分散大容量メモリ DLM の WAN 接続クラスタ群への適用
― クラスタ･サーバ自動選定システムの提案 ―

鈴木 悠一郎† 緑川 博子 †

1. まえがき

筆者らは，クラスタの遠隔ノードのメモリを利用し，仮想的に大容量のメモリ空間を逐次プログラムに提供するシステム，DLM(Distributed Large Memory)を構築，評価してきた[1]．DLM のマルチクライアント向けシステム DLM-M[2]では，ローカルメモリサイズを超えるプログラムの実行を複数のクライアントが同時にすることが可能である．しかし，ユーザが各クライアントで使用する遠隔ホスト名(メモリサーバ)とその遠隔ホストでの使用メモリサイズを明示的に指定するため，複数クライアント間の調整は行われておらず，特定のメモリサーバへの負荷の集中などが起こる可能性があった．そこで，筆者らはユーザ間でクラスタ内のメモリサーバの割付や負荷の分散などを効率的に行うために，クラスタ内の負荷状況を判断し適切なメモリサーバ自動割付を行う，動的メモリ提供システムを提案した[3]．本報告ではそのシステムを実装，評価した．さらに，１つのクラスタ内でのメモリサーバの選択だけでなく，WAN に接続されたクラスタ群から，各クラスタのメモリサーバ･計算ノード情報をもとに，条件の良いクラスタを自動選択し，さらにそのクラスタ内の適切なメモリサーバノードと計算ノードも自動選択する機構を提案し，構築した．

2. 稼動環境

提案システムは，以下の環境を持つ WAN で接続されたクラスタ群での実行を前提とする．ただし，メモリサーバと計算ノードは同一クラスタ内に割り付けられて，クライアントプログラムを実行する．

１．クラスタ間でのユーザアカウントは同一．
２．同一クラスタ内のノードはファイルを共有．
３．クラスタ内のノードはグローバル IP を持ち，WAN で接続された遠隔からのアクセスが可能．
４．各ノードの CPU は同一ファミリーでバイナリプログラムの互換性がある．

この条件を満たす実験環境として，多くの大学・研究所のクラスタを結合した分散コンピュータシステムである InTrigger[4]を本報告では用いた．

3. クラスタ内のメモリサーバ自動選定システム

クラスタ内のクライアントプロセスの要求により管理プロセスがメモリサーバを自動で割り付けする，メモリサーバ自動選定システムを実装した[図 1]．管理プロセスは，クライアントからの問い合わせに対し，各メモリサーバが現在サービスしているクライアント数をもとに負荷が分散するようにメモリサーバを割り当てる．ここで用いるクライアント数とは１つのクライアントプログラムが使用する全メモリサイズに対する各メモリサーバに割り当てられたメモリサイズで重み付けを行っている．

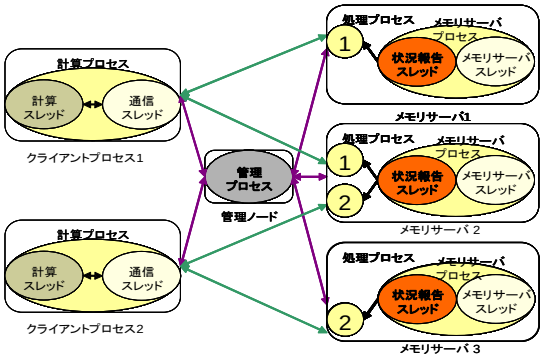

図１．メモリサーバ自動選定システム

4. クラスタ内のメモリサーバ自動選定システム評価

クラスタ内の逐次プログラムをメモリサーバ自動選定システムの評価は InTrigger 内の hongo クラスタで行った．

メモリサーバ自動選定システムの性能評価として，姫野ベンチマークと NAS Parallel Benchmarks を用いた．ここでは，姫野ベンチマ

†成蹊大学　理工学部情報科学科
Department of Computer and Information Science, Seikei University

