

計算機クラスタ用通信解析ツールの開発

3 Z B - 0 4

菊池 康祐、 緑川 博子、 飯塚 肇
成蹊大学大学院工学研究科情報処理専攻

1. はじめに

最近では計算機クラスタを用いた並列処理が盛んに行われるようになってきた。このような計算機クラスタでの並列プログラム開発では、性能向上のために、各計算機間の通信パターンを解析し通信量を下げる工夫をすることが重要にである。しかし、従来のネットワーク監視ソフトウェアは、長時間の定常的トラフィック監視を主体としており、計算機クラスタの並列処理解析のために、短い実行時間内の詳細なデータとるには不向きで、並列処理解析用の通信解析ツールが望まれている。

このような用途に適した計算機クラスタ向き通信解析ツールを設計し、実装を行ったので、報告する。

2. 特徴

・本ツールは、並列処理実行時における本来の通信パターンを乱さないために、解析ログを各計算機で採っておき、並列処理実行後、すべての計算機のログファイルを使ってオフラインで解析を行うことを前提としている。ログファイルとしては、多くの UNIX 系 OS で使用できる `tcpdump` の出力ログとした。複数のログファイルの時刻は、各計算機の時間のずれにも対応できるように、基準計算機 (最初にプログラムを起動する計算機) とそれぞれの計算機との最初の通信を対応づけることにより、相対時間に統一してすべてのログファイルのデータを一つのファイルに統合している。

・OS による `tcpdump` の出力形式の違いに対応できるように、出力形式の違いを記述した OS 定義ファイルにより、ログファイルの中から規定の要素のみをとりだし、様々な OS でのログファイル結果を解析できるようにしている。

・複数計算機間の通信パターンをできるだけ視覚的に解析できるように、通信の元のデータや統計量などを表形式に出力するだけでなく、ユーザが指定した様々な指標で即座に結果をグラフ化すること、さらに通信パターンの時系列変化を直感的に理解できるような時系列通信再生モニタなどを提供する。

・本ツールは現在、LINUX 上で稼働するが、UNIX 系 OS への移植が可能である。

3. 機能

本ツールは大きく分けて 3 種類のウインドウから構成され、以下のような機能を備えている。

(1) 表形式ウインドウ

・通信量に関する情報

ユーザが指定した単位時間で送受信メッセージ数と通信バイト数を、時系列に各計算機ごとに表示する。また、それらの統計量を各計算機ごとに表示する。

・メッセージ長に関する情報

ユーザが指定したメッセージ幅で、送受信メッセージ長の度数分布を、各計算機ごとに表示する。また、それらの統計量を各計算機ごとに表示する。

(2) グラフウインドウ

・(1)の情報を棒グラフで表示する。図 2 は、縦軸に通信バイト数、横軸に単位時間を指定した時系列グラフである。

Communication Traffic Analyzer for Computer Cluster

Kosuke Kikuchi, Hiroko Midorikawa, Hajime Iizuka

Department of Information Sciences, Seikei University

(3) 通信モニタウインドウ

並列プログラム実行時の時系列通信パターンを視覚的に理解しやすくするために、各プロセッサ間の通信量（バイト）を濃淡に対応させて2次元表示し、再現できる。また、時系列通信パターンを連続再生できるだけでなく、ステップ再生、巻き戻し、一時停止などの機能を備えており、さらに細かい通信解析が行える。

(4) CSV形式出力

解析を行ったデータをCSV形式で出力する。それによって、解析データを汎用の表計算ソフトウェアに容易に利用できる。

4. ツールの評価

本システムの有効性を評価するため、表1に示す計算機クラスタにおいて、数種類の並列プログラムを動かし、ログファイルの解析を行った。単位時間は0.05秒である。

- PC 8台
- OS TurboLinuxServer6.1
- Network 1Gbps EtherNet

表1 計測環境

図1から図3は、解析を行った一例である。図1は、各計算機ごとにユーザが指定した単位時間あたりの送受信メッセージ数を示している。この図から実行時間の最初の部分は、どの計算機でもほとんど通信が行われていないことが分かる。

