



Ferramenta para Automatizar o Processo de Avaliação de Desempenho em Aplicações Paralelas

Vinicius Meirelles Pereira, Dalvan Griebler (co-orientador), Luis Gustavo Leão Fernandes (orientador)

*Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS)
Faculdade de Informática (FACIN), Grupo de Modelagem de Aplicações Paralelas (GMAP)*

Resumo

Na programação paralela, a principal meta é diminuir o tempo de execução de programas através de sua divisão em threads ou processos, e aumentar a eficiência de cada processador na execução do algoritmo. As duas principais maneiras de se medir o desempenho de aplicações paralelas são o *speedup* (utilizado para medir o quão mais rápido o novo método é em relação ao algoritmo sequencial) e a eficiência (que verifica o quão bem utilizados foram os processadores nessa tarefa).

Um dos problemas na avaliação do desempenho de algoritmos paralelizados é o volume de dados gerado nos arquivos de log. Isso ocorre porque cada teste deve ser executado pelo menos 30 vezes, de forma que se obtenha um intervalo de confiança aceitável no cálculo das médias. Muitas informações são geradas nos logs, porém os únicos dados relevantes para o cálculo de *speedup* e eficiência são os tempos de execução e o número de threads ou processos.

Para que se obtivesse os resultados desejados, era necessário coletar e filtrar manualmente esses dados em cada teste. Posteriormente, o tempo médio de cada um dos experimentos era calculado, e com ele calculava-se o *speedup* e a eficiência. Esses resultados ainda precisavam ser organizados de forma que a ferramenta de geração de gráficos escolhida pudesse utilizá-los. O GMAP, por exemplo, faz uso da ferramenta Gnuplot, que necessita que os dados estejam em um arquivo “.dat” (arquivo de texto) dividido em colunas.

A fim de acelerar a manipulação da informação gerada pelos testes, foi desenvolvido um programa na linguagem C++ capaz de organizar e calcular todos os dados de forma automática. Primeiramente, os logs são analisados pelo programa, que roda no terminal do

Linux. O usuário insere uma palavra-chave como parâmetro de filtragem, e todos os valores numéricos que a sucedem nos logs são salvos, assim como o número de threads de cada arquivo. Com base nessas informações, o programa calcula o tempo médio de cada teste e gera um arquivo “.dat” contendo-os. Então, com esse arquivo, calcula-se primeiro o *speedup* e depois a eficiência, gerando seus respectivos arquivos “.dat”. As colunas de todos os arquivos são organizadas de forma que o Gnuplot sempre seja capaz de interpretá-las.

O programa criado é uma ferramenta científica bastante útil para testes de desempenho, pois elimina horas de trabalho repetitivo na coleta, análise e cálculo dos dados obtidos. Ele já está sendo utilizado por membros do laboratório, e gerou gráficos para diversos testes.