バックプレッシャ機能を有する ATM LAN スイッチの性能評価

～ 平均パケット遅延時間とパケット廃棄率 ～

大崎 博之　若宮 直紀　村田 正幸　宮原 秀夫

大阪大学基礎工学部

〒560 大阪府豊中市待兼山町 1-1

あらまし　次世代高速 LAN として、現在注目されている ATM LAN におけるトラヒック制御方式として、網が輻輳に陥ってから制御を施す輻輳適応型制御方式が有効であると考えられる。本稿では、その実現方式の 1 つとして、入出力双方にバッファを有する ATM スイッチにバックプレッシャ機能を付加した方式を対象とした性能評価を行なう。ここでは、対象とするスイッチにバースト性を持つトラヒックを収容した時のパケット遅延時間分布およびパケット廃棄率を数学的解析手法を用いて導出し、その評価結果を示す。

和文キーワード　ATM LAN、入出力バッファ型スイッチ、バックプレッシャ機能、バーストトラヒック

Performance Analysis of an ATM LAN Switch with Backpressure Function

～ Average Packet Delay and Pakcet Loss Probability ～

Hiroyuki OHSAKI, Naoki WAKAMIYA, Masayuki MURATA, Hideo MIYAHARA

Faculty of Engineering Science, Osaka University

Toyonaka, Osaka 560, Japan

Abstract　Reactive congestion control is thought to be effective in an ATM-based LAN. In this paper, we focus on an ATM LAN switch with input and output buffers equipped with back-pressure function to investigate its performance. The packet delay distributions and approximate packet loss probabilities are derived using a mathematical analytic technique against bursty traffic which is essential to be supported in ATM LANs .

英文 key words　ATM LAN, Input and Output Buffer Switch, Backpressure Function, Bursty Traffic

1 はじめに

ATM (Asynchronous Transfer Mode) 方式は、データ・音声 画像など様々なマルチメディア情報を非同期に扱うことによって、高速広帯域通信網 B-ISDN (Broadband Integrated Services Digital Network) を実現する。ATM 方式の有用性は、様々なメディアのトラヒックを、セルと呼ばれる固定長のパケットに分割し、多重化することにより網資源を有効に利用できることにある。これまで ATM 方式については、公衆網を対象とした研究開発、および標準化が行なわれてきた。

一方、コンピュータが急速に高速化、マルチメディア化されるにつれ、より高速なコンピュータ間通信の実現が要求されている。そこで、従来の Ethernet / FDDI に替わる次世代マルチメディア LAN として、これまで公衆網を対象としてきた ATM 方式を応用した、ATM LAN が注目を浴びている [1, 2]。

ATM LAN においては、網が輻輳に陥いってから制御を施す輻輳適応型制御方式 (Reactive Congestion Control) が有効であると考えられる。その実現方式の 1 つとして、 入出力双方にバッファを有する ATM スイッチにバックプレッシャ機能を付加する方式が提案されている [3]。バックプレッシャ機能とは、出力バッファ内のセル数がある閾値をこえた時に、入力バッファから出力バッファへの転送を止めることにより、一時的な輻輳を防ぐというものである。最近、文献 [4] においてバックプレッシャ機能を ATM LAN スイッチに適用することを目的として、その実現方式が提案されている。また、ATM LAN において、データ伝送系サービスへの適用を考えると、バースト性を考慮したトラヒックのモデル化が不可欠である。バックプレッシャ機能を有する ATM LAN スイッチに、バースト性を有するトラヒックが到着する場合の最大スループットについては文献 [5] で報告した。本稿では、平均パケット遅延時間およびパケット廃棄率を数学的解析手法により導出し、その評価結果を示す。

以下、2章において、本稿において扱う ATM スイッチの評価モデルを示す。3章において、定常状態における状態確率を求める。これを用いて、4章で平均パケット遅延時間を導出し、5.1章でパケット廃棄率を導出する。

