

長距離超高速インターネット (3) — 特性解析 —

1U-4

天海良治*, 村上健一郎*, 釘本健司**, 岡敦子**

NTT 基礎研究所*, NTT ソフトウェア研究所**

1. はじめに

長距離超高速インターネット (Long Fat Pipe, LFP) 研究の一環として、超高速専用線を使用したインターネット接続の実験を行なった。東京・京都間約 880Km を SDH の STM-1, 速度 155.520Mbit/s の専用線で結び、これを分割した 45Mbit/s の通信路を用意した。この上で、計算機プロセス間で TCP/IP^[1] によるネットワーク接続を構築し、1本のストリームの (多重化していない) 通信の特性を解析した。その結果、物理的な応答遅延の影響により、回線速度に見合った通信速度が得られないことが明らかになった。

本論文では、接続実験の結果について説明し、伝送遅延の影響を示す。特に LFP シンドロームと呼ぶ現象についての解析結果を示す。

2. 長距離超高速インターネット接続実験

実験の環境、理論値については [4],[5] を参照のこと。

2.1 測定条件

測定は、2台のワークステーション間に TCP/IP による接続を確立し、この1本のストリーム上でメモリ間データ転送を行なった。実験には我々で作成したプログラム rttcp を使用した。これは送り側、受け側の計算機でそれぞれプロセスを生成し、いくつかのパラメータを互いで設定したのち、指定量のデータを転送し、転送速度などの情報を表示するものである。

今回の測定では、約 4Kbyte のデータを 20000 回のシステムコールで送り出した。ウィンドウサイズによるフロー制御に従ってパケットが送り出される。rttcp が表示する統計情報と共に、転送パケットを FDDI プロトコルアナライザによって収集し、RTT の値、Seq 番号、Ack 番号の推移を解析した。実験のパラメータを示す。

1. ウィンドウサイズ: 8Kbyte から 520Kbyte
2. MTU (Maximum Transmission Unit): 4352byte
3. 1回の転送量: 約 80Mbyte
4. ホストは 100Mbit/s の FDDI に接続。ルータ間は 45Mbit/s であり、これが TCP ストリームが使え

る最大帯域。

5. 光ファイバー長距離片道約 880Km

2.2 測定結果と解析

ウィンドウサイズとスループットの関係を図 1 に示す。これから、ウィンドウサイズが 100Kbyte 以下の領域、100Kbyte から 170Kbyte の領域、170Kbyte 以上の領域の 3つの領域で状態が異なることがわかる。

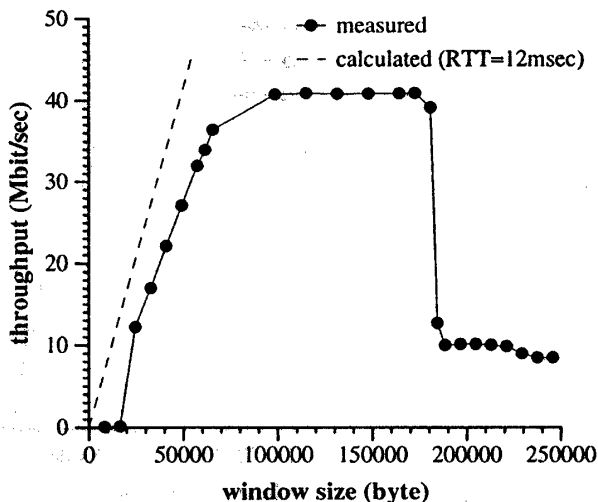


図 1. ウィンドウサイズとスループット

1. ウィンドウサイズが 100Kbyte 以下の領域

この領域では、ウィンドウサイズが帯域と RTT の積より小さいので、ウィンドウサイズの増加に従ってスループットは線形に大きくなる。スループット増大の割合は $1/RTT$ である。プロトコルアナライザで測定したワークステーション間の RTT の実測値は約 12 ミリ秒であった。この値による計算値を図の破線で示した。

2. ウィンドウサイズが 100K から 170Kbyte の領域

この領域では安定した転送が行なわれている。転送速度もほぼ回線の帯域と同じである。確認応答のパケットがちょうど送信パケットと同じタイミングで戻ってくる。パイプライン効果が最大限に働いている。

3. 170Kbyte 以上の領域 (LFP シンドロームの発生)

この領域では、ウィンドウサイズが帯域と RTT の積より大きくなる。ホストは回線帯域以上にデータを送りこもうとする。だが、TCP プロトコルは回線の帯域を知らない。今回の実験は、ワークステーション

A Long Fat Pipe Internet - An analysis of Behavior -
Yoshiji Amagai*, Ken-ichiro Murakami*, Takeshi Kugimoto**
Atsuko Oka**

