

GRID でのネットワークスループット予測

秋岡明香 村岡洋一
早稲田大学大学院 理工学研究科

世界中に散在する遊休計算機を活用する GRID に関する研究が近年盛んに行なわれているが、これらの計算機資源を有効活用するには負荷分散を行なうことが必須である。しかし GRID では通信遅延が大きく、計算機やネットワークの負荷状況が時々刻々と変化するため、適切な負荷分散のためにはこれらの負荷の変化を予測する必要がある。そこで本稿では、GRID 上でネットワークスループットを予測する手法を提案する。提案手法の特徴は、複数の予測アルゴリズムから次回最も良い予測値を算出すると期待される手法を隠れマルコフモデルにより選択し、ネットワークスループットの予測値を算出する点である。また提案手法を実環境にて性能評価を行なった結果を報告する。

Predicting Network Throughput on Computational GRID

SAYAKA AKIOKA AND YOICHI MURAOKA
LargeGraduate School of Science and Engineering,
Waseda University

In this paper, we propose an algorithm that predicts network throughput on Computational GRID. Recently, there are many studies on Computational GRID. Computational GRID means a big cluster that consists of many computers scattered all over the world. In order to make these computational resources work well, load distribution or load balancing is required. On GRID, however, cpu load and network status are always changing. Moreover, delay in communication is serious. That is why we need to predict network throughput. Our algorithm is characterized by the point that the most promising predicting method is chosen among several ones using Hidden Markov Model. In addition, our algorithm doesn't disturb other applications in their work.

1 はじめに

近年、世界中に散在する計算機をネットワーク接続し、ひとつの巨大なクラスタとして見なすことで遊休計算機を有効活用しようとする GRID に関する研究が盛んに行なわれている。これらの計算機資源を有効活用するためには、タスクを最適に配置する負荷分散システムが必要となる。負荷分散システムがタスク配置を決定する際には、各計算機の負荷情報や計算機間

のネットワークの状況を知ることが必要だが、GRID では通信遅延が大きい上に、計算機負荷やネットワークの状況は時々刻々と変化している。従って、有効な負荷分散を実現するためには、計算機負荷やネットワーク状況を予測し、これらの予測値に基づいて負荷分散を行なう必要がある。そこで本稿では、ネットワークスループットを予測するアルゴリズムを提案する。また、提案アルゴリズムを用いて実際にネットワークスループットをリアルタイムに予測する

実験を行なった結果を報告する。

以下では、第2章で本研究の目的を明らかにし、第3章で関連研究の紹介を行なう。第4章で提案アルゴリズムの紹介を行ない、第5章で提案アルゴリズムの評価実験とその結果に対する考察を行なう。最後に第6章で本稿のまとめと今後の課題を提示する。

2 本研究の目的

ネットワークの動向を把握・予測する研究は多く行なわれてきたが、GRIDでの負荷分散を補助することを目的とした手法は少ない。予測を負荷分散に役立てるためには、予測手法は以下のような要求に答える必要がある。

1. 予測値の算出に膨大な時間を費さない。
2. 予測値を算出するために、cpu やネットワークを占有して他のアプリケーションの実行を妨げることがない。
3. 予測値の算出に膨大な過去データを必要としない。
4. 複数ネットワークを比較検討することが可能な精度で予測値を提供する。

1., 2., 3. については、予測値計算サーバのような専門サーバを設置する場合には問題にならないが、専門サーバを設置することで拡張性が損なわれる恐れがある。一方 1., 2., 3. を満たし、専門サーバを必要としない分散型システムでの運用に耐える手法であれば、拡張性を損なうことはない。

そこで本稿では、これらの条件を満たしつつ負荷分散システム等にネットワークスループットの予測値を提供する、ネットワークスループット予測アルゴリズムの開発とそれを実装したシステムの構築を目指す。

3 従来研究

本研究と同様な、GRIDでの負荷分散を援助するためのネットワークスループット予測研究

の代表例としては、Network Weather Service (NWS)[1]を挙げることができる。NWSでは、過去の測定値の平均値及び中間値、自己回帰モデルの各手法により常に3種類の予測値を求めておき、前回最も誤差の小さい予測値を与えた手法により今回得られた値を、予測値として採用する方法をとっている。

