

ATM-LAN 環境と ATM-WAN 環境の転送特性の比較

松井 康範[†] 内海 秀介[‡] 船渡 大地[†] 中村 嘉志^{††} 成田 多良[‡] 細川 達己[‡] 徳田 英幸^{††}

慶應義塾大学大学院政策・メディア研究科[†]

慶應義塾大学大学院理工学研究科[‡]

電気通信大学大学院情報システム学研究科^{††}

慶應義塾大学環境情報学部^{††}

慶応大学で運用を開始した ATM ネットワーク上でのパフォーマンス計測の結果、物理層の速度に比べ TCP, UDP/IP の実効速度が低く、またキャンパス間 ATM バックボーンを介した場合と ATM-LAN 環境でパフォーマンスにかなりの差が生じることも認められた。セルレベルのトラフィック特性を調べることでこの原因について考察をする。それを踏まえ、ATM 上での効率的な転送プロトコル実現に向けての議論をする。

A Comparison of Transmission Characteristics between ATM-LAN and ATM-WAN environment

Yasunori Matsui[†] Shuusuke Utsumi[‡] Daichi Funato[†] Yoshiyuki Nakamura^{††}
Tomio Narita[‡] Tatsumi Hosokawa[‡] and Hideyuki Tokuda^{††}

Graduate School of Media and Governance, Keio University[†]

Faculty of Science and Technology, Keio University[‡]

Graduate School of Information Systems, The University of Electro-Communications^{††}

Faculty of Environmental Information, Keio University^{††}

The performance of TCP/IP over ATM-WAN network and ATM-LAN environment is evaluated. The result showed significant difference in throughput, both of which are far lower than expected from ATM physical layer performance. By conducting cell-level traffic measurement and analysis, we explain their causes. From that standpoint, required characteristics of efficient upper layer transmission protocol over ATM is discussed.

1 はじめに

ATM は次世代の大容量通信のための基盤技術としての地位を確立しつつあり、キャンパスネットワークのバックボーンとして ATM を導入する大学が相次いでいる。慶応大学でも「慶応大学情報スーパーハイウェイ計画」の一環として首都圏に点在している 5 つのキャンパス (三田, 日吉, 矢上, 信濃町, 湘南藤沢) を NTT の ATM 公衆回線により接続し、遠隔教育や遠隔医療の実験などを進めている。

ATM は主にデータリンク層のスイッチング技術・規格である。任意のデータを、ヘッダ 5 byte + データ 48 byte の固定長セルに分割して転送する。このためスイッチング処理を単純化でき、高速にデータを転送することができる。運用時は、その上に AAL5 など、さらに TCP, UDP/IP といった上位層の転送プロトコルをか

ぶせて利用することになる。そのため、ATM セルの制御は上位プロトコルのパフォーマンスが最大になるように行なわれなければならない。

ところで ATM ネットワーク導入初期の計測の結果、現在コンピュータ間通信で最も利用されているプロトコルである TCP/IP および UDP/IP の実効速度が、使用しているデータリンク層 OC-3c の速度 155Mbps に比べて非常に低いことが測定された。また、少数の LAN 用 ATM スイッチとホストによる ATM-LAN 環境と、NTT 網によるキャンパスバックボーンでの TCP/IP のパフォーマンスにかなりの差があることが観測された。

これらの上位プロトコルの転送特性は下位の ATM セルの転送特性および上位プロトコルと下位プロトコルの間のマッピングから決定されると考えられる。本論文ではセルレベルでのセルロス, RTT の計測を通じて

ATM 上での上位プロトコルの効率向上に対する考察を与える。

本論文の構成は次の通りである。第 2 章は実験環境を述べる。第 3 章では ATM 上に IP パケットなどの上位プロトコルを乗せる方法について説明する。第 4 章で TCP, UDP のパフォーマンス計測の結果を、第 5 章ではセルレベルでのセルロス, RTT の計測について述べる。第 6 章で計測結果についての考察を与える。関連研究は第 7 章にまとめた。

2 実験環境

実験に使用したマシン環境は次の通りである。ホストマシンは Sun SS-20 (1 CPU SuperSparc 60 MHz), SunOS 4.1.4, ATM カードは FORE 社製 SBA-100 TAXI 100Mbps, LAN-ATM スイッチは FORE 社製 ASX-200 および ASX-200BX である。TCP, UDP/IP の計測は FORE 社製ドライバ (バージョン 3.0.2) を、セルレベルの計測には自作のドライバを使用した。