ークの LARGE でのシステム導入前後での相対性能を示す[図 2]．図 2 はメモリサーバ数 4，クライアント数 8 で動作させた場合の図である．アクセスが集中した場合(サーバ 1 対クライアント 8)と，新システムの導入により負荷分散がされた場合(サーバ 1 対クライアント 2)では，どのローカルメモリ率[1]でも，均一に性能が向上している．

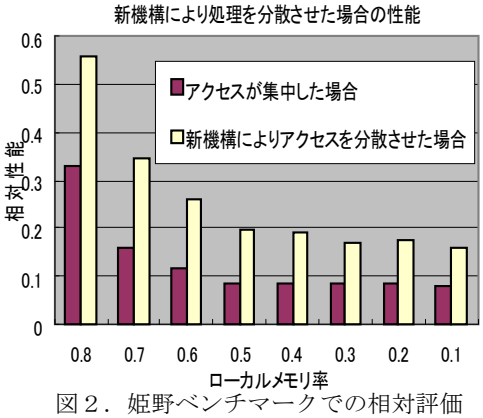

図 2．姫野ベンチマークでの相対評価

5. WAN への拡張

4 節のメモリサーバ自動選定システムの管理プロセスはメモリサーバの負荷情報のみを管理していたが，計算ノード情報も組み込み，担当クラスタ全体の管理を行うプロセス(DLM-LAN Admin)に拡張する．さらに，WAN 環境において複数のクラスタの DLM-LAN Admin 情報をもとに，クライアントプログラムの実行に最適なクラスタを自動選択するクラスタ自動選択プロセス(DLM-WAN Admin)を構築，導入した[図 3].

DLM-WAN Admin は以下のことを可能にする．

1．ユーザが指定したクライアントプログラムの必要メモリ量を元に，各クラスタの利用可能メモリ量と計算ノードの情報を表示.
2．各クラスタ情報を元に，必要メモリを提供でき負荷の少ないクラスタと計算ノードをユーザが選択.
3．DLM-WAN Admin が最適なクラスタと計算ノードを自動で選択.
4．上記の 2 または 3 で選択されたクラスタの算ノードにプログラムを転送し，遠隔実行.

現在，初期実装を行い，実行動作確認をしている．図 4 は hongo クラスタ内ノードから tsukuba クラスタが自動選択され，実行された様子を示す.

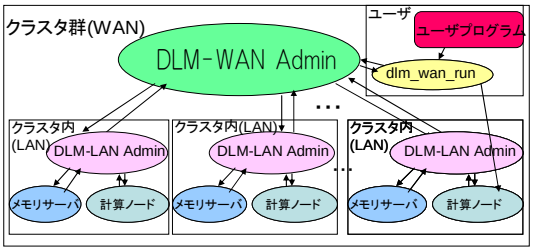

図 3．WAN 環境での DLM システム

図 4．DLM-WAN Admin へのクライアントの実行

6. おわりに

今回はメモリサーバ自動選定システムの評価と WAN 環境への拡張の提案を行った．これによりユーザには見えない形で，適切な資源を選び効率的な実行が可能となる．今後は，プログラムに必要な入出力ファイルの転送や，多様な処理を記述可能なシェルスクリプトによる実行形態なども構築予定である.

参 考 文 献

[1] 緑川,黒川,姫野: "遠隔メモリを利用する分散大容量メモリシステム DLM の設計と 10GbEthernet における初期性能評価", 情報処理学会論文誌 ACS, Vol.1,No 3,pp 136-157（2008,12）

[2] 齋藤, 緑川, 甲斐:"マルチクライアント向け分散型大容量メモリシステム DLM-M の設計と実装",FIT2008 論文集,C－003,pp.199-200, （2008,9）

[3] 三浦,緑川,甲斐:" クラスタをメモリ資源として利用するための動的メモリ提供システムの提案",FIT2009 論文集,B-029, pp.421-422,（2009,9）

[4] 田浦: "InTrigger : オープンな情報処理・システム研究プラットフォーム", 情報処理 Vol.49 No.8,pp939-944,(2009.8)