

289626

Arquitetura de Computadores -
Processamento distribuído SBO/II

CNPq 1.03.03.006

Uma proposta de escalonamento de recursos para *Pervasive Computing*

Rodrigo Araujo Real¹

Adenauer Yamin²

Iara Augustin³

Luciano Cavalheiro da Silva⁴

Gustavo Frainer⁵

Cláudio Fernando Resin Geyer⁶

Resumo

Os próximos anos serão caracterizados por elevados níveis de mobilidade, heterogeneidade e interação entre dispositivos conectados a redes de abrangência global. Estas redes interligadas utilizarão tanto conexões cabeadas como sem fio. As primeiras pesquisas envolvendo sistemas distribuídos em redes de larga escala responderam a diversas questões pertinentes ao acesso a recursos, contudo existem lacunas no que diz respeito ao gerenciamento dinâmico dos mesmos. Neste trabalho, é introduzida uma proposta direcionada a *Pervasive Computing* para escalonamento de recursos no Projeto ISAM, empregando uma estratégia baseada em Redes de Bayes.

1 Introdução

A disseminação crescente da Internet, somada ao aumento da velocidade operacional das redes de computadores e das suas interconexões, levam uma perspectiva de uso unificado dos recursos distribuídos, o qual pode ser realizado a partir de qualquer equipamento das redes interconectadas. Esta nova forma de processamento em rede assume diferentes perspectivas: *Grid Computing* (FOSTER; KESSELMAN, 1999), *Internet Computing* (REAL; FILHO; YAMIN, 2001) e mais recentemente *Peer-to-Peer Computing* (O'REILLY & ASSOCIATES, 2001).

Por sua vez, a Computação Móvel é uma nova perspectiva que amplia o conceito de rede-sem-fio. Nesta perspectiva, o usuário portando dispositivos móveis como *palmtops* e

¹rreal@inf.ufrgs.br Apoio CAPES/DELL/FINEP

²adenauer@inf.ufrgs.br bolsista CAPES

³august@inf.ufrgs.br bolsista CAPES

⁴lucc@inf.ufrgs.br bolsista CNPq

⁵frainer@inf.ufrgs.br bolsista CNPq

⁶geyer@inf.ufrgs.br

notebooks, independentemente da sua localização física, terá acesso a uma infra-estrutura de serviços (NOBLE, 2000).

Integrando estas duas visões, observa-se um movimento em direção à *Pervasive Computing* (SATYANARAYANAN, 2001), criando aplicações com novas funcionalidades, as quais tentam aumentar sua usabilidade e atender às necessidades dos usuários que se deslocam. *Pervasive Computing* é a proposta de um novo paradigma computacional, que permite ao usuário o acesso ao seu ambiente computacional a partir de qualquer lugar, a qualquer tempo, usando vários tipos de dispositivos (móveis ou não). Em contraste com a premissa dos sistemas distribuídos de fornecer transparência da distribuição para os usuários, esta nova classe de aplicações é *mobility-aware* e tenta tirar vantagem desta informação (AUGUSTIN, 2002). Para alcançar esta consciência, a aplicação ou o ambiente de execução pró-ativamente monitoram e controlam as condições do ambiente. A aplicação ou o sistema reagem às alterações no ambiente através do processo de adaptação. Este processo requer a existência de múltiplos caminhos de execução para uma aplicação, ou configurações alternativas, as quais exibem diferentes perfis de utilização (históricos) dos elementos computacionais.

Esta visão apresenta uma série de novos (e renovados) desafios, que começam a ser abordados, oriundos do dinamismo e heterogeneidade do ambiente, além dos novos requisitos das aplicações no estilo *follow-me* da *Pervasive Computing*. Tanto a aplicação quanto o sistema de execução operam em um ambiente cujas condições na disponibilidade e no acesso aos recursos são variáveis no tempo e no espaço. Como consequência, novos tipos de aplicações estão aparecendo, as quais têm um comportamento determinado pela sua sensibilidade à variação nas condições de alguns elementos do ambiente. O ambiente é definido por elementos computacionais que podem ser medidos, como largura de banda, latência da rede, consumo de energia, localização do usuário, preferências do usuário, entre outros (AUGUSTIN, 2002). Davies (DAVIES, 1997) acredita que é somente através de um processo de adaptação, fornecendo às aplicações informações gerenciadas sobre trocas na sua infra-estrutura de suporte, que se pode operar eficientemente em um ambiente distribuído altamente dinâmico.

