



P2P分散システムXtremWeb 上でのGrid RPCシステム OmniRPC/XW

筑波大学システム情報工学研究科
中島佳宏・佐藤三久



発表の概要

- 研究の背景・目的
- グリッド環境向けRPCシステムOmniRPC
- P2P分散環境上のGrid RPCシステム
OmniRPC/XW
 - 設計
 - XtremWebの概要
 - OmniRPC/XWの実装
 - 簡単な性能評価
- まとめ



研究の背景

- インターネットをはじめとする広域ネットワークや分散コンピューティング技術の進展
 - ネットワーク上にある計算資源やデータ共有、並列プログラミングを可能とするグリッドコンピューティング
 - それらを利用するための面倒なプログラミング
- 高性能なコモディティPCの普及
 - 巨大な計算リソースが有効利用されていない
 - CF. SETI@homeでは数百万台



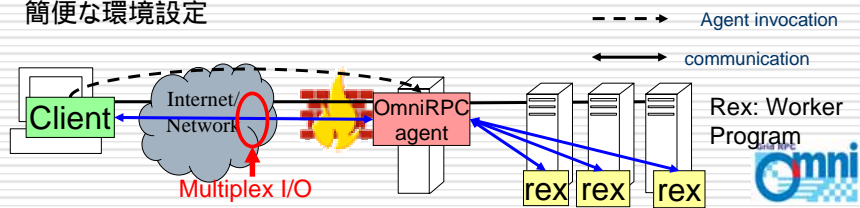
私が思うGridとPeer to Peer

- Grid Computing
 - ノード数は少ない 1000台程度
 - 信頼性があるノード
 - ネットワークがよりStatic
 - DHCP, NAT, Firewallの環境はあまり考慮されていない
 - Fault-toleranceの研究はあるがVolatileまで考えていない
 - Pushモデルのコンピューティング
- Peer to Peer コンピューティング
 - ノード数が多い 100万台に達することも
 - ノードは信頼できないかつVolatile
 - より Dynamic なネットワーク環境を想定
 - DHCP, NAT, Firewall 等
 - Pullモデルのコンピューティング

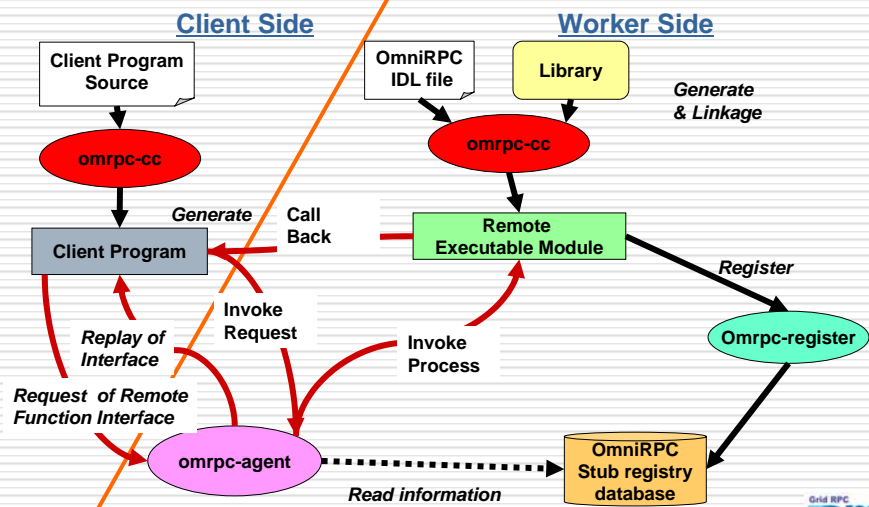


グリッド環境向けRPCシステムOmniRPC

- Ninf [msato *et al.* 97]のAPIをベースにしたthread-safeなRPCシステム
- ローカルなクラスタからグリッドでのマルチクラスタまでのシームレスなプログラミング環境
 - Globus ToolkitのGSI認証のほかにもSSHも利用可能に
 - 通信の多重化 (MXIO)
 - 1000台規模まで対応
 - FireWallやプライベートアドレスで構成されたクラスタ
- Master/Workerの並列プログラムを対象
 - OmniRPC 自動再初期化モジュール機能
- 簡便な環境設定



OmniRPCの概要



グリッド環境向けのRPCシステムの課題

- OmniRPCは150台程度までならスケールアップ
 - さらに1000台以上の計算機を活用
 - 常設されたサーバーやクラスタはそれほど多くない
 - 個人用PCは強力かつ大量に存在
 - いくつかの問題
 - 実験中ノードが計算途中で止まるなどの耐故障性
 - プログラムのデプロイもノード(使用するクラスタ)が増えるに従い煩雑な手間
- ➡ 個人用のPCを計算リソースとして使用できる
Peer to Peer分散環境上でのプログラミング
環境の構築



