L-029

アプリケーションから見たネットワーク遅延とネットワーク帯域の関係
Relation of network latency and bandwidth from application view

大山義仁 † 猪俣敦夫 ‡
Yoshihito Oyama Atsuo Inomata

概 要

ブロードバンドネットワークの整備が一段落した現在、ネットワーク遅延が問題となりつつある。大容量のデータを伝送する場合にはネットワーク帯域が重要な要素となるが、データ容量が小さい場合にはネットワーク遅延が問題となってくる。ここでいうネットワーク遅延とはデータを転送する時間やパケット生成オーバヘッドなどを含めたデータ送信時間全体のことである。本研究では、科学技術計算などの遅延時間に敏感なアプリケーションに注目して、ネットワーク遅延とネットワーク帯域の関連を数理モデルを用いて解析する。モデルでは転送するデータ容量を主要パラメータみて解析を実施する。さらに実際のネットワークを用いた測定結果とモデルによる解析の結果を比較検討する。本発表では以上の解析結果について報告する。

1. はじめに

昨今、ADSL や光ファイバを用いたブロードバンド通信が広く普及し、ネットワークゲームや P2P によるデータの交換を行う利用者数が増加している。しかし、ブロードバンド通信を用いたアプリケーションが不足しているため、多くの利用者はブロードバンド通信を手軽に利用することができていない。そのため、新しいアプリケーションの研究や開発が急務となっている。

一方で, 科学技術の発展にはコンピュータを用いた科学技術計算が大きな役割を果たしている. 昨今はコンピュータの性能が著しく向上し, ハードウェアが安価になってきたため, 簡単に複雑な科学技術計算が行えるようになった. こうしたことから, より複雑な科学技術計算を簡単に実施する環境の整備が求められている.

このような背景のもとブロードバンドネットワークを活用した分散数値計算プラットフォームの研究開発が行われている. PC クラスターを用いる方法や GRID コンピューティングを用いた方法などが検討されている. こうした研究開発の成果として MPI[1] 等のライブラリーの提供や Globus toolkit による分散コンピューティング環境の整備があげられる. これらライブラリやプラットフォームではブロードバンド通信が必須なため、ブロードバンド通信の新たなアプリケーションとしての期待がある.

科学技術計算を行う場合, 目的の計算が短時間で終了することが常に求められる. 大容量のデータを処理する場合や複雑な方程式の計算では, 計算が終了するまで長時間を要することが頻繁にある. より複雑な現象を解析しようとすれば計算量も膨大となり, 計算時間の短縮化は大きな課題である.

†北海道大学, Hokkaido University
‡独立行政法人科学技術振興機構, JST

計算時間の短縮を目的として分散数値計を行う場合, 通信帯域よりもネットワーク遅延が問題となってくる. ブロードバンド通信環境の整備が整いつつある現在, 今後はネットワーク遅延の縮小に注目したシステムの研究開発が必要になってくる. そこで、科学技術計算という視点から, データ容量と通信帯域、通信オーバヘッドの関係を明らかにすることでデータ転送処理時間の検討を行う.

2. 数理モデル

計算機クラスタにおいてアプリケーションプログラムはネットワークオーバヘッドに敏感であることが報告されている [4]. そこで, ネットワークオーバヘッドに注目し, ネットワークを用いてデータを転送する際に要する時間であるデータ転送処理時間を用いて, ネットワークオーバヘッドと通信帯域の関係をモデル化する. はじめに, 帯域, ネットワークオーバヘッド, データ転送処理時間をそれぞれ次のように定義する [2]. 続いて, モデルを用いて理論的視点からネットワークオーバヘッドと帯域の関係を検討する.

Definition 2.1
帯域 B は, 各デバイスが単位時間当たりに転送できるビット数とし, 単位は bit/sec で表す.

Definition 2.2
ネットワークオーバヘッド oh_{nw} は, ネットワークを経由してデータを送信する時間で, パケット生成オーバーヘッドやインターフェース遅延を含めたメッセージ送信時間全体をいう. 受信の場合も同様に考える.