2 バックプレッシャ機能を有する ATM スイッチとその評価モデル

本稿では、輻輳適応型のトラヒック制御法であるバックプレッシャ機能を有する $N \times N$ の ATM スイッチの性能を評価する。評価の対象とするスイッチは、入力側、出力側の両方にバッファを持ち、バッファサイズをそれぞれ、N_I、N_O とする。入力バッファは、セルを到着順にバッファに入れる FIFO 方式とする。また、交換速度は回線速度の N 倍とする。すなわち、入力バッファから出力バッファに、1 スロットで最大 N 個のセルが転送可能であるとする。

バックプレッシャ機能は、出力バッファ内の待ちセル数があるスレシホールド値を越えた時に、出力ポートから入力ポートへ制御信号を送り、入力側から出力側へのセルの転送を停止するものである [4]。この制御によって、出力バッファにおけるセル損失を無くすことが可能になる (図 1参照)。特に、このスイッチを ATM LAN に適用し、データ伝送系サービスに適用する場合を考えると、トラヒックモデルとして上位層からのパケット到着を考慮した評価が必須になる。そこで、本稿では、以下のような解析モデルを評価の対象とする。

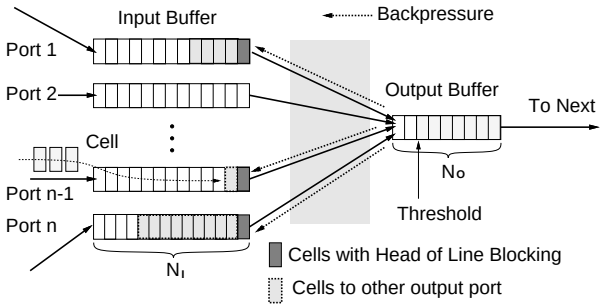

図 1: 解析モデル

入力ポートには、同一パケットを構成するセルが連続的に到着し、その個数は平均$\overline{BL}$ の幾何分布に従うとする。あるスロットにセルが到着した時に、次のスロットでも連続して同じパケットを構成するセルが到着する確率を p とすると、

$$\overline{BL} = \sum_{i=1}^{\infty} i p^{i-1} = \frac{1}{1-p}$$

なる関係がある。

出力バッファのスレシホールド値については、$N_O - N$ とするのが現実的であると考えられるが [4]、ここでは理想的な場合を考え、出力バッファに空きがなくなるまで入力バッファからランダムに選んで転送するものとする。

3 定常状態確率の導出

性能解析を行なうために、一般性を失なうことなく、ある出力ポートのみを考える。k 番目のスロットにおいて、入力ポートの先頭に並んでいる、着目する出力ポート向けのセル (これを HOL セルと呼び、HOL セルによる仮想的な待ち行列を HOL 待ち行列と呼ぶ) の数を確率変数 H_k であらわし、出力バッファ内セル数を確率変数 Q_k であらわす。ただし Q_k には、サービス中のセルも含むとする。

各スロットの状態は、確率変数 H_k、Q_k の組を用いて表すことができる。今、$k \to \infty$ として定常状態を考え、さらに、$N \to \infty$ とし、入出力ポート数を無限大とした場合、各スロットでの定常状態確率

$$r_{n,m} = P[Q = n, H = m]$$

求められる。$r_{n,m}$ の導出方法は [5] に述べてあるため、ここでは省略する。

4 パケット遅延時間

本節ではパケット遅延時間分布を求める。パケット遅延時間は、パケットを構成する先頭のセルが入力ポートに到着してから、最後のセルが出力ポートより送出されるまでに要する時間であり、以下、次の 3 つの時間に分けて考える。

1. W_I : パケットを構成する先頭のセルが入力ポートに到着してから、HOL 待ち行列に並ぶまでに要する時間。
2. W_S : パケットが HOL 待ち行列に並んでから、パケットを構成するすべてのセルを出力バッファに転送し終えるまでに要する時間。
3. W_O : パケットを構成する先頭のセルが出力側ポートに到着してから、リンクへ送出されるまでに要する時間。

