

ATM網におけるAvailable Bit Rate (ABR)上のTCPの 動作解析

阿野 茂浩

長谷川 亨

加藤 聡彦

国際電信電話株式会社 研究所

Available Bit Rate (ABR)は、ATM回線の使用効率を高くするために、送受信側の端末およびATMスイッチの協調動作により、セルレベルのレート制御を行う方式である。ABRは主にデータ通信への適用が検討されており、データ通信で広く利用されているTCPを、ABR上で動作させることが多いと考えられる。しかし、TCPをABR上で動作させる場合は、TCPレベルのフロー制御とセルレベルのABRレート制御が独立に動作するため、それらの間にどのような相互作用があるかを検討する必要がある。そこで、本稿では、ABR上でのTCPの動作を解析し、以下の問題点を明らかにした。まず、ATMネットワークの輻輳時には、セルが損失がしないにもかかわらず、TCPがセグメントの再送を行うことがある。次に、TCPがスロースタート・輻輳回避アルゴリズムを用いてセグメントを再送する場合、TCPの送信ウィンドウの上昇が、ABRにおける送信レートの上昇に伴わないことがある。さらに本稿では、これらの問題点について、通信シーケンス例を用いて述べる。

An Analysis on TCP communication using Available Bit Rate (ABR) in ATM Network

Shigehiro Ano

Toru Hasegawa

Toshihiko Kato

KDD R&D Laboratories

In Available Bit Rate (ABR), the sending rate is controlled by the coordination among ATM switches and sending/receiving terminals. Since ABR, which realizes the high ATM network utilization, is suitable for data communication, it becomes popular that TCP which is widely used as data transport protocol is used over ABR. TCP over ABR, however, may not achieve the high throughput due to the incompatibility between TCP flow control and ABR rate control. We, therefore, analyze the behavior in detail and point out a couple of problems. First, when ATM network is congested, TCP retransmission timer may expire even if no TCP segments are dropped. Second, after the TCP retransmission timeout, the increase of TCP send window may not be as fast as that of sending rate of ABR, due to the slow start mechanism of TCP. This paper describes the problems using examples of communication sequences.

1. はじめに

Available Bit Rate(ABR)^[1]は、高いATM回線使用効率を実現するために、ATMスイッチと送受信端末が協調して、送信セルの速度を制御する方式である。ABRは主にデータ通信への適用が考えられており、その上位プロトコルとしてはTCP^[2]が使用されることが多いと考えられる。しかし、ABR上のTCPにおいてはセルレベルのABRレート制御とTCPレベルのフロー制御とが独立に動作するため、その相互作用に関する動作解析が必要となる。このため、シミュレーションを用いたABR上のTCPの性能評価が広く行われている^[3]。

これに対し本稿では、TCPのセグメントやATMセルの送受信シーケンスを詳細に解析することにより、ABR上のTCPの動作解析を行う^[4]。この結果、TCPのフロー制御とABRのレート制御の相互作用により、スループットが低下する可能性があることを示すとともに、このような不整合が発生する通信シーケンス例を明らかにする。

2. ABRとTCPの概要

2.1 ABRレート制御

ABRは、ATMスイッチまたは受信側端末の負荷に応じて、送信側端末が送信可能なセル速度(ACR: Actual Cell Rate)を制御する方式である。送信側端末は、図1に示すように、以下の規則に従って、フォワードリソースマネジメント(FRM)セルを送出する。

- ① ユーザセル(ユーザデータを含むセル)をNrm個送出する毎
- ② 最後のFRMセル送出後、2個以上Nrm個未満のユーザセルを送出しており、さらにTrm時間経過した時

受信側端末は、FRMセルを受信すると、内容はそのままにして、バックワードリソースマネジメント(BRM)セルとして折り返す。

送信側端末は、通信を開始すると、輻輳を通知しないBRMセルの到着毎に、ACRを、セル速度の初期値(ICR: Initial Cell Rate)からPCR(Peak Cell Rate)まで、RIF(Rate Increase Factor)*PCRずつ増加する。通常、Nrmとして32が、Trmとして100msが推奨されている。

逆に、ネットワークの輻輳を通知するBRMセルが到着すると、以下の方法でACRを削減する。

- ① 輻輳発生を通知するCI(Congestion Indication)ビットが設定されている場合、ACRを現在のACR* RDF (Rate Decrease Factor)分削減する。