セルの RTT 計測には 125nsec の精度のクロックをもつ MAGMA 社製の SBus 用タイマーボードを利用した。ATM ドライバにこのタイマーボードのドライバルーチンを組み込み、コンテキストスイッチの影響などを受けないようにした。

測定のための ATM ネットワークトポロジは以下の 3 種類がある。まず、スイッチなどの影響がない状態で計測を行なうためのクロス接続環境 (図 1)。本実験ではホスト側の ATM セル処理に与える影響を除くため、この構成を計測の基準とした。ATM-LAN 環境として図 2 にあるように 1 つの ATM スイッチを介した 2 つのホストを接続した。これに対して、ATM-WAN を介した環境を図 3 のように構築した。慶応大学の ATM バックボーンはまだ本格運用に入っていないためトラフィックは比較的少なく、実験に与える影響は小さいと考えられるが、実験は日をかえて複数回行なうようにした。また計測はすべて PVC で行った。

3 ATM 上での TCP, UDP/IP について

本章では IP パケットを ATM セルにマッピングする方法について述べる。

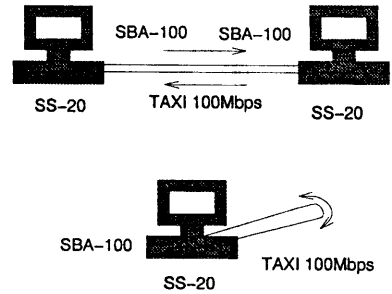


図 1: ATM クロス接続環境 (下は RTT 計測時)

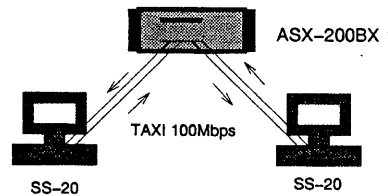


図 2: ATM-LAN 環境

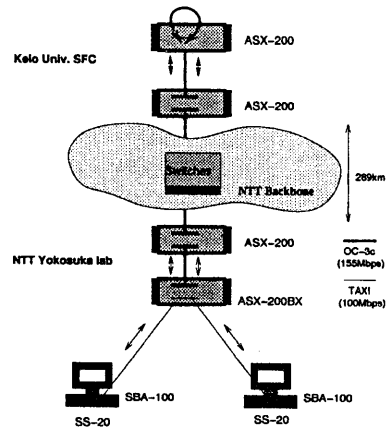


図 3: ATM-WAN 環境

TCP, UDP はデータ転送の単位として IP パケットを使用する。IP パケットは可変長であるが、通常物理インターフェースによって定められる MTU (Maximum Transmission Unit) の長さ以上のデータは別々の IP パケットに分割される。Ethernet の場合は通常 1500byte 程度であり、ATM の場合は RFC によって 9180 bytes が推奨されている [1]。FORE 社のデフォルト設定は 9188 bytes である。これにヘッダ・トレーラがつき、48 bytes ごとにセルに分割される。

ATM レイヤーはデータリンク層技術であるので、フロー制御などは行なわない。このため、ATM 上で信頼性のある通信を行なう場合、信頼性保証を行なう機構が別に必要となる。通常は SSCS など上位プロトコルレベルで再送処理などを行ない、信頼性を保証する。TCP/IP over ATM では TCP 層が再送処理を引き受ける。このとき再送の判断は IP パケットレベルで行われる。すなわち、パケットを構成する ATM セルが一つでも落ちたとき、その IP パケット全体が再送の対象となる。そのため MTU の長さで TCP の転送効率には密接な関係がある。MTU が長いと、一つの IP パケットが多くの ATM セルから構成されることになるため、わずかのセル落ちからパケット全体の再送が発生しやすくなる。逆に MTU が短いと、IP パケット処理のオーバーヘッドが大きくなり、全体の転送効率が低下する。

ATM 上で TCP/IP の効率を上げる研究については関連研究のところで述べる。

4 TCP, UDP のパフォーマンス計測

本節では TCP, UDP/IP over ATM の計測結果を述べる。計測には `ttcp` を用い、ソケットバッファサイズのみを変化させ計測をした。各スイッチに帯域を予約せずに (PCR などが unlimited の状態) PVC を張り、転送パスを作成した。

ATM-LAN 環境と ATM-WAN 環境で行なった実験の結果は図 4 のようになった。グラフよりはっきりと読みとれるように、ATM-LAN 環境に対して、ATM-WAN を介すると 1/4~ 1/5 程度のパフォーマンスしかでない。

