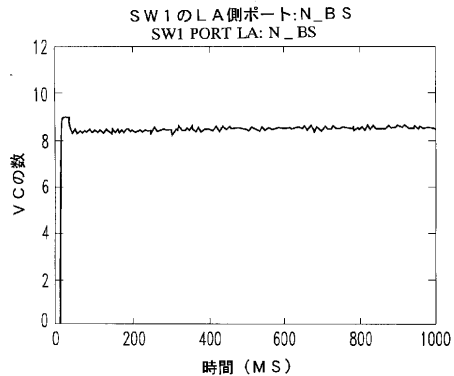
【図4】



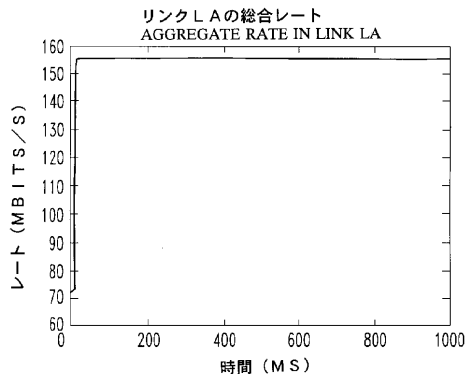
【図5】



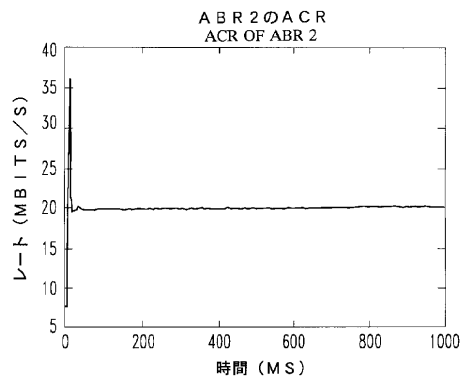
【図7】



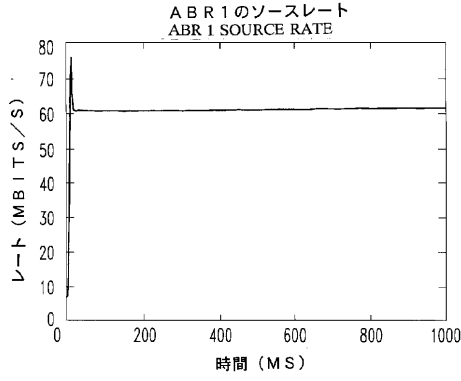
【図6】



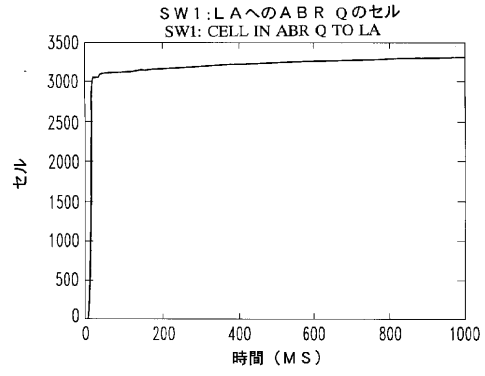
【図8】



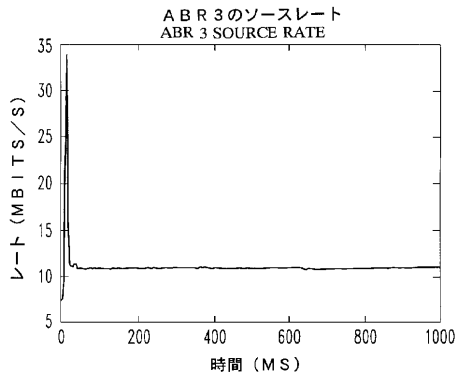
【図 9】



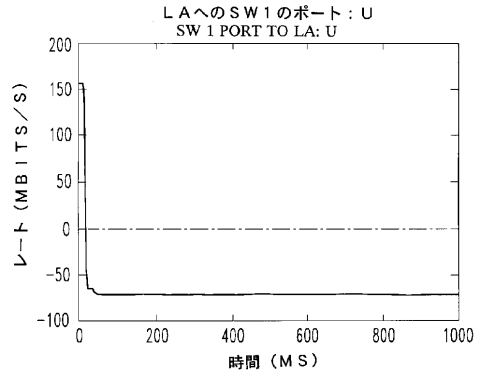
【図 11】



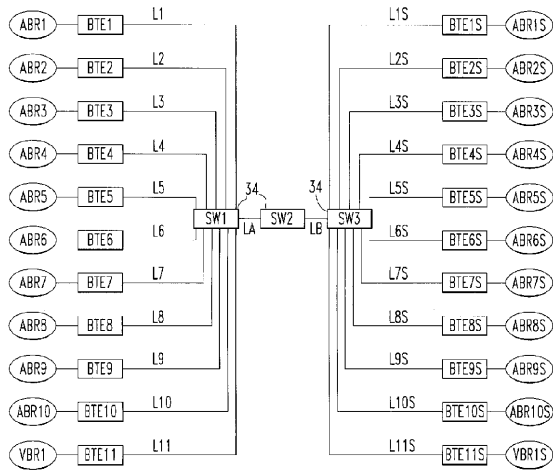
【図 10】



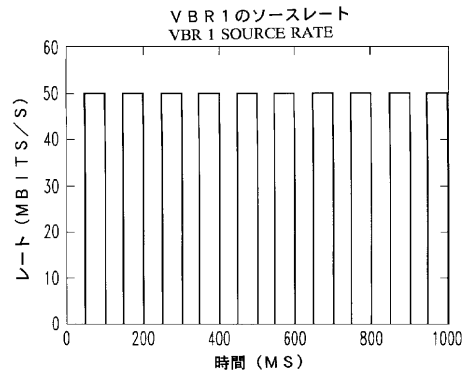
【図 12】



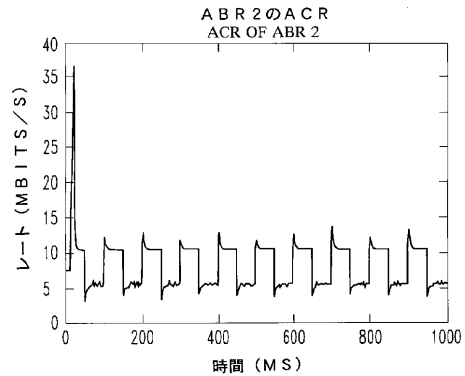
【図 13】



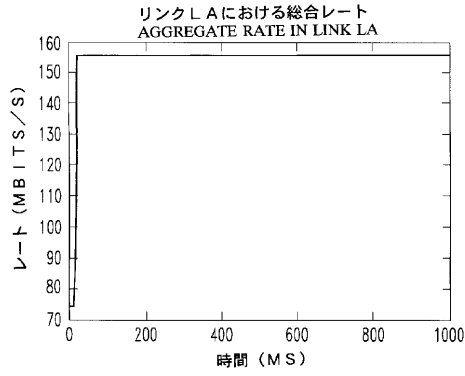
【図 14】



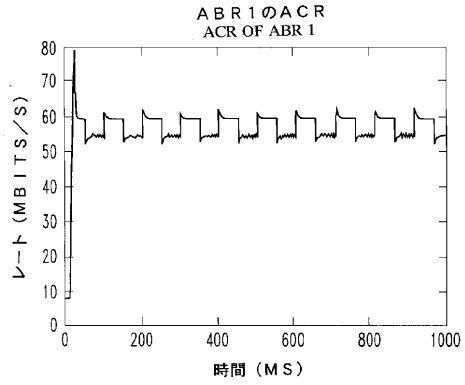
【図 15】