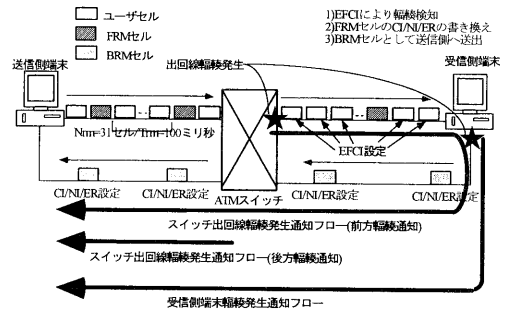


図1 ABRにおける輻輳通知方法

- ② NI (No Increase) ビットが設定されている場合、ACRを増加しない。
- ③ ER (Explicit Rate) が設定されている場合、ACRを直接ERに削減する。

一方、受信側端末およびATMスイッチは、以下の方法で送信側端末に輻輳を通知する。

受信側端末は、ユーザセルの順方向の輻輳表示ビットであるEFCI (Explicit Forward Congestion Indication) により輻輳を検知すると、次に送出するBRMセルに対して、CIビットあるいはNIビットの設定、もしくは、ERの設定のどれかを行う(図1中の受信側端末輻輳発生通知フロー)。この選択は、受信側端末に任されている。

ATMスイッチは輻輳を検出すると、EFCIビットの設定にくわえて、以下の方法を用いて送信側端末のACRを制御する(図1参照)。

- ① 次のFRMセルに対して、CI、NIあるいはERを設定し、受信側端末に送出する。このFRMセルは受信側端末で折り返され、BRMセルとして、送信側端末に通知される。
- ② 受信側端末が送信してきた、次のBRMセルの内容を書き換えることにより同様な設定を行い、直接、送信側端末に送出する。
- ③ 受信側端末がBRMセルとは独立に、BRMセルを生成し、送信側端末に送出する。ただし、このBRMセルの生成は1秒に10回以下に制限される。

ただし、ABRでは最小限の送信セル速度MCR (Minimum Cell Rate) を保証しており、ERあるいはCIにより、MCR未満のセル速度に削減されることはない。

2.2 TCPフロー制御

TCPでは、送信側端末は通常は受信側端末から通知されるウィンドウでセグメントを送出する。しかし、セグメントの損失などにより、送出セグ

メントの受信確認(ACK)がRTO (Retransmission TimeOut value)ミリ秒受信されないと(再送タイムアウト)、送信側端末は送信ウィンドウを一旦減少させてから徐々に増加させる。まず、送信ウィンドウの約1/2までは指数関数的に増加させ(スロースタート)、それ以降は線型に増加させる(輻輳回避アルゴリズム)。

受信側端末では2最大セグメントサイズ(MSS)を受信するか、またはTCPの200ミリ秒タイマが発火するとACKを返送する。なお、RTOはACK受信時の往復遅延時間から、500ミリ秒単位で計算される。

また、TCPコネクションの設定直後では、送信側と受信側の端末が異なるサブネットワークに属する場合、送信側端末はスロースタートに従って、セグメントを送出し始める。

3. ABRレート制御とTCPフロー制御の不整合

ABRではセルのレート制御により、輻輳によるセル損失を防ぐことが可能である。従って、ABR上にTCPを実現する場合、ACRが十分大きい限り、再送タイムアウトが発生しないと考えられる。しかし、ABRのレート制御とTCPのフロー制御の相互作用により、次のような場合に問題が生ずる可能性がある。

(1) BRMセルによりACRが減少されると、ATM網でセルが損失しないにもかかわらず、セグメントの伝送速度が減少したために、再送タイムアウトが発生することがありうる。

(2) ACRの増加と、TCPのスロースタートおよび輻輳回避アルゴリズムによる送信ウィンドウの増加が独立に行われ、協調が取れない。

以下では、この2つの問題点について検討する。

3.1 前提

検討を簡略化するため、ATMコネクションとTCPコネクションを1対1に対応させている。TCPについては、再送タイムアウト発生後のスロースタートおよび輻輳回避アルゴリズムを適用中は、セグメント毎にACKが返送されることを仮定している。また、送信側におけるTCPの送出処理時間は、考慮の対象外とする。ABRについては、送信側は正しくレート制御に従い、ATM網ではセルが損失しないことを仮定している。