UDP に関しては、SBA-100 と附属のドライバを利用した環境では転送のロスが非常に大きく、意味のある

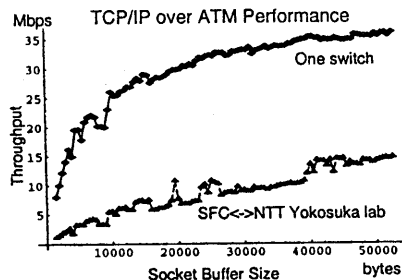


図 4: TCP/IP のスループット

データがとれなかった。そのため、SBA-200 ボードによって計測を行なったところ、ATM-WAN, ATM-LAN 環境のスループットに差は見られなかった¹。

5 ATM セルレベルのセルロス、RTT 計測

ATM 層の予約帯域を越えてセルを流した場合は、予約パラメータに適合しないセルがスイッチにより廃棄されセルロスが発生する。上位レイヤーのトラフィック特性の原因を調べるためには、これら ATM セル層のトラフィック特性を知る必要がある。そこで、ATM セルレベルでのセルロスの計測、RTT の計測を ATM-WAN, ATM-LAN 環境で行なった。

5.1 セルロスの計測

シーケンス番号を振った 25000 個のセルを次々に転送することで、セルが失われる様子を観察した。

予約帯域以下の速度でセルを送出した場合、セルロスは ATM-LAN 環境および ATM-WAN 環境共にまったく観測されなかった。より多くのセルを流した場合の参考資料として、回線品質測定のために他のトラフィックをすべて遮断して行なわれた ATM アナライザを用いた同様の実験の結果を表 1 に示す。NTT 網における慶

¹SBA-200 で TCP を測定した場合、スループットは SBA-100 よりも向上している (最大で 45Mbps 程度)。しかし ATM-WAN と LAN 環境の差は SBA-100 のときよりも小さいが (WAN は LAN の 60% 程度) SBA-200 でも観測された。

試行	送出速度 (Mbps)	受信セル数	セルロス数
1 回目 (30min)	134.05159	569.08693E6	0
2 回目 (30min)	134.05159	569.08694E6	0

表 1: アナライザ (HP E-4200A) による ATM-WAN のセルロス観測実験

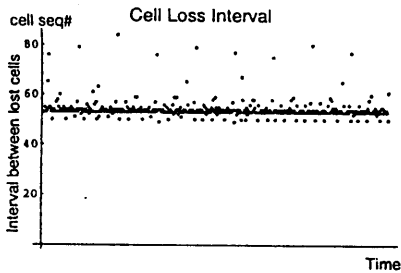


図 5: 予約帯域よりも多くのセルを流した場合のセル落ちパターン

応大学の予約帯域は 135Mbps であるので、ほぼ帯域一杯で送った場合もセルは落ちていない。

次に帯域を PCR (Peak Cell Rate) によって制限した場合のセルロスのパターンについて観察をした。ASX-200BX スイッチで PCR による帯域制限を行ない、それよりもわずかに速くセルを送出して観測したものが図 5 である。図は、あるセルが落ちてから次のセルが落ちるまでの間隔を、セル番号の差によって計測し、プロットしている²。

予約帯域よりも多いセルを定常的に流すと、ほぼ周期的にセルが落されていることがわかる。これはスイッチにおけるセルの流量制御アルゴリズム Leaky Bucket Algorithm[3] によって発生する。Leaky Bucket Algorithm はバッファと等間隔でセルを送出する部分からなっており、セルの到着がセルの送出間隔 (PCR) 以下の場合にはセルはそのまま流れるが、到着間隔が送出間隔よりも速い場合、バッファに溜めることのできるセル数以上は廃棄をしてしまう。つまり、一時的に PCR を越えるセルがスイッチに流れ込んだ場合は制限値に合う

²NTT 網側は、慶応大学全体が利用している VP に対して 135Mbps のシェーピングをかけているが、それ以外の設定を行なうことはできない。

ようにトラフィックシェーピングがかけられるだけであるが、一定時間以上流入セルが多い状態がつづくるとセルロスが発生する。バッファが一杯の状態では、PCR の間隔でのみセルが通過できる。このセル落ちパターンが TCP などの上位層の転送効率に重大な影響を与えると考えられる。この問題については第 6 章で議論をする。

5.2 RTT (Round Trip Time) の計測

セルロスの計測と同様に各セル毎にシーケンス番号を振り、次々に転送する。転送開始時にタイマーボードから時間を得、帰ってきたセルに対してもタイマーボードから時間を得、その差から RTT を算出した。

