社団法人 情報処理学会 研究報告
IPSJ SIG Technical Reports

2007－ＱＡＩ－23（８）
2007／5／30

グリッドネットワークサービス機構の実装と評価

中村 勝一† 池永 全志† 山崎 克之‡ 尾家 祐二†

†九州工業大学 〒804-8550 北九州市戸畑区仙水町 1-1

‡長岡技術科学大学 〒940-2188 長岡市上富岡町 1603-1

E-mail: †nakamura@infonet.cse.kyutech.ac.jp, ike@ecs.kyutech.ac.jp

あらまし ネットワーク上に存在する計算資源を統合し仮想的な高性能計算機として利用するグリッド環境を構築するためのグリッドミドルウェアの開発が進んでいる．グリッド環境において実行されるアプリケーションがネットワーク資源を効率良く利用するためには，グリッドミドルウェアと連携するネットワークサービスが必要である．本研究では，グリッド環境においてタスク割当てのスケジューリング機能を有するグリッドミドルウェアであるスーパースケジューラとの連携機構を提案し，その機構により実現可能となるネットワークサービスについて検討した．さらに，連携機能を試作し，スケーラビリティおよび応答性を評価し，その有効性を示す．

キーワード グリッド，ネットワークサービス，ネットワーク制御，ネットワーク管理

Implementation and evaluation of grid network service architecture

Katsuichi NAKAMURA† Takeshi IKENAGA† Katsuyuki YAMAZAKI‡ and Yuji OIE†

† Kyushu Institute of Technology 1-1 Sensui-cho, Tobata-ku, Kitakyushu, 804-8550 Japan

‡ Nagaoka University of Technology 1603-1 Kamitomioka-cho, Nagaoka, 940-2188 Japan

E-mail: †nakamura@infonet.cse.kyutech.ac.jp, ike@ecs.kyutech.ac.jp

Abstract In a grid computing environment, the grid middleware handles the grid resource management, such as computing resource and network resource and so on. The grid middleware has been proposed and developed for the efficient and distributed processing of grid applications. On the other hand, the network service architecture for the grids has also proposed and developed for managing routing and network resource allocation and scheduling based on network monitoring information while cooperating with the grid middleware. Therefore, we investigated the grid network functions, which are required from grid middleware, and we designed the grid network service architecture. In this paper, we introduce our proposed architecture and developed grid network management system. Moreover, we evaluate the scalability and response time of the network service function cooperating with the super scheduler, which serves to allocate tasks of grid applications.

Keyword Grid, Network Services, Network Control, Network Management

1. はじめに

近年，ネットワークの広域・高速化及び，計算機の高性能化が進み，ネットワーク上に存在する計算機資源を相互接続し，仮想的な高性能計算環境を実現するグリッド環境の構築が進められており，様々な科学技術計算において利用され始めている[1]，[2].

グリッド環境において，アプリケーションが計算機資源及び，ネットワーク資源を効果的に利用して高速に実行されるためには，最適な計算機資源とネットワーク資源を選択する仕組みや，アプリケーション実行中においてもグリッド資源の利用状況に柔軟に適応してジョブを実行する仕組みが必要となる[3]．これらのグリッド環境を支える基盤ソフトウェアはグリッドミドルウェアと呼ばれている．グリッドミドルウェアとしては，例えば，特定のリソースに関する情報を収集し，複数の情報プロバイダに提供するグリッドリソース情報サービスや，実行要求のあったアプリケーションを適切な計算機資源に割り当てるためのスケジューラ（スーパースケジューラ）などが含まれる．

このようなグリッドミドルウェアの研究開発が，「最先端・高性能汎用スーパーコンピュータの開発利用プロジェクト」におけるサイエンスグリッド NAREGI（National Research Grid Initiative）プログラム[4]や，欧州における EGEE[5]などにおいて活発に進め