P2P分散環境上でのプログラミング環境

- P2P分散環境上での並列プログラミングのためのモデルはまだ確立されていない
 - Condor (Wisconsin), BOINC (Berkley), XtremWeb (INRIA)
 - ファイルを基本としたバッチ処理を行うシステム
 - 並列プログラミングをサポートした機構はない
 - P3(AIST)
 - P2P上でのプログラミング環境を提供している
 - Native Codeのサポートは重要



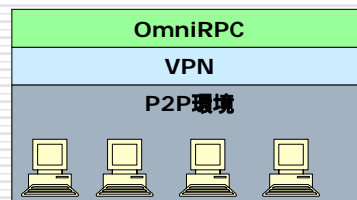
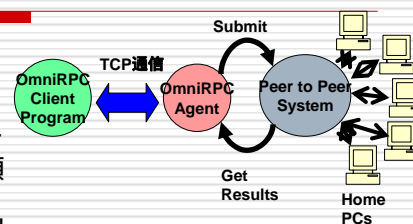
研究の目的

- P2P分散環境上でのGrid RPCによる並列プログラミングを簡便に行えるシステムの開発
 - 実行環境をP2P分散システムまで広げ、さらなる計算機を利用可能な並列RPCシステムの開発
 - P2P分散環境ではジョブのサブミッションだけ
 - P2P分散環境上に並列プログラミングモデルを与える

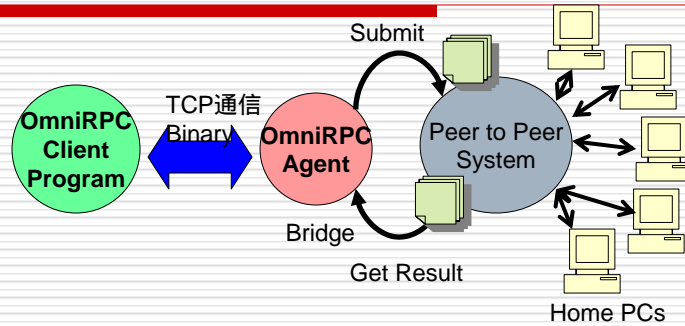


P2P環境でのOmniRPCシステムの設計

1. OmniRPC Agentを使ってOmniRPCとP2Pシステムを中継するブリッジ機能を付加することにより実現
 - ⇒ AgentでRPCをジョブに変換してP2Pシステムに計算実行を依頼
 - ⇒ ドキュメントベースにする
 - ⇒ 耐故障性はP2Pシステムに任せることができる
 2. P2P環境上でVPNを作りその上でOmniRPCシステムを実現する
 - ⇒ ワーク側の障害に対する処理をクライアントが処理
 - ⇒ システムが複雑になる
- 1の方法を選択



P2P分散システム上での OmniRPCシステムの概要



- XtremWebやBOINCなどP2P環境を想定
 - Pullモデルのシステム
 - (OmniRPCはPushモデルのシステム)
- コネクションにより行っていた計算の依頼, 結果の転送を, ファイルを介して行う. (ドキュメントベース)



P2P分散計算システムを利用することによるPros. & Cons.

□ Pros.

- ユーザは、大規模な計算資源を利用可能
 - クラスタだけではなくオフィスや家庭にある個人PCも対象

□ Cons.

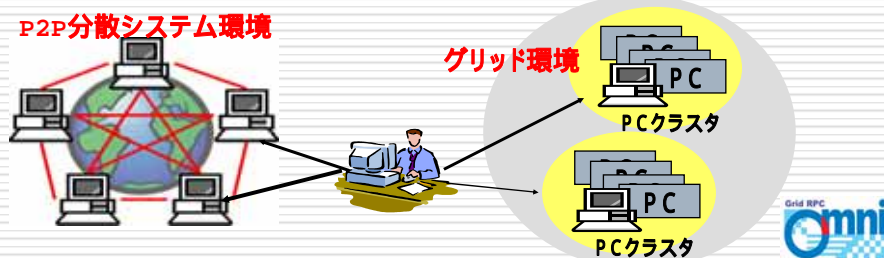
- 性能がOmniRPC単体で使うよりも劣る
 - ドキュメントベースでの処理(ファイルのコピーなど)
 - 断続的な通信

