C-024

マスタの多重化による信頼性のある並列処理システムの提案
A Proposal of Dependable Parallel Processing System Using Multiplexing Master

高津　麻季子†　　柳瀬　龍郎†　　田村　信介†
Makiko Takatsu　Tatsuro Yanase　Sinsuke Tamura

1　はじめに

ワークステーションクラスタや多種多様な資源を共有したグリッドコンピューティングなど，分散環境が大規模であるほど，アプリケーションを実行する際に，さまざまな要因によりシステムに故障が発生する可能性がある．このような環境で並列プログラムを実行するためには，信頼性は重要な要素であり，信頼性に重点を置いたシステムについてさまざまな研究がなされている．

また，大規模な分散環境では，故障のほかに，システム実行中に資源が新しく追加されることも考えられる．新しく追加された資源も動的にシステムに追加可能であれば，より柔軟性のあるシステムとなり，多くのアプリケーションに適応可能になる．

これまでの研究では並列プログラム実行システムにおいて，マスタワーカ型のシステムについて多くの研究がされてきた．特にワーカを多重化し，階層構造を用いてシステムを構成したものが多いが，本研究ではマスタワーカ型システムにおいて，ワーカと同様に，マスタも多重化するシステムを提案する．従来のマスタワーカ型システムではマスタ１台に複数台のワーカの管理を任せることで，並列処理を行ってきたが，この方法だと，マスタ自身が故障した場合に対処できなくなる可能性があった．そこで，マスタを１台から複数台に多重化することで，マスタが相互に自身のデータをバックアップ可能になり，システムの信頼性をあげることができるのではないかと考えた．

2.1　リング型接続によるマスタの多重化

マスタを多重化するためには，マスタ間で接続する必要があるが，本研究では図１のようにマスタをリング型に接続することにした．これは従来のマスタワーカ型と比べ，マスタの通信負荷が一極集中になるのを防ぎ，通信負荷が減ることによって高速化が行えるのではないかと考えられるためである．また，リング型にすることでマスタ１台に対しマスタ２台で故障の検出が可能になる．

各マスタはそれぞれの担当ワーカ，ジョブを管理するためにワーカリスト，ジョブリストを持つ．ジョブリストは未処理ジョブ，処理中ジョブ，処理済ジョブのリストがあり，各マスタはワーカリストとともにバックアップデータとしてそれぞれのバックアップ先マスタに送信する．

マスタ間，マスタワーカ間の通信にはTCP/IPソケットを用いる．マスタワーカともに常にソケットを接続しておくことによって，故障時の切断信号を受信する．故障時の処理については後ほど述べる．

†　福井大学大学院工学研究科情報・メディア工学専攻

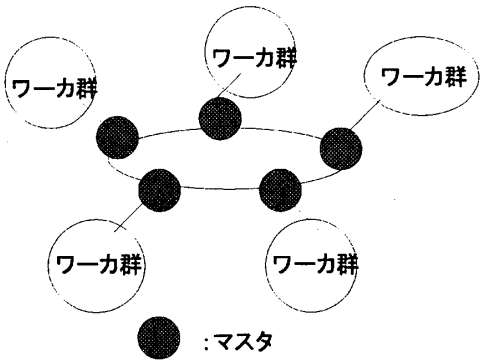

図１　リング型接続マスタ

2.2　マスタ，ワーカの割り当てとジョブの割り当て

システムを起動する際，ユーザが操作するプロセッサをメインマスタとする．システムを起動すると，メインマスタ以外のプロセッサはまずメインマスタに接続し，メインマスタは接続してきたプロセッサをプロセッサリストに登録する．登録されたリストから，マスタ，ワーカの割り当てを行う．このとき，１台のマスタに対するワーカの台数をあらかじめ設定しておき，その台数を元にメインマスタは割り当てを行う．割り当てが行われたあと，メインマスタは他のマスタと同様な処理を行う．

ジョブは各マスタのジョブリストによって管理される．マスタは未処理ジョブリストに入っているジョブを，ワーカに割り当てる．ジョブを割り当てられたワーカはジョブを実行し，終了したらマスタに結果を返す．ジョブを処理中のワーカが処理を終了せずに故障してしまった場合の処理については３章で詳しく述べる．

3　マスタ，ワーカの故障

マスタは自身のジョブリスト，ワーカリストなどの情報を接続したマスタの一方方向に対して送信し，バックアップをとることで故障に備える．マスタが故障した際の処理は以下のような流れになる．

１．バックアップ依頼マスタは故障を検出したらバックアップ依頼先を変更
２．故障したマスタに接続している２台のマスタが故障を検出
３．バックアップ担当マスタは故障マスタのバックアップデータリストから故障マスタの担当していたワーカを引き受ける
４．バックアップデータを自身のデータに加え，通常の処理を再開する

このときバックアップ先マスタは故障マスタが担当し

ていたワーカを全て引き受けるためマスタ間でワーカの不均等が起こる可能性がある．ワーカの不均等を防ぐため，ワーカの台数分散処理を行う．ワーカの台数分散処理については後ほど詳細を述べる．