られている．

グリッド環境において効率良くアプリケーションを実行するためには，分散配置された計算機資源を接続するネットワーク特性についても考慮した上でジョブの割り当てやグリッド資源選択を行う必要がある．さらに，グリッド環境では管理主体が異なる実際の組織（PO: Physical Organization）を跨ってグリッド資源を共有する必要があるため，グリッド資源保有者が，自らが保持するグリッド資源に対して利用制限を設けつつ，共通の目的を持つ利用者との間でグリッド資源を共有する仮想的な組織（VO: Virtual Organization）を構成する．そのため，ネットワーク資源の情報に関しても，VO 毎に管理する必要が生じる[6]．

本研究では，グリッドアプリケーションに対して良好な通信環境をコスト効率良く提供するための機構として，グリッドミドルウェアと連携し，グリッド資源の一つであるネットワーク資源の管理，計測，制御機能を提供する機構を備えたグリッドネットワークサービスのフレームワークを提案する．さらに，グリッド環境にて VO 単位で利用可能なネットワーク資源の管理を可能とする VO を考慮したグリッドネットワーク管理システム[6]において，グリッドミドルウェアとの連携機構を試作した．

以降，2 節でグリッドミドルウェアの機能について説明した後，3 節にて本研究で提案するグリッドネットワークフレームワークを示した後，4 節で評価対象としたスーパースケジューラとの連携機能の実装，5 節で実装した機能の評価について記述する．最後に 6 節でまとめる．

2. グリッドミドルウェアの機能

仮想的な高性能計算機環境であるグリッド環境において，アプリケーションが計算機資源及び，ネットワーク資源を効果的に利用して高速に実行されるためには，最適なグリッド資源（計算機資源とネットワーク資源）を選択する仕組みや，アプリケーション実行中においてもグリッド資源の利用状況に柔軟に適応してジョブを実行する仕組みが必要となる[3]．これらのグリッド環境を支える仕組みを備えた基盤ソフトウェアがグリッドミドルウェアと呼ばれている．以下に，グリッド資源に関連するグリッドミドルウェアについて説明する．

- グリッドリソース情報サービス

  計算機の IP アドレス，OS 名称やバージョン，CPU に関する情報，物理・仮想メモリサイズや空き状況など，計算機資源に関するリソース情報と，ネットワークパス，可用帯域，推定遅延や，ホスト，スイッチ，ルータなど，ネットワークトポロジ情報など，ネットワークに関するリソース情報といった，特定のリソースに関する情報を収集し，複数の情報プロバイダに提供するサービスが提唱されている[8]．

- スーパースケジューラ

  実行要求のあったアプリケーションを適切な計算機資源に割り当てるためにスケジューラが必要であるが，グリッド環境においては，膨大な数の計算機資源から実行要求アプリケーションが必要とする計算機資源を検索し，アプリケーションの配置を決めるため，単一スケジューラで大規模なグリッド環境で対応することは，非常に困難である．そこで，計算機クラスタシステムなどの比較的小規模な単位をローカルスケジューラとして管理し，これらローカルスケジューラを束ねて管理するスーパースケジューラが提唱されている[7]．

3. グリッドネットワークサービスのフレームワーク

ネットワーク資源の管理，計測，制御機能を実現するため，グリッドミドルウェアに対するネットワークサービスのフレームワークを提案する．これは，図 1 に示す通り，ネットワーク資源を取り扱う複数の機能をグリッドネットワーク管理システムとして統合し，スーパースケジューラ等のグリッドミドルウェアとの間で適切なインタフェースを介してサービス提供を行うものである．本フレームワークで提案するグリッドネットワークサービスを以下に示す．

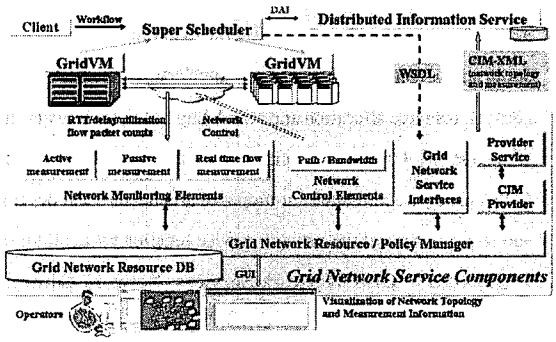

図 1．グリッドネットワークサービスの概要

3.1. ネットワークモニタリングサービス

グリッドネットワーク管理システムにおいては，アプリケーションが利用可能な物理的資源および網構成を運用管理者が登録する情報によって把握する必要がある．また，多様なアプリケーションが並行して実行されるグリッド環境では，一つのアプリケーションが通信を行うことによってネットワーク資源の状態が変化すると，その変化が資源を共有する他のグリッドアプリケーションにも影響を与える．そのため，グリッ