ここで、はスロット間では FIFO、スロット内ではランダムに HOL セルの転送を行うものとする。以下、これらの導出を示す。

4.1 W_S の導出

まず W_S を導出する。そのために、パケットを構成するセルが実際に出力バッファに並んで行くのではなく、セルが仮想的に待ち行列を形成するものと考える。

今、あるパケットに着目し、そのパケットが HOL 待ち行列に到着した直後の状態を考える。着目するパケットを含めて、HOL 待ち行列に並んでいるパケット (セル) の個数が m である確率を u_m とすると、

$$u_m = \sum_{n=0}^{N_O} \sum_{j=1}^{m} r_{n,m-j} a'_j$$

ただし、a'_j は着目するパケットが、同時に到着する j 個のパケットの集団に含まれる確率で、

$$a'_j = \frac{j a_j}{\sum_{k=1}^{\infty} k a_k} = \frac{j a_j}{\lambda_p}$$

である。

HOL 待ち行列に並んでいるセルをすべて出力バッファに転送するのに要する時間をサイクルと呼ぶ。今、あるサイクルの最初に m 個の HOL セルがあり、次のサイクルの最初に m' 個の HOL セルがある時のサイクル長の確率分布を $T_{m,m'}(k)$ と定義する。

サイクルの最初にあった、着目するパケットを除く $m-1$ 個のパケットのうち j 個がさらに送るべきセルを持っていて、サイクルの間に $m'-1-j$ 個のパケットが新たに HOL 待ち行列に到着した場合、HOL 待ち行列内の HOL パケットの数が $j+k+1$ であるから、この時の遷移確率 $t_{m,m'}$ は、

$$t_{m,m'} = \sum_{j=0}^{m'-1} b_{m-1,j} a^m_{m'-1-j}$$

となる。ここで、a^m_k は m スロットの間に k 個のパケットが HOL 待ち行列に到着する確率であり、

$$a^m_k = \frac{(\lambda_p m)^k e^{-\lambda_p m}}{k!}$$

で与えられる。これより、$T_{m,m'}(k)$ は、

$$T_{m,m'}(k) = \begin{cases} t_{m,m'}, & \text{if } k = m \\ 0, & \text{otherwise} \end{cases}$$

次に、複数にまたがったサイクル長の確率分布を求める。サイクル数を l とした時のサイクル長の確率分布 $T^l_{m,m'}(k)$ は、

$$T^l_{m,m'}(k) = \sum_{j=1}^{\infty} [T^{l-1}_{m,j} \otimes T_{j,m'}](k)$$

ここで $\otimes$ は確率分布の畳み込み記号を示し、二つの確率分布 $y_1(k)$、$y_2(k)$ に対して、

$$[y_1 \otimes y_2](k) \equiv \sum_{j=0}^{k} y_1(j) y_2(k-j)$$

と定義する。

パケットを構成する最後のセルの、サイクル内での遅延時間分布を $U_m(k)$ とする。今、サイクル内ではランダムに HOL セルを選び、出力バッファに転送すると仮定すると、

$$U_m(k) = \begin{cases} 1/m, & \text{if } 0 \le k \le m-1 \\ 0, & \text{otherwise} \end{cases}$$

で与えられる。

着目するパケットが HOL 待ち行列に到着した場合に、m 個の HOL パケットが存在した時に、パケットを構成する最後のセルが出力バッファに転送されるまでに要する時間の確率分布 $W_m(k)$ は、

$$\begin{aligned} W_m(k) &= (1-p) U_m(k) \\ &\quad + \sum_{l=1}^{\infty} p^l (1-p) \sum_{j=1}^{\infty} [T^l_{m,j} \otimes U_j](k) \end{aligned}$$

従って W_S は、

$$W_S = \sum_{m=1}^{\infty} \sum_{k=1}^{\infty} k W_m(k) u_m$$

4.2 W_I の導出

次に W_I を導出する。そのために、入力バッファの先頭にパケットが並んでから、パケットを構成するすべてのセルが出力バッファに転送されるまでに要する時間を求める。その導出方法は W_S の導出方法と同様であるが、ここでは HOL 待ち行列の状態に加えて、出力バッファの状態も考慮する必要がある。