*NTT Basic Research Laboratories

**NTT Software Laboratories

は100Mbit/sのFDDIに接続されているが、経路には45Mbit/sの回線が含まれている。

実際には、TCPの下位のIPレイヤでもってIPパケット単位で複数の経路を通して送ることもある。さらに、同一回線上の他の接続によるトラフィックもあるので、TCPパケット処理時に実際の利用可能な帯域を計ることは不可能である。そこで、TCP処理ではパケットの消失の検知をもって、利用可能な帯域を超えたときとみなす。さらに、ストリームごとに利用可能な帯域をシミュレートしている輻輳回避アルゴリズム (Congestion window) とこれを使った輻輳回避アルゴリズム [2][3]、利用可能帯域の最大値を探る Slow-start アルゴリズム [2] により、回線の状況に応じた最大の帯域を求め、効率を上げる。だが、実験の結果が示すようにLFPではこれがうまく働かず、スループットが8Mbit/s程度にまで低下した。

2.3 LFP シンドロームの解析

LFP シンドロームの発生している状況をアナライザで収集した情報によって解析した。送信パケットのTCP seq番号と確認応答パケットに含まれるAck番号の推移を図2に示す。

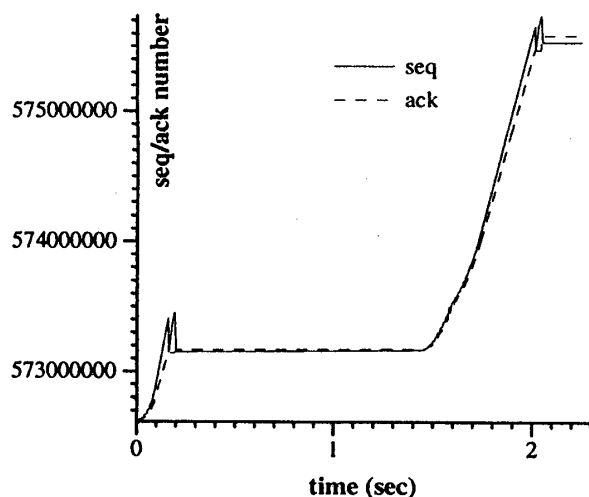


図2. Seq/Ack 番号の推移

図2で経過時間0.2秒の点から1.5秒の点まではパケットはまったく送られていない。このためトータルのスループットが低下した。これは、輻輳回避アルゴリズムと Slow-start アルゴリズムがうまく働いていないことを表わしている。アナライザによる情報によれば、0.2秒の時点で、seq番号とack番号の差は約270000である。この時点でRTTは約300ミリ秒になっていた。これだけのデータ量が伝送路、および、途中のルータのバッファに保持されていた。また、seq番号の増加の傾きは約65Mbit/s、Ack番号の増加の傾きは約44Mbit/sであった。伝送路の途中の帯域は45Mbit/sであるが、ワークステーションには100Mbit/sのFDDIが接続されているので、送り側がデータを大量に送り

こみ、輻輳を招いた。問題は、輻輳が起きたあとの送信の戦略アルゴリズムがLFPに適應できなかったことである。

LFP シンドローム発生の原因は以下のようにまとめられる。

- 1 LFPではRTTが大きいので、パケットの消失を検知するまでのタイムアウトが長くなり、長い間パケットの送信が停止してしまう。図2では転送が0.2秒継続したあと、確認応答が返ってこないためウィンドウが尽きて送信を停止するが、消失を検知して転送を再開するまで1.3秒かかっている。
- 2 RTTが大きいので消失を検知するまでに多くのパケットを送り出してしまう。
- 3 パケット消失が発生すると、TCPプロトコルの再転送アルゴリズムにより、消失したパケット以降のすべてのパケットが再転送される可能性がある。LFPでは、ウィンドウサイズが大きいので、再転送されるパケットも多くなり、スループットが低下する (TCPの選択的確認応答オプションはここでは仮定しない)。

これらの問題を根本的に解決するには、予測による先行処理、確認応答を要しないエラー回復、互いが協調して予測することによる通信をしない通信、といった新手法を作り出す必要がある。これは今後の課題である。

3. おわりに

本論文では、LFPでは、従来のTCPの送信戦略では長距離超高速の通信に対応できず、スループットが大幅に低下するLFPシンドロームが発生することを示し、その原因について考察した。現在、実験を継続しており、LFPに適したアルゴリズムとその効果を検証中である。

[文献]

- [1] J. Postel: Transmission Control Protocol, RFC793, 1981
- [2] V. Jacobson: Congestion Avoidance and Control, SIGCOMM'88, pp.314-329, 1988
- [3] L. Zang, S. Shenker and D. D. Clark: Observations on the Dynamics of a Congestion Control Algorithm - The Effect of Two-Way Traffic -, SIGCOMM'91, pp.133-147, 1991
- [4] 岡 敦子, 釘本健司, 天海良治, 村上健一郎: 長距離超高速インターネット (1) - 実験概要について -, 第50回情報処理全国大会, 1995.
- [5] 釘本健司, 村上健一郎, 天海良治, 岡 敦子: 長距離超高速インターネット (2) - スループット -, 第50回情報処理全国大会, 1995.
- [6] 村上健一郎, 天海良治, 釘本健司, 岡 敦子: 長距離超高速インターネット (4) - ボトルネック -, 第50回情報処理全国大会, 1995.