ドネットワーク管理システムが備えるネットワークモニタリングサービスは，特定のフローを識別してモニタリングすることが可能な計測機能が必要である．さらに，計測機能によって収集したフロー情報，経路情報等を網構成情報と関連付けたうえでデータベースに保存し，グリッドミドルウェアにおけるリソース情報サービス機構に対して提供する機能が求められる．

3.2. ネットワークリソース・スケジューリングサービス

グリッド環境においては，実行要求のあったアプリケーションを適切な計算機資源に割り当てるためにスケジューラが必要であり，膨大な数の計算機資源から実行要求アプリケーションが必要とする計算機資源を検索し,アプリケーションの配置を決める必要がある．グリッドネットワーク管理システムが備えるネットワークリソース・スケジューリングサービスは，グリッドミドルウェアであるスーパースケジューラが資源割り当てを行う際に，ネットワーク資源に関する制約条件等を考慮する場合を想定し，網構成情報やネットワーク資源利用状況等にもとづいたスケジューリングの支援を行う機能を備えるものとする．

グリッドミドルウェアであるグリッドリソース情報サービスに各種ネットワーク情報を，スケジューリングを行う情報として活用する手法もあるが，大規模化するネットワークトポロジや多様なネットワーク計測情報など，グリッドリソース情報サービスに格納された，これら大量の情報に基づいて，様々な観点からスーパースケジューラにおいてスケジューリングする事は今後,困難になると考えられる．そこで，スーパースケジューラからの要求に対して，ネットワーク資源の利用状況や,ネットワークの構成情報等に基づいて，タスクのスケジューリングを行うために，ネットワークの観点からみたタスク実行計算機の候補となる計算機の絞込みを行う機能が必要と考えられる．

さらに，ネットワークモニタリングサービスと連携し，グリッドアプリケーション実行中に，その処理性能がネットワークの要因で劣化した場合に再配置を可能とする機能を備えるものとする．

3.3. ネットワークリソース・アロケーションサービス

スーパースケジューラは，利用者からの要求を受けグリッドアプリケーションの処理に適した計算資源を予約した後に，各計算資源へのデータ送信を効果的に行うためにネットワーク資源の確保を行う必要がある．グリッドネットワーク管理システムが備えるネットワークリソース・アロケーションサービスは，グリッドアプリケーションの処理に伴い発生するデータ送受信に必要な要求帯域情報を，スーパースケジューラが最初に利用者からのグリッドアプリケーション処理要求と共に受け取り，グリッドアプリケーションが必要とする通信特性に応じた適切なネットワーク制御を行う機能を備えるものとする．

4. グリッドネットワーク管理システムの実装

前節で提案したフレームワークに従って，グリッドミドルウェアとネットワーク管理システムが連携して動作するために必要な機能を実装した．グリッドリソース情報サービスとの連携インタフェースの実装については[9]に，グリッドリソース・アロケーションサービスに関連するネットワーク制御，経路制御については[6]および[10]に示している．本稿では，スーパースケジューラとの連携インタフェースについての評価について示す．

4.1. スーパースケジューラとの連携機能

グリッドネットワーク管理システムとスーパースケジューラとのシステム連携図を図２に示す．これにより，グリッドネットワーク管理システムが有するネットワーク資源情報を用いて，スーパースケジューラによるスケジューリングを支援する機能を提供する．機能連携にあたっては，Web Service Description Language (WSDL) によるインタフェース定義を用いる．

