

Design and Implementation of Synchronization Primitives for λ Computing Environment

大阪大学 大学院情報科学研究科
村田研究室 博士前期課程 2年 井本 舞

研究の背景

4. グリッドコンピューティング

- ネットワークを介して複数の計算機を接続し、計算資源、ストレージを共有
 - 広域で大規模な計算を行う
 - 高速な大容量データ転送が求められる
 - 高速かつ高信頼なネットワークが必要となる

アプリケーション層	統合
TCP	
IP	
Ethernet	物理層

4. 通信オーバーヘッドの影響が大きい

- 複数階層によるプロトコルオーバーヘッド
- 従来の TCP/IP におけるオーバーヘッド
 - パケット処理によるオーバーヘッド
 - 輻輳制御による転送レート低下

高速かつ、高信頼な通信パイプをエンドユーザに提供する新たな技術が必要

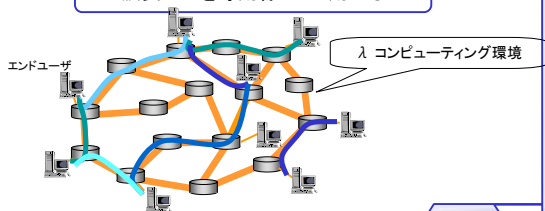
2007/2/16

2

λ コンピューティング環境の提案

- 4. 計算機、ルータを光ファイバで接続する
- 4. 波長パスを設定し、波長パスを通信の最小粒度とする

波長パスを専用線として用いる



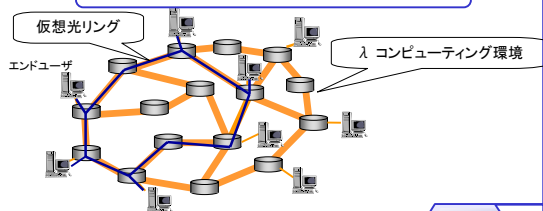
2007/2/16

3

λ コンピューティング環境の提案

- 4. 計算機、ルータを光ファイバで接続する
- 4. 波長パスを設定し、波長パスを通信の最小粒度とする

波長パスをつないで仮想光リングを構成



2007/2/16

4

研究の目的

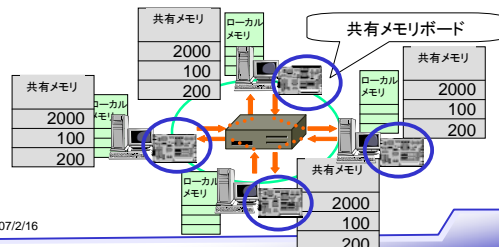
- 4. λ コンピューティング環境において高速な分散並列計算を実現する
 - NTT フォトニクス研究所が開発した AWG-STAR システムを利用
- 4. AWG-STAR システムの共有メモリの特性を活用し、並列計算アプリケーションを動作させ、評価する
 - 並列計算時に必要となる同期プリミティブの設計、実装
 - ロック制御機能
 - バリア同期機能
 - 動的メモリ割り当て機能

2007/2/16

5

AWG-STAR システム

- 4. ノードは波長ルータ (AWG) に接続し光リングネットワークを構成
- 4. ノードは共有メモリボードを搭載
 - 各共有メモリボードは同じデータを保持

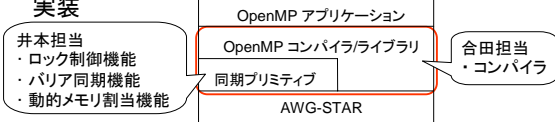


2007/2/16

6

OpenMP のための同期プリミティブ

- 4 OpenMP は複数プロセス間で共有メモリをもつことを前提とした並列計算用プログラミングAPI
 - AWG-STAR システムの共有メモリを活用
 - 1つのマスタープロセスと複数のワーカプロセスで実行
- 4 AWG-STAR システムに応じた OpenMP の設計と実装



現行 AWG-STAR の同期プリミティブを設計、実装
次期 AWG-STAR の同期プリミティブを設計

2007/2/16

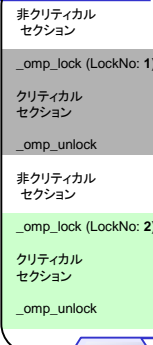
7

ロック制御機能の設計 (1/2)

- 4 クリティカルセクションを排他的に実行するための機能
- 4 それぞれのクリティカルセクションに対して Lock番号 をつける
- 4 共有メモリ上のインデックスでロック/アンロックを管理
 - マスターのみがインデックスを更新

ロックのインデックス

LockNo	ロック/アンロック
1	ノード1がロック
2	ノード2がロック
3	アンロック

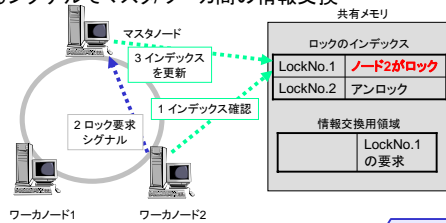


2007/2/16

8

ロック制御機能の設計 (2/2)

- 4 ロックするためにワーカはマスターにロックを要求
 - マスターに要求が届いた順にロックする
- 4 共有メモリに確保した情報交換領域と AWG-STAR が提供するシグナルでマスター/ワーカ間の情報交換



2007/2/16

9

アプリケーションプログラムによる実機評価

- 4 ノード数4台で実行
 - 1台につき1つの OpenMP プロセス実行
- 4 マンデルブロー集合計算
 - データサイズを自由に設定できる
 - 大きいデータサイズのときは分割して計算
 - 計算を容易に並列化できる
 - 計算結果を収集する際に最も共有メモリアクセスが発生する
- 4 計算の実行状況を視覚的に確認できるアプリケーションを作成

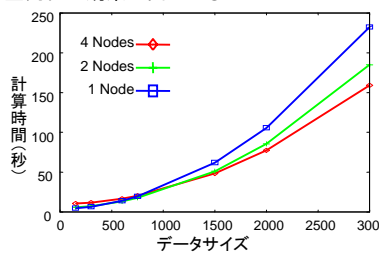


2007/2/16

10

実行時間による評価

- 4 データサイズが大きくなるにつれて、ノード数が多いほど計算時間が短い
 - 並列化の効果が出ている

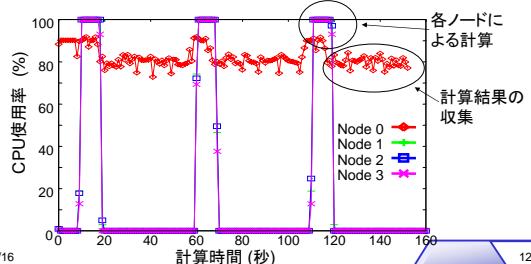


2007/2/16

11

CPU使用率による評価

- 4 計算結果を収集している部分の時間が長い
 - 2/3 が共有メモリからの結果読み込み、1/3 が画像描画にかかる時間
 - 次期 AWG-STAR では共有メモリアクセス速度の向上が期待される



2007/2/16

12

まとめと今後の課題

- 4 AWG-STAR システムを用いて λ コンピューティング環境を構築した
- 4 OpenMP アプリケーションが動作するために必要な AWG-STAR システムの同期プリミティブの設計と実装を行った
- 4 アプリケーションで性能を評価をした
 - 並列化の効果を確認
 - 共有メモリへのアクセス速度がボトルネックである
 - 次期 AWG-STAR システムのメモリアクセス速度の向上により性能改善
- 4 今後の課題
 - 次期 AWG-STAR システムによる λ コンピューティング環境の構築

2007/2/16

13