また、Schopfらは正確な値を予測することが困難である点に注目し、上限値と下限値を持った幅のある予測値を提供することで、運用上の予測の精度を高めることを試みている[2]。

4 提案手法

以下では、提案アルゴリズムについて述べる。提案手法の特徴は、複数の予測手法から、隠れマルコフモデルにより次回最も有効であると期待される手法を選択し、その手法により算出される予測値を正式な予測値として採用する点である。隠れマルコフモデルは時間の経過と共に最適化されるため、提供する予測値の精度が向上することが期待できる。また、複数の性質の異なる手法を常に比較することにより、様々なネットワークの状況変化に柔軟に対応することができる。

提案アルゴリズムの大まかな処理の流れを図1に示す。提案手法では、常に複数の予測手法によって複数の予測値を求める。これらのうち、次回最も良い予測値を得ることが期待される手法を隠れマルコフモデルにより選択し、その手法により求められた予測値を正式な予測値として採用する。その後、計測値が得られた際に、実際にはどの手法が最も良い予測値を得たかを調べ、その結果に基づいて隠れマルコフモデルの確率分布を再計算する。次回予測は、この新しい確率分布に基づいた隠れマルコフモデルを用いて行なわれる。

ネットワークスループットの指標としては様々なものが考えられるが、今回はSimple Network Management Protocol (SNMP)[3]により遠隔地にある計算機からcpu負荷の1分間の平均値を取得するのに要した時間を用いた。将来的に、計算機負荷情報の収集や予測を行

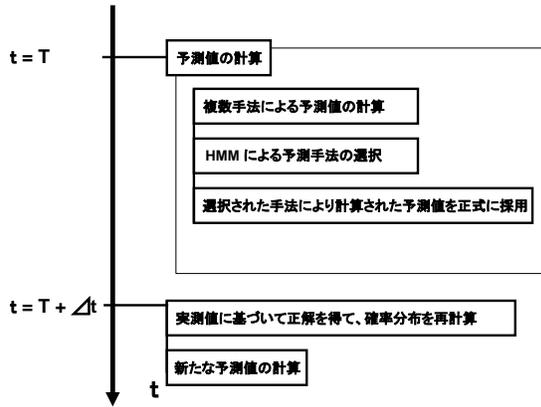


図 1: 提案アルゴリズムの処理の流れ

なうことを考慮した場合、ネットワークスループットの指標をこのように定めることで、ネットワークスループット測定のための無駄なトラフィックを発生させる必要がなくなる。

4.1 予測手法

今回、予測値の計算手法として採用したのは以下の 6 手法である。本稿では、他のアプリケーションの実行を妨げるのを避けるため、各予測手法は実行時間が短く、異なる特徴を持つ手法を中心に選んだ。以下では、時刻 t におけるネットワークスループットの計測値を $v(t)$ 、時刻 t におけるネットワークスループットの予測値を $p(t)$ とする。

DV 直前の測定値を予測値とする手法。ネットワークスループットが大きく変化せず、安定した状況に強い。

$$p(t+1) = v(t)$$

GD 前回と前々回の測定値から直線による近似を行ない予測値を求める手法。ネットワークスループットが単調に増加、もしくは減少する場合に有効である。

$$p(t+1) = (v(t) - v(t-1)) + v(t)$$

SG 統計的最急勾配法により予測値を求める手法。TCP/IP には、本手法によるエン

ドノード間のラウンドトリップタイムの予測が組み込まれているため、提案手法でも採用した。なお、今回は $g = 0.5$ とした。

$$p(t+1) = (1-g) * p(t) + g * v(t)$$

$$(0 < g < 1)$$

SW 過去の測定値の平均値を予測値とする手法。最近数件の計測値の平均値を予測値とすることで、最近の傾向を予測値に反映させることができる。また、直前値をそのまま予測値として出力するよりも、広い範囲に渡る傾向が反映できる。以下では i を $err_i(t)$ を最小にする値とする。今回は $1 \leq K(t) \leq 10$ とした。