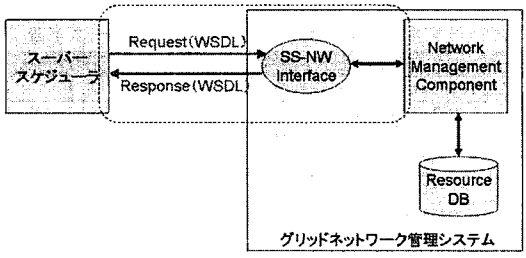

図２．スーパースケジューラとの連携

4.2. スーパースケジューラとの連携機能の実装

本研究では，ネットワーク管理システムとスーパースケジューラの連携によるネットワークサービス実現の例として，ネットワーク資源情報にもとづくタスク実行計算機候補の絞込みを行う機能の実装例を示す．これは，スーパースケジューラからの要求にもとづいて，ネットワーク資源の利用状況や，ネットワークの構成情報等を調査し，要求に合致する計算機群をタスク実行計算機の候補として回答するものである．

このような機能を実現するために，WSDL により，以下に示す GroupType の組合せにより計算機の絞込みを行うインタフェースを定義した．

- GroupType=localsite
  ネットワークトポロジ情報に基づいて，同一ネットワーク機器インタフェースに接続されたクラスタを絞込む
- GroupType=samerouter
  ネットワークトポロジ情報に基づいて，同一ネッ

トワーク機器に接続されたクラスタを絞込む

- GroupType=bandwidth
  ネットワーク資源の計測情報に基づいて，クラスタ間を結ぶ経路の帯域が Boundary 値以上のクラスタを絞込む
- GroupType=delay
  ネットワーク資源の計測情報に基づいて，クラスタ間を結ぶ経路の遅延が Boundary 値以下のクラスタを絞込む
- GroupType=moninodes
  補完的な GroupType で，クラスタのノード数が Boundary 値以上のクラスタを絞込む

さらに，スーパースケジューラとのインタフェース定義を以下のように策定した．

「スーパースケジューラからの要求インタフェース定義」

GroupConditionArray	グルーピング条件リスト
- GroupCondition	グルーピング条件
- GroupType	グルーピング条件種別
- Boundary	グルーピング境界値
ClusterGroup	候補クラスタリスト
- Cluster	クラスタ情報
- IPAddress	クラスタ代表 IP アドレス
- NodeNum	クラスタ収容ノード数

「スーパースケジューラへの応答インタフェース定義」

ClusterGroupArray	クラスタグループリスト
-ClusterGroup	クラスタグループ
-Cluster	クラスタ情報
-IPAddress	クラスタ代表 IP アドレス
-NodeNum	クラスタ収容ノード数

以上の WSDL で記述されたインタフェース定義に基づいて実際にスーパースケジューラとグリッドネットワーク管理システムとの間で行われる通信の例を図３，図４に示す．

```
<SOAP-ENV:Body>
  <tns:GetGroupsRequest
   xmlns:tns="http://www.naregi.org/ws/ogsa/ghpn/gnis">
    <tns:GroupConditionArray>
      <tns:GroupCondition>
        <tns:GroupType>localsite</tns:GroupType>
        <tns:Boundary>0</tns:Boundary>
      </tns:GroupCondition>
    </tns:GroupConditionArray>
    <tns:ClusterGroup>
      <tns:Cluster>
        <tns:IPAddress>10.1.1.11</tns:IPAddress>
        <tns:NodeNum>11</tns:NodeNum>
      </tns:Cluster>
      ...計算機リソースの繰り返し...
      <tns:Cluster>
        <tns:IPAddress>10.10.2.110</tns:IPAddress>
        <tns:NodeNum>110</tns:NodeNum>
      </tns:Cluster>
    </tns:ClusterGroup>
  </tns:GetGroupsRequest>
</SOAP-ENV:Body>
```

