

## 安価なコンピュータを用いた実験・教育用並列計算機環境の構築

†信田 圭哉 †長谷川 明生

†中京大学情報科学研究科

## 1 はじめに

ハイパフォーマンスコンピューティングの分野において、MPIを用いた分散メモリ型の並列計算機に対する注目が高まっている。教育や簡単な実験目的での市販のスーパーコンピュータの利用には、金銭的成本、管理コストの問題がある。

Coxらは、一般的なパーソナルコンピュータ(以下、PC)よりもはるかに安価なRaspberry Piを使用し、低コストなコンピュータクラスタが構築できることを示した[1]。Coxらは、64台のRaspberry Piをネットワーク接続し、その上でMPIを使用した並列計算機を実現している。

本研究では、Coxらの手法に基づき24台のRaspberry Piを用いてMPIの並列計算機を構築した。その上で、並列計算環境の構築の一部を自動化して管理コストを減らし、コストの面と扱いやすさの点から初学者向けの環境構築を目指す。

## 2 Raspberry Piを使用したクラスタの構築

構築したクラスタの写真を図1に示す。図に示すように、24台のRaspberry Piを、アクリル板を加工して作成した箱に並べた。

Raspberry Piはクレジットカード程度の大きさのコンピュータであり、Raspberry Pi Model Bは700MHzのARM1176プロセッサをコアに、512MBのRAMを搭載している。

Raspberry Pi上ではLinuxベースのOSが動作し、市販のPCと同様に使える。そのため、Beowulf型と呼ばれるコンピュータクラスタの実験に利用できる。

一般的なPCは専用の並列計算機に比べ安価であるが、1台数万から数十万円はかかるPCを複数台揃えるには、金銭的成本が問題になる。教育や簡単な実験を行う目的であれば、金銭的成本はできるだけ低くしたい。Raspberry Piは日本で購入する場合1台4500円程で、一般的なPCと比較し安価であるため、金銭的

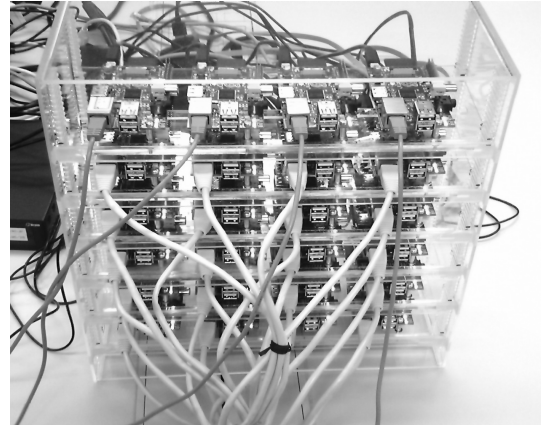


図1 Raspberry Pi 24台を使用したクラスタ

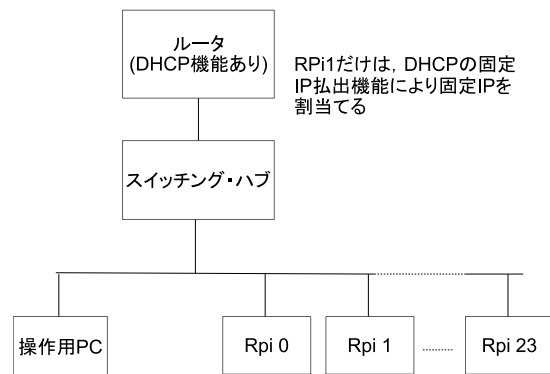


図2 ネットワーク構成

コストを低くできる。

ストレージにはSDカードを使用し、Raspberry Pi用にビルドされたOSのイメージをSDカードに書き込むことで使用できる。OSにはRaspbian wheezy(バージョン2013-02-09)を使用した。これはRaspberry Pi FoundationのWebページ[2]からダウンロードできる。電力供給はmicroUSBケーブルを使用し、消費電力は $5V \times 700mA = 3.5W$ 程であり、電力の面でも低コストであることが期待できる。10/100 Base-TのEthernetソケットがあり、ネットワーク接続に使用する。

ネットワーク構成の概要を図2に示す。クラスタを構成するのはRaspberry Piのみで、図にはRpi  $n(n=0 \sim 23)$ で示している。これらを全て同じサブネットに接続する。操作作用PCをクラスタを構成するマシンと同じネットワークに接続する。

Construction of parallel computer for experiments and education using cheap computer