$$p(t+1) = SW(t, K(t))$$

$$SW(t, K(t)) = \frac{1}{K(t)+1} \sum_{i=t-K(t)}^t v(i)$$

$$K(t+1) = K(t) + i$$

$$err_i(t) = (v(t) - SW(t, K(t) + i))^2$$

MF メディアンフィルタにより予測値を求める手法。本手法により最近の測定値の中間値を選ぶことで、一時的な急変に引きずられて大きく狂った予測値を出力することを防ぐことができる。以下では、 S_K は最近 K 個の測定値をソートした数列を表わし、 $S_K(l)$ は数列 S_K の l 番目の要素を表わす。ただし i は $err_i(t)$ を最小にする値とする。

$$p(t+1) = MF(t, K(t))$$

$$MF(t, K) = \begin{cases} S_K((K+1)/2) & (K : \text{odd}) \\ \frac{S_K(K/2) + S_K(K/2+1)}{2} & (K : \text{even}) \end{cases}$$

$$K(t+1) = K(t) + i$$

$$err_i(t) = (v(t) - MF(t, K(t) + i))^2$$

AR 自己回帰モデルにより予測値を求める手法。近年の研究により，ネットワークの状況は ARIMA モデルにより表現できることが知られている [5][6]。しかし，ARIMA による予測には非線型方程式を解く必要があるため，今回は解くのが容易な自己回帰モデルを採用した。時刻 t における M 次自己回帰モデルは次式により表現できる。今回は最大次数を 10 とした。

$$v(t) = \sum_{m=1}^M a(m)v(s-m) + err(t)$$

$$1 - a(z) - a(z^2) - \dots - a(M)z^M = 0$$

4.2 隠れマルコフモデル

提案アルゴリズムで用いた隠れマルコフモデルを図 2 に示す。図 2 で $S1$ は前回と前々回で最適な予測値を与えた手法が同じである状態， $S2$ は前回と前々回で最適な予測値を与えた手法が異なる状態をそれぞれ表わし， $P(a|b)$ は状態 a から状態 b へ遷移する状態遷移確率を表わす。 $P(m|a \rightarrow b)$ は状態が a から b へ遷移した際に m を出力する確率を表わす。 m は予測手法を表わす。

予測値を決定する際には，現在の状態からどの状態へ遷移し，何を出力する確率が最も高いかを隠れマルコフモデルから計算し，次回最も良い予測値を与えることが期待できる手法を選択する。

予測後は，複数手法から求めた予測値と，それに対応する測定値を比較し，最も良い予測値を与えた予測手法を調べ，その結果に基づいて図 2 の隠れマルコフモデルの確率分布を計算し直す。この作業を繰り返すことによって，隠れマルコフモデルが最適化され，予測精度が向上することが期待できる。

5 性能評価

提案アルゴリズムを C++ で実装し，予測精度と運用面からの性能評価を行なった。なお，SNMP ライブラリには ucd-snmp[7] を用いた。

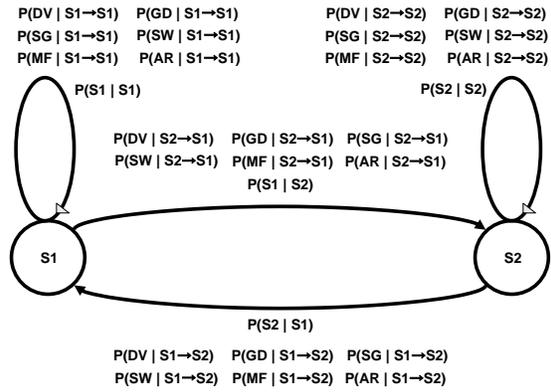


図 2: 隠れマルコフモデル

5.1 予測精度の評価

表 1 の環境で提案手法を用いてネットワークスループットの予測実験を行なった。実験は 2001 年 10 月 4 日午前 2 時 30 分から午前 7 時 30 分の間に行ない，SNMP プロトコルによる遠隔計算機の負荷情報の取得は約 1 分間隔で行なった。2 つの計算機はスイッチングハブを介して 100BASE-TX で接続されている。実験結果を図 3，図 4 に示す。なお，予測誤差の測定値に対する割合が 10% 以下である場合は，実験結果全体の 79.9% であった。また，各測定値について最も近い予測値を与えた手法を調べ，手法ごとの出現回数についてまとめた結果を図 5 に示す。この結果から，ネットワークスループットの予測は単一の手法で実現することは非常に困難であり，複数手法の組み合わせが予測精度の向上に繋がる可能性が高いことが分かる。

