

# Técnica de Caracterização de Carga de Trabalho de Máquinas Paralelas para Extração de Informações Utilizadas por um Escalonador Reconfigurável de Tarefas

Lesandro Ponciano dos Santos<sup>1</sup>, Luís Fabrício W. Góes<sup>2</sup>  
Pontifícia Universidade Católica de Minas Gerais  
lesandrop@yahoo.com.br<sup>1</sup>, lfwgoes@pucminas.br<sup>2</sup>

## Resumo

Neste trabalho propomos e implementamos uma técnica de caracterização de carga de trabalho para extrair informações utilizadas pelo escalonador reconfigurável de tarefas RGSA.

Como estudo de caso, utilizamos um log de um supercomputador real e obtivemos um modelo da carga analisada. Além disso, identificamos vários padrões de comportamento da carga de trabalho, que são essenciais para a escolha de uma política de escalonamento.

## 1. Introdução

Na atualidade, o aumento de desempenho das arquiteturas paralelas e distribuídas tem se tornado um objetivo fundamental, pois, com o surgimento de aplicações mais complexas, as tarefas têm consumido cada vez mais recursos de processamento, de entrada e de saída [1,2,8].

A pesquisa em arquiteturas paralelas e distribuídas, como aglomerados de computadores e grades computacionais, têm procurado obter o máximo de desempenho computacional através da otimização de vários componentes de software e hardware, tais como: processadores, memórias, redes de comunicação, protocolos de rede etc. [4]. Dentre estes componentes, destacamos o escalonador de tarefas. Ele é responsável por alocar cada processo de uma tarefa para cada um dos processadores disponíveis na arquitetura paralela. O escalonamento pode ser realizado com [9] ou sem [8] informação sobre a carga de trabalho (conjunto de tarefas) [3].

O escalonamento com informação pode possibilitar maior desempenho do sistema [3,4]. No entanto, a obtenção de informação confiável constitui um problema, pois o uso de informações fornecidas pelo usuário é um método não confiável, ou seja,

informações erradas estimadas pelos usuários podem prejudicar o desempenho do escalonador de tarefas [7]. Através caracterização de logs (históricos de execução de tarefas) é possível construir um modelo de desempenho das tarefas e utilizá-lo como informação confiável para o escalonador de tarefas [1].

O RGSA (*Reconfigurable Gang Scheduling Algorithm*) é um escalonador que possui um conjunto de políticas de escalonamento (configurações), onde ele seleciona e executa a que apresenta maior desempenho para cada carga de trabalho específica. Para a seleção das políticas, ele utiliza informações sobre o dia, hora, tempo de submissão, número de processos e tempo de execução dos *jobs* ou tarefas [4,9].

Neste contexto, o problema deste trabalho é a obtenção de informações confiáveis, por meio de caracterização de carga, para a seleção de políticas de escalonamento do RGSA. Como estudo de caso, caracterizamos um log do supercomputador Thinking Machines CM-5 [5].

Assim, o objetivo principal deste trabalho é a proposta e implementação de uma técnica de caracterização de carga de trabalho de máquinas paralelas para extração de informações utilizadas por um escalonador reconfigurável de tarefas (RGSA).

Este artigo está assim organizado: na seção 2 é apresentado o estudo dos trabalhos relacionados, na seção 3 a técnica de caracterização de carga, na seção 4 os resultados experimentais, na seção 5 a conclusão e trabalhos futuros e, por fim, na seção 6 as referências bibliográficas.

## 2. Trabalhos Relacionados

Neste tópico destacamos alguns trabalhos relacionados [1,2,4]. Em [1] são analisados diversos modelos de carga de trabalho baseados em logs de supercomputadores e formulações matemáticas.

Um análise da carga de trabalho do supercomputador iPSC/860 é apresentada em [2]. São consideradas, no estudo, as tarefas executadas no quarto trimestre de 1993. Diversos aspectos foram analisados nesse estudo, entre eles estão: o grau de utilização do sistema durante do dia, distribuição do tempo de execução e dos recursos utilizados, a tarefa submetida em função tempo de execução, o número de tarefas submetidas e executadas em relação aos recursos utilizados.