まず、着目するパケットが HOL 待ち行列に到着した直後の状態を考える。着目するパケットを含めて、HOL 待ち行列に並んでいるパケットの数が m、かつ出力バッファに並んでいるセルの数が n である確率を u_{nm} とすると、

$$u_{n,m} = \begin{cases} \sum_{j=1}^{m}(r_{0,m-j} + r_{1,m-j})a'_j, & \text{if } n = 0 \\ \sum_{j=1}^{m} r_{n+1,m-j}a'_j, & \text{otherwise} \end{cases}$$

サイクルの最初に状態が (n, m) であった時に、次のサイクルの最初に状態が (n', m') になる時の、サイクル長の確率分布を $C_{n,m,n',m'}(k)$ とする。なお、ここではサイクルの定義が先と異なり、HOL 待ち行列で見た時の時間である。この値は、以下のようにして求めることができる。

- $m \leq N_O - n$ の時

  すべての HOL セルは同時に出力バッファに転送されるため、サイクル時間は 1 である。出力バッファの状態は、$n' = n + m$ となる。

  一方、HOL 待ち行列については、サイクルの最初にあった、着目するパケットを除く $m-1$ 個の HOL パケットのうち j 個がさらに送るべきセルを持っていて、サイクルの間に k 個のパケットが新たに到着する場合を考えると、$m' = j + k + 1$ になる。従って、

$$C_{n,m,n',m'}(k) = \begin{cases} \sum_{j=0}^{m'} b_{m-1,j}a_{m'-1-j}, \\ \text{if } k = 1 \text{ and } n' = n + m \\ 0, \text{otherwise} \end{cases}$$

- $m > N_O - n$ の時

  出力バッファが一杯になるまで、すなわち、$N_O - n$ 個のセルは 1 スロットで出力バッファに転送される。その後、残りの $m - (N_O - n)$ 個のセルが 1 スロットずつ送られるから、サイクル長は

$$1 + m - (N_O - n)$$

  になり、出力バッファの状態は $n' = N_O$ となる。また、サイクルの最初にあった $m-1$ 個のパケットのうち j 個がさらに送るべきセルをもっていて、サイクルの間に k 個のパケットが新たに到着する場合を考えると、$m' = j + k + 1$ になる。従って、

$$C_{n,m,n',m'}(k) = \begin{cases} \sum_{j=0}^{m'} b_{m-1,j}a_{m'-1-j}^{m-(N_O-n)+1}, \\ \text{if } k = m - (N_O - n) \text{ and } n' = N_O \\ 0, \text{otherwise} \end{cases}$$

l 個の複数サイクルにまたがったサイクル長の確率分布は W_S の場合と同様にして、

$$C^l_{n,m,n',m'}(k) = \sum_{n''=0}^{N_O}\sum_{m''=1}^{\infty}[C^{l-1}_{n,m,n'',m''} \otimes C_{n'',m'',n',m'}](k)$$

パケットを構成する最後のセルの、サイクル内での遅延分布を $U_{n,m}(k)$ とする。今、サイクル内ではランダムに HOL セルを選び、出力バッファに転送すると仮定すると、

- $m \leq N_O - n$ の時

$$U_{n,m}(k) = \begin{cases} 1, & \text{if } k = 0 \\ 0, & \text{otherwise} \end{cases}$$

- $m > N_O - n$ の時

$$U_{n,m}(k) = \begin{cases} (N_O - n)/m, & \text{if } k = 0 \\ 1/m, & \\ 0, & \text{if } k \leq m - (N_O - n) \\ & \text{otherwise} \end{cases}$$

従って、パケットが HOL 待ち行列に到着してから、パケットを構成する最後のセルが HOL 待ち行列から出て行くまでに要する時間 W_H は以下のようになる。

$$W_H(k) = \sum_{n=0}^{N_O}\sum_{m=1}^{\infty} u_{n,m}\Bigg[(1-p)U_{n,m}(k) + \sum_{l=1}^{\infty} p^l(1-p) \sum_{n'=0}^{N_O}\sum_{m'=1}^{\infty}[C^l_{n,m,n',m'} \otimes U_{n',m'}](k)\Bigg]$$