Definition 2.3
ネットワークによるデータ転送処理時間 T_{nw} を以下のように定義する.

$$T_{\mathrm{nw}}(n) = \frac{8n}{B_{\mathrm{nw}}} + oh_{\mathrm{nw}} \tag{1}$$

ここで, n は処理データサイズ (Unit: [MBytes]), B_{nw} は論理ネットワーク帯域 (Unit: [Mbps]).

Theorem 2.4
２つのネットワークを考える. ネットワーク 1 は, 通信帯域 B_1 とネットワークオーバヘッド oh_1 によって定義される. ネットワーク 2 はネットワーク 1 と比較して, 通信帯域 B_2 が小さく, ネットワークオーバヘッド oh_2 も小さいものとする. すなわち,

$$B_1 > B_2, \quad \text{かつ} \quad oh_1 > oh_2$$

の関係が成り立つものとする.

表 1: 実験環境

	LAN	CAN	Internet
帯域 B	1000Mbps	1000Mbps	1000Mbps
ルータ数	0	3	14
サーバ	XserveG5	onyx300	IA server

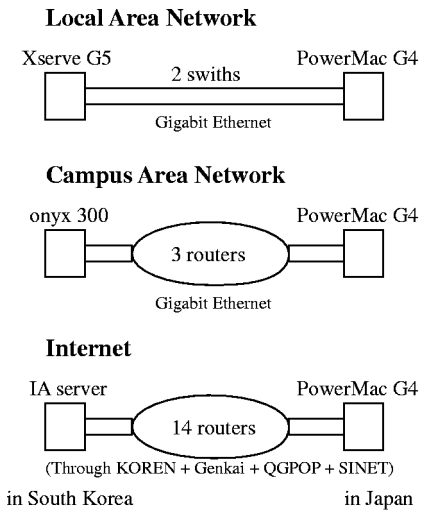

図 1: 実験ネットワーク構成

このとき、ネットワーク 2 を用いた方がネットワーク 1 を用いるよりもデータ転送処理時間が短くなる条件は以下のように表せる．

$$oh_1 - oh_2 \geq 8n\left(\frac{1}{B_2} - \frac{1}{B_1}\right). \qquad (2)$$

式 (2) は，ネットワーク 1 のデータ転送処理時間 T_1 とネットワーク 2 のデータ転送処理時間 T_2 の間の関係 $T_1 \geq T_2$ より求めることができる．

式 (2) において n が小さな場合，オーバヘッドの差 $oh_1 - oh_2$ も小さくなる必要が出てくるため，帯域 B とオーバヘッドのバランスが重要になってくる．

3. 計測実験

この章では式 (1) を確認するために 3 つのネットワーク環境を用いて計測実験を行う．はじめに，実験環境について述べる．実験環境は表 1 の 3 通りを用いる．LAN 環境は研究室内の LAN 環境を用いる． Campus Area Network(CAN) は大学内のネットワークで間に 3 台のルータが入った状態である．Internet は日本と韓国を結んで実験を行う．間にはおよそ 14 台のルータを経由している．すべて端末は PowerMacG4 を使用した (図 1)．