3.2 再送タイムアウトの発生

TCPにおける再送タイムアウトの条件は式(i)により与えられる。

$$500 \text{ ミリ秒} * (\text{ACK 返送までの} 500 \text{ ミリ秒} \text{ タイマ発火数}) > \text{RTO} \text{ -----(i)}$$

ここで、式(i)のACKが返送されるまでの時間は、以下の2つの内の短い方である。

$$\begin{aligned} &2 \text{セグメント送出時間} + \text{往復遅延時間} \text{ -----(ii)} \\ &1 \text{セグメント送出時間} + \text{最大} 200 \text{ ミリ秒} \\ &\quad + \text{往復遅延時間} \text{ -----(iii)} \end{aligned}$$

従って、MSSが大きく、往復伝送遅延時間が長く、輻輳を通知するBRMセルの受信が長く続き、さらにACRが急減しセグメント送出時間が増加すると、式(i)が成立することがありえる。ACRが急減する場合としては、(1) ERによるACRの急な減少、(2) CIによる段階的な減少、(3) 低いICRの設定が考えられる。

以下では、この3つの場合について、再送タイムアウト発生の可能性について検討する。

(1) ERによるACRの減少時

式(ii)の2セグメントをATM回線へ送出する時間は、それまでの時間とセグメントの残りをERで送出する時間との和である。一方、式(iii)の送出時間の内、1セグメントの送出時間も同様に決まる。このため、ERが非常に小さく、かつATM網の輻輳が続くために、ERが設定されたBRMセルを連続して受信する場合、セグメントの残りを送出する時間が増加し、式(i)を満たすことがある(図2のケース1)。

例えば、MSSが512バイトで、往復遅延時間が200ミリ秒の場合に、ACKが必要なセグメントを新たに送信する直前にERが30セル/秒のBRMセルを受信し、さらに送信側と受信側の端末の間に1個もFRMセルおよびBRMセルが存在しないと、(iii)の方が小さく約600~800ミリ秒となる。この間に500ミリ秒タイマが2回発火し、RTOが500ミリ秒であれば、再送タイムアウトが発生する。

一方、輻輳している時間が短い場合、再送タイムアウトが発生する可能性は低下する。例えば、ERが30セル/秒のBRMセルの到着後すぐに、輻輳

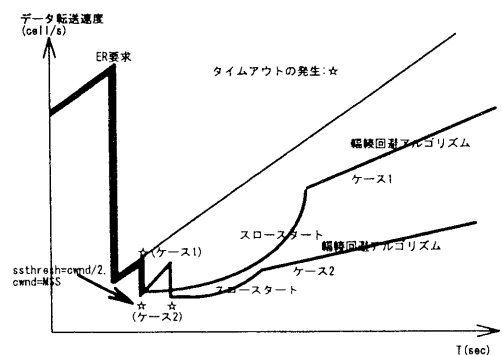


図2: ABRにおけるTCPフロー制御の動作

を通知しないBRMセルが到着すると、ACRもすぐに増加する。この時、式(ii), (iii)で表される送出時間は、上述の場合よりも短くなる。

(2) CIが設定されたBRMセル受信時のACRの段階的減少時

現在のACRが低く、RDFが大きい場合、さらに再送タイムアウトが発生しやすい。例えば、MSSが9148バイト、現在のACRが640セル/秒、RDFが最も小さい1/4の場合、減少されたACRが480セル/秒となる。ここで、輻輳が継続し、NIビットあるいはCIビットが設定されたBRMセルを400ミリ秒受信し続けると、(I)と同じ条件で1セグメントを送出する時間は400ミリ秒となる。従って、(iii)の条件により再送タイムアウトが発生する可能性がある。

この場合も(1)と同様に、輻輳が長く継続しない場合は、再送タイムアウトが発生する可能性は低下する。

(3) TCPコネクション設定直後

TCPコネクション設定直後では、ICRが低い値であっても、再送タイムアウトが発生する可能性は低い。例えば、MSSが9188バイト、往復遅延時間が200ミリ秒、RIFの値が推奨パラメータの最小値16で、ICRが30セル/秒の場合を考える。この場合、まず、最初の32セルの送出に約1秒かかる。次に、この時送出したFRMセルが折り返されたBRMセルを約1200ミリ秒後に受信し、ACRは480セル/秒になる。このため、ACKの返送に必要な2セグメントの残りを送出する時間は、最大で約330ミリ秒である。従って、2セグメント送出に約1530ミリ秒かかるが、コネクション設定直後の再送タイム値の2000ミリ秒以下であり、再送タイムアウトは発生しない。

