

8. Conclusões

Neste capítulo são apresentadas as principais ponderações e conclusões sobre o trabalho desenvolvido nesta tese de doutorado e sugestões de trabalhos futuros para a continuidade da pesquisa.

8.1 *Considerações Finais*

Esta tese propôs uma maneira viável de equacionar alguns problemas existentes na atividade de escalonamento de processos em plataformas distribuídas. O problema central focado foi a definição de mecanismos para a avaliação adequada do índice de carga a ser utilizado nas decisões do escalonamento, quando utilizando sistemas computacionais distribuídos heterogêneos.

Visando a um escalonamento de processos em ambientes heterogêneos mais flexível, dinâmico e transparente, foi proposto nesta tese um índice de desempenho que incorpora o índice de carga e as características da plataforma computacional.

A análise da bibliografia apresentada no capítulo 3 demonstrou a real necessidade em se apresentar novas métricas para obtenção de carga do sistema e de recursos em ambientes distribuídos heterogêneos, uma vez que a literatura contempla basicamente índices específicos para cada recurso, não caracterizando apropriadamente a plataforma avaliada.

Ao visualizar a necessidade de uma nova métrica para a obtenção da carga de trabalho, a heterogeneidade presente no ambiente computacional distribuído, foi outro fator amplamente abordado e avaliado nesta tese. Uma nova métrica para a caracterização da homogeneidade e heterogeneidade dos sistemas computacionais distribuídos foi proposta e validada. Essa métrica avalia não somente o sistema computacional como um todo, como pode ser estendida para avaliar recursos individuais.

A definição da métrica de heterogeneidade (capítulo 5) teve como princípios básicos flexibilidade e simplicidade. A flexibilidade em se permitir que tanto o sistema global quanto os recursos individuais possam ser caracterizados; como simplicidade o fato da nova métrica ser fácil de se entender e calcular.

Partindo dessa nova métrica, pode-se notar que mesmo em sistemas homogêneos observam-se momentos em que a capacidade do sistema disponível para a execução de novos processos indique, claramente, uma situação de heterogeneidade. Isto se dá uma vez que um processo em execução em uma máquina consome parte dos recursos disponíveis (como, por exemplo, CPU, memória, entre outros), deixando uma outra parte que pode ser utilizada por novos processos. De modo análogo, pode ocorrer, de um sistema heterogêneo se apresentar como homogêneo em um dado instante (Branco et al, 2003a; Branco et al, 2003c).

Dessa investigação surgiu como uma contribuição desta tese o conceito de homogeneidade/heterogeneidade temporal. Por heterogeneidade temporal entende-se que em um determinado instante, mesmo em sistemas arquiteturalmente e configuracionalmente homogêneos as diversas máquinas podem ser vistas como temporalmente heterogêneas, levando-se a um sistema heterogêneo.

Diversos foram os estudos de casos efetuados para que as contribuições de ambas as métricas (de heterogeneidade e do índice de desempenho), no contexto de escalonamento de processos em sistemas computacionais distribuídos, fossem comprovadas.

Para testá-las foi construído um programa de simulação baseado em um modelo de redes de fila que, submetido a testes, permitiu comprovar as contribuições alcançadas com o novo índice de desempenho, que incluem: os baixos tempos médios de resposta dos processos e a estabilidade do índice quanto aos diferentes tipos de aplicações submetidas. Para validar os resultados da simulação, foram efetuados testes com o AMIGO (*dynAMical flexible scheduling environment*) (Souza, 2000), que confirmaram os resultados de simulação e, conseqüentemente, as contribuições do novo índice.

Ao final desta tese se conclui que aplicar corretamente o escalonamento de processos não é uma tarefa fácil, principalmente quando se fala de escalonamento em ambientes heterogêneos. Essa dificuldade se deve aos vários fatores que influenciam o bom desempenho do escalonamento de processos, sendo um dos principais a escolha e determinação do índice de carga apropriado.