▶ 欠点を補うぐらいの、大量のPCを使えることからのメリットが大きい



P2P分散環境上でのGrid RPC OmniRPCシステムの目標

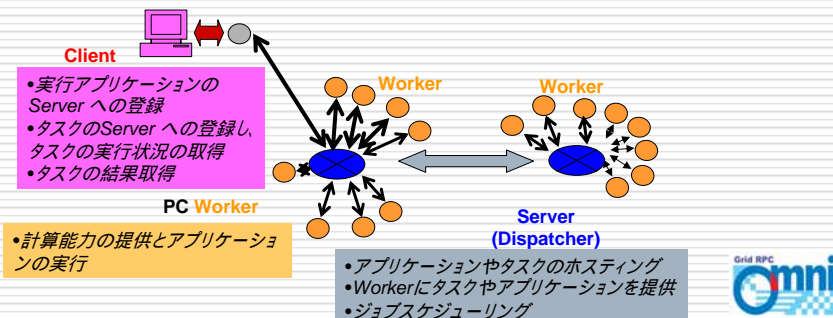
- P2P環境における並列プログラミング環境の提供
- P2P環境における耐故障性を備えた Grid RPCシステムの構築
- ソースプログラム変更なしにクラスタ環境からP2P環境のPCなどを使用可能に
- 実行プログラムのDeploymentの自動化



P2P分散システム XtremWeb



- インターネット上での大規模分散処理を目的としたミドルウェア @INRIA, Univ. of Paris Sud.
- インターネットに接続された遠隔計算資源(PC, Workstation, Server) やLAN 内部にある計算資源プールを活用
- Client / Server (Dispatcher) / worker の3構成
- Jobを投入するとJobの終了するまでシステムがマネージメント

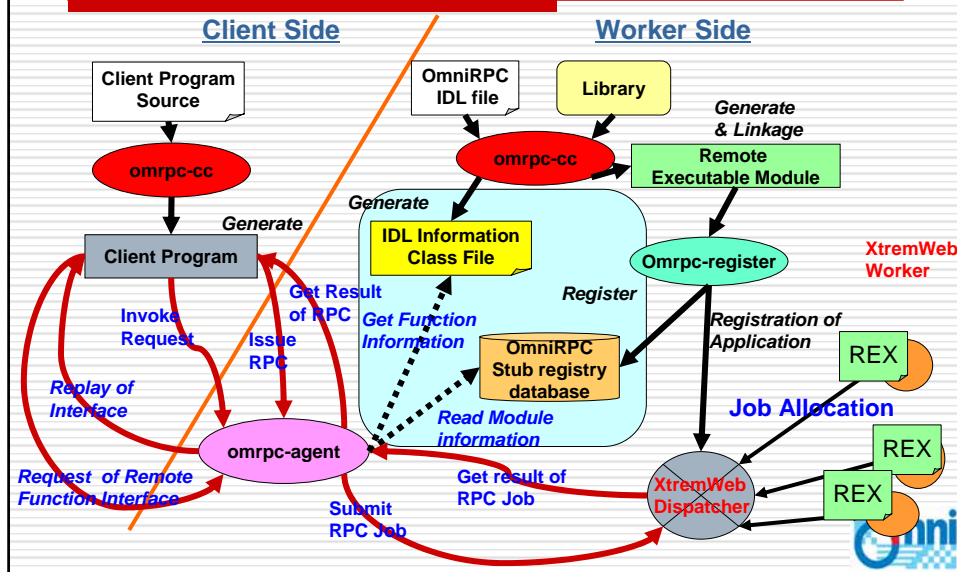


OmniRPC/XWの設計

- 提案するシステムに向けてのOmniRPCの拡張
 - OmniRPC のAgent にXtremWeb のServer とOmniRPC のクライアントプログラム間のプロトコルブリッジを行う機能 (OmniRPC/XW Bridge Agent)
 - OmniRPC のリモート実行モジュールでデータ入出力をファイルから行う機能
 - XtremWeb へのレジストリ情報の登録とリモート関数のインタフェース情報の登録
- XtremWebの対故障性機能を使用し、OmniRPCシステムのFault-Tolerantを保障する



OmniRPC/XWの概要



OmniRPC/XWの実装

- OmniRPC Protocol (ClientとAgent間のプロトコルの変更)
 - クライアントがリモート実行プログラムに送るデータの終端を通知する機構
- リモート実行モジュールに代わりAgentがリモート実行モジュールのスタートアップ時のデータ授受を行う
 - バイトオーダーやリモート実行モジュールの情報
- Agentがリモート関数のインターフェース情報の転送を代行
 - リモート実行モジュールが有する関数のインタフェース情報を保持するクラスファイルを使用
 - リモート実行プログラムに送られるデータをAgentがチェックして判別

