Thales Alexandre Zirbel Hubner

Abstração da Topologia de Rede para Aplicações Distribuídas

Florianópolis

Thales Alexandre Zirbel Hubner

Abstração da Topologia de Rede para Aplicações Distribuídas

Monografia submetida ao Programa de Graduação em Ciência da Computação para a obtenção do Grau de Bacharel.

Universidade Federal de Santa Catarina

Departamento de Informática e Estatística

Ciência da Computação

Orientador: Vinicius Marino Calvo Torres de Freitas Coorientador: Márcio Bastos Castro

> Florianópolis 2018

FOLHA DE APROVAÇÃO DE PROPOSTA DE TCC Acadêmico(s) Thales Alexandre Zirbel Hubner Título do trabalho Abstração da Topologia de Rede para Aplicações Distribuídas (subtítulo) Ciência da Computação ou Sistemas de Informação /INE/UFSC

Sistemas de Computação/Arquitetura de Sistemas de

Instruções para preenchimento pelo ORIENTADOR DO TRABALHO:

Computação

Curso

Área de Concentração

- Para cada critério avaliado, assinale um X na coluna SIM apenas se considerado aprovado. Caso contrário, indique as alterações necessárias na coluna Observação.

Critérios		Apro			
		Sim Parcial Não		Não se aplica	Observação
O trabalho é adequado para um TCC no CCO/SIN (relevância / abrangência)?	X				
2. O titulo do trabalho é adequado?	X				
O tema de pesquisa está claramente descrito?	×				
O problema/hipóteses de pesquisa do trabalho está claramente identificado?	X		SE CONTRACTOR OF THE PERSON OF		
5. A relevância da pesquisa é justificada?	X				
6. Os objetivos descrevem completa e claramente o que se pretende alcançar neste trabalho?	X				
7. É definido o método a ser adotado no trabalho? O método condiz com os objetivos e é adequado para um TCC?	X				
8. Foi definido um cronograma coerente com o método definido (indicando todas as atividades) e com as datas das entregas (p.ex. Projeto I, II, Defesa)?	X	33.4 30.5 37.7 37.7 37.7 37.7 37.7 37.7 37.7 37			
9. Foram identificados custos relativos à execução deste trabalho (se houver)? Haverá financiamento para estes custos?	X				
10. Foram identificados todos os envolvidos neste trabalho?	X				
11. As formas de comunicação foram definidas (ex: horários para orientação)?	X				
12. Riscos potenciais que podem causar desvios do plano foram identificados?	X				
13. Caso o TCC envolva a produção de um software ou outro tipo de produto e seja desenvolvido também como uma atividade realizada numa empresa ou laboratório, consta da proposta uma declaração (Anexo 3) de ciência e concordância com a entrega do código fonte e/ou documentação produzidos?	X				

Avaliação	⊠ Aprovado	□ Não Aprovado
Professor Responsável	Márcio Bastos Castro	03/07/18
Orientador externo	Vinicius Marino Calvo Torres de Freitas	03/07/2014 Mich State

Resumo

A repartição de trabalho em ambientes distribuídos é um problema a ser tratado, pois aplicações como simulações sísmicas, dinâmica molecular e previsão de tempo possuem comportamento dinâmico, o que gera desbalanceamento de cargas de trabalho no sistema. Uma das maneiras de resolver esse problema é a utilização de balanceadores de carga dinâmicos, cuja função é reduzir o tempo de execução da aplicação através de uma distribuição de tarefas mais homogênea. A execução destes balanceadores gera um sobrecusto relacionado à comunicação para o balanceamento e à migração de tarefas. Poucos balanceadores consideram a topologia do sistema de maneira dinâmica, devido à dificuldade do balanceador de acessar as informações completas sobre a rede e de levá-las em consideração para o remapeamento de tarefas.

Este trabalho visa a implementação de uma abstração da topologia de rede, facilitando o acesso à métricas mais complexas da rede e a utilização destas em aplicações distribuídas. Por cima dessa abstração, será implementado uma estratégia de balanceamento de carga que utilizará a informação de custos heterogêneos de comunicação na rede, reduzindo o sobrecusto de comunicação através do uso de partes mais velozes da rede. Este balanceador será comparado com outros balanceadores de carga que não utilizam desta informação, verificando sua efetividade em cenários heterogêneos e seu sobrecusto em cenários homogêneos.

