MAC-5701: Tópicos em Ciência da Computação

IME - Instituto de Matemática e Estatística - USP - Universidade de São Paulo

Plano de Estudos

Alexandre Cesar Tavares Vidal - N.o USP: 4791925
Orientador - Prof. Fábio Kon

08 de Abril de 2004

1 Introdução

Este Plano de Estudos contém a descrição das atividades de pesquisa que deverei realizar como parte da disciplina MAC5701 - Tópicos em Ciência da Computação. O tema da pesquisa se insere na área de sistemas distribuídos e explora a extensão do `Repositório de Aplicações` do InteGrade, sistema de computação em grade em desenvolvimento no IME-USP. A realização destas atividades propiciará um envolvimento mais aprofundado com os detalhes técnicos do projeto, com as técnicas avançadas de programação distribuída adotadas no projeto, e deverá, possivelmente, integrar o corpo da minha qualificação. O resultado concreto dessas atividades de pesquisa é a geração de código e documentação da extensão do `Repositório de Aplicações` que é parte do InteGrade.

1.1 Motivação

O projeto InteGrade (`http://gsd.ime.usp.br/integrade/`) foi motivado principalmente pela necessidade de aproveitamento de recursos computacionais compartilhados (computadores pessoais, estações de trabalho), a maioria com predomínio de períodos ociosos, para execução de aplicações com grande demanda de recursos. Em particular, as características do InteGrade incluem suporte para uma faixa abrangente de aplicações paralelas e mecanismos que tornem imperceptível para os proprietários dos recursos compartilhados, qualquer possível perda de qualidade de serviço [K^+04b].

Atualmente, uma implementação prototípica do InteGrade está em desenvolvimento, e a nossa atividade de pesquisa é parte dessa evolução. Na 6a. reunião de *Checkpoint* do InteGrade, realizada em setembro de 2003, foi apontada a necessidade de "refinamento do Repositório de Aplicações, para permitir categorização de aplicações, entre outras características" [K^+04a].

A Seção 2 descreve uma visão geral do InteGrade e seu atual estágio de implementação. A Seção 3 descreve o módulo Repositório de Aplicações, as funcionalidades já implementadas e as funcionalidades desejadas para futuras versões. A Seção 4 descreve o plano de trabalho e o escopo da implementação desta atividade de pesquisa. A Seção 5 contém uma conclusão sintética deste plano de estudos.

2 O InteGrade

A grade InteGrade se constitui de `aglomerados` (*clusters*) de computadores organizados de forma hierárquica e escalável. Os principais componentes de um aglomerado são:

Gerenciador do aglomerado: nó responsável pelo gerenciamento do aglomerado e pela comunicação com gerenciadores de outros aglomerados.

Nó de usuário: pertence ao usuário da grade que submete aplicações à grade.

Nó fornecedor de recursos: tipicamente uma estação de trabalho, disponibiliza parte de seus recursos para execução de aplicações dos usuários da grade.

Nó dedicado: reservado para computação na grade.

O *Local Resource Manager* (LRM) e o *Global Resource Manager* (GRM) cooperam no gerenciamento intra-aglomerado, em um processo conhecido como *Information Update Protocol*. O LRM é executado em cada nó do aglomerado, coletando informações sobre o estado do nó (uso de CPU, disco, memória e recursos de rede) e as envia periodicamente ao GRM. Este último executa no nó gerenciador do aglomerado e usa essas informações enviadas pelos LRMs para escalonamento de máquinas para execução das aplicações submetidas à grade.

LRM e GRM também colaboram no *Resource Reservation and Execution Protocol*. O GRM seleciona nós candidatos para executar uma aplicação submetida pelo usuário da grade, com base nos requisitos da aplicação e na disponibilidade de recursos. Estas últimas informações coletadas pelos LRMs, nos nós fornecedores de recursos.

De modo semelhante à cooperação LRM/GRM, os módulos *Local Usage Pattern Analyzer* (LUPA) e *Global Usage Pattern Analyzer* (GUPA) também trabalham cooperativamente na análise de padrões de uso das máquinas para contribuir para um escalonamento eficiente na grade [G.04].

O módulo *Node Control Center* (NCC) permite aos proprietários de recursos configurar as condições de compartilhamento, como, por exemplo, o percentual de CPU e memória física disponibilizados para a grade, ou definir quando tais recursos podem ser considerados ociosos e portanto disponíveis.

O módulo *Application Submission and Control Tool* (ASCT) permite aos usuários do InteGrade submeter as aplicações para execução na grade. As aplicações registradas no ASCT são armazenadas em um Repositório de Aplicações e recuperadas pelos LRMs para execução.

Nossa atividade de pesquisa se deterá na extensão das funcionalidades do Repositório de Aplicações atualmente implementado.

3 O Repositório de Aplicações

O Repositório de aplicações implementado atualmente no Integrade é muito simples e sua utilização é satisfatória apenas como prova de conceito. No entanto, o InteGrade pretende dar origem a uma vasta grade de computação com potencial para milhões de máquinas [G.04]. Nesse contexto, o Repositório de Aplicações deve atender a demandas mais sofisticadas.

O Repositório de Aplicações oferece atualmente apenas duas funcionalidades básicas:Registrar uma aplicação e Recuperar uma aplicação. Essas funcionalidades são usadas pelo módulo ASCT quando um usuário da grade deseja executar uma aplicação. O módulo LRM também participa desse caso de uso, intermediando a execução da aplicação. Além de limitadas, tais funcionalidades ainda apresentam problemas na identificação das aplicações registradas.

4 Plano de Trabalho

A abordagem para estender o Repositório de Aplicações será dirigida principalmente para prover qualidade à essas funcionalides básicas. Uma característica desejável seria a utilização de nomes compostos no registro das aplicações, possivelmente provendo uma interface de usuário para navegação pelo espaço de nomes (diretórios) contendo aplicações registradas e armazenadas provavelmente em um sistema de arquivos. Uma outra característica importante é o suporte para execução das aplicações em diferentes sistemas operacionais. Essas duas extensões serão o foco da implementação nessa atividade de pesquisa.

Também serão investigadas as possibilidades de extensões mais sofisticadas, tais como suporte à segurança, controle de versões e mecanismos de replicação do repositório de aplicações. Tais extensões, por serem mais complexas, estão além do escopo possível de implementação para a presente atividade de pesquisa.

4.1 Atividades

1. Estudo do código fonte do repositório de componentes do 2K;
2. Estudo do código fonte do repositório de componentes do GAIA;
3. Estudo do repositório de componentes do GLOBUS;
4. Definição da Interface do Repositório de Aplicações do InteGrade;
5. Implementação do Repositório de Aplicações do InteGrade.

5 Conclusão

Esse plano de estudo apresentou uma visão geral do sistema InteGrade e seu estado atual de implementação, em particular, do Repositório de Aplicações. O foco da atividade de pesquisa foi estabelecido sobre extensões para melhoria desse módulo do InteGrade, contribuindo portanto para a qualidade do sistema como um todo.

Referências

[G.04] Goldchleger A. Kon F. Goldman A. Finger M. Bezerra G. Integrade:oject-oriented Grid middleware leveraging idle computing power of desktop machines. *Concurrency and Computation: Practice and Experience*, 4, 2004.

[K+04a] Fábio Kon et al. Integrade: checkpoints. `http://gsd.ime.usp.br/integrade/documentation/internal/checkpointings/%`, 2004.

[K+04b] Fábio Kon et al. Integrade: Open Research Topics. `http://gsd.ime.usp.br/integrade/documentation/internal/openTopics.html%`, 2004.