その n 次モーメント $W_H^{(n)}$ は次のようにして得られる。

$$W_H^{(n)} = \sum_{k=1}^{\infty} k^n W_H(k)$$

最後に、入力バッファに到着したパケットを構成する全てのセルが HOL 待ち行列に加わるまでに要する時間 W_I を求めるために、サービス時間の平均が $W_H^{(1)}$、その 2 次モーメントが $W_H^{(2)}$ であるような Geom/G/1 待ち行列モデルを考える。この時、W_I は以下のようにして与えられる。

$$W_I = \frac{\lambda_p W_H^{(2)}}{2(1 - \lambda_p W_H^{(1)})}$$

4.3 W_O の導出

W_O は出力バッファにおける先頭セルの平均遅延時間であるから、

$$W_O = 1 + \sum_{n=1}^{N_O}\sum_{m=0}^{\infty} n r_{n,m}$$

4.4 数値例

図 2、3に、それぞれ平均パケット長$\overline{BL} = 1, 2$ の場合の、負荷と平均パケット遅延時間の関係を示す。

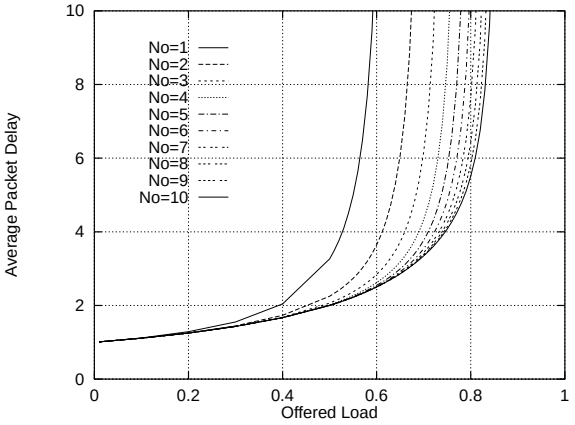

図 2: 平均パケット遅延 $(\overline{BL} = 1)$

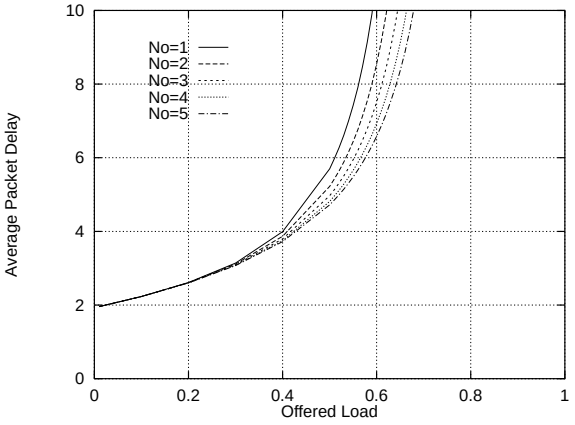

図 3: 平均パケット遅延 $(\overline{BL} = 2)$

これより、平均パケット遅延時間はスイッチに与える負荷が大きくなると急激に増加し、負荷が最大スループットになる時、無限大に発散することがわかる。また、出力バッファサイズを適切に設定することにより、平均パケット遅延時間の増大をある程度抑えることが可能である。

図 4に、解析結果と共に、シミュレーションで得られた、平均パケット長を変化させた時の、負荷と平均パケット遅延時間の関係を示す (出力バッファサイズ $N_O = 1$)。本解析ではポート数 N を無限大に仮定しているため、平均パケット遅延時間はポート数が有限の場合の上限を与えていることがわかる。ただし、シミュレーションの場合には入力バッファから出力バッファへの転送をサイクリックに行なうため、多少の違いが見られ、負荷の高い時には逆転が生じている。