Palavras-chaves: topologia de rede. balanceamento de carga. aplicações distribuidas.

Lista de abreviaturas e siglas

PE Processing Element, unidade de processamento

LB Load Balancer, balanceador de carga

Sumário

1	PROJETO	Ĉ
1.1	Introdução	Ĉ
1.2	Objetivos	C
1.3	Método de Pesquisa	1
2	PLANEJAMENTO 1	3
2.1	Cronograma	3
2.2	Recursos Humanos	5
2.3	Custos	5
2.4	Comunicação	5
2.5	Riscos	7
	REFERÊNCIAS	Ç

1 Projeto

Este capítulo apresenta o tema de pesquisa do Trabalho de Conclusão de Curso, o escopo no qual o problema em questão será tratado e a justificativa do projeto. Na Seção 1.1 é apresentado o contexto e a motivação para a realização do trabalho. Na Seção 1.2 são introduzidos os objetivos gerais e específicos, as restrições e premissas existentes, a lista de marcos e os critérios de aceite do projeto. Na Seção 1.3, são abordados os métodos de pesquisa envolvidos para alcançar a solução proposta.

1.1 Introdução

Aplicações no âmbito científico e industrial incluem cada vez mais detalhes, demandam precisão ainda maior e/ou possuem uma complexidade muito elevada, criando uma demanda cada vez maior por velocidade e poder de processamento (PILLA, 2014). Uma das maneiras utilizadas para suprir esta demanda é com computação paralela, que alcança a solução destes problemas através da repartição de trabalho entre unidades de processamento (*Processing Elements* ou PE's) e a execução destas parcelas simultâneamente. Este método permite que um aumento no número de PE's leve a uma redução no tempo da aplicação. Contudo a melhoria no tempo é limitada pela parcela não paralelizável do programa e pelos recursos do sistema (AM-DAHL, 1967).

Supercomputadores que utilizam da computação paralela foram criados para alcançar processamento massivo e por décadas estas máquinas cresceram. Elas alcançaram um ponto onde se percebeu que velocidade e poder de processamento não eram as métricas adequadas para desempenho destas máquinas. Estes computadores consomem uma quantidade tremenda de energia e dissipam muito desta energia, se tornando grandes geradores de calor (FENG; CAMERON, 2007). A geração de calor excessiva é um problema em sistema pois temperaturas altas aumentam sua chance de falha e tempo fora de operação. Como contramedida é necessário a implantação de um sistema eficiente de redução de temperatura, que traz outro grande sobrecusto de energia. Com estes problemas em mente a métrica pensada para estas máquinas é a eficiência de processamento, que leva em conta a energia gasta no sistema, a velocidade e o poder de processamento (HSU; FENG; ARCHULETA, 2005).

Uma das opções para obter eficiência de processamento é o uso de um número maior de núcleos menos potentes (SNIR, 2011). A organização de um número tão grande de PE's acaba levando a um sistema com memória distribuída, que operam com o uso de diversas máquinas e utilizam alguma rede para a interligação e comunicação. A escolha destes sistemas é devida ao seu custo baixo e a sua grande escalabilidade através do aumento no número de máquinas. O contraponto destas vantagens é o desenvolvimento destas aplicações, que se torna mais com-

10 Capítulo 1. Projeto

plexo pois têm de levar em consideração fatores impactantes como sincronização, dependência e distribuição de dados, balanceamento de carga e custos de comunicação (PILLA, 2014).