3.3 再送タイムアウト後のTCPの動作

再送タイムアウトの発生後、TCPの送信側端末は、送信ウィンドウcwndを制御することにより、送信レートを制御する(図2のケース1)。まず、cwndを1MSSに減少してから、ACK受信の度にMSS単位でしきい値ssthreshまで指数関数的に増加する(スロースタート)。この時、一度ACRが式(i)を満たさなくなると、以降ACRが上昇する限り、タイムアウトが再度起こる可能性はない。

次に、ssthreshに達すると、輻輳回避アルゴリズムによりcwndは、ACK受信の度に以下の式に基づき計算される。

$$cwnd = cwnd + (\text{セグメントサイズ}) ** 2 / cwnd + (\text{セグメントサイズ}) / 8$$

このため、セグメントサイズが512バイトで、

cwndが12288バイト(12Kバイト)の場合、cwndはACK受信毎に約85バイト分しか増加せず、ACRの増加割合に比較して、TCPの送信レートの増加は非常に小さくなってしまう。

また、再送タイムアウトの発生後も、BRMセルによりACRの増加が押さえられると、再び再送タイムアウトが発生する(図2のケース2)。この他にも、輻輳通知の解除後に、BRMセルの到着毎にRIF*PCR単位に増加するACRを考慮に入れても、引き続き式(i)を満たせば、再び再送タイムアウトが発生する可能性がある。

再送タイムアウトの連続的な発生は、再送タイムアウト毎にssthreshを1/2に減少させるため、ケース1と比較して輻輳回避アルゴリズムが低いcwndの時から適用されるようになる。このため、AIRの増加割合に比較して、TCPの送信レートの増加がさらに小さくなり、ABRにより提供可能なATM回線の使用率をさらに悪くしてしまう。

4. 再送タイムアウト時の動作例

本章では、TCPの再送タイムアウトが発生する場合について、再送タイムアウト後にスロースタートを行うシーケンス、および行わないシーケンスを示す。

4.1 スロースタートを行う場合

図3に、以下の条件でERが60セル/秒のBRMセルを受信した時のTCPとABRのシーケンスを示す。

(ABR) PCR=65536セル/秒(27.79Mbps),

MCR=60セル/秒(25.44Kbps), Nrm=32, RIF=16

(TCP) 送受信ウィンドウサイズ=24キロバイト,

MSS=512バイト

(往復伝送遅延時間) 200ミリ秒

ここで、NrmとRIF値としては、ATMフォーラムの推奨パラメータの中でACRの増加が最も緩やかなものを用いている。またMSSは、再送タイムアウトの発生しにくい512バイトを用いている。

まず、ACKが必要な新たなセグメント(図3の#1)を送出する直前にACRがER(60セル/秒)に減少し、さらに300ミリ以上ACRを増加させるBRMセルを受信しないと、最初の2セグメント(図3の#1,2)に対応するACKの受信前で、3セグメント目を送信中に、再送タイムアウトが発生することになり、3セグメント分の不必要なデータを送信する。