8.2 Contribuições desta Tese

Os trabalhos descritos nesta tese abordaram vários aspectos tanto da área de escalonamento de processos como da computação heterogênea, propondo uma maneira viável para que os ambientes computacionais distribuídos heterogêneos possam ser utilizados eficientemente. As principais contribuições apresentadas são:

- análise crítica da bibliografia existente, fornecendo um contexto geral para a área de escalonamento de processos em plataformas computacionais distribuídas no que diz respeito aos diversos índices de carga utilizados e que se encontram pulverizados na literatura. O texto também permite que os pesquisadores iniciantes previnam-se das divergências existentes, aprofundem posteriormente seus estudos de uma maneira correta e mais fácil, além de permitir que encontrem as informações condensadas em um único lugar e não pulverizadas na literatura;
- levantamento das métricas de heterogeneidade apontando as falhas existentes e a proposição de uma nova métrica que permite a avaliação não somente da velocidade das máquinas, mas de qualquer recurso que se queira avaliar;
- flexibilidade e facilidade de cálculo da nova métrica de heterogeneidade;
- definição da heterogeneidade Temporal ou Dinâmica, que contempla os sistemas computacionais distribuídos;
- projeto e modelagem da métrica de carga denominada índice de desempenho (*Vector for Índices of Performance – VIP*) que:
 - (1) vem ao encontro dos problemas da área destacados nesta tese;
 - (2) viabiliza a utilização de ambientes de escalonamento não somente em ambiente computacionais distribuídos homogêneos, mas também em ambientes computacionais distribuídos heterogêneos;

-
- (3) permite a união das mais diferentes propostas, maximizando as vantagens de cada uma e minimizando as suas desvantagens;
 - (4) torna a atividade de escalonar processos mais eficiente ao usuário.
 - contribui com a proposição de um novo índice de desempenho que é flexível e simples (cálculo simples), ocasionando baixa sobrecarga no sistema;
 - proposição de um modelo para simulação de índices de carga e desempenho. A utilização da simulação para desenvolvimento e análise de escalonamento de processos em sistemas distribuídos permite a avaliação de diferentes plataformas sem que seja necessária a implementação e a utilização de ambientes de escalonamentos reais. Outras características relevantes na concepção do modelo e que podem ser citadas como contribuições relevantes desta tese de doutorado são:
 - formalização e descrição do modelo para simulação de escalonamento de processos;
 - facilidade para testar outros índices ou tipos de aplicações;
 - possibilidade de se testar a heterogeneidade sem que seja necessário o uso de máquinas reais.
 - consolidação de linha de pesquisa em escalonamento de processos no grupo de Sistemas Distribuídos e Programação Concorrente do ICMC/USP, sendo que novos trabalhos de mestrado estão sendo desenvolvidos com base nos resultados deste trabalho.

8.3 Lista de Publicações

As publicações geradas tendo como base os resultados obtidos no desenvolvimento desta tese são:

- Branco, K. R. L. J. C., Santana, M.J., Santana, R. H. C. (2003a). A Novel Metric for Evaluation of Computer System Heterogeneity. *Proceedings of The International Conference on*

Parallel and Distributed Processing Techniques and Applications (PDPTA'2003). p. 437 – 441 V. 1, Las Vegas – Nevada – USA, June.

- Branco, K. R. L. J. C., Santana, M.J., Santana, R. H. C. (2003b). A Novel Performance Metric for Evaluation of Computer System Heterogeneity. *Proceedings of The International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS'2003)*. p. 292 – 301, V. 34, n. 04 SCS. Montreal – Canadá, July.
- Branco, K. R. L. J. C., Santana, M.J., Santana, R. H. C. (2003c). A Novel Metric for Checking Levels of Heterogeneity in Distributed Computer Systems. *Proceedings of The Fourth Congress of Logic Applied to Technology (LAPTEC'2003)*. p. 148 – 155, V. 101, IOS Press. Marília – São Paulo – Brazil, Novembro.