Message Number	gpc01	gpc02	gpc03	gpc04	gpc05	gpc06	gpc07	gpc08
Unit 1	0	2	0	0	1	0	0	0
Unit 2	0	1	0	0	0	0	0	0
Unit 3	7	4	0	1	0	1	0	1
Unit 4	50	83	26	16	18	29	11	15
Unit 5	147	147	144	147	146	151	153	154
Unit 6	157	158	163	162	155	159	153	152
Unit 7	162	163	152	156	159	158	160	162
Unit 8	164	160	159	163	163	164	162	162
Unit 9	160	162	162	167	167	163	165	162
Unit 10	159	158	155	154	155	159	152	155
Unit 11	158	162	163	161	163	161	164	161
Unit 12	165	167	170	167	162	163	167	168
Unit 13	163	160	161	160	163	165	164	163
Unit 14	165	164	164	162	165	168	164	163
Unit 15	163	164	163	164	160	155	164	164
Unit 16	157	155	156	156	160	161	156	160
Unit 17	159	162	160	160	163	161	159	157
Unit 18	163	163	163	162	159	159	159	159
Unit 19	161	162	152	155	164	162	163	164
Unit 20	160	157	157	158	164	161	165	163
Unit 21	161	164	163	163	160	160	162	162
Unit 22	162	162	163	162	162	160	156	158
Unit 23	165	161	161	161	164	165	166	166
Unit 24	159	160	161	162	157	156	154	154
Unit 25	167	166	166	166	169	156	169	169
Unit 26	140	163	140	124	117	156	151	167

図1

図2は、ある計算機の単位時間あたりの通信パ

イト数を時系列グラフで示している。グラフのバーの上の部分は送信バイト数を示しており、バーの下部分は受信バイト数を示している。このグラフから、送信バイト数と受信バイト数がほとんど等しいことが分かり、通信がループ回数に対応して定期的起こっていることが分かる。

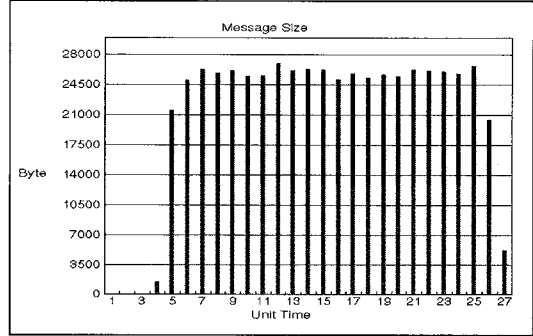


図2

図3は、通信モニタの表示例である。単位時間あたりの各計算機間の通信バイト数を色の濃淡で示している。縦軸が受信側計算機で横軸が送信側計算機である。この図からは単位時間11においてgpc31からgpc34の間やgpc32からgpc33の間に通信が集中していることが分かる。

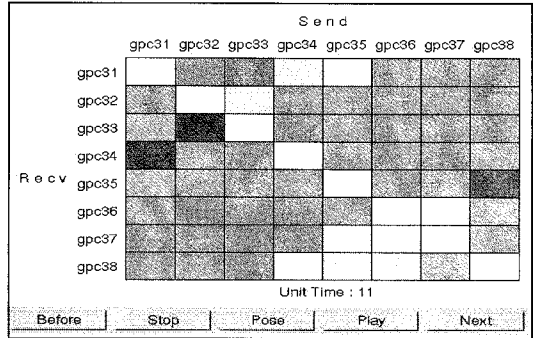


図3

5. おわりに

本研究では、計算機クラスタ向き通信解析ツールを設計し、実装を行った。その結果、並列処理プログラムの解析が容易になった。今後、ツールを使用し解析を行っていく上で必要な機能があれば追加していきたい。