図３　スーパースケジューラからの要求(XML)

```
<SOAP-ENV:Body>
  <tns:GetGroupsResponse
   xmlns:tns="http://www.naregi.org/ws/ogsa/ghpn/gnis">
    <tns:ClusterGroupArray>
      <tns:ClusterGroup>
        <tns:Cluster>
          <tns:IPAddress>10.10.3.11</tns:IPAddress>
          <tns:NodeNum>11</tns:NodeNum>
        </tns:Cluster>
        <tns:Cluster>
          <tns:IPAddress>10.10.3.12</tns:IPAddress>
          <tns:NodeNum>12</tns:NodeNum>
        </tns:Cluster>
      </tns:ClusterGroup>
    </tns:ClusterGroupArray>
  </tns:GetGroupsResponse>
</SOAP-ENV:Body>
```

図４．スーパースケジューラへの応答(XML)

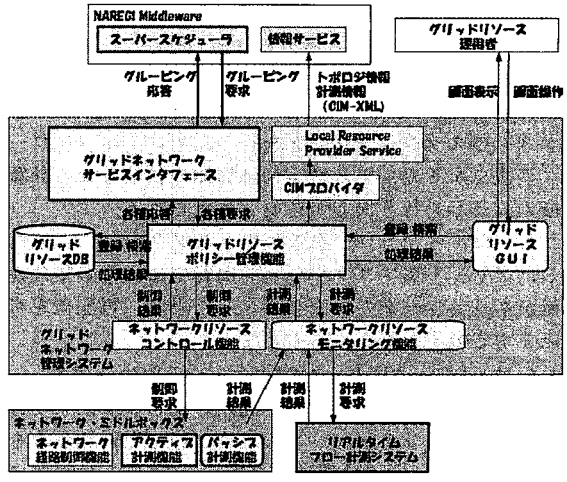

図５．グリッドネットワーク管理システムにおけるスーパースケジューラ連携機能

4.3. スーパースケジューラ連携機能の動作

グリッドネットワーク管理システムにスーパースケジューラ連携機能を実装したシステム全体図を図５に示す．グリッドネットワーク管理システムは，前項で示したスーパースケジューラ連携インタフェースを

備えたグリッドネットワークサービス・インタフェースを介し，スーパースケジューラからの要求がグリッド・リソース・ポリシー管理機能に渡される．

グルーピング条件がネットワークトポロジに関する条件であれば，グリッドネットワーク管理システムのグリッドリソース DB が保持しているネットワークトポロジ情報により候補クラスタ絞込み処理が行われ応答をスーパースケジューラに返す．グルーピング条件がネットワーク計測結果を必要とする条件の場合は，各種計測情報収集機能を備えた外部装置からネットワーク・リソース・モニタリング機能を介して，グリッドリソース DB に蓄積された計測情報を取得した後に，候補クラスタ絞込み処理が行われる．

5. スーパースケジューラとの連携機能の評価

グリッドネットワーク管理システムが備えるスーパースケジューラとの連携機能の応答性とスケーラビリティに関して，試作システムを用いて性能評価を行った．

5.1. 評価環境

性能評価を行うにあたり，試作システムを用いて，図６に示す環境を構築した．

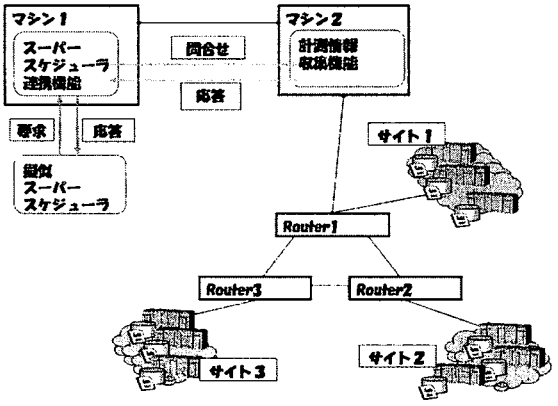

図６．評価環境

図中におけるマシン１は，前節に示したグリッドネットワークサービス・フレームワークを実現するための，スーパースケジューラ連携機能を備えたグリッドネットワーク管理システムである．マシン２は各サイト間の可用帯域や遅延等を計測する計測情報収集機能を備えた装置である．擬似スーパースケジューラは，評価のため開発した擬似ツールであり，スーパースケジューラからの要求を擬似的に発生させ，その応答を受信するものである．ネットワークは３台のルータで構成しており，各ルータは複数のクラスタ群を有するサイトに接続されているものとする．