データ転送処理時間を測定するため TCP ストリームを転送データサイズを変化させながら計測を行った．なお TCP/IP はチューニング等は行っていない．

図 2 は測定結果である．インターネットを使用した場合のデータ転送処理時間が大きくなっているため，LAN

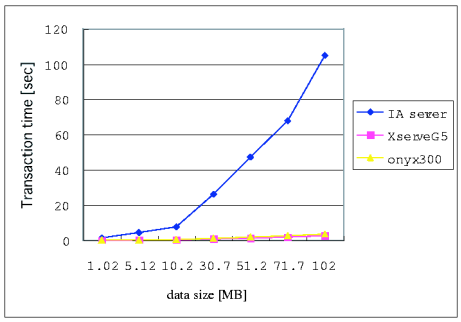

図 2: データ転送処理時間: Internet, Campus area network, LAN

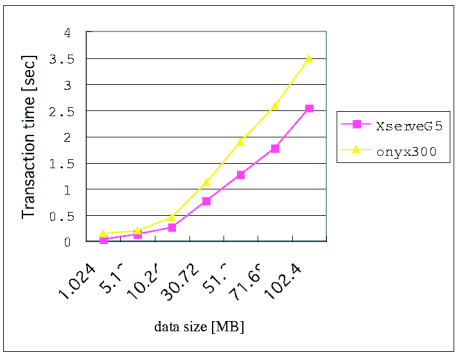

図 3: データ転送処理時間: Campus Area Network, LAN

と Canpus Area Network での結果を図 3 に改めてプロットしている．図 2 と図 3 は同じデータである．図 2 は同じデータをプロットしている．

測定結果を見ると 10[MB] の近傍で接戦の傾きが変化していることが見てとれる．これは計算機のバッファの影響によるものと考えられる． 図 4 は，本実験でのスループットを表示したものである．インターネットを用いた場合のスループットが他の場合と比べて低下していることがこのグラフからも見てとれる．データ容量が小さい場合のスループットが低下しているのは，同様に OS でのバッファによる影響と考えられる．

図 2 と図 3 よりデータサイズに対して線形にデータ転送処理時間が大きくなっていることが分かる．これは式 (1) が実験結果と一致することを示している．そこで各環境下での式 (1) を算出してみるとおよそ以下のようになる．なお n は転送データサイズ [MB] を表すものとする．

$$T_{\mathrm{LAN}}(n) = \begin{cases} 0.0245n + 0.00952 & n < 10 \\ 0.0247n + 0.00747 & n > 10 \end{cases}$$

$$T_{\mathrm{CAN}}(n) = \begin{cases} 0.0331n + 0.0294 & n < 10 \\ 0.0327n + 0.0220 & n > 10 \end{cases}$$

$$T_{\mathrm{Internet}}(n) = \begin{cases} 0.6886n + 0.6580 & n < 10 \\ 1.0150n + 2.6844 & n > 10 \end{cases}$$

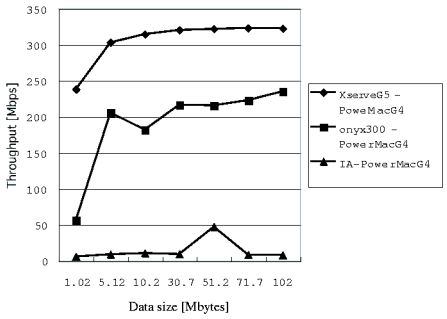

図 4: Throughput

この結果から, 転送データサイズが小さい場合は, ネットワークオーバヘッドの影響が大きくなることが分かる. 小容量のデータを頻繁に転送し合うようなアプリケーションでは注意が必要である. 帯域を大きくするだけではなくネットワークオーバヘッドを小さくする必要がある.

4. まとめ

計測実験の結果、データ転送処理時間は 1 次式で適切に表現することが分かった. それをもとに数理モデルを構築した. 構築したモデルを用いて理論解析を行ってみると, 転送データ容量が小さい場合は, ネットワーク帯域が大きくなれば良いのではなく, 通信オーバヘッドも考慮しないとデータ転送処理時間は短くならないことが分かった. 科学技術計算のようにプロセス処理時間の短縮化が望まれるアプリケーションにおいては、ネットワーク帯域と通信オーバーヘッドのバランスが重要になってくることが分かった.

参考文献

[1] Message Passing Interface(MPI), http://www.mpi-forum.org/

[2] B. Wilkinson, M. Allen, "Parallel Programming: Techniques and Applications Using Networked Workstations and Parallel Computers", *Prentice-Hall Inc., (1999)*

[3] K. Ranganathan, I. Foster, "Simulation Studies of Computation and Data Scheduling Algorithms for Data Grid", *J. Grid Computing 1*, p53-62, 2003

[4] Martin, R. P., A. M. Vahdat, D. E. Culler, and T. E. Anderson, "Effects of Communication latency, Overhead, and Bandwidth in a Cluster Architecture", *Proc 24th Ann. Int. Symp. Comput. Arch.*, ACM, pp.85-97.