Sistemas distribuídos tradicionais são construídos com suposições sobre a infra-estrutura de execução, como conectividade permanente e disponibilidade dos recursos necessários. Porém, essas suposições não são válidas na *Pervasive Computing* (SATYANARAYANAN, 2001).

Neste artigo será introduzido o mecanismo para escalonamento de recursos baseado no uso de redes bayesianas proposto para o ISAM. Seus objetivos são manter reduzidos os custos pertinentes ao escalonamento, sem inviabilizar a qualidade dos mapeamentos gerados.

2 O Contexto da Pesquisa

Para tratar das questões relativas a mobilidade lógica e física da *Pervasive Computing*, o projeto de pesquisa ISAM (ISAM, 2002), em andamento no II/UFRGS, propõe um *middleware* voltado para o gerenciamento de recursos em redes heterogêneas, com suporte a mobilidade de software e hardware, adaptação dinâmica e com as aplicações modeladas uti-

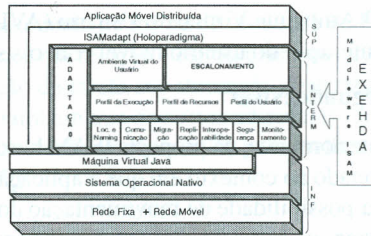


Figura 1: Arquitetura de Software do ISAM

lizando componentes. A estratégia consiste de um ambiente integrado: (i) que oferece um paradigma de programação direcionado aos objetivos e seu respectivo ambiente de execução; e (ii) que gerencia o processo de adaptação através de um modelo colaborativo multinível, no qual tanto o ambiente de execução, como a aplicação, participam das decisões adaptativas. Os dois focos que norteiam os trabalhos são: (i) como construir interfaces de programação que isolem o usuário da complexidade dos contextos de execução modernos, simplificando ao máximo o esforço de programação; (ii) como deve ser um middleware para potencializar o desempenho do processamento de uma aplicação modelada com estas interfaces de programação, considerando a elevada dinamicidade dos contextos de execução atuais.

O ponto de equilíbrio entre estes dois focos é uma questão central da pesquisa em andamento no projeto ISAM (Infra-estrutura de Suporte às Aplicações Móveis Distribuídas).

2.1 A Arquitetura de Software ISAM

A arquitetura proposta é organizada em camadas com níveis diferenciados de abstração e está direcionada para a busca da manutenibilidade da qualidade de serviço, oferecida ao usuário móvel, através do conceito de adaptação. Uma visão organizacional desta arquitetura é apresentada na figura 1. Salientam-se dois pontos: (i) a adaptação que permeia todo o sistema e por isto está colocada em destaque; e (ii) o escalonador que é o núcleo da arquitetura (*middleware* ISAM). A camada superior (SUP) da arquitetura é composta pela aplicação móvel distribuída. A construção desta aplicação baseia-se nas abstrações do Holo-paradigma (BARBOSA, 2002), as quais permitem expressar mobilidade, acrescidas de novas abstrações para expressar adaptabilidade (ISAMadapt).

Camada Intermediária - Primeiro Nível

As atividades de pesquisa pertinentes a camada intermediária (INTERM) estão sendo desenvolvidas no escopo da proposta EXEHDA, compreendendo três níveis de abstração. O primeiro nível é composto por dois módulos de serviço à aplicação: Escalonamento e Ambiente Virtual do Usuário. O Escalonamento, por sua vez, é o componente chave da adaptação na arquitetura ISAM. Fornecer heurística para atuação do escalonamento é o ponto focal deste

trabalho para a arquitetura. O Ambiente Virtual do Usuário (AVU) compõe-se dos elementos que integram a interface de interação do usuário móvel com o sistema.