A repartição de trabalho em ambientes distribuídos é um problema a ser tratado, pois aplicações como simulações sísmicas (DUPROS et al., 2010) (TESSER et al., 2014), dinâmica molecular (BHATELÉ; KALÉ; KUMAR, 2009) ou previsão de tempo (RODRIGUES et al., 2010) possuem comportamento dinâmico, ou seja suas cargas mudam ao longo da execução da aplicação, levando um mapeamento homogêneo de tarefas a um estado desbalanceado do sistema, resultando em uma subcarga de alguns PE's. Esta subcarga resulta no sub-uso de processadores e na possibilidade de alguns PE's não terem trabalho para executar, reduzindo o desempenho da aplicação. Para consertar tais desvios pode-se usar um balanceador de carga (*Load Balancer* ou LB) dinâmico, cujo trabalho é remapear as tarefas entre os PE's, buscando um estado mais balanceado do sistema. Este rearranjo garante a utilização dos PE's de maneira mais homogênea e leva a uma redução do tempo da aplicação.

Os PE's em sistemas distribuídos se comunicam usando diversas topologias de rede, como malhas, tori ou *fat-trees*. Os PE's utilizam de *links* de conexão da rede para efetuar sua comunicação. O compartilhamento destes pode resultar num congestionamento da rede devido ao sobreuso de alguns *links* (BATHELE, 2011). Outro fator é a possibilidade da rede não ter ligações com custos iguais comunicação (BHATELE et al., 2016), tendo variações de latência e velocidade em partes distintas da rede. O impacto disto é que os caminhos menos custosos entre PE's da rede nem sempre são os menores em questão de *links* percorridos.

Os caminhos utilizados dentro de uma rede têm impacto no tempo de execução de um programa, pois caminhos mais lentos afetam o sobrecusto de comunicação. É possível que um LB leve em conta os custos de comunicação entre os PE's e os custos de movimentação destas tarefas dentro da rede. Tal abordagem pode levar a uma redução do custo de comunicação, evitando a troca de informação e tarefas através de partes lentas da rede.

1.2 Objetivos

Os objetivos deste trabalho são de desenvolver uma abstração da topologia de rede, facilitando seu uso para outras aplicações e propor uma estratégia de balanceamento de carga ciente desta abstração. Para este fim, seguem os objetivos específicos:

- 1. Desenvolver um método de abstração dinâmica da topologia de rede para a plataforma Charm++ (ILLINOIS, 2017);
- Desenvolver um LB que utilize esta base, visando a otimização para redes com custos de hops heterogêneos;

3. Testar e comparar este balanceador com outros LB's, dentro de sistemas com topologia de rede com custos de hop heterogêneo.

Premissas:

- Um computador estará disponível para realização do trabalho;
- O orientador terá disponibilidade para reuniões periódicas;
- Disponibilidade de energia e internet.
- Acesso a uma máquina com uma topologia diferenciada para a realização dos testes.

Marcos:

- Entrega do resumo em TCC I: 4ª semana de Novembro/2018;
- Entrega da abstração da topologia: 4ª semana de Novembro/2018;
- Entrega da implementação do balanceamento de carga: 4ª semana de Abril/2019;
- Entrega da primeira versão da monografia em TCC II: 2ª semana de Maio/2019;
- Defesa da monografia: 2ª semana de Junho/2019;
- Entrega da versão final da monografia em TCC II: 4ª semana de Junho/2019.

Critérios de aceite:

- Aprovação da banca avaliadora;
- Aprovação do orientador;
- Aprovação do coorientador;
- Conformidade da monografia com as normas definidas pela instituição;
- Prazos cumpridos.

1.3 Método de Pesquisa

O ínicio do trabalho terá duas partes: uma de cunho teórico, onde será estudada a base e o estado da arte para o projeto, e outra mais prática, que envolve o aprendizado do *framework* Charm++. A primeira parcela será constituída por um estudo sobre topologia de rede e balanceamento de carga em aplicações paralelas, criando uma base para o projeto. A segunda parte

12 Capítulo 1. Projeto

terá foco no entendimento da plataforma e suas nuances, visando facilitar a implementação da abstração.

A criação da abstração da topologia de rede utilizará a linguagem de programação C++ e será realizada dentro do *framework* de programação paralela Charm++. A plataforma será utilizada pois têm uma base sólida para a criação, utilização e teste de balanceadores de carga e outras aplicações paralelas (PILLA; MENESES, 2015). Além disto Charm++ contém somente abstrações de topologia fixas, carecendo de uma abstração elaborada da topologia de rede, proposta neste trabalho.