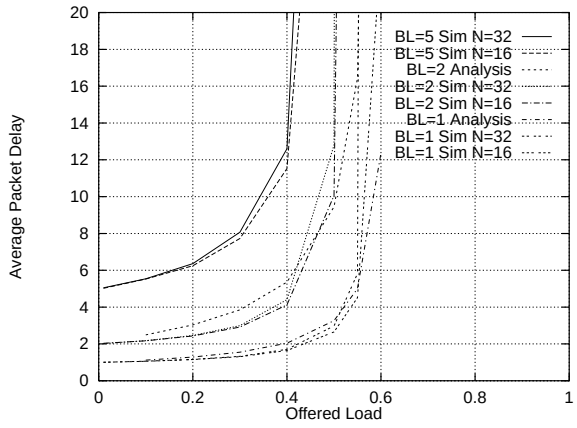

図 4: シミュレーションとの比較 $(N_O = 1)$

5 パケット棄却率

5.1 ガウス近似によるパケット棄却率の導出

本章ではパケット棄却率を導出する。前章までの解析では、入力バッファに到着順にセルを格納する場合を考えてきたが、それ以外に、入力ポートごとに、各出力ポートに対応したバッファを設け、HOL ブロッキングを防ぐ方式が提案されている [4]。ここでは、これら 2 方式 (前者を FIFO 方式、後者を RIRO 方式と呼ぶ) のパケット棄却率を近似解析により導出する。

今、離散時間系 $\mathrm{Geom/G/1}$ 待ち行列を考える。パケットの到着が平均 λ_p の幾何分布に従う場合、その z 変換は、

$$\Lambda(z) = 1 - \lambda_p + \lambda_p z$$

さらに、サービス時間の分布の z 変換を $B(z)$ とし、

$$b^{(i)} \equiv \frac{d^i B(z)}{dz^i}$$

とする。このような待ち行列システムの残余仕事量の z 変換は、

$$U(z) = \frac{(1-\rho)(1-z)\Lambda[B(z)]}{\Lambda[B(z)] - z}$$

で与えられる [6]。

今、残余仕事量をバッファ内セル数と考えれば、バッファサイズを L とした時のパケット棄却率は、ガウス近似を用いて、

$$P_L \cong Pr[U > L] = \int_L^\infty \frac{1}{\sqrt{2\pi V[U]}} e^{-\frac{(y-E[U])^2}{2V[U]}} dy$$

となる。ただし、$E[U]$、$V[U]$ は残余仕事量の平均および分散である。

5.2 FIFO 方式の場合

4節の解析によって得られた W_H をサービス時間とする待ち行列システムを考える。すなわち、

$$\begin{aligned} b^{(1)} &= W_H^{(1)} \\ b^{(2)} &= W_H^{(2)} - W_H^{(1)} \\ b^{(3)} &= W_H^{(3)} - 3W_H^{(2)} + 2W_H^{(1)} \end{aligned}$$

として、5.1節の結果を適用すればよい。

5.3 RIRO 方式の場合

今、ある入力バッファに着目する。入力バッファには、出力ポートに対応して N 個の $\mathrm{Geom/G/1}$ 待ち行列があり、それぞれ独立にサービスされると仮定する。これは、入力バッファ内セルを出力バッファに転送するスケジューリングを適切に行なえば、ほぼ実現可能である [4]。均一なセル到着を仮定すると、出力ポート j 向けのパケットの到着率 $\lambda_j = \lambda_p/N$ となり、その z 変換は

$$\Lambda_j(z) = 1 - \lambda_j + \lambda_j z$$

となる。

今、V_j を入力バッファにおいて出力ポート j 向けパケットの待ち行列の残余仕事量を表す確率変数とする。入力バッファを占有するのを防ぐために、スレシホールド値を設けることが考えられるが、その値を L とし、それによる棄却率を P_L とすると、

$$P_L \cong Pr[V_j > L]$$

入力バッファにおけるパケット転送時間のモーメントは、FIFO 方式と同様、前章の結果を用いることができ、それを 5.1節の結果に適用すればよい。

次に、入力ポート全体の残余仕事量 U_N を考える。

$$U_N = \sum_{j=1}^{N} V_j$$

その z 変換 $U_N(z) = V_j(z)^N$ を用いて、平均 $E[U_N]$、および分散 $V[U_N]$ が直ちに求められる。全体のバッファサイズ N_{in} を越える確率 P_N は

$$P_N \cong Pr[\lim_{N\to\infty} \sum_{j=1}^{N} V_j > N_{in}] = Pr[\lim_{N\to\infty} U_N > N_{in}]$$

