

長距離超高速インターネット (2)

1U-3

- スループット -

釘本 健司[†] 村上 健一郎[†] 天海 良治[†] 岡 敦子[†]

NTT ソフトウェア研究所[†] NTT 基礎研究所[†]

1 はじめに

我々は、東京大学医科学研究所と京都大学化学研究所の間を結ぶ、ヒトゲノム解析計画のためのインターネットである GenomeNet[1] のバックボーンを、両大学の協力を得て高速化した。このバックボーンでは、155 Mbit/sec の STM (Synchronous Transfer Mode)-1 による総延長 880km の SDH[2] (Synchronous Digital Hierarchy) 長距離専用線を用いている。

現在、このバックボーン上でコンピュータ間の通信の様々な特性の計測を行ない、理論値との比較を行っている。TCP/IP を用いた長距離高速ネットワーク (LFP, Long Fat Pipe) での通信では、高速な計算機を用いただけでは高いパフォーマンスの通信を実現できない。既存の TCP[3] で用いられるアルゴリズムやパラメータが、伝送路が長距離になることによって起こる大きな伝送遅延を考慮したものとなっていないためである。本論文では、LFP におけるボトルネックについて説明した後、応答時間およびスループットの理論値を示す。

2 長距離高速ネットワークの特性

2.1 伝送遅延の要素

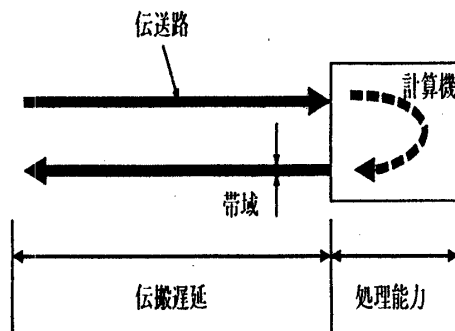


図 1: 伝送遅延の要素

図 1 に示すように、伝送遅延には、帯域、伝搬遅延、計算機の処理能力の三つの要素が考えられる。ここでは、まず帯域と伝搬遅延だけについて考える。パケットがネットワークに一方の口から入ってからも

う一方の口から完全に出るまでの時間、すなわち伝送遅延は以下のようになる。

$$T = \frac{L}{c} + \frac{8N}{B} \quad (1)$$

$L(m)$ はファイバの距離、 c はファイバ中の光の速度 ($1.8 \times 10^8 m/s$)、 B は帯域 (bit/sec) であり、また、転送メッセージ長は $N(\text{bytes})$ である。

ここで、155 Mbit/sec の伝送路を使って、128 byte のメッセージを 1000 km 転送する場合を考えると、上式の第一項は 5.5 msec、第二項は 6.61×10^{-3} msec となる。ネットワークが高速になってくると第二項はさらに小さくなるが、第一項を小さくする方法はない。すなわち長距離高速ネットワークにおいては、伝送遅延は光速と距離に支配される。

2.2 平均応答時間

次に、伝搬遅延と計算機の処理速度の関係について考える。LFP における平均応答時間 R を $M/M/1$ 待ち行列モデルを用いて求めてみる [4]。

ヘッダの処理にかかる時間は、メモリ間コピーや構造体へのアクセスにかかる時間などの総和である。Kay[5] の測定によれば、メッセージの大きさが 128 Kbytes 程度の場合、処理時間がメッセージの長さに依存するのは、これらの要素のうちチェックサム計算だけである。20 MIPS (Million instructions per second) の計算機では、チェックサム計算は 1 byte あたり 3.4 μ sec かかり、それ以外の処理には、合計 930 μ sec の時間がかかる。すなわち処理時間は合計 $(930 + 3.4N) \mu$ sec となる。

この数値をもとに、今日の 200 MIPS の高速 CPU の処理時間として、この 10 分の 1 の値、すなわち平均 $(93 + 0.34N) \mu$ sec の指数型サービス時間を想定すると次式を得る。

$$R = \frac{(93 + 0.34N) \times 10^{-6}}{1 - \rho} + \tau \quad (2)$$

ここで、メッセージ長 $N = 128$ byte、窓口使用率 $\rho = 0.9$ とする。前節で述べたように、LFP 環境では式 (1) の第二項が無視できるので、往復の遅延時間は $\tau = \frac{2L}{c}$ であるとし、 $L = 1000$ km とすれば、第二項は 11.1 msec となり、第一項は 1.37 msec となる。平均応答時間は第二項、すなわち伝搬遅延に支配されることがわかる。

したがって、伝送距離が大きくなれば平均応答時間が増大するため、計算機間での処理の速度が低下すると考えられるが、これを解決するには、並列ストリームを用いた並列処理などのパラダイムの転換が必要となる。