Após a criação e teste da abstração da topologia de rede, será desenvolvido um LB que utiliza desta abstração para reduzir o custo de comunicação dentro de uma rede com custos de comunicação heterogêneos. O balanceador também será realizado em C++ na plataforma Charm++. Serão executados testes de funcionamento e avaliação de desempenho.

Na última etapa será realizado uma análise de desempenho do LB em comparação com outros LB's, verificando sua efetividade e seu sobrecusto como balanceador de carga. Para tal, serão realizados experimentos utilizando algum *benchmark* que simule aplicações reais. O teste será realizado em um cenário onde a topologia tem custos de comunicação heterogêneos e outra com custos homogêneos. Também será realizado um teste para verificar o sobrecusto do LB em um cenário onde não há benefícios em balanceamento de carga.

2 Planejamento

Este capítulo contém o planejamento do projeto onde cada parte do plano de gerenciamento se encontra em seções diferentes. A Seção 2.1 apresenta as atividades planejadas e seu cronograma a ser seguido ao longo da realização deste trabalho. A Seção 2.2 apresenta os recursos humanos envolvidos. A Seção 2.3 indica os custos estimados para a execução do projeto. A Seção 2.4 dispõe o gerenciamento da comunicação entre as partes envolvidas. A Seção 2.5 apresenta os riscos identificados e suas respectivas estratégias.

2.1 Cronograma

As atividades previstas no projeto estão descritas abaixo:

- A1: Estudo da fundamentação teórica. Nesta parte inicial será realizada a revisão de artigos e materiais relacionados a computação paralela, topologia de rede e balanceamento de carga.
- **A2:** Familiarização com a plataforma charm++. Processo de familiarização com a plataforma para programação paralela a ser utilizada.
- A3: Revisão do estado da arte e prática. Nesta etapa será tratada a revisão do estado da arte envolvido no contexto do trabalho, a fim de fortalecer a base do conhecimento necessária para a realização do mesmo.
- A4: Elaboração da proposta. Nesta etapa será apresentada a proposta de solução do problema em questão, bem como suas estratégias de implementação.
- A5: Escrita do relatório do TCC I. Nesta etapa será realizada a escrita do relatório do TCC I. A entrega deste documento está prevista para a quarta semana do mês de novembro.
- A6: Implementação da abstração da topologia de rede. Nesta parcela será realizada a implementação da abstração proposta assim como a verificação do funcionamento desta.
- A7: Teste da abstração da topologia de rede. Nesta parcela será realizada a verificação do funcionamento da abstração.
- A8: Desenvolvimento de um balanceador de carga. Nesta etapa será efetuada o desenvolvimento de um balanceador de carga que utilize a abstração de rede criada.
- **A9: Testes e comparações de desempenho.** Nesta etapa será efetuado testes de desempenho do balanceador de carga a a comparação deste com outros balanceadores existentes.

- A10: Escrita do rascunho do TCC II. A escrita do rascunho do TCC II será realizada neste período. A entrega deste documento está prevista para a segunda semana do mês de maio.
- A11: Preparação da defesa pública. Nesta etapa será realizada a preparação da apresentação oral e visual do conteúdo deste trabalho para a defesa pública.
- A12: Defesa pública. Nesta etapa será realizada a defesa do projeto desenvolvido. Pretendese realizar a defesa pública do trabalho na primeira semana do mês de junho.
- A13: Correções e entrega da versão final do TCC. Nesta etapa serão realizadas as correções e os ajustes da monografia e a entrega final do documento. A entrega da versão final está prevista para a quarta semana do mês de junho.

A Figura 1 apresenta o cronograma previsto para a realização das atividades descritas anteriormente. As atividades estão distribuídas ao longo do primeiro e segundo semestres de 2018 e o primeiro semestre de 2019.