である。これより、パケット棄却率 P は

$$P \cong \max(P_N, P_L)$$

として得ることができる。

5.4 数値例

図 5に、平均パケット長 $\overline{BL} = 1$ の場合の、負荷と平均パケット廃棄率との関係を、FIFO 方式、RIRO 方式、出力バッファ型スイッチのそれぞれの場合について示す $(N_I + N_O = 30)$。FIFO 方式、RIRO 方式どちらの場合も、スイッチに与える負荷が最大スループットを越えるあたりでパケット廃棄率が急激に高くなっている。FIFO 方式は RIRO 方式に比べて、同じ出力バッファサイズの時のパケット廃棄率は大きいが、出力バッファサイズをある程度大きくすることによって性能の向上が可能である。また、RIRO 方式はスイッチの構成が複雑であるが、高負荷時においてもはるかに良い性能を示している。出力バッファ型スイッチと比較しても、 ある程度出力バッファサイズを大きくすることによって、より優れた性能が得られることがわかる。

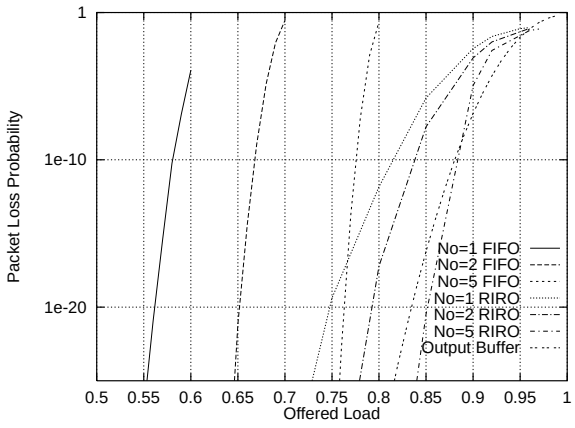

図 5: パケット廃棄率 $(\overline{BL} = 1)$

謝辞

本報告に関して貴重な御討論を頂いた日本電気株式会社 鈴木洋氏、および同社 樊瑞雪氏に深謝いたします。

参考文献

[1] 清水 洋, 西田 竹志, 鈴木 洋, “ATM-LAN へ至る道 (上) LAN の高速化、マルチメディア化,” 日経コミュニケーション, vol. 165, pp. 88–90, 1 1994.

[2] 西田 竹志, 鈴木 洋, 清水 洋, “ATM-LAN へ至る道 (下) LAN の高速化、マルチメディア化,” 日経コミュニケーション, vol. 166, pp. 88–90, 1 1994.

[3] I. Iliadis, “Performance of a packet switch with input and output queueing under unbalanced traffic,” in *Proceedings of IEEE INFOCOM '92*, vol. 2, (Florence, Italy), pp. 743–752 (5D.4), May 1992.

[4] 樊 瑞雪, 鈴木 洋, 山田 健治, “バッファ容量拡張可能な ATM スイッチ : *XATOM*,” 電子情報通信学会技術研究報告, vol. SSE-93-6, pp. 31–36, 1993.

[5] 大崎 博之, 若宮 直紀, 村田 正幸, 宮原 秀夫, “バックプレッシャ機能を有する ATM LAN スイッチの性能解析,” 電子情報通信学会技術研究報告, vol. SSE-94-4, pp. 19–24, 1994.

[6] H. Takagi, “Queueing analysis volume 3: Discrete-time systems,” *North-Holland*, 1993.