Takeshi Kugimoto, Ken-ichiro Murakami,
Yoshiji Amagai, Atsuko Oka

NTT Software Laboratories, NTT Basic Research Laboratories

2.3 TCPのスループット

計算機間の通信では、信頼性確保のためにACK (Acknowledgement) と呼ばれる確認応答を要するが、送り側が一つのメッセージを送るごとにACKを待っていたのでは高速転送が望めない。そこでTCPでは、ウィンドウ制御(図2)と呼ばれるパイプライン処理を用いてデータを先送りすることにより転送速度の向上をはかっている。ACKを受け取るまでの間に先送りできる最大のデータ量を表すものがウィンドウサイズである。

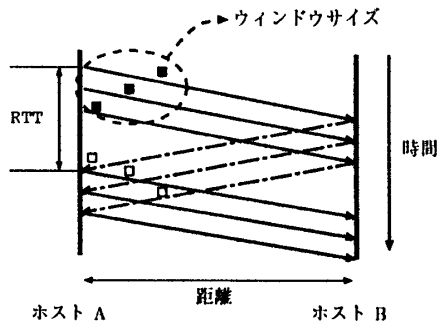


図2: ウィンドウ制御

伝送路の帯域を効率良く使用するには、このパイプラインを継続的に動かす必要がある。そのためにはウィンドウサイズ W (byte) が以下の式を満たせばよい。

$$8 \times W \geq RTT \times B \quad (3)$$

ここで、 RTT (sec) はラウンドトリップ時間 (Round Trip Time) であり、 B (bit/sec) は伝送路の帯域である。

逆に、ウィンドウサイズ W が決まっている場合のスループット T (bit/s) は、次式の通りとなる。

$$T = \min\left(\frac{W}{RTT}, B\right) \quad (4)$$

式(4)から、ウィンドウサイズが帯域と遅延の積よりも小さい場合には、伝送路の帯域をいかに大きくしてもスループットはウィンドウサイズによって制限されることがわかる。

この式を元に計算した、二つのBSD-UNIX間でのスループットと距離の関係を表1に示す。

表1: 伝送距離とスループット

メディア	距離 (km)	RTT (msec)	スループット (bit/sec)
ファイバ	1000	11.1	2.8 M
ファイバ	3000	33.3	960 K
ファイバ	14000	155.6	206 K
通信衛星	60000	400	80 K

BSD-UNIXのdefault window sizeは4Kbytesである。1000kmの伝送距離では光ファイバによる伝送

の場合でもRTTは11.1msであり、光速度の限界があるため、これよりも小さくすることはできない。この時のスループットは2.8Mbit/secである。さらに伝送距離が日米間の直線距離である14000kmになると、RTTは155.6msであり、このときのスループットは206Kbit/secしか得られない。このことは、TCPを用いた長距離の計算機間通信では、いくら伝送路が高速になっても、ウィンドウサイズが小さければ高スループットが得られないことを示している。

以上の議論から、LFPにおいては、単純にウィンドウサイズを大きくすることによりスループットが向上すると予想された。しかし、従来のTCPアルゴリズムとの不整合によって、スループットが予想以上に低下する領域があることを実験により確認した。この現象については今大会の別の論文[7]を参照していただきたい。

3 おわりに

TCP/IPプロトコルを用いた長距離高速ネットワーク、すなわち長距離高速インターネットでは、これまでのインターネットとは様相が異なることを示した。そして、既存のアルゴリズムとパラメータでは広い帯域に見合ったプロセス間通信が困難であること、その原因が光の速度の限界に由来する伝搬遅延にあることを明らかにした。

参考文献

- [1] 星田昌紀他: 遺伝子情報処理への挑戦, 共立出版, 1994.
- [2] C.G. Omidyar and A. Aldridge: Introduction to SDH/SONET, IEEE Communications, Vol.31, No.9, pp.30-33, 1993
- [3] J. Postel: Transmission Control Protocol, RFC793, 1981
- [4] L. Kleinrock: The Latency/Bandwidth Trade-off in Gigabit Networks, IEEE Communication Magazine, Vol.30, No.4, pp.36-40, 1992
- [5] J. Kay and J. Pasquale: The Importance of Non-Data Touching Processing Overhead in TCP/IP, SIGCOMM'93, pp.259-268, 1993
- [6] 岡敦子, 村上健一郎, 天海良治, 釘本健司: 長距離超高速インターネット(1) 実験概要, 情報処理全国大会(1995, 春).
- [7] 天海良治, 村上健一郎, 岡敦子, 釘本健司: 長距離超高速インターネット(3) 特性解析, 情報処理全国大会(1995, 春).
- [8] 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, 情報処理全国大会(1995, 春).
- [9] 村上健一郎, 天海良治, 釘本健司, 岡敦子, 伊藤正樹, 後藤滋樹, 伊藤光恭: 長距離超高速インターネット, NTT R&D, Vol. 43, No. 9, 1994.