2K-8

Ships1におけるネットワーク性能の測定と評価

岡本恵介† 松原裕人† 大谷真†
湘南工科大学情報工学科†

1. はじめに

Ships1(Shonan Institute of Technology Parallel System 1)は、安価な複数の CPU を繋げたコストパフォーマンスに優れる並列計算機である。一般に複数の CPU でネットワークを構成するには、性能と機能の面からどのようなネットワーク構成が可能であり、最適であるかを考えなければならない。Ships1 においても同様に最適な構成を見つけ出す必要がある。

本研究では、様々な条件でネットワークの実効性能を測定し、評価を行った。さらに測定結果を元に Ships1 におけるハブなどのハードウェア構成の最適化の検討を行った。

2. Ships1 のネットワーク構成

Ships1 の各ノードと管理コンピュータは 2 種類の内部的な LAN で接続される。一つは全ノードを接続し、並列処理のためのデータ交換に使用する主ネットワークである。もう一つは各ノードを管理・制御するために、各ノードと管理用のマシンを接続する管理ネットワークである。両ネットワークとも 1Gbps の LAN とである。図 1 に Ships1 の論理ネットワーク構成を示す。

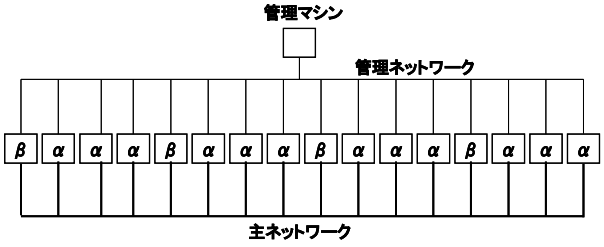

図 1. Ships1 の論理ネットワーク構成

実際のネットワークの構築は、スイッチングハブを用いて実現する。

本研究では、特に性能が要求される主ネットワークについて、OS によるオーバヘッドを考慮したアプリケーション間の実効性能や、スイッチングハブの構成やスイッチングによる遅延を考慮した場合の実効性能の測定を行った。さらに、測定結果から Ships1 におけるスイッチングハブを用いた最適なネットワーク構成の検討を行った。

Measurement of Network Performance in Ships1
Keisuke Okamoto, Hiroto Matsubara and Makoto Oya
Department of Information Science, Shonan Institute of Technology

3. ネットワークの実効性能

ネットワークでは様々な要因により理論通りの性能が発揮されない。その原因には大きくわけてソフトウェアによる要因とハードウェアによる要因がある。ソフトウェアによる要因としては、OS のオーバヘッドによる遅延がある。Ships1 においてもデータ交換の際に、OS(Linux2.6)のオーバヘッドによる遅延が発生し、理論通りの性能が発揮されない。

ハードウェアによる要因としては、スイッチングハブのスイッチング機能やパケットの衝突によって生じる遅延がある。データを送る際、あて先を決定するためにハブのスイッチング機能が用いられる。Ships1 における並列処理のためのデータ交換では何台ものノードからデータが送信され、ハブのスイッチングやパケットの衝突が頻繁に発生する。このスイッチングによって遅延が発生し理論通りの性能が発揮されない。よって、ハブを 1 台でネットワークを構築する場合、ハブにかかる負荷が増加し実効性能が低下するが、ハブを複数台用いてネットワークを構築することで 1 台のハブにかかる負荷が減少し、実効性能が向上すると考えられる。

ネットワークの性能を測定するためには、これらの様々な要因を考慮して測定を行う必要がある。Ships1 においても最も性能を引き出すことができるネットワーク構成を見つけ出すためには、OS のオーバヘッドによる遅延や、ハブの構成とスイッチングやパケットの衝突による遅延を含んだアプリケーション間の実効性能の測定を行わなければならない。

4. 測定方法

測定のために、異なるマシン上の 2 つのアプリケーション間で一定の長さのデータを TCP/IP を使って転送するプログラムを作成した。

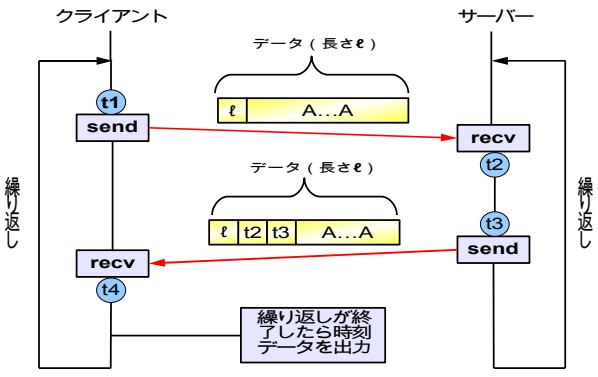

図 2. 測定プログラムの動作図

次の点に留意する必要があった。
・2つのマシンの時計は完全には同期していない。したがって一方の時刻と他方の時刻の差分をとるのでは正確な転送時間は測定できない。
・ハブのスイッチング時間を含む測定にはマイクロ秒より小さい粒度の測定が必要である。

1つ目の点については、1方向の転送ではなくラウンドトリップ時間（往復時間）をクライアント側の時計を使って計測する方法とした。これを実現するためにサーバ内部で発生する測定範囲外オーバヘッドを減算するため、オーバヘッド部分の開始時刻と終了時刻を戻り電文内に含める方式にした。2つ目については、繰り返し測定を行いその平均をとるとともに、Linux2.6 の gettimeofday()を使いマイクロ秒精度で時刻取得を行うようにした。