ワーカは担当マスタによってジョブを割り当てられ，割り当てられたジョブが終了したらその結果をマスタに返す．マスタはワーカとジョブの組をリストで管理する．現在ジョブを処理中のワーカの故障を検出した際には以下の処理を行う．

1．処理中のジョブを未処理ジョブリストの先頭に加える
2．故障したワーカとジョブの組をリストからはずす
3．故障したワーカをワーカリストからはずす

このとき，故障したワーカは一旦システムから外す．ワーカが故障から復帰し，再度システムに参加する場合にはもう一度ワーカをマスタのリスト登録し直す必要がある．この場合はワーカの新規参加と同様な処理を行う．

4.1 ワーカの新規参加

新しくワーカがシステムに参加するには，ワーカはマスタに参加依頼をしなければならない．参加依頼を行うマスタはシステムに参加しているマスタならどれでも良いが，通常はメインマスタに依頼を行う．ワーカから参加依頼を受けたマスタは自身のワーカリストに参加依頼をしてきたワーカを追加する．この時点で，ワーカはシステムに参加したことになり，これ以降は通常のワーカと同様な処理を行う．

4.2 ワーカの台数分散

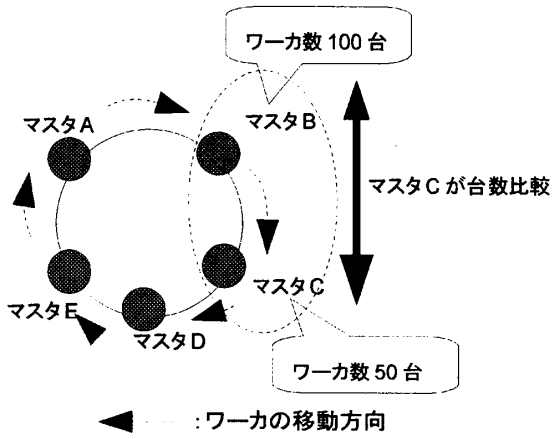

図２　ワーカの台数分散

各マスタはワーカの台数について常にチェックを行い，その情報を元にマスタ間でワーカの台数の比較を行う．比較は，自身とバックアップデータを送ってくるマスタとの間で行う．ワーカの台数チェックはワーカがジョブの処理を終えた時点や，ワーカの台数に変化があった時点で行う．

図２のようにB，C双方のワーカの台数に不均等があった場合，マスタCはマスタBに，ワーカを移動するようメッセージを送信する．メッセージを受け取ったマスタBは，Cにワーカを移動させるための処理を行う．この例ではマスタBからCへ２５台のワーカが移動を行う．ただし，ワーカがジョブを処理してる間は移動できないため，マスタからのジョブ待ち状態のワーカを順次移動させる方法をとる．

このように常にチェックを行い，ある一部のマスタの担当するワーカの台数が他のマスタと比べて多いなどの不均等は無くなると思われる．また，ワーカの台数分散を行うことによって，マスタが故障した際にワーカをバックアップ先のマスタに移動しても最終的にワーカの不均等を無くすことができると考えられる．

5 評価方法

システムの信頼性，動的なプロセッサの台数変化への適応性について評価を行う．評価についてはシミュレーションを行い，システムの有用性を調べることにする．

シミュレーションでは，以下のような事柄について調べる．

・故障の有無での信頼性と実行時間
・マスタ，ワーカ数を変化させたときの信頼性と実行時間
・通信にかかるコスト
・MTTFを変化させたときの信頼性と実行時間

これらの条件を変えてシミュレーションすることにより実際のシステムでの理想値を考察し，システムの信頼性を確認する．

また，マスタを１台で実行した場合と複数台で実行した場合での信頼性を比較する．このとき，マスタ，ワーカの区別無しにランダムに起こるプロセッサの故障の有無によってそれぞれのシステムの実行状況がどのように変化するかを調べることによって，２つのシステムでの信頼性を比較することにする．

6 まとめ

本研究では，信頼性と，動的なプロセッサの台数変化への適応性を考慮した並列実行システムを提案した．今回提案したシステムでは，マスタをリング型で接続し多重化を行ったが，もし仮に全く同時に連続した複数台のマスタが故障することがあれば，システムが正常に終了するのは困難になると予想される．それは，バックアップ先のマスタが１台しかないためであり，連続した２台に同時に故障が起これば，１台のバックアップデータが消滅してしまい，復元不可能になってしまうためであるそのため，リング型以外に，自身以外の全マスタと接続するようなスター型や，リング型で複数台と接続するような接続の方法についても検討を行う予定だ．

今後の課題としては，提案したシステムのシミュレーション評価，連続した複数台のマスタが同時に故障した場合のシステムダウン回避方法の検討があげられる．

参考文献

[1] Hai JIN, Xuanhua SHI, Weizhong QIANG,and Deqing ZOU, DRIC:Dependable Grid Computing Framework, IEICE TRANS. INF. & SYST.,VOL.E89-D,NO.2 FEBRUARY 2006

[2] 松田大樹 Javaによるジョブの多重実行支援環境, 福井大学大学院情報メディア工学専攻平成17年度修士論文