マシン１とマシン２のスペックを以下に示す．

ＯＳ：RedHatLinux9

メモリ：１GB

ＣＰＵ：Pentium(R) 4 CPU 3.06GHz

マシン１で動作させるプロセスを以下に示す．

- Apache&Tomcat
- PostgreSQL
- Globus
- スーパースケジューラ連携 SOAP サーバ

5.2. 結果および考察

前項で示した環境において，スーパースケジューラとの連携機能に関する評価を行った結果を図７，図８，図９に示す．図７に，グリッドネットワーク管理システムが備えるスーパースケジューラとの連携インタフェースの応答特性を示す．ここでは，ネットワーク内に 90 個のクラスタが存在する環境で評価を行った．横軸にスーパースケジューラからの要求メッセージに含まれるグルーピングの候補となるクラスタ数，縦軸にグリッドネットワーク管理システムにおいて絞込みを行った結果の平均応答時間を示す．点線のグラフは，グルーピング条件として帯域幅 30Mb/s 以上と指定された場合，実線のグラフは，グルーピング条件として同一ネットワーク内と指定された場合の特性である．

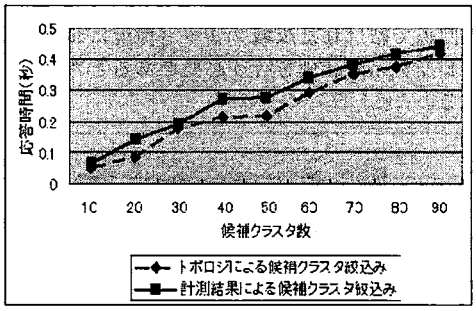

図７．連携インタフェースの応答時間

図７に示す通り，グリッドネットワーク管理システムが保持しているネットワークトポロジ情報を用いた候補クラスタ絞込みは，ネットワーク計測結果を必要とする候補クラスタ絞込みに比べて約数％程度応答速度が速くなっている．これは，ネットワーク計測結果がサイト間の経路情報にリンクされているために，条件を満たす経路の探索に必要な処理時間が増えているためであると考えられる．

ここで想定している機能は，スーパースケジューラが予め，数千オーダのクラスタ群から，クラスタの実行環境や稼動状況により適切な候補クラスタを絞り込んだ後，計算資源に関する情報だけではそれ以上にクラスタを絞り込めない状況下でネットワーク資源を考慮して候補クラスタを絞り込むことを想定している．図７の結果より，スーパースケジューラから与えられ

るジョブ割当て可能な候補クラスタ数が 90 の場合でも，0.5 秒以内に応答可能であり，ネットワーク資源の情報をスケジューリングに反映可能であるといえる．

図８および図９に，グリッドネットワーク管理システムが備えるスーパースケジューラとの連携インタフェースのスケーラビリティを示す．横軸にネットワーク内に存在する総クラスタ数を，縦軸にグリッドネットワーク管理システムにおいて，絞込みを行った結果の平均応答時間を示す．スーパースケジューラからの要求としてグルーピング条件に従いグルーピングの候補となるクラスタ数を 10, 20, 30 と変化させた場合の特性を示している．図８は，グルーピング条件として同一ネットワーク内と指定された場合，図９はグルーピング条件として帯域幅 30Mb/s 以上と指定された場合のグラフである．

図８，図９で示す通り，グリッドネットワーク管理システムが保持しているネットワークに存在する総クラスタ数が増加しても，ある程度一定の応答時間で，スーパースケジューラからの要求に対する応答が可能である．しかしながら，本評価環境では，総クラスタが少ない事もあり，スケーラビリティ評価が十分とは言えない．今後，次世代学術情報ネットワーク[11]や国際回線で結ばれた広域グリッド環境において，スケーラビリティの評価を行う必要があると考え，計画中である．