RTT 計測時にはホスト側のオーバーヘッドを考慮しなければならない。これには、ボードのレジスタへの書き込み、読みだし時間、インタラプトを発生するための時間、インタラプトハンドルーチンへの移行時間などが含まれる。まずクロス環境により測定を行いこの時間を評価したところ、この値は一定であることが確認された。本研究は ATM ネットワーク自体の特性を調べるのが目的であるため、クロス環境の計測を元にこれらのホスト側の影響を差し引くようにした。測定の結果は図 6,7 の通りである。

SFC ↔ NTT 横須賀通研の往復 RTT は平均 3880 μ sec である。TCP の再送アルゴリズムで用いられる RTT はこの値を利用する。これが TCP のスループットを悪化させる原因の一つになると考えられる。この点については第 6 章で議論をする。

6 議論

以上の実験の結果および以下に述べる考察から次のような ATM の特性を見ることができると考えられる。

- 予約帯域以内であれば、ATM-LAN, ATM-WAN 環境を問わずセル落ちはほとんど発生しない。予約

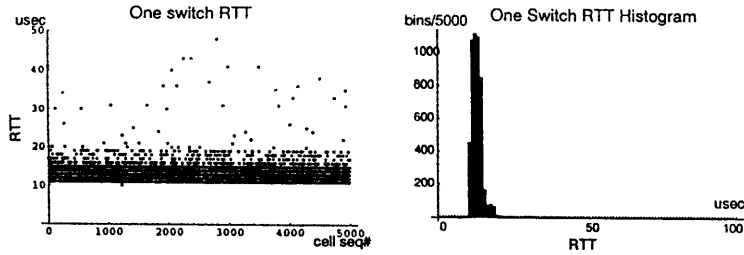


図 6: ATM-LAN 環境におけるセルレベル RTT (平均 13 μ sec)

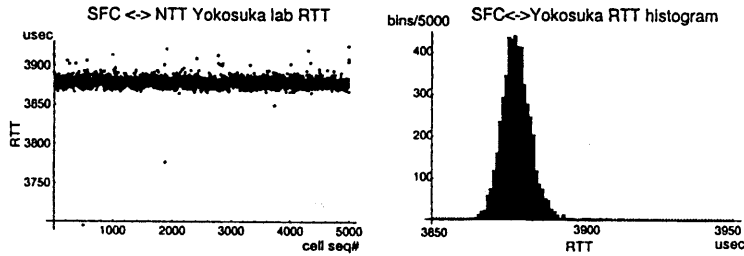


図 7: ATM-WAN におけるセルレベル RTT (平均 3880 μ sec)

したパラメータを守るようにセルを送出することが重要である。

- 予約帯域以上のセルを流すと、セル落ちが発生する。わずかなセル落ちも上位層の転送効率に重大な影響を与える。
- RTT の大きさと帯域幅を考えた場合、フィードバックによる流量制御はトラフィックの状態を予測しながら行なわなければならない。

セル落ちには、特に予約帯域以上のセルを流した場合に問題が生じる。ATM セルレベルのトラフィックを規定するパラメータとして、PCR、SCR、BT、MBS³がある。ATM 上でデータ転送を行なう場合は、上位プロトコルとこれらのパラメータをマッピングしなければならない。特に TCP のようにバースト的に best-effort で転送を試みる上位プロトコルを ATM に流す場合、平均的には予約帯域以下のトラフィックしか流さない場合

³PCR (Peak Cell Rate), SCR (Sustained Cell Rate), BT (Burst Tolerance), MBS (Maximum Burst Size), これらは ATM Forum UNI 標準で規定されている

でも一時的に PCR、SCR を越えるトラフィックの発生によってセル落ちが生じる可能性がある。そのため、トラフィックを発生するホスト側で、上位プロトコルレベルのトラフィック制御だけでなく、ATM セルレベルで指定パラメータを守るような制御を行わなければならない。

前述したように、予約した帯域以上のセルが定常的に流れた場合、スイッチの流量制御アルゴリズム (Leaky Bucket Algorithm) は等間隔でセルを落す。このとき、上位層から見ると、全部のバケット、フレームに対してデータ欠損が生じることになり、すべてが廃棄の対象となる。つまり、このときセルは大量に流れているにもかかわらず、データは全く流れなくなる。現在の TCP over ATM ではセル落ちが生じてから初めて輻輳として検出され、ACK が送り手側に返されなくなることで流量制御が行なわれる。すなわち、輻輳が起こると全く無駄に帯域を消費している状態が一定時間生じてしまう。