予測誤差の測定値に対する割合が 10% 以下である場合が実験結果全体の 79.9% に留まった原因として最も大きいと思われるのは，ネットワークスループットが大きく変動する場合の予測精度の悪さである。今回採用した予測手法の他に，大きな変動に強い特性を持つ予測手法を追加することで精度は改善できると考えられる。

また，実測値を用いた実験 [8] に対して精度が低かった原因として，学習期間の短さが挙げられる。[8] では約 30 秒ごとに 24 時間に渡って実験を行なったのに対し，今回は予測回数が 300 回程度と極端に少ない。いずれも実測値を

表 1: 提案アルゴリズムの予測精度の評価環境

Host A	Intel Celeron 800MHz 512MB Linux (kernel 2.4.2)
Host B	Intel Pentium 4 1.7GHz 512MB Linux (kernel 2.4.2)

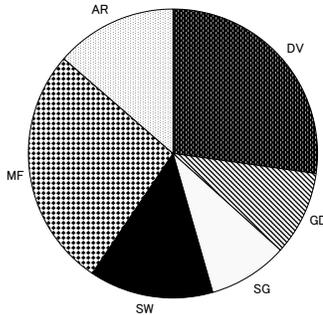


図 5: 各予測手法の出現率

用いた実験であることから、今回の実験では隠れマルコフモデルの最適化が十分に行なわれないうちに実験が終了してしまった可能性がある。

5.2 運用面からの評価

SNMP プロトコルによる遠隔計算機への負荷問い合わせから隠れマルコフモデルの確率分布の再計算までの、予測に関する一連の作業に要する実行時間の測定を行なった。実験には Intel Celeron 800MHz, Memory 512MB, Linux kernel 2.4.2 の計算機を用いた。一連の予測作業を 10000 回繰り返して 1 回当たりの予測に要する実行時間を求めたところ、一回当たり 1.42 ミリ秒であることが分かった。この結果より、提案アルゴリズムは非常に短時間で処理を終えることができ、分散型システムのような他のアプリケーションと同時に実行する必要がある場合にも、他のアプリケーションの実行を妨げることなく予測値の計算が行なえることが分かった。

6 まとめと今後の課題

本稿では、GRID においてネットワークスループットを予測するアルゴリズムを提案し、提案アルゴリズムを実装、評価実験を行なった結果を示した。提案アルゴリズムの特徴は、複数の予測手法から隠れマルコフモデルによって次回最も最適な予測値を算出すると期待できる手法を選択し、予測値を求める点である。

評価実験では、予測誤差が測定値に対して 10%以下であった場合が全体の 79.9%を占めており、隠れマルコフモデルを十分に学習させることや、特徴の異なる他の予測手法を追加することで、更なる精度向上が期待できることが分かった。また、提案アルゴリズムは実行に要するコストが非常に低く、拡張性の高い分散型システムに適用しても、他のアプリケーションを邪魔することなく予測値を提供することができることも分かった。これにより、提案アルゴリズムは非常に広く活用できると言える。

今後は、更なる精度の向上の他、cpu 負荷の予測への応用やネットワークポロジ情報の追加などに取り組んでいく予定である。また、提案アルゴリズムを分散型システムで実現する場合を考慮して、計算機間での効率の良い通信方法についても研究していきたい。

参考文献

- [1] Rich Wolski, Neil T. Spring, and Jim Hayes. The Network Weather Service: A Distributed Resource Performance Forecasting Service for Metacomputing. The Journal of Future Generation Computing Systems, 1999.
- [2] Jennifer M. Schopf and Francine Berman. Using Stochastic Information to Predict Application Behavior on Contended Resources. International Journal of Foundations of Computer Science, Special Issue on Parallel Distributed Computing, June 2001.