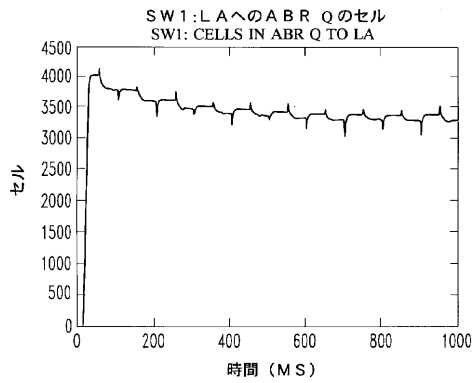
【図16】



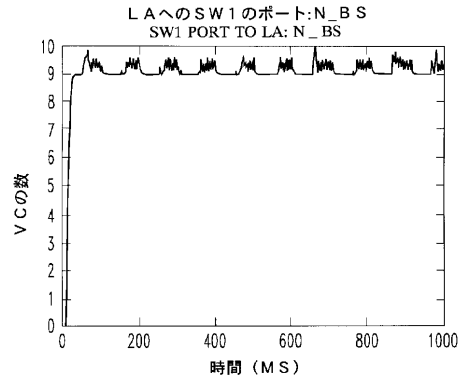
【図18】



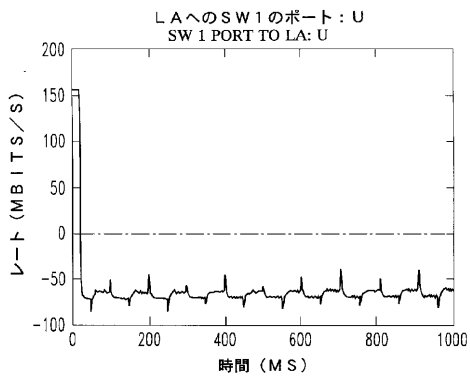
【図17】



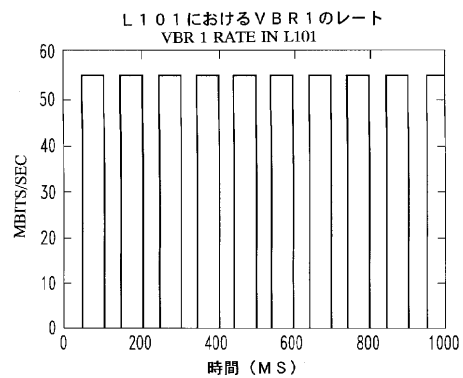
【図19】



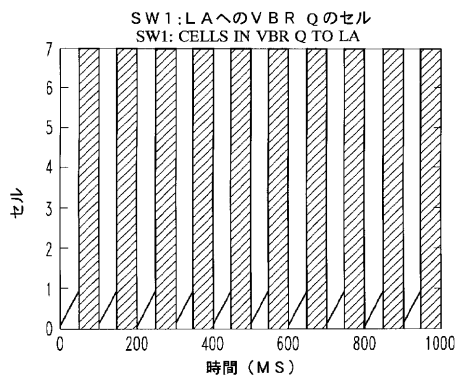
【図20】



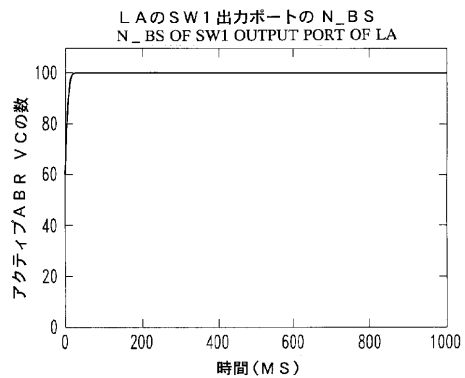
【図22】



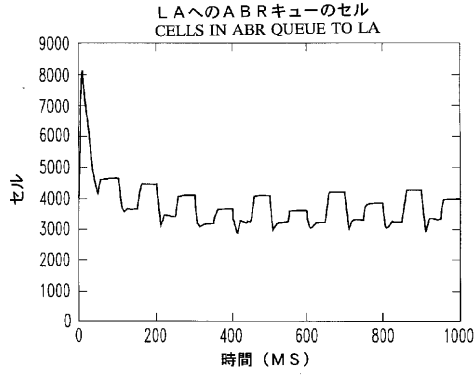
【図21】



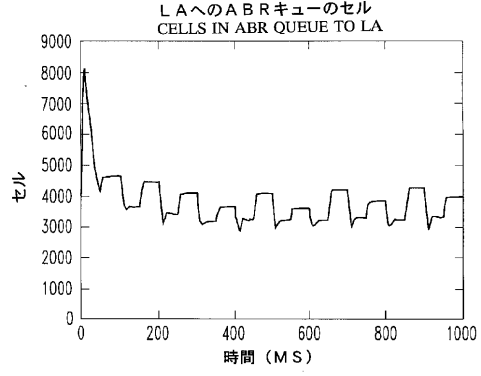
【図23】



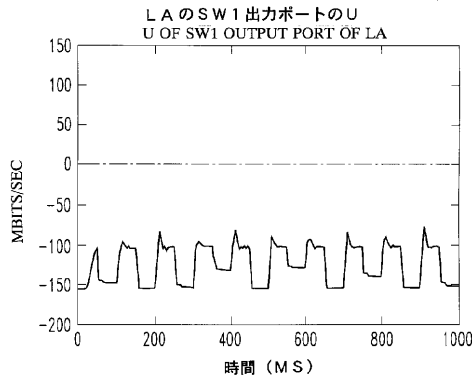
【図 24】



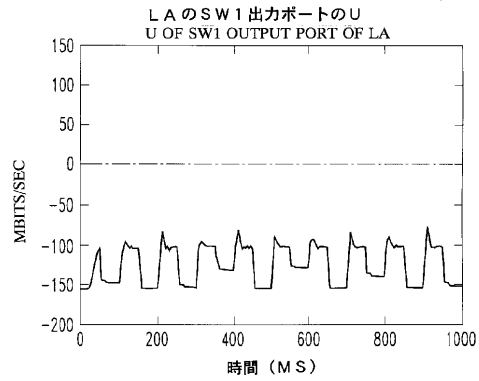
【図 26】



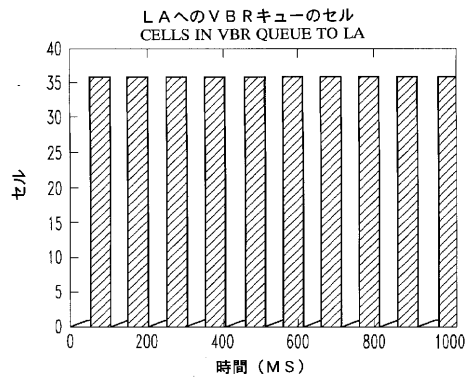
【図 25】



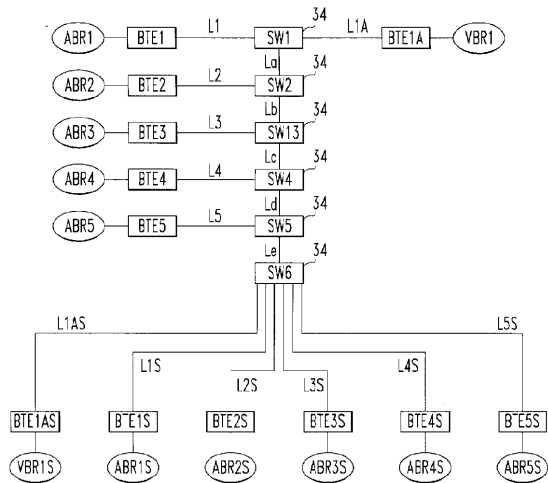
【図 27】