RTT の大きさは ATM が大容量のデータ転送を行なうことを考慮すると大きな問題になってくる。TCP では ACK の帰ってくる速度とバッファサイズによって転

送効率が決まってくる。ATM-WAN と ATM-LAN 環境のパフォーマンスの違いはこの点によるものと考えられる。50Mbps の帯域を使用しているとすると、今回の実験で得られた、NTT 横須賀研究所 ← 慶応大学湘南藤沢キャンパスの RTT 値約 4ms の間に、約 25 Kbyte のデータが流れる。ACK やフィードバックによって流量制御を行なうことを考える場合、セル廃棄が始まってからフィードバックをかけたのではセルが無駄に流れることになる。このためセル落ちが発生する前にフィードバックをかけるメカニズムが必要になる。これは特に ABR を用いる場合の大きな問題となることが予想される。フィードバック制御を用いる場合は、帯域の変動をかなり正確に予測した上での制御が要求される。

7 関連研究

セル廃棄の低減に関しては、これまで様々なメカニズムが提案されているが、大きくフィードバック方式と帯域予約方式との2通りに分類できる [6]。フィードバック方式では、ネットワーク内での輻輳の情報をホストにフィードバックして、動的に転送速度をコントロールする。一方の帯域予約方式では、コネクション設定時にあらかじめネットワーク資源を予約し、輻輳の可能性を減らす方式である。フィードバック方式は best-effort 型のサービスに向き、帯域予約方式はサービス保証型のサービスに向くと考えられている。

TCP/IP を ATM 上で利用する際の問題点はいくつか指摘されている。例えば、ウィンドウサイズの問題や、slow start アルゴリズムの問題、輻輳回復アルゴリズムの問題などである。TCP Vegas [2] は TCP の輻輳回避のアルゴリズムを改良する事により、高速ネットワークに対応している。Vegas では RTT を細粒度に測定することによりスループットを計算し、その変化から、輻輳を予測している。これにより、パケットロス未然に防ぐ事ができ、スループットの極端な低下を防ぐ事が出来る。また、TCP Reno の再送機構を更に強化して実装することにより、再送待ち時間を減少させ、スループットの低下を防いでいる。

また、Flow-Controlled Virtual Channels (FCVC) というセル廃棄の低減を狙った方法も提案されている [4]。FCVC では各スイッチで VC 毎にバッファを予約し、バッファの使用状況を制御している。そしてこの情報を

送信側へフィードバックすることにより、フロー制御を行い帯域を無駄なセルで浪費することを防いでいる。

これまでの多くの実験・研究は ATM-LAN 環境下で行なわれたものが多かった。本研究は実際の運用を目的とした ATM-WAN 環境における性能評価であり、今後のさまざまな研究の基礎となると考えられる。

8 最後に

本論文では実際に運用を開始した ATM 網を利用してその特性の基礎的なデータを収集し、特に TCP/IP のパフォーマンスについて議論をした。現在、この実験から得られた ATM の特性に関する知識をもとに、ATM に適合した TCP/IP の実現や連続メディアに適したネットワークアーキテクチャについて研究を進めている。

謝辞 NTT 情報通信研究所の南部明氏、光澤敦博士、木原誠司氏、稲村浩氏、NTT 光ネットワークシステム研究所の藤井哲郎博士には実験環境の提供や貴重な助言を頂きました。また本研究は慶応大学 ATM グループの方々との多くの議論に依っています。この場を借りてお礼申し上げます。

参考文献

- [1] Atkinson, R. *Default IP MTU for use over ATM AAL5*. RFC1626, 1994
- [2] Brakmo, L. S. O'Malley, S. W. Peterson, L. L. *TCP Vegas: New Techniques for Congestion Detection and Avoidance*. ACM SIGCOMM '94, 1994
- [3] *Traffic Control and Congestion Control in B-ISDN*. ITU-T Recommendation I.371, 1993
- [4] Morris, R. Kung, H.T. *Impact of ATM Switching and Flow Control on TCP Performance: Measurements on an Experimental Switch*. In Proceedings of GLOBECOM 1995, 1995
- [5] Romanow, A. Floyd, S. *Dynamics of TCP Traffic over ATM Networks*. ACM SIGCOMM '94, 1994
- [6] Zhang, H. *Service Disciplines For Guaranteed Performance Service in Packet-Switching Networks*. to appear in Proceedings of the IEEE, 1995