これ以降は、再送タイムアウトも発生せず、順調にACKが返送されるため、新たなセグメントがスロースタートに従って送信される。この例

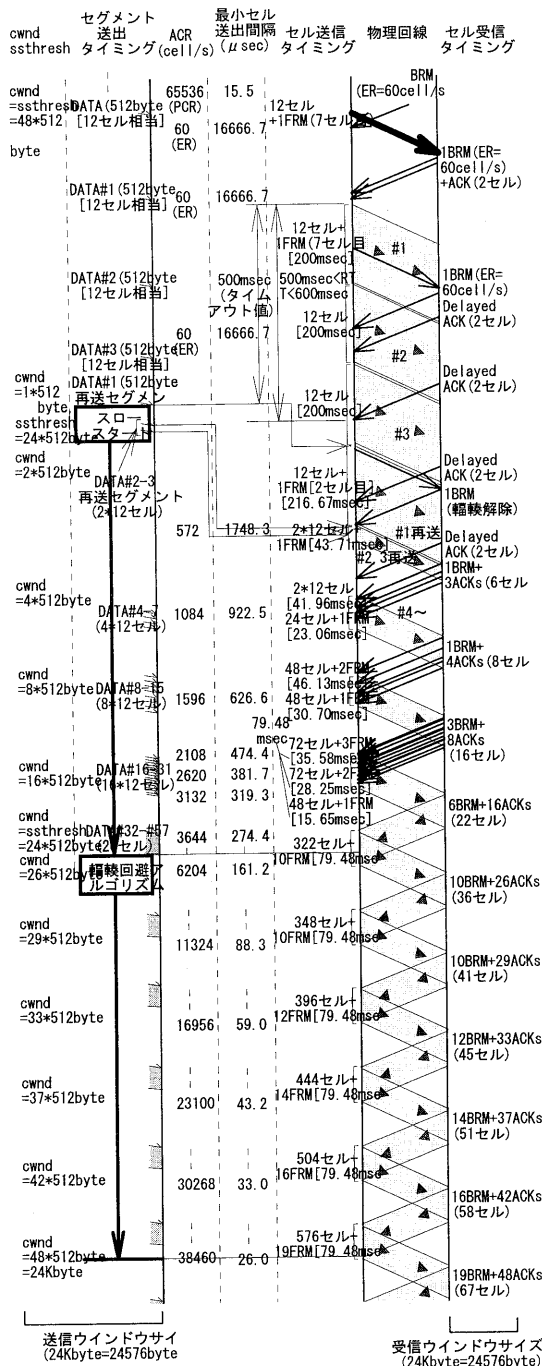


図3：再送タイムアウト時の動作 (スロースタート実行時)

では、cwnd が 1 に相当する最初の再送セグメント (#1 再送) を送出するのに、約 216.67ms かかる。次に、2 番目のセグメントの再送(#2 再送)直前に、輻輳解除を通知する BRM セルが到着し、ACR が通常通り増加しはじめる。このため、cwnd が 2 に相当する 2 番め、3 番目の再送セグメントの送出には約 43.7ms かかるだけである。この時点で、送信ウィンドウ分のセグメントの送出時間が往復遅延を下回っている。すなわち、TCP の cwnd の増加が ACR の増加に追従できず、送信ウィンドウが閉じてしまう。これ以降も ACK の受信により cwnd が増加しても、ACR の増加ほどではなく、cwnd 内のセグメントの送出時間は常に往復遅延 200 ミリ秒以下である。従って、往復遅延毎に、cwnd 分のセグメントを送出し、その ACK を待つ動作を繰り返す。

約 5 回の往復遅延でスロースタートが終了し、次に、輻輳回避アルゴリズムが開始される。輻輳回避アルゴリズムでは cwnd の増加の割合はスロースタートより少ないため、同様な動作を繰り返す。約 5 回の往復遅延で cwnd が 24K バイトになった時点で終了する。この時、タイムアウトの発生から、輻輳回避アルゴリズムが終了するまでの時間は、約 2080 ミリ秒である。

4.2 スロースタートを行わない場合

図4に、4.1節と同じ条件で、再送タイムアウト発生後にスロースタートを行わない場合のシーケンスを示す。

再送タイムアウト後、再送対象の3セグメントを送出する。この送出時間は往復遅延の200ミリ秒以上であるため、送信ウィンドウ内のセグメントの全てを送出し終える前に、受信側端末から最初の2セグメントに対するACKが到着する。このため、送信待ちセグメントが存在する状態となり、セグメントの送出が中断されない。

しかし、すぐにACRが増加し、送信ウィンドウ分のセグメントの送出時間が往復遅延の200ミリ秒以下となる。この例では、4つめのセグメントの送出以降は、この状態となっている。この時、2セグメントの送出時間は短くなっても、そのACKは往復遅延の200ミリ秒後まで戻ってこない。従って、これ以降は、200ミリ秒の間に2セグメントの送出とACKの受信を24回繰り返すことになる。