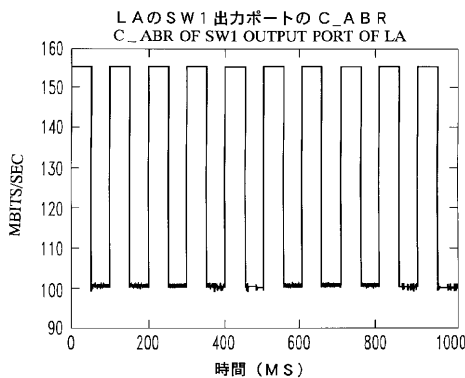
【図 28】



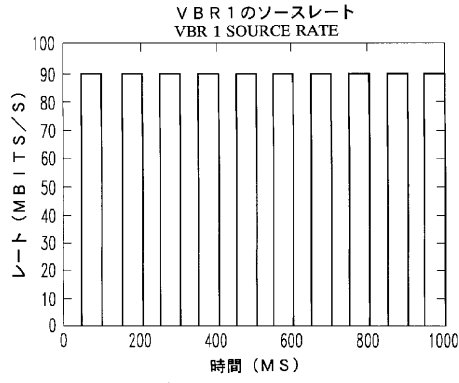
【図 30】



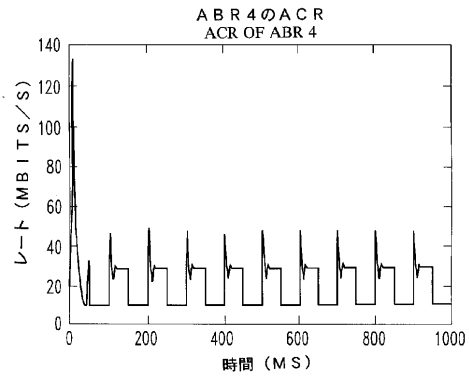
【図 29】



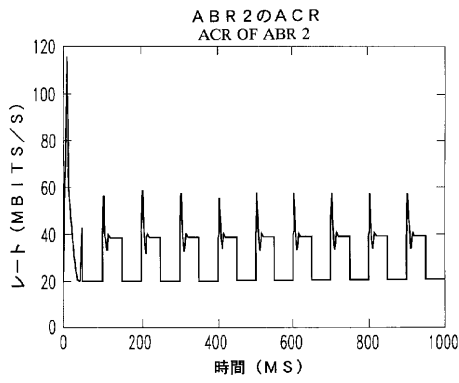
【図 3 1】



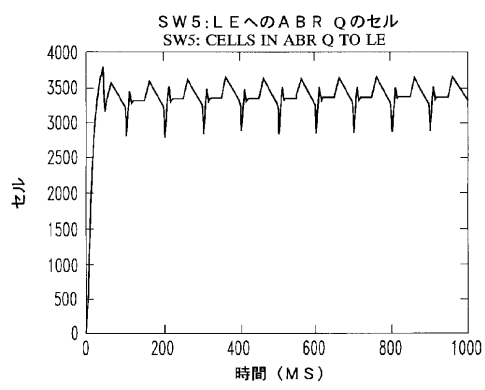
【図 3 3】



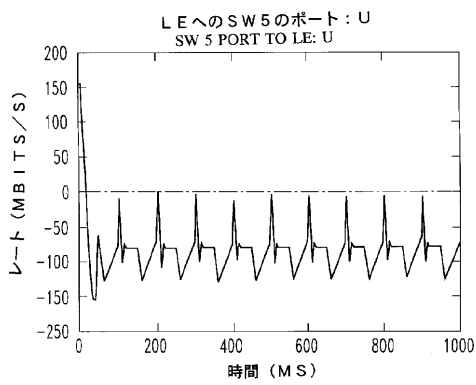
【図 3 2】



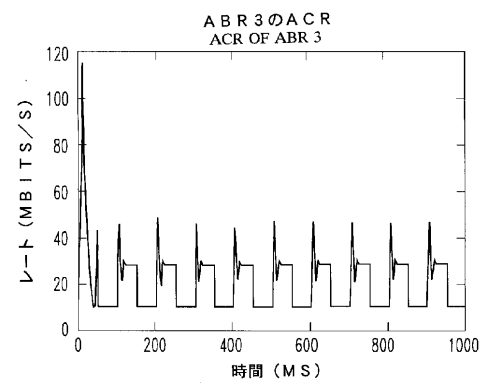
【図 3 4】



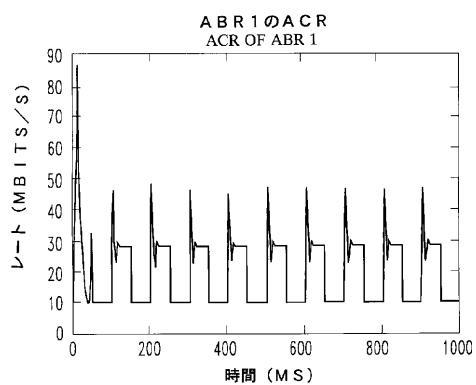
【図 3 5】



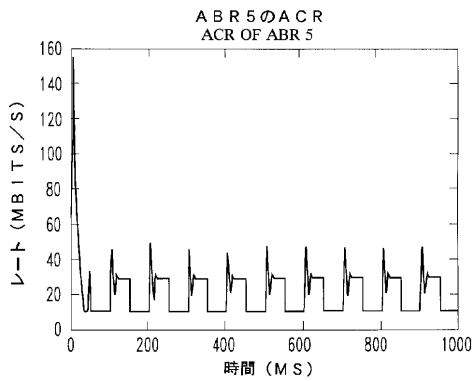
【図 3 7】



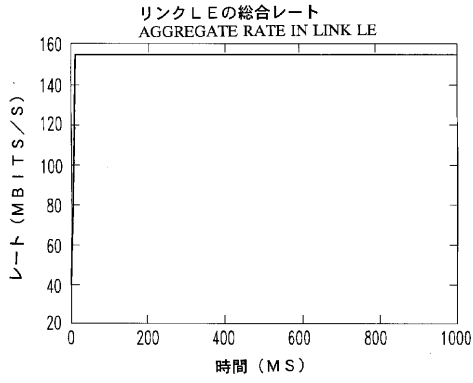
【図 3 6】



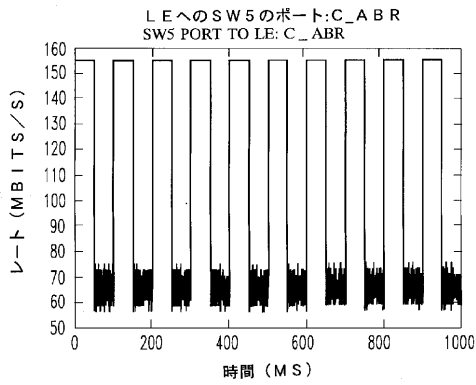
【図 3 8】



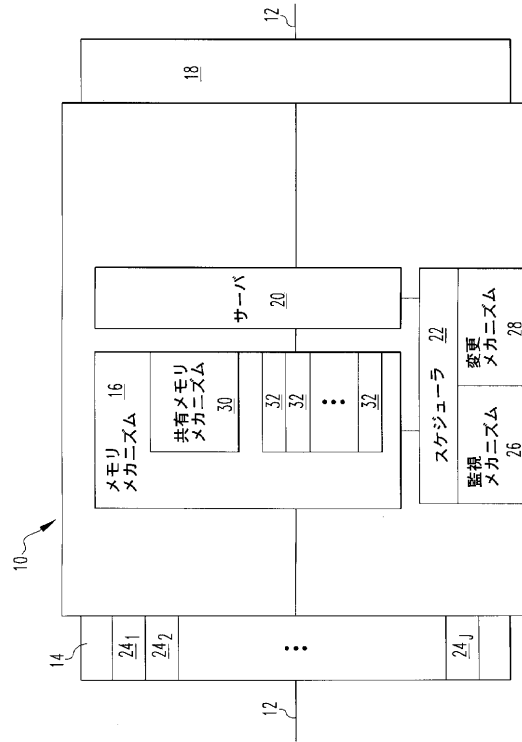