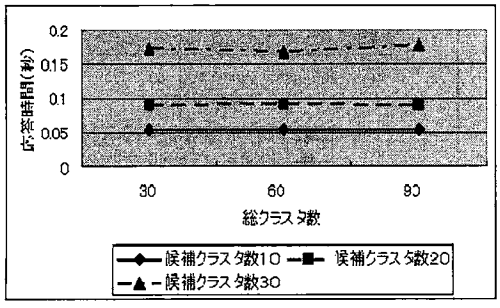

図８．グルーピング条件が同一ネットワーク

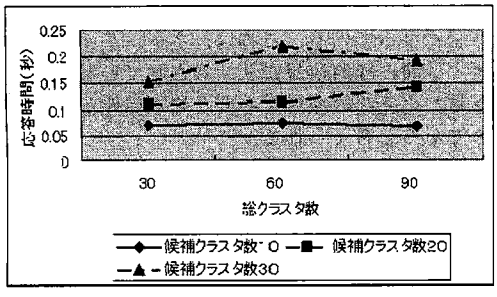

図９．グルーピング条件が帯域幅 30Mbps 以上

6. まとめ

本研究では，グリッドアプリケーションに対して良好なネットワーク環境を提供するため，ネットワーク資源の管理，計測，制御機能を実現するグリッドミドルウェアに対するネットワークサービスのフレームワークを提案し，グリッド環境においてタスク割当てのスケジューラ機能を有するグリッドミドルウェアであるスーパースケジューラとの連携機構を提案した．その機構により実現可能ある機能を検討および一部機能を試作して，スケーラビリティおよび応答性を評価し，その有効性を示した．グリッドアプリケーションが動作するグリッド環境は，広域化，大規模化が進んでおり，グリッドネットワークサービスは，十分なスケーラビリティを必要とされているので，広域な（国際的な）グリッド環境において，スケーラビリティの評価を今後実施する予定である．

謝辞 本研究の一部は，文部科学省の「最先端・高性能汎用スーパーコンピュータの開発利用プロジェクト」におけるサイエンスグリッド NAREGI（National Research Grid Initiative）プログラムの支援を受けている．ここに記して謝辞を表す．

文　献

[1] I. Foster and C. Kesselman, "The Grid Blueprint for a New Computing Infrastructure," Morgan Kaufmann Publishers, 1998.

[2] I Foster, C. Kesselman and S. Tuecke, "The Anatomy of the Grid: Enabling Scalable Virtual Organizations," International Journal of Supercomputer Applications, 15(3), 2001.

[3] Tommaso Coviello and Tiziana Ferrari, et al, "Bridging Network Monitoring and the Grid," http://egee-jra4.web.cern.ch/egee-jra4/cesnet2006-NPM-paper.pdf.

[4] NAREGI(National Research Grid Initiative) Project, http://www.naregi.org/index_e.html

[5] EGEE (The Enabling Grids for E-sciencE) Project, http://public.eu-egee.org/

[6] 池永全志，中村勝一，尾家祐二，"Grid 環境における VO を考慮したネットワーク制御システムの開発，" 信学技報，NS2003-365，IN2003-320，pp.367-372，2004 年 3 月

[7] J. Schopf, "Ten actions when superscheduler," GGF Documents, GFD.4, 2001.

[8] Beth Plale, et al, "Key Concepts and Services of a Grid Information Service," In 15th International Conference on Parallel and Distributed Computing Systems (PDCS 2002), pp. 437-442, Sep. 2002.

[9] 中村勝一，青柳好織，池永全志，尾家祐二，"グリッドミドルウェアと連携するネットワークサービスの開発"信学技報，vol.106, no.578, IN2006-260, pp.477-482, 2007 年 3 月

[10] 西繁則，山本寛，池永全志，尾家祐二，"Grid ミドルウェアと連携するネットワーク制御手法，" 信学技報，IN2004-333，pp.441- 446, 2005.

[11] 次世代学術情報ネットワーク (SINET3) http://www.kuins.kyoto-u.ac.jp/news/55/sinet3.html