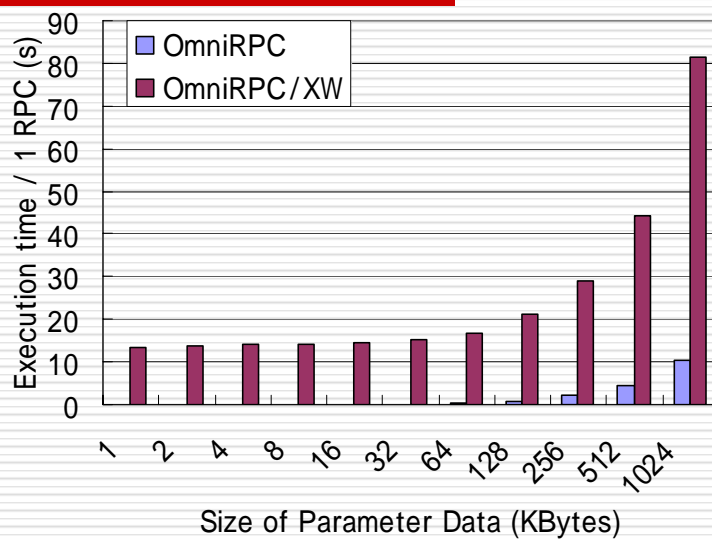


簡単な性能評価

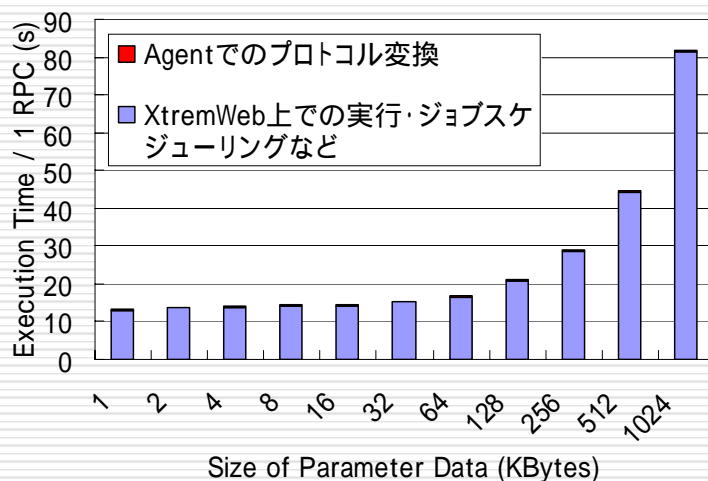
- OmniRPC/XWの1回の遠隔手続き呼び出しにかかるオーバーヘッドがどのくらいかを調べる
 - ローカルホストにおいてパラメータのデータ量を増やすことによる1回の遠隔手続き呼び出し時間を計測
 - 1つの計算機を用いてローカルにて実験を行う
 - 計算機環境
 - IBM ThinkPad T40
 - CPU: Pentium M 1.6GHz Memory 1.5G Bytes
 - OS: Linux Kernel Version 2.6.11-rc1 + Andrew Morton patch
 - Java VM: SUN J2SE(TM) Development Kit 5.0
 - OmniRPC Version 1.2.1
 - XtremWeb Version 1.3.4
 - MySQL Version 4.0.23



パラメータのデータ長による1RPCの実行時間



OmniRPC/XWにおける1RPC実行時間の詳細



実験に関する考察

- OmniRPCとOmniRPC/XWの性能差は大きい
 - パラメータデータ量が増えると性能差は小さくなるが依然性能差は大きい
 - OmniRPC/XW Bridge Agentによるプロトコル変換の処理は小さい
 - XtremWebにジョブをサブミットしてから計算結果の取得にかかわる時間がほとんどをしめる
 - XtremWeb Dispatcher側でのジョブスケジューリング
 - XtremWeb Workerが常にジョブのQueueを調べているわけではない(Workerの休止時間が存在する)
- 想定しているアプリケーションにおいては、1RPCが1時間以上かかるものを想定しているので問題なし？



OmniRPC/XWの現状

- XtremWeb側のJob Submissionが正常に動作せず正常に動かないことを発見！！
- XtremWebのDebug中



まとめ

- Peer to Peer環境においてVolatileな計算機をGrid RPCプログラミングモデルで利用可能にするためのXtremWeb上でのOmniRPCシステムについて提案
- OmniRPC/XW Bridge Agent でのプロトコル変換の処理のオーバーヘッドは少なく、XtremWeb 側が原因で素のOmniRPC と比べたときに性能が低い



今後の計画

- XtremWebは使えなかったもので(コードが汚い・論文に書いてあることが行えない etc.)、新しくちゃんとしたP2P向けのプラットフォームをつくります
- 2006年までにはリリース予定