†Keiia Nobuta, Hasegawa Akiumi, Chukyo University

MPI プログラムにおいて、プロセスのランクに応じてマスタとスレーブという呼び方がある。マスタプロセスを実行するマシンのことを、ここではマスタマシン、それ以外をスレーブマシンと呼ぶ。マスタマシンの操作は、操作用 PC から SSH でログインして行う。

ネットワーク接続には 24 ポート以上接続するだけのスイッチングハブが必要で、ルータには DHCP サーバ機能が備わっているものを使用した。Raspberry Pi 側では固定 IP の設定はせず、DHCP により自動的に IP を取得するようにした。DHCP サーバ側で、マスタマシンとなる Raspberry Pi と操作用 PC に対し固定の IP アドレスを割り当てるようにした。

本研究では MPICH2(バージョン 1.4.1p1) のソースコードを、Web ページ [3] からダウンロードし使用する。ソースコードのビルドは Raspberry Pi 上で行った。MPICH の configure オプションは、インストールディレクトリのみを指定した。MPICH を特に指定せずにビルドすると、通信を SSH で行うよう設定される。

クラスタを構成する全てのマシンに同じライブラリをインストールする必要があるが、1 台のマシンにインストールした後、残りのマシンはその OS イメージをコピーして使用すればよい。

全てのマシンにイメージをコピーした後、MPICH に付属している円周率を求めるプログラムを使用し、PC クラスタとして動作していることを確認した。

### 3 並列計算環境の自動構築

並列計算機を手軽に扱うためには管理コストをできるだけ減らしたいという要求がある。管理コストには、ハードウェアやソフトウェアプラットフォームのセットアップ、クラスタの構成等が挙げられる。並列計算環境の管理コストのうち、ハードウェアのセットアップは手作業で行う必要がある。ソフトウェアプラットフォームについてはマシン 1 台分を手作業で行い、他のマシンについてはそのコピーを利用することができるため、台数を増やしてもその作業分ずつしかコストは増加しない。

クラスタを構成するには、クラスタを構成するマシンのネットワーク上のアドレスをマスタマシンが全て知っている必要がある。MPICH を使用する場合、使用する全てのマシンの IP アドレスのリストファイルを、管理者がマスタマシンに用意する。

リストは 1 度作れば同じ環境を利用し続けられるが、例えばクラスタにマシンを追加したときや、マシンの台数を適宜変化させて実験を行いたい時などにはリストを再度作成する必要がある。

そこで、リストを自動作成するシステムを作成した。このシステムでは、マスタマシンは自身と全てのスレーブマシンの IP アドレスのリストを作成する。マスタマ

シンとスレーブマシン上のプログラムの動作について説明する。

マスタマシンのプログラムは、起動後に自身の IP アドレスを IP アドレスリストの先頭に書き込む。スレーブマシンのプログラムはマスタマシンに対し、自身の IP アドレスをリストに追加する要求を TCP のメッセージとして送る。マスタマシンのプログラムは、メッセージを受け取ったら接続してきたスレーブマシンの IP アドレスを記録する。その IP アドレスが既にリストに存在しなければリストに追加し、リストに追加したというメッセージを返信する。スレーブマシン側は、メッセージを受け取ったらプログラムを終了する。

これらプログラムは、OS 起動時に自動起動するように設定した。全ての Raspberry Pi の電源を入れた後、マスタマシン上に、マスタマシンと全てのスレーブマシンの IP アドレスリストが作成されていることを確認した。

### 4 今後の展望

今回作成したシステムは簡易的なもので、IP アドレスをリストに追加した後にその IP アドレスを持つマシンをネットワークから切断した場合、リストの再構成は行われぬ。クラスタを構成するマシンを動的に変化させるには、マスタマシンはリストの中の IP アドレスが使用可能であるか定期的に検査する必要がある。

また、リストに対する操作を、操作用 PC において GUI で行うことも考えられる。各マシンをネットワークから切断せずにリストの内容を変更することによって、クラスタを構成するマシンの台数を動的に変化させることが可能になる。

### 5 まとめ

Cox らの手法に基づき、24 台の Raspberry Pi を使用することで、安価なコンピュータクラスタを構築した。クラスタを構成するマシンの IP アドレスを自動でリスト化することで、クラスタ構築の手間を軽減した。

### 参考文献

- [1] Simon J. Cox, James T. Cox, Richard P. Boardman, Steven J. Johnston, Mark Scott, Neil S. O'Brien "Iridis-pi: a low-cost, compact demonstration cluster" Cluster Computing, DOI: 10.1007/s10586-013-0282-7, 2013
- [2] <http://www.raspberrypi.org/>
- [3] <http://www.mpich.org/>