

Em Direção a Soluções Distribuídas para Balanceamento de Carga Ciente de Comunicação

Vinicius M. C. T. de Freitas¹, Laércio L. Pilla¹, Márcio Castro¹

¹Departamento de Informática e Estatística
Universidade Federal de Santa Catarina (UFSC) – Florianópolis – SC – Brazil

vinicius.mct.freitas@gmail.com, {laercio.pilla,marcio.castro}@ufsc.br

***Resumo.** A distribuição homogênea de carga nos grandes sistemas paralelos atuais é um desafio ligado às aplicações de comportamento imprevisível. O re-escalamento global de tarefas é uma possível solução para este problema, contudo diversos fatores podem ser levados em conta ao realizá-lo. Este trabalho se propõe a explorar tanto o paralelismo das plataformas, quanto carga de tarefas e a comunicação entre elas.*

1. Introdução

A crescente complexidade de simulações computacionais cria uma demanda por soluções paralelas cada vez mais eficientes. Para atingir uma escalabilidade forte dessas aplicações, uma série de fatores deve ser levada em conta. Entre eles, o balanceamento de carga das aplicações paralelas é crucial para garantir a eficiência do sistema. Algumas simulações computacionais possuem carga de trabalho variável e imprevisível [Menon et al. 2013], que tornam mapeamentos estáticos pouco eficientes.

Conforme o tamanho das plataformas computacionais cresce, o desempenho de soluções de escalonamento global centralizadas diminui, uma vez que estas não exploram o paralelismo das plataformas. Estratégias hierárquicas e distribuídas surgem como uma forma de mitigar esse problema. Utilizando informação local e buscando se aproveitar do paralelismo do sistema, essas abordagens tornam-se mais eficientes em larga escala [Menon et al. 2013].

Uma distribuição homogênea da carga de trabalho é o principal objetivo das atuais soluções do estado da arte. Quando a topologia e custos de comunicação são levados em conta, simulações com uma alta taxa de comunicação podem ter o seu desempenho melhorado. Aplicações científicas como simulações de propagação de onda, dinâmica molecular e refinamento adaptativo de malha podem se beneficiar de um balanceamento que não interfira muito em seus custos de comunicação.

Existem diferentes formas de se considerar o problema da comunicação de tarefas. Algumas estratégias de escalonamento utilizam bibliotecas de particionamento de grafos [Bhatele et al. 2011], enquanto outras calculam os benefícios de migrações [Pilla et al. 2012]. Contudo, dentre estas estratégias do estado da arte, nenhuma é capaz de tanto aproveitar o paralelismo do sistema quanto contabilizar diferentes níveis de comunicação.

2. Proposta

Propomos um balanceador paralelo capaz de ponderar carga e comunicação na escolha de carga de trabalho para migração. Essa estratégia deverá determinar quais os melhores

candidatos para migração e entidades de processamento a recebê-las. Novas heurísticas devem ser desenvolvidas para realizar essas tomadas de decisão, baseadas nas métricas apresentadas a seguir.

Nos maiores supercomputadores do mundo, a comunicação deve ser levada em conta através da topologia de rede dessas máquinas. Topologias de árvore e de malha são diferentes entre si, mas ambas podem se beneficiar do cálculo da média de *hops* por byte [Bhatele et al. 2011] para medir custos de comunicação.

Além de fatores simples como *hops/byte* e avaliação de qualidade da migração, fatores mais complexos podem ser aliados a estes. Informações específicas de cada ambiente como largura de banda e *jitter* da rede podem ser usados para avaliar custos de comunicação e migração; assim como informações individuais da aplicação (e.g., número de mensagens enviadas entre nós).

O desenvolvimento dessa nova estratégia de escalonamento global será realizado num sistema paralelo com alto grau de escalabilidade. Isso viabiliza comparações com o estado da arte e uso nativo em *benchmarks* para avaliação de desempenho. O Charm++ [Acun et al. 2016] apresenta o potencial de escalabilidade desejado, além de prover uma série de balanceadores do estado da arte e aplicações para compará-los.

3. Conclusão

A localidade dos dados e processos em uma aplicação paralela é crucial para garantir desempenho, evitando gargalos de execução atrelados à troca de mensagens em arquiteturas multi-computadas [Hoeffler et al. 2014]. Sendo assim, um mapeamento de tarefas só é realmente eficiente se garante uma boa distribuição das cargas de trabalho sem aumentar demais os custos de comunicação.

Apesar do avanço no estado da arte, o contínuo crescimento das plataformas computacionais ainda pode se beneficiar de soluções distribuídas sensíveis aos padrões de comunicação das aplicações e às topologias das plataformas. Nossa nova proposta de escalonamento distribuído ciente de comunicação está inserida no contexto de uma dissertação de mestrado com início em Março de 2018.

Referências

- Acun, B. et al. (2016). Power, reliability, and performance: One system to rule them all. *Computer*, 49(10):30–37.
- Bhatele, A. et al. (2011). Heuristic-based techniques for mapping irregular communication graphs to mesh topologies. In *2011 IEEE International Conference on High Performance Computing and Communications*, pages 765–771.
- Hoeffler, T. et al. (2014). *An Overview of Topology Mapping Algorithms and Techniques in High-Performance Computing*, pages 73–94. John Wiley Sons, Inc.
- Menon, H. et al. (2013). A distributed dynamic load balancer for iterative applications. In *International Conference for High Performance Computing, Networking, Storage and Analysis, SC '13*, pages 15:1–15:11, New York, NY, USA. ACM.
- Pilla, L. L. et al. (2012). A hierarchical approach for load balancing on parallel multi-core systems. In *International Conference on Parallel Processing*, pages 118–127. IEEE.