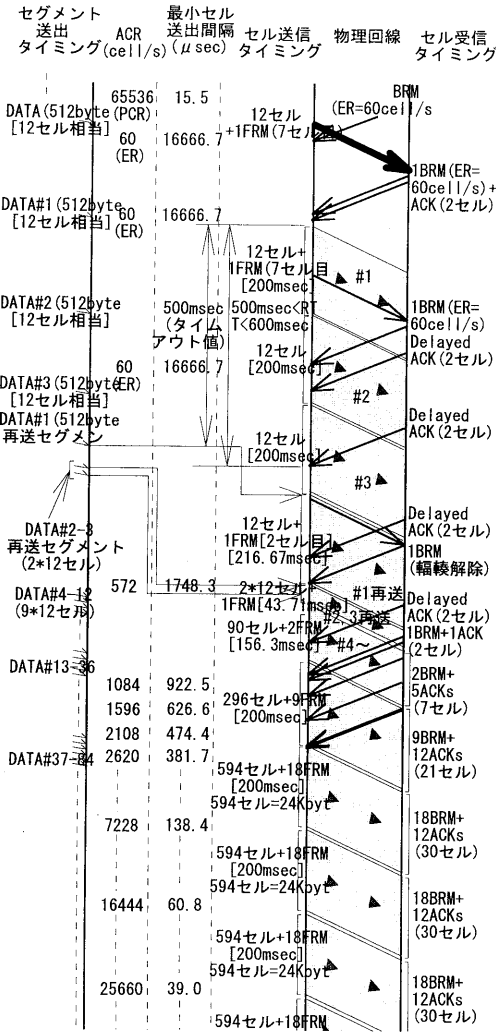


図4：再送タイムアウト時の動作
(スロースタートを行わない時)

5. 考察

(1) 4章の例では、TCPの再送タイムアウトが発生するのに、ERによるACRの減少要求が約300ミリ秒以上継続しているが、この程度の輻輳は頻繁に発生する可能性がある。ABRを提供するATMスイッチでは、物理回線速度で送信して数十ミリ秒以上のセルのバッファを有することが考えられる。送信側と受信側の端末の間に数台のATMスイッチが存在し、その内の何台かで同時に輻輳が発生すると、それらのATMスイッチのバッファ内の

セルが送出されて、輻輳が解除されるのに数百ミリ秒かかることがある。

(2) 4章の例において、スロースタートを行う場合、送信ウィンドウが24KバイトになるまでにTCPが送出したデータ量(再送分を含まない)は、スロースタートを行わない場合に同じ時間で送出したデータ量の約60%である。スロースタートおよび輻輳回避アルゴリズムのために、ABRの提供するACRを活用できていない。

(3) 4章の例では、再送タイムアウトが1回だけ発生する例を検討したが、例えば、輻輳解除を通知するBRMセルの到着が、4番目のセグメントを送出し終えるまでないと(1.6秒後)、再び再送タイムアウトが発生する可能性がある。再送タイムアウトが2回発生すると、ssthreshが 12×512 バイト(=6キロバイト)となる。この例では、cwndが24キロバイトとなるまでの時間は約3110ミリ秒となり、再送タイムアウトが1回の場合と比較して、約1.5倍となってしまう。

6. おわりに

本稿では、ABR上でTCPを実現する場合に、ABRレート制御とTCPフロー制御との相互作用により、以下の問題が発生する可能性を示唆した。まず、BRMセルによりACRが急激に減少されると、送信待ちのセグメントの送出時間が増加し、受信側からのACKの受信も遅れるため、そのセグメントが損失しなくても、再送タイムアウトが発生することがある。また、一般的に、TCPのスロースタート/輻輳回避アルゴリズムによる送信ウィンドウの増加割合は、ACRの増加割合と比較して小さいため、再送タイムアウト発生後の回線容量を十分に有効活用できないことがある。そして、その具体例についての解析を行い、さらにTCPのスロースタート/輻輳回避アルゴリズムがない場合との比較を行なった。最後に、日頃ご指導頂くKDD研究所浦野所長、鈴木次長に感謝します。

参考文献

- [1]:The ATM Forum Technical Committee, "Traffic Management Specification Version 4.0", atm95-0013R10, February 1996.
- [2]:W.Richard Stevens, "TCP/IP Illustrated, Volume 1 [The Protocols]", Addison-Wesley, ISBN 0-201-63346-9, February 1994.
- [3]:長谷川他, "ATM網におけるTCP over ABRサービスの性能評価とパラメータ調整," 信学技法, SSE95-117, Dec. 1995.
- [4]:阿野他, "ATM上のAvailable Bit Rate (ABR)を用いたTCPの実現法に関する一検討," 第60回情処全大, 1-151, Sept. 1995.