



PROJETO DE PESQUISA

IDENTIFICAÇÃO DO PROJETO

Nome do candidato: Vítor Alano de Ataides

Título do Projeto: Um estudo sobre escalonamento consciente de energia em sistemas heterogêneos na nuvem

Linha de Pesquisa a que se insere o Projeto*: Computação paralela e distribuída

Palavras-chave (até 3): Escalonamento consciente de energia Computação em Nuvem sistemas heterogêneos

*Deve ser informada a Linha de Pesquisa de maior prioridade indicada no Formulário de Inscrição.

JUSTIFICATIVA E MOTIVAÇÃO

O escalonamento conciente de energia em computação recentemente se tornou popular não somente para sistemas embarcados com a finalidade de aumentar a vida das baterias, mas também em grandes sistemas constituídos de várias unidades de processamento com o objetivo de reduzir o consumo de energia também associado ao custo de resfriamento. Visto que processadores consomem um grande percentual da energia em sistemas computacionais, especialmente em sistemas embarcados, muito tem se trabalhado no gerenciamento do consumo de energia para processadores. Baseados na técnica **Dynamic Voltage Scaling**, sistemas de gerenciamento de energia para processadores de um núcleo tem sido intensamente explorados. Poucos trabalhos tem focado em gerenciamento de energia em sistemas tempo-real paralelos e distribuídos. A motivação desse trabalho é focar nesse tipo de sistema, mais especificamente em sistemas distribuídos na nuvem.

A Computação em Nuvem é um paradigma de computação distribuída que tem conquistado um significativo espaço e destaque dentre as tecnologias que dão suporte e promovem o atual cenário da tecnologia da informação. E isso ocorre principalmente em virtude de duas principais necessidades desse cenário: redução de custos e diminuição da complexidade na manutenção e administração das infraestruturas que o suportam. A Computação em Nuvem busca atender essas necessidades através da otimização do uso de recursos computacionais [1].

O modelo de negócio oferecido pela Computação em Nuvem utiliza-se das premissas e tecnologias da Computação em Grade e da Computação Utilitária, propõe-se a fornecer infraestrutura computacional, plataforma de desenvolvimento e software aplicativo na forma de serviços altamente escaláveis, dinâmicos e entregues sob demanda aos usuários finais [2]. O acesso a esses serviços pode, então ser cobrado de seus consumidores, estabelecendo assim relação de consumo regulada através de Acordos de Nível de Serviço, o qual estabelece os níveis aceitáveis de qualidade de um serviço entregue a um consumidor por um prestador de serviço [3].

Para garantir que esse modelo consiga atender as expectativas de seus usuários quanto a escalabilidade e a disponibilidade de seus serviços, a Computação em Nuvem apoia-se em mecanismos que contribuam para a flexibilidade de operação e manutenção do ambiente que dão suporte aos seus serviços. Um exemplo disso é o uso da virtualização como estratégia de provimento de capacidade computacional e armazenamento. Com ela os recursos podem ser rapidamente alocados e desalocados conforme a demanda, fornecendo assim a noção de elasticidade e possibilitando expandir serviços para grandes escalas, mesmo com o rápido aumento na demanda.

Com essas premissas e com amadurecimento de suas tecnologias, a Computação em Nuvem tem se expandido e atraído a atenção de usuários e empresas de várias áreas, buscando nela reduzir os custos associados a operação e manutenção das infraestruturas que dão suporte a atual tecnologia de informação. Além disso, a comunidade científica também tem mostrado interesse na Computação em Nuvem, não só pela redução de custos, mas também pela capacidade de expansão do poder computacional disponível, em um cenário que precisa lidar com um grande volume de informações.

No entanto, sua natureza compartilhada e dinâmica em conjunto com as necessidades de respeitar o acordo de nível de serviço e de atender os requisitos de qualidade de serviço, tem contribuído para gerar diversos desafios quanto ao uso e gerenciamento de um ambiente típico de nuvem, principalmente aqueles relacionados ao uso mais eficiente de seus recursos e o consumo de energia.

Fala um pouco de escalonamento antes, cita que hardware pode ser heterogeneo (Intel Xeon Phi e GPUs), depois: Desta forma, torna-se importante o desenvolvimento de técnicas de escalonamento para sistemas heterogeneos em nuvem bla bla bla

ESTADO DA ARTE

Muitas técnicas tanto para hardware quanto para software tem sido propostas para reduzir o consumo de energia em sistemas distribuídos, como por exemplo desligar componentes que não estão sendo utilizados e projetos de circuitos de baixa energia. Com a tecnologia CMOS, o poder de processamento é dominado pela dissipação dinâmica de energia que é determinada pela tensão do processador e pela frequência do clock [3]. Reduzindo a frequência do clock e a tensão, reduzimos o consumo de energia perdendo performance. Processadores que utilizam da técnica Dynamic Voltage Scaling estão sendo comercializados atualmente [4]. Existe uma troca muito interessante entre consumo de energia e performance dos processadores, especialmente para sistemas tempo-real que as vezes a alta performance é necessária para atingir as restrições de tempo.

Para sistemas com um processador, Weiser et al. inicialmente discute o problema de agendar tarefas para reduzir o consumo de energia em processadores [5]. Yao et al. Descreveu um algoritmo de agendamento off-line para tarefas independentes executadas com velocidades diferentes, assumindo o pior caso de tempo execução [6]. Kunar et al. fez previsão do tempo de execução de tarefas baseado nas estatísticas colhidas das execuções anteriores da mesma tarefa [7].