Camada Intermediária - Segundo Nível

Como já caracterizado anteriormente, na proposta ISAM busca-se um conceito flexível de adaptação, o qual está relacionado ao contexto em que a aplicação está inserida. Por sua vez, a mobilidade física introduz a possibilidade de movimentação do usuário durante a execução de uma aplicação. Dessa forma, os recursos disponíveis podem se alterar, tanto em função da área de cobertura e heterogeneidade das redes, quanto em função da disponibilidade dos recursos devido à alta dinamicidade do sistema. Assim, a localização corrente do usuário determina o contexto de execução, definido como toda informação, relevante para a aplicação, que pode ser obtida e usada para definir seu comportamento (AUGUSTIN, 2001).

As informações que dirigem as decisões do escalonador e dão suporte à decisão de adaptação da aplicação são advindas de quatro fontes: perfil da execução, perfil dos recursos, perfil do usuário e da aplicação (ISAMadapt). O módulo de monitoramento do ISAM, situado no terceiro nível da Camada Intermediária permite determinar a evolução histórica e quantitativa das entidades monitoradas. A interpretação destas informações estabelece o perfil do usuário e das aplicações. Desta forma, as aplicações móveis ISAM poderão se adaptar à dimensão pessoal, além das dimensões temporal e espacial presentes nos demais sistemas móveis.

Camada Intermediária - Terceiro Nível

No terceiro nível da Camada Intermediária estão os serviços básicos do ambiente de execução da plataforma ISAM, os quais provêm a funcionalidade necessária para o segundo nível. É no terceiro nível que estão os serviços disponibilizados que cobrem aspectos relativos as tarefas de instanciação remota e migração de objetos, comunicação e monitoramento (SILVA, 2001). Integram ainda o terceiro nível da Camada Intermediária serviços que tratam da localização de recursos, replicação, interoperabilidade etc.

A integração das funcionalidades providas pelo escalonador à plataforma ISAM dá-se através do EXEHDA, o qual tem associadas, num sentido mais amplo, as tarefas relativas à gerência da execução distribuída propriamente dita (YAMIN, 2002a, 2002c).

3 Motivação

O escalonamento ótimo em sistemas distribuídos é conhecido no caso geral como um problema NP-completo (KWOK; AHMAD, 1999), este problema se potencializa no cenário de *Pervasive Computing*. Um dos motivos que faz com que o problema não tenha uma solução ótima é a limitada capacidade de manutenção do conhecimento preciso do estado do ambiente, além disto, as necessidades da aplicação também em geral não são conhecidas de forma precisa.

Os resultados apresentados por Santos (SANTOS, 2001) indicam que estratégias de escalonamento dinâmico baseadas em sensores com amostragem dinâmica e que incluem um

modelo do ambiente a ser gerenciado na forma de uma rede de Bayes, podem obter significativas melhorias de performance em relação a estratégias mais tradicionais, como distribuição uniforme do trabalho, alocação de tarefa sob demanda e estratégias determinísticas baseadas em sensores. Santos mostra também que os custos adicionais impostos pelo mecanismo mais sofisticado de tomada de decisão são compensados pela melhor qualidade do escalonamento de recursos que pode ser obtido.

4 Objetivos Perseguidos

O trabalho se encontra em andamento com a parte de revisão bibliográfica pertinente a Redes Bayesianas e estratégias para escalonamento em *Pervasive Computing* encaminhadas.

É objetivo geral do trabalho propor um módulo de escalonamento para o projeto ISAM baseado em um mecanismo de Redes Bayesianas.

Os objetivos específicos contemplam: (i) utilizar Redes Bayesianas como um mecanismo probabilístico para tratar a incerteza com relação ao estado do ambiente de execução; (ii) avaliar o uso de estratégias de *reinforcement learning* para aumento da eficiência da heurística de escalonamento; (iii) considerar na heurística as chances de fracasso das alternativas de decisão a serem tomadas.