図2に作成したプログラムによる測定の様子を示す。クライアント側では送信直前の時刻 t1 と受信直後の時刻 t4 をクライアントマシンの時計を使って計測する。サーバ側ではデータを受取った直後の時刻 t2 とデータ送信の直前の時刻 t3 をサーバマシンの時計で計測する。1回のラウンドトリップ時間 T を

T ＝ t4 － t1 －（t3 － t2）

で求める。これを繰り返し、T の平均値を求める。

また、1:1 ではなく、複数の測定を異なるマシン間で同時に行う場合には、測定開始時刻にばらつきが発生するため、t2 または t3 値を使って同時に動作している区間の値だけを計測値とすることにした。このために、t1，t2，t3，t4 は一旦ファイルに出力し、計測期間の決定と平均値の計算は別の解析プログラムで行うようにした。

測定プログラムはクライアントとサーバを 1:1 で接続するため、測定では1つのノードで複数の測定プログラムを動かす。測定の構成は 1:1，1:n，n:n でノードを接続したものとする。測定の論理構成を図3に示す。1:n と n:n の実際の測定では、スイッチングハブを組み込んで図3に示す構成を実現する。

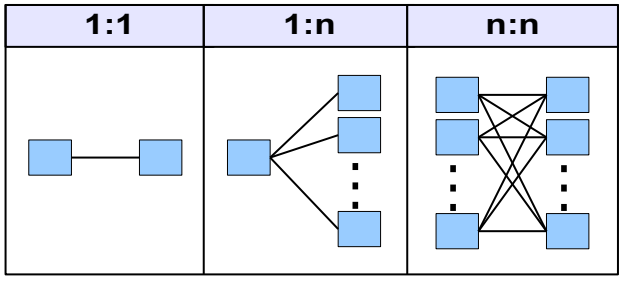

図3. 測定の論理構成

5. 測定の構成

1:1 の測定では、主に考慮すべきは OS のオーバヘッドによる遅延のみである。よって測定の物理構成はハブを組み込まずにノード間を直接 LAN ケーブルで接続する。1:1 での測定からは、OS のオーバヘッドによる遅延だけを考慮した実効性能と理論値である 1Gbps との差がどれだけあるかがわかる。

1:n の測定では、ノードを接続するのにスイッチングハブを用いるため、OS のオーバヘッドに加えてハブのスイッチングやパケットの衝突による遅延も考慮して測定を行う必要がある。よって、様々なハブの構成で測定を行い実効性能の高いハブの構成を調べることにした。測定を行う物理構成の例を図4に示す。図4に示す構成以外にノードの数やハブの数を変えた構成についても測定を行う。

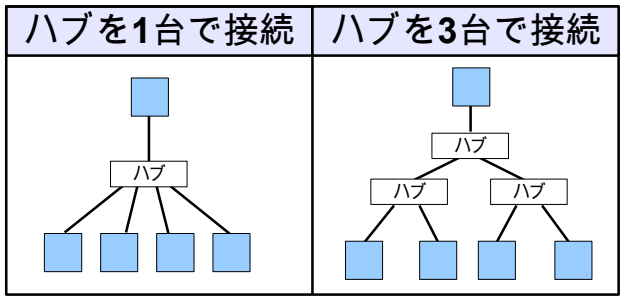

図4. 1:n での測定の物理構成例

n:n の測定では、1:n の測定と同様に OS のオーバヘッドとハブのスイッチングやパケットの衝突による遅延を考慮して測定を行う必要がある。よって、n:n の測定に関しても様々なハブの構成で測定を行う。測定を行う物理構成の例を図5に示す。図5に示す構成以外にノードの数やハブの数を変えた構成についても測定を行う。

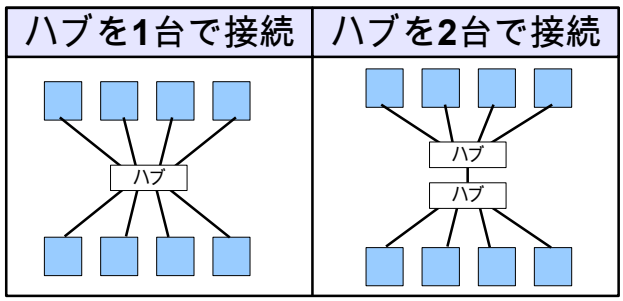

図5. n:n での測定の物理構成例

測定結果から、スイッチングハブをどのような構成にした時に最も高い実効性能が発揮されるかを見つけ出し Ships1 における最適なネットワーク構築を行う。

6. まとめ

1:1 での測定の結果、スループットは 823Mbps であった。理論値である 1Gbps と比べて 82.3%の実効性能が発揮されている。今後は、さらに測定を進め Ships1 における最適なネットワーク構成を見つけ出し、報告する予定である。

参考文献

[1]大谷真，松原裕人，櫻井一欽，加藤悠，中小規模並列コンピュータ Ships1 の開発，湘南工科大学紀要，Vol.41，No.1，2007
[2]松尾成志，岡本恵介，大谷真，中小規模並列コンピュータ Ships1 の開発，情報処理学会第 69 回全国大会，予稿集，2007