

# Utilização de Infra-Estruturas de Grade para Computação de Algoritmos de Bioinformática

Débora Pandolfi Alves, Luciano Paschoal Gaspar, Ney Lemke

Programa Interdisciplinar de Pós-Graduação em Computação Aplicada - Universidade do Vale do Rio dos Sinos  
Av. Unisinos 950, São Leopoldo  
debora@exatas.unisinos.br, paschoal@unisinos.br, ney@unisinos.br

## Resumo

O desenvolvimento da Bioinformática trouxe um novo desafio à área de computação: a execução de algoritmos que requerem alto poder computacional para lidar com uma grande quantidade de dados. Embora existam equipamentos capazes de desempenhar esse papel, o alto custo dos mesmos restringe seu acesso, dificultando o desenvolvimento da pesquisa na área. Este problema pode ser minimizado com a utilização de tecnologias alternativas como a computação em grade, que permite tirar proveito da infra-estrutura já existente nas instituições e que em grande parte do tempo se encontra subutilizada ou ociosa. Neste trabalho descrevemos o esforço realizado no desenvolvimento de algoritmos de Bioinformática para execução em infra-estruturas de grades computacionais.

## 1 Motivação

A necessidade de executar algoritmos que requerem alto poder computacional é um problema que se agrava à medida que mais informações são geradas. Ao mesmo tempo em que é preciso executar experimentos de computação intensa e que lidam com uma grande quantidade de dados, os recursos necessários para tanto não são de fácil acesso pois apresentam custos elevados.

Uma alternativa viável neste caso é aproveitar os recursos ociosos ou subutilizados já disponíveis na instituição, utilizando grades computacionais. Assim é possível reunir os recursos disponíveis na instituição sem que os mesmos tenham que ficar dedicados exclusivamente a essa tarefa.

## 2 Simulador Metabólico

Neste contexto foi desenvolvido o YAMONES (*Yet Another MOlecular NEtwork Simulator*), uma ferramenta para simulação de redes metabólicas. Essa ferramenta simula o comportamento celular a partir de um modelo que pode seguir diferentes caminhos, dependendo das condições do ambiente, processando assim uma grande quantidade de dados.

A simulação é feita calculando-se as probabilidades das reações metabólicas na célula. Esta computação pode ser separada em tarefas que podem ser executadas independentemente, tendo os resultados recuperados ao final da execução. Devido à inexistência de comunicação entre as tarefas da simulação, essa computação é facilmente paralelizável e bastante adequada para execução em grades computacionais.

Para realizar a simulação foi necessário adaptar o YAMONES e configurar uma estrutura de grade que foi montada utilizando duas ferramentas: no primeiro estágio foi utilizado o OurGrid versão 3, e em seguida o Globus Toolkit versão 4.

A execução da simulação sob o OurGrid, requer a criação de arquivos JDF (*Job Description File*), cujo formato é definido pela arquitetura do OurGrid. Estes arquivos descrevem como devem ser executadas as tarefas. No caso do YAMONES, foi necessário descrever quais arquivos deveriam ser transferidos para os nós da grade, quantas simulações deveriam ser feitas (com seus respectivos parâmetros de entrada) e o arquivo que deve ser recuperado ao final da simulação. O OurGrid se encarrega então de transmitir os arquivos, iniciar a simulação em cada nó e distribuir as tarefas para os mesmos.

O Globus, no entanto, requer a criação de um *webservice* que disponibilize as funcionalidades do simulador. Neste caso, o *webservice* já deve estar disponível em cada estação que compõe a grade antes da requisição do serviço. Assim todos os arquivos e bibliotecas necessárias são copiados para as estações onde deseja-se executar a simulação e então o serviço pode ser disponibilizado. Com o auxílio de uma aplicação cliente, é feita a requisição da execução de uma tarefa, a partir da invocação do *webservice* que recebe os parâmetros da tarefa a ser executada.

### 3 Execução da Simulação em Grade

O YAMONES foi executado em uma grade composta por 16 estações Intel Pentium 4 de 1,8GHz com 256MB de RAM do laboratório do Programa de Pós-Graduação em Computação Aplicada da UNISINOS, e um *cluster* computacional formado por 13 estações Intel Xeon 2,4GHz com 1GB de RAM.

Embora a maior parte das tarefas tenha sido executada em nós do *cluster* as estações do laboratório executaram uma porção significativa das tarefas, aproximadamente 25%. Um bom resultado, considerando que o laboratório é de uso comum dos alunos, por tanto raramente foi possível utilizar todas as estações simultaneamente.

### 4 Conclusões e Trabalhos Futuros

Os experimentos realizados confirmaram a eficiência da utilização da grade para a simulação. Devido à natureza das simulações executadas pelo YAMONES, foi possível observar que a execução utilizando o OurGrid foi capaz de aproveitar melhor os recursos da infra-estrutura, uma vez que as tarefas não possuem comunicação. Já no caso do Globus, muitos dos recursos por ele disponibilizados não precisaram ser utilizados. Entretanto o Globus apresenta um potencial de extensão muito maior, sendo possível orquestrar a execução de tarefas muito mais complexas, o que no OurGrid, se tornaria uma tarefa mais complicada devido a necessidade de uma implementação muito complexa.