【図 39】



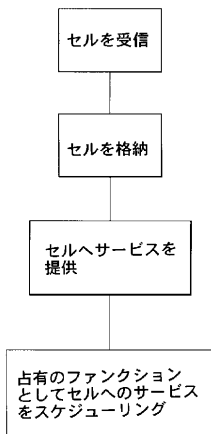
【図 40】



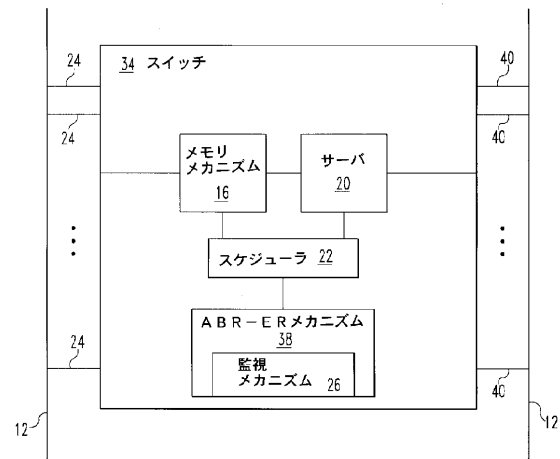
【図 41】



【図 42】



【図 43】



【 図 4 4 】

A Pseudo-code of SERIA

```

output port initialization
init()
{
   $N_{bs} = 0; N_{bs}^n = 0; U = +L; n_g = 0; Q_n = 0; Q_p = 0; q_i = 0 \forall i;$ 
}

cell arrival event handler
cell_arrival( VC i )
{
   $q_i = q_i + 1$ 
  if ( ith VC is ABR )
  {
     $Q_n = Q_n + 1$ 
    if (  $q_i == T$  )
       $N_{bs}^n = N_{bs}^n + 1;$ 
  }
}

cell departure event handler
cell_service_completion( VC i )
{
   $q_i = q_i - 1$ 
  if ( ith VC is CBR/VBR )