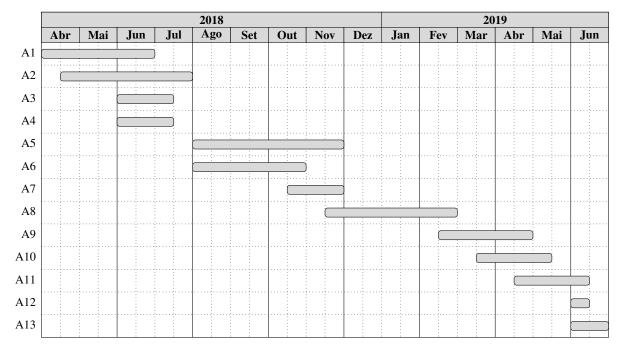


Figura 1 – Cronograma de atividades.

2.2. Recursos Humanos 15

2.2 Recursos Humanos

Papel	Nome
Orientador	Vinicius Marino Calvo Torres de Freitas
Coorientador	Márcio Bastos Castro
Coordenador	Renato Cislaghi
Membro da Banca I	
Membro da Banca II	
Autor	Thales Alexandre Zirbel Hubner

2.3 Custos

Estimativas para Recursos Humanos						
Nome	Data Início	Data Fim	Hora/Mês	Valor/Hora	Custo Total	
Autor	01/04/2018	01/06/2019	40	R\$ 15,00	R\$ 8400,00	
Orientador	01/04/2018	01/06/2019	4	R\$ 20,00	R\$ 1120,00	
Coorientador	01/04/2018	01/06/2019	4	R\$ 71,00	R\$ 3976,00	
Coordenador	01/04/2018	01/06/2019	1	R\$ 102,00	R\$ 1428,00	
Membro da Banca I	01/05/2019	01/06/2019	1	R\$ 60,00	R\$ 60,00	
Membro da Banca II	01/05/2019	01/06/2019	1	R\$ 60,00	R\$ 60,00	
Subtotal estimativas para recursos humanos					R\$ 15044,00	

Estimativas para Recursos Não Humanos						
Descrição	Data Início	Data Fim	Quantidade	Valor Unitário	Custo Total	
Computador para uso	04/18	06/19	1	R\$ 2500,00	R\$ 2500,00	
CDs para o código desenvolvido	06/19	06/19	2	R\$ 4,00	R\$ 8,00	
Impressão para o relatório	12/18	12/18	3	R\$ 20,00	R\$ 60,00	
Subtotal estimativas para recursos não humanos					R\$ 2568,00	

2.4 Comunicação

O que precisa ser comunicado	Reuniões periódicas com o Orientador e Coorientador
Emissor	Thales Alexandre Zirbel Hubner
Receptor	Vinicius Marino Calvo Torres de Freitas, Márcio Bas-
	tos Castro
Comunicação	Reuniões periódicas com o Orientador para acompa-
	nhamento do projeto
Forma de comunicação	Pessoalmente
Frequência ou Quando	Quinzenalmente

O que precisa ser comunicado	Entrega da Proposta
Emissor	Thales Alexandre Zirbel Hubner
Receptor	Renato Cislaghi
Comunicação	Entrega da proposta completa do TCC
Forma de comunicação	Sistema de TCC
Frequência ou Quando	Única vez

O que precisa ser comunicado	Entrega do Relatório em TCC I
Emissor	Thales Alexandre Zirbel Hubner
Receptor	Renato Cislaghi
Comunicação	Entrega da primeira parte da monografia
Forma de comunicação	Sistema de TCC
Frequência ou Quando	Única vez

O que precisa ser comunicado	Entrega da primeira versão da monografia completa
	em TCC II
Emissor	Thales Alexandre Zirbel Hubner
Receptor	Renato Cislaghi
Comunicação	Entrega da primeira versão da monografia completa
Forma de comunicação	Sistema de TCC
Frequência ou Quando	Única vez

O que precisa ser comunicado	Defesa do TCC		
Emissor	Thales Alexandre Zirbel Hubner		
Receptor	Vinicius Marino Calvo Torres de Freitas, Márcio Bas-		
	tos Castro, Membro da Banca I, Membro da Banca II		
Comunicação	Realização da defesa do TCC aos membros da banca.		
Forma de comunicação	Pessoalmente		
Frequência ou Quando	Única vez		