Para atingir os objetivos perseguidos, este trabalho pode ser dividido em três grandes etapas: (i) o estudo das redes bayesianas buscando aplicações em estratégias de escalonamento aplicáveis em ambientes *pervasive*; (ii) elaboração e implementação de uma estratégia de escalonamento; (iii) e integração da proposta com o ISAM, buscando avaliação de desempenho e a validação do módulo de escalonamento.

Referências

AUGUSTIN, I. et al. Towards a taxonomy for mobile applications with adaptive behavior. In: *International Symposium on Parallel and Distributed Computing and Networking (PDCN02)*. Innsbruck, Austria: International Association of Science and Technology for Development (IASTED), 2002.

AUGUSTIN, I. et al. Requisitos para o projeto de aplicações móveis distribuídas. In: *VIII CACIC Congreso Argentino de Ciencias de la Computación*. Santa Cruz, Argentina: [s.n.], 2001.

BARBOSA, J. L. V. *Holoparadigma: Um Modelo Multiparadigma Orientado ao Desenvolvimento de Software Distribuído*. Tese (Tese (Doutorado em Ciência da Computação)) — Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2002. 213p.

- DAVIES, N. et al. Limbo: A tuple space based platform for adaptive mobile applications. In: *Joint International Conference on Open Distributed Processing and Distributed Platforms (ICODP/ICDP '97)*. Toronto, Canada: Chapman and Hall, 1997.
- FOSTER, I.; KESSELMAN, C. (Ed.). *The Grid: Blueprint for a New Computing Infrastructure*. San Francisco, CA: Morgan Kaufmann, 1999.
- ISAM. *Projeto ISAM*. 2002. WWW. Disponível em <http://www.inf.ufrgs.br/~isam>. Acesso em dezembro 2002.
- KWOK, Y.-K.; AHMAD, I. Benchmarking and comparison of the task graph scheduling algorithms. *Journal of Parallel and Distributed Computing*, v. 59, n. 3, p. 381–422, 1999. Disponível em: <citeseer.nj.nec.com/kwok99benchmarking.html>.
- NOBLE, B. System support for mobile, adaptive applications. *IEEE Personal Communications*, fev. 2000. Disponível em: <citeseer.nj.nec.com/noble00system.html>.
- O'REILLY & ASSOCIATES, I. *Peer-to-peer: Harnessing the Benefits of a Disruptive Technology*. [S.l.]: O'Reilly & Associates, 2001.
- REAL, R. A.; FILHO, N. D. L.; YAMIN, A. C. *Unicluster: Uma Proposta de Internet Computing Platform para Processamento de Alto Desempenho*. Trabalho para conclusão do curso de Engenharia de Computação. Rio Grande, Dec 2001.
- SANTOS, L. P. P. dos. *Application Level RunTime Load Management: A Bayesian Approach*. Tese (Doutorado) — Universidade do Minho, Braga, Portugal, 2001.
- SATYANARAYANAN, M. Pervasive computing: Vision and challenges. *IEEE Personal Communications*, p. 10–17, ago. 2001. Disponível em: <citeseer.nj.nec.com/satyanarayanan01pervasive.html>.
- SILVA, L. da et al. Mecanismos de suporte ao escalonamento em sistemas com objetos distribuídos Java. In: *VIII CONGRESO ARGENTINO DE CIENCIAS DE LA COMPUTACIÓN (CACIC 2001)*. Santa Cruz, Argentina: Universidad Nacional de la Patagonia Austral, 2001.
- YAMIN, A. et al. Collaborative multilevel adaptation in distributed mobile applications. In: *XXII International Conference of the Chilean Computer Science Society*. Atacama, Chile: IEEE-CS, 2002.
- YAMIN, A. C. et al. A framework for exploiting adaptation in high heterogeneous distributed processing. In: *IEEE. XIV Symposium on Computer Architecture and High Performance Computing (SBAC-PAD 2002)*. Vitória - Brazil, 2002.