Para SOCs (system-on-chip) com dois processadores executando a níveis fixos e diferentes de voltagem, Yeang et al. propôs um agendador de duas-fases que minimiza o consumo de energia enquanto alcança as restrições de tempo escolhendo diferentes opções de agendamento em tempo de compilação [8]. Em [9] Zhang et al. propôs um mapeamento baseado em prioridade, agendamento para um mapa de tarefas fixas e formulou o problema de escalonamento de voltagem como um problema de programação linear.

Tem que falar do estado-da-arte em escalonamento de cloud. Tens refs? Senão te passo

OBJETIVOS E RESULTADOS

Os objetivos desse projeto são:

- Estudo sobre escalonamento consciente de energia.
- Estudo dos sistemas heterogêneos.
- Estudo das técnicas atuais para escalonamento de energia consciente em sistemas heterogêneos.
- Aplicação de técnicas estudadas em sistemas heterogêneos na nuvem.
- Analise dos resultados

objetivo principal: desenvolver escalonamento consciente de energia para nuvens que considere o hardware heterogêneo, inclusive aceleradores como GPU e Phi

O resultado esperado é um amplo estudo sobre as técnicas de escalonamento de energia em sistemas heterogêneos. Bem como uma análise dessas técnicas aplicadas em sistemas na nuvem.

MÉTODOS E MATERIAIS

Para o desenvolvimento desse trabalho inicialmente será feita uma revisão bibliográfica onde será estudado técnicas atuais de escalonamento de energia, sistemas heterogêneos e computação em nuvem. Em seguida será feita uma análise de quais técnicas de escalonamento de energia poderão ser utilizadas em sistemas heterogêneos na nuvem. Por fim essas técnicas serão implementadas e então sua eficiência será testada para conclusão do estudo.

Para que esse trabalho possa ser desenvolvido será necessário um computador capaz de realizar computação heterogênea. A alternativa que mais se destaca é a utilização de GPUs (Graphics Processing Units) para acelerar cálculos altamente paralelizáveis. Para isso seria necessário uma placa gráfica capaz de executar tarefas de propósito geral, temos vários exemplos como os modelos da NVIDIA que utilizam a plataforma CUDA (Compute Unified Device Architecture) e também temos as placas gráficas da AMD que fazem uso da plataforma OpenCL.

Além disso seria necessário a utilização de um serviço na nuvem para realização dos testes em ambiente real.

podes colocar alguma coisa de ambiente experimental, Openstack e CUDA.

REFERENCIAL BIBLIOGRÁFICO

- [1] Foster, I.; Kesselman, C.; Tuecke, S. *The Anatomy of the Grid – Enabling Scalable Virtual Organizations. International Journal of Supercomputer Applications*, [S.l.], v.15, p.2001, 2001.
- [2] Fox, A.; Griffith, R. et al. *Above the clouds: A Berkeley view of cloud computing. Dept. Electrical Eng. and Comput. Sciences, University of California, Berkeley, Tech. Rep. UCB/EECS, [S.l.], v.28, 2009.*
- [3] FOX, A.; GRIFFITH, R. et al. *Above the clouds: A Berkeley view of cloud computing. Dept. Electrical Eng. and Comput. Sciences, University of California, Berkeley, Tech. Rep. UCB/EECS, [S.l.], v.28, 2009.*
- [3] T. D. Burd and R. W. Brodersen. *Energy efficient cmos microprocessor design. In Proc. of The HICSS Conference, pages 288–297, Maui, Hawaii, Jan. 1995.*
- [4] <http://developer.intel.com/design/intelxscale/>.
- [5] M. Weiser, B. Welch, A. Demers, and S. Shenker. *Scheduling for reduced cpu energy. In Proc. of The First USENIX Symposium on Operating Systems Design and Implementation, pages 13–23, Monterey, CA, Nov. 1994.*
- [6] Yao, A. Demers, and S. Shenker. *A scheduling model for reduced cpu energy. In Proc. of The 36 th Annual Symposium on Foundations of Computer Science, pages 374–382, Milwaukee, WI, Oct. 1995.*
- [7] P. Kumar and M. Srivastava. *Predictive strategies for low-power rtos scheduling. In Proc. of the 2000 IEEE International Conference on Computer Design: VLSI in Computers and Processors, Austin, TX, Sep. 2000.*
- [8] P. Yang, C. Wong, P. Marchal, F. Catthoor, D. Desmet, D. Kerkest, and R. Lauwereins. *Energy-aware runtime scheduling for embedded-multiprocessor socs. IEEE Design & Test of Computers, 18(5):46–58, 2001.*
- [9] Y. Zhang, X. Hu, and D. Z. Chen. *Task scheduling and voltage selection for energy minimization. In Proc. of The 39 th Design Automation Conference, pages 183–188, New Orleans, LA, Jun. 2002.*

(assinatura)

/ /
(data)

Informações complementares:

- O Projeto de Pesquisa será considerado no contexto da Linha de Pesquisa indicada, mesmo quando o candidato for selecionado para uma linha indicada com menor prioridade.
- A análise deste Projeto tem por objetivo identificar a vocação e os conhecimentos do candidato para estudos em nível de mestrado.
- O Projeto de Pesquisa não deve ultrapassar TRÊS páginas. A fonte utilizada deve ter tamanho entre 10 e 12 pontos e espaçamento simples.
- Para elaboração do Projeto de Pesquisa, os docentes do PPGC poderão ser consultados. No entanto, isso não implica que o candidato, caso selecionado, irá desenvolver no seu curso de mestrado o projeto apresentado durante o Processo Seletivo nem que será orientado pelo(s) docente(s) consultado(s).

Outras informações, contactar:

E-mail: secretaria-ppgc@inf.ufpel.edu.br – Web: <http://ppgc.ufpel.edu.br>
Telefone: (53) 39.21.13.27