O que precisa ser comunicado	Entrega da versão final da monografia		
Emissor	Thales Alexandre Zirbel Hubner		
Receptor	Renato Cislaghi		
Comunicação	Entrega da monografia com os ajustes feitos após a		
	defesa		
Forma de comunicação	Sistema de TCC		
Frequência ou Quando	Única vez		

2.5. *Riscos*

2.5 Riscos

Nome	Probabilidade	Impacto	Exposição	Estratégia	Ações de Prevenção
Perda de arquivos	Baixa	Alto	Média	Mitigar	Realizar backups com frequencia semanal e utilizar armazenamento em nuvem.
Alteração no cronograma	Baixa	Alto	Média	Aceitar	-
Alteração no escopo	Baixa	Alto	Média	Aceitar	-
Problemas de saúde do Orientando	Média	Baixo	Baixa	Aceitar	-
Indisponibilidade do Orientador/Coorientador	Média	Médio	Média	Aceitar	-
Indisponibilidade da plataforma para testes	Média	Alto	Alta	Mitigar	Contato com responsáveis pela plataforma e agendamento de execuções.

Referências

- AMDAHL, G. M. Validity of the single processor approach to achieving large scale computing capabilities. *AFIPS Conference Proceedings*, AFIPS, Atlantic City, N.J., v. 12, n. 3, 1967.
- BATHELE, A. Topology aware task mapping. In: PADUA, D. (Ed.). *Encyclopedia of Parallel Computing*. Boston, MA: Springer US, 2011. p. 2025–2025. ISBN 978-0-387-09766-4.
- BHATELE, A. et al. Analyzing network health and congestion in dragonfly-based supercomputers. In: 2016 IEEE International Parallel and Distributed Processing Symposium (IPDPS). Chicago, IL, USA: IEEE, 2016. p. 93–102. ISSN 1530-2075.
- BHATELÉ, A.; KALÉ, L. V.; KUMAR, S. Dynamic topology aware load balancing algorithms for molecular dynamics applications. In: *Proceedings of the 23rd International Conference on Supercomputing*. New York, NY, USA: ACM, 2009. (ICS '09), p. 110–116. ISBN 978-1-60558-498-0.
- DUPROS, F. et al. High-performance finite-element simulations of seismic wave propagation in three-dimensional nonlinear inelastic geological media. *Parallel Computing*, v. 36, p. 308–325, 2010.
- FENG, W. c.; CAMERON, K. The green500 list: Encouraging sustainable supercomputing. *Computer*, v. 40, n. 12, p. 50–55, Dec 2007. ISSN 0018-9162.
- HSU, C. H.; FENG, W. C.; ARCHULETA, J. S. Towards efficient supercomputing: a quest for the right metric. In: *19th IEEE International Parallel and Distributed Processing Symposium*. Denver, CO, USA: IEEE, 2005. p. 8 pp.—. ISSN 1530-2075.
- ILLINOIS, U. of. *Parallel Languages/Paradigms: Charm*++. 2017. Disponível em: http://charm.cs.illinois.edu/research/charm.
- PILLA, L. L. *Topology-Aware Load Balancing for Performance Portability over Parallel High Performance Systems*. Tese (Doutorado) Université de Grenoble, 2014.
- PILLA, L. L.; MENESES, E. Programação Paralela em Charm++. 2015.
- RODRIGUES, E. R. et al. A comparative analysis of load balancing algorithms applied to a weather forecast model. In: *2010 22nd International Symposium on Computer Architecture and High Performance Computing*. Petropolis, Brasil: IEEE, 2010. p. 71–78. ISSN 1550-6533.
- SNIR, M. Distributed-memory multiprocessor. In: PADUA, D. (Ed.). *Encyclopedia of Parallel Computing*. Boston, MA: Springer US, 2011. p. 574–578. ISBN 978-0-387-09766-4.
- TESSER, R. K. et al. Improving the performance of seismic wave simulations with dynamic load balancing. In: 2014 22nd Euromicro International Conference on Parallel, Distributed, and Network-Based Processing. Torino, Italy: IEEE, 2014. p. 196–203. ISSN 1066-6192.