     $n_g = n_g + 1;$ 
  else if ( ith VC is ABR )
  {
     $Q_n = Q_n - 1$ 
    if (  $q_i == T - 1$  )
       $N_{bs}^n = N_{bs}^n - 1;$ 
  }
}

measuring period event handler: executed every N slots
N_slot_tick()
{
  /* Exponentially average the number of active ABR VCs */
   $N_{bs} = N_{bs} + \alpha_n(N_{bs}^n - N_{bs});$ 

  /* Compute total leftover bandwidth */
   $C_{abr} = (N - n_g)R;$ 
}

```

図 4 5 へ続く

【 図 4 5 】

図 4 4 から続く

```

/* Compute change in ABR buffer occupancy */
 $\Delta Q = Q_n - Q_p;$ 
 $Q_p = Q_n;$ 

/* Compute ABR buffer occupancy weighting */

 $\gamma = Q_n / \beta B;$ 
if (  $\Delta Q < 0$  )
   $\gamma = \max(1 - \gamma, 0);$ 

/* Update the additional bandwidth parameter */
if (  $Q_n < N$  )
   $U = +L;$ 
else if (  $Q_n > (B - N)$  )
   $U = -L;$ 
else
{
   $U = U - \alpha_u \gamma \Delta Q R;$ 
   $U = \min(U, +L);$ 
   $U = \max(U, -L);$ 
}

/* Compute Fair Share */
 $FS = \max(C_{abr} + U, 0) / \max(N_{bs}, 1);$ 

/* Reset number of VBR/CBR cells in period measurement */
 $n_g = 0;$ 
}

backward RM cell arrival event handler
bRM_cell_arrival()
{
  /* MCR plus equal share fairness criteria */
  cell.ER = min( cell.ER, cell.MCR + FS );
}

```

フロントページの続き

- (72)発明者 ミーナラチャガン ビシュヌ
アメリカ合衆国 1 6 0 6 6 ペンシルベニア, クランベリー, レキシントン ドライブ 1 1 7
- (72)発明者 デバシス バサク
アメリカ合衆国 1 5 1 0 1 ペンシルベニア, アリソン パーク, プレジデンシャル ドライブ
9 8 0 4, アpartment 2 0 8
- (72)発明者 ヒョン エス.キム
アメリカ合衆国 1 5 2 1 5 ペンシルベニア, ピッツバーグ, グレンガリ ドライブ 6 2 5

審査官 石田 紀之

- (56)参考文献 特開平10 - 1 3 5 9 7 5 (J P , A)
特開平10 - 0 0 4 4 1 9 (J P , A)
特開平10 - 0 3 2 5 8 8 (J P , A)
M.Shinohara et al. , Multiclass LargeScale ATM Switch with QoS Guarantee , 1997 IEEE International conference on Communications , 米国 , IEEE , 1 9 9 7 年 1 1 月 1 4 日 , 第 1 卷 , 第 4 5 3 - 4 6 1 ページ

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