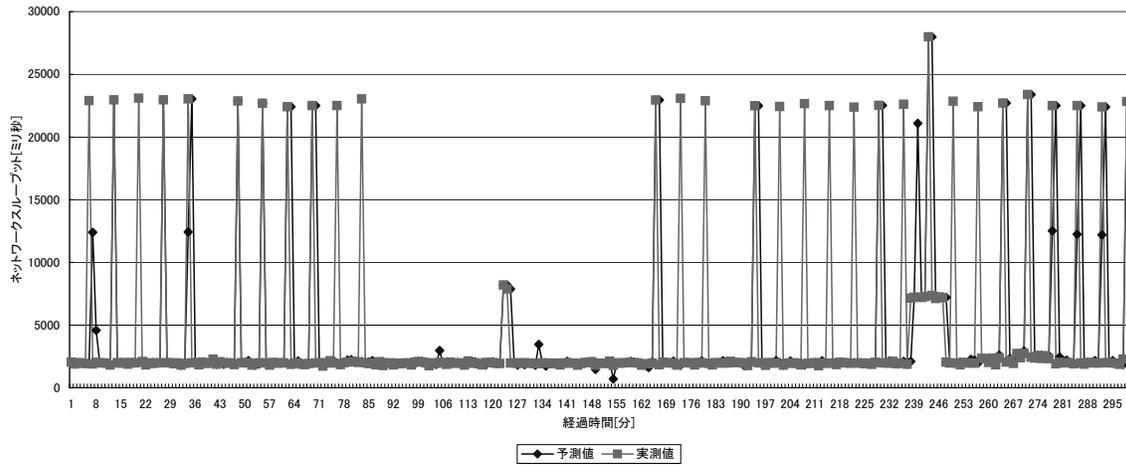


図 3: 予測値と測定値の比較

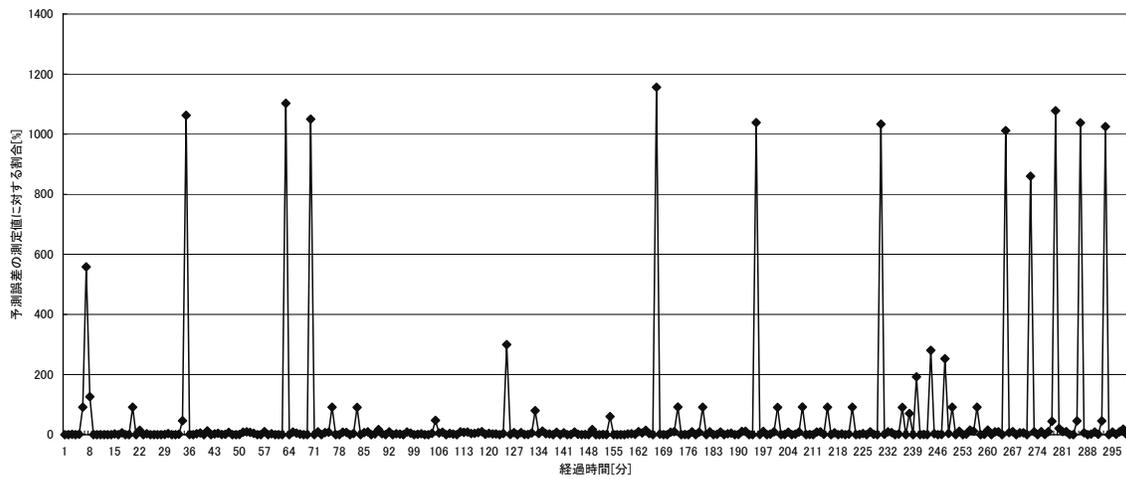


図 4: 予測誤差の測定値に対する割合

- [3] A Simple Network Management Protocol(SNMP). ARPA Working Group Requests for Comment DDN Network Information Center, SRL International, Melno Park, CA, RFC1157.
- [4] Transmission control protocol specification, 1981. ARPA Working Group Requests for Comment DDN Network Information Center, SRL International, Menlo Park, CA, RFC-793.
- [5] S. Basu, A. Mukherjee, and S. Kilvan-sky. Time series models for internet traffic. Technical Report GIT-CC-95-27, Georgia Institute of Technology, 1996.
- [6] N. Groschwitz and G. Polyzos. A time series model of long-term traffic on the nsfnets backbone. In Proceedings of the IEEE International Conference on Communications (ICC '94), May 1994.
- [7] ucd-snmp. <http://net-snmp.sourceforge.net/>.
- [8] 秋岡明香, 村岡洋一. GRID における負荷分散のためのネットワークスループット予測. 第 63 回全国大会, 情報処理学会, 2001.