8.4 Sugestões para Trabalhos Futuros

Devido à grande necessidade de bons índices de cargas e de desempenho, e da presença desses em ambientes heterogêneos, diferentes características e/ou situações podem ser investigadas. Como continuidade a este trabalho tem-se:

- investigar o impacto das políticas e não somente dos índices de carga quando do escalonamento de diferentes tipos de aplicações. Isso permitirá investigar com maiores detalhes o impacto dos diferentes tipos de aplicações no escalonamento;
- implementação de técnicas de redes neurais para a escolha da política de escalonamento apropriada, permitindo que a mesma seja feita de maneira inteligente e específica para os seus usuários;
- implementação de uma política inteligente que se adapte ao ambiente de escalonamento. Com uma política inteligente o sistema pode adequar-se de acordo com o seu estado, caso a aplicação a ser escalonada seja CPU-Bound e a carga do sistema seja leve, o sistema poderá ser dado como heterogêneo e, a partir de então, a política determina a eliminação das máquinas mais lentas de modo a torná-lo

homogêneo (se este for o caso), de maneira a explorar as máquinas mais rápidas do sistema, por exemplo;

- avaliar Heterogeneidade de outras formas, por exemplo, enquadrando os conjuntos de máquinas em funções de distribuição e efetuando a normalização em função das características e especificidades destas funções de distribuição;
- usar *lógica fuzzy* para a obtenção de graus de heterogeneidade mais fiéis;
- fazer uso de *neurofuzzy* para obtenção de índices de desempenho. Usar uma rede neural em conjunto com modelagem *fuzzy* para que se possa conseguir índices de desempenho levando em consideração o não determinismo do sistema;
- através dos estudos efetuados pode-se chegar à conclusão da real necessidade em se separar, em um escalonador, as políticas dos mecanismos. Em muitos softwares de escalonamento, como por exemplo o AMIGO, a obtenção da carga se encontra dentro das políticas, o que faz com que uma política seja específica para determinado índice de carga. Uma maneira de separar as políticas dos mecanismos seria a criação de um módulo para obtenção dos índices de carga que poderiam ser especificados e implementados através da utilização de agentes (o que ocasionaria uma baixa sobrecarga no sistema) possibilitando a obtenção dos índices de carga e tendo assim seu uso facilitado pelas políticas;
- desenvolver novos benchmarks que levem em conta os principais recursos do sistema (CPU, memória, rede e disco). A proposição de padrões que determinem especificidades dos recursos e permitam que *benchmarks* mais apurados possam ser executados nas máquinas, levando a obtenção de um índice de desempenho mais fiel;
- uma vez que esta tese não aborda máquinas com memória compartilhada, sugere-se investigar a possibilidade de desenvolver uma ferramenta para monitorar o desempenho de máquinas com memória compartilhada;

- fazer avaliações mais profundas a respeito dos diversos índices de carga existentes na literatura no que diz respeito à apresentação de uma maneira sistemática, mais clara e cientificamente provada, dos limitantes (*thresholds*) de cada índice de carga para cada recurso;
- estudos mais refinados sobre os índices de carga específicos que podem compor a função de índice de desempenho proposta neste trabalho;
- investigar a possibilidade de reduzir todo e qualquer sistema em uma coleção de recursos homogêneos dada as considerações tecidas sobre heterogeneidade temporal tratadas nos capítulos 5 e 6;
- definir uma meta-linguagem para a definição e incorporação de novos índices de carga no ambiente de simulação;
- gerar uma interface gráfica para o modelo de simulação proposto;
- demonstrar que o novo índice de desempenho proposto é adaptativo através da inserção de diferentes tipos de aplicação ao mesmo tempo e da variabilidade das aplicações com o decorrer do tempo;
- estudos investigativos mais detalhados podem ser realizados de modo a demonstrar a existência da heterogeneidade temporal, principalmente quando observados os resultados da simulação em ambientes homogêneos no qual o índice de desempenho apresenta melhor resultado que o algoritmo round-robin, o qual distribui igualmente as aplicações entre os recursos disponíveis.