Em [4] é proposto e desenvolvido o algoritmo reconfigurável de escalonamento paralelo de tarefas RGSA (*Reconfigurable Gang Scheduling Algorithm*). O RGSA visa obter um escalonamento de tarefas que se adapta às variações das arquiteturas paralelas e carga de trabalho. Para isso, o algoritmo necessita de informações confiáveis dos *jobs* para entrada de dados do escalonador.

Assim, o nosso trabalho relaciona-se aos apresentados em [1,2] pela análise e caracterização da carga de trabalho de supercomputadores de modo a obter informações confiáveis. No entanto, neste trabalho focamos na criação do caracterizador de carga de trabalho para o escalonador paralelo de tarefas RGSA [4].

### 3. Técnica de Caracterização de Carga de Trabalho

A técnica de caracterização de carga deste trabalho realiza o agrupamento dos *jobs* segundo três fatores: tempo de submissão, número de processos e tempo de execução dos *jobs*.

O principal objetivo da técnica é agrupar os *jobs* de uma carga de trabalho, determinando a porcentagem das classes de jobs em cada período do dia e identificar mudanças no comportamento da carga de trabalho ao longo das horas e dias da semana. Para cada carga existe uma política de escalonamento mais adequada, ou seja, quanto menos mudanças na carga ou menor o número de grupos gerados, menor será o número de trocas de políticas de escalonamento necessárias para garantir um maior desempenho para a máquina paralela.

**Tabela 1 – Classes de Jobs**

Classe	Número de Processos	Tempo de Execução
LL (Low-Low)	Baixo (Low)	Baixo (Low)
LH (Low-High)	Baixo (Low)	Alto (High)
HL (High-Low)	Alto (High)	Baixo (Low)
HH (High-High)	Alto (High)	Alto (High)

Primeiramente, a mediana dos fatores número de processos e tempo de execução é calculada como um *threshold*, para classificar os *jobs* de acordo com as classes definidas em [4]. O RGSA divide os jobs em 4 classes [4] que refletem as possíveis combinações de tempo e espaço (número de processos) mostrados na Tabela 1.

Os *jobs* são classificados de acordo com o seguinte algoritmo:

```

se job.n_processos ≤ mediana_n_processos
se job.tempo_exe ≤ mediana_tempo_exe
então job.classe = LL;
senão job.classe = LH;

senão se job.tempo_exe ≤ mediana_tempo_exe
então job.classe = HL;
senão job.classe = HH;

```

Depois de classificados, os *jobs* são agrupados usando-se o tempo de submissão, de acordo com a hora do dia (de 0 à 23) e o dia da semana (de domingo à sábado). Para cada dia da semana, a técnica de caracterização gera uma Tabela de Agrupamento como descrita na Tabela 2, na qual:

$i$  – Índice que representa um grupo de *jobs* agrupados pelo tempo de submissão, número de processos e tempo de execução;

$LI_i$  – Limite Inferior do grupo  $i$ : é a parte inteira da hora em que foi submetido o *job* de menor tempo de submissão contido no grupo  $i$ ;

$LS_i$  – é a parte inteira da hora em que foi submetido o *job* de maior tempo de submissão contido no grupo  $i$ ;

$\%HH_i$  – é o percentual de *jobs* contidos no grupo  $i$  e que possuem alto tempo de execução (H) e alto número de processos (H) em relação ao total de *jobs* do grupo  $i$ . Para  $\%HL$ ,  $\%LL$  e  $\%LH$  segue-se o mesmo método de cálculo de  $\%HH$  considerando-se os *jobs* dos níveis L e H descritos.

**Tabela 2 – Estrutura da Tabela de Agrupamento**

Dia da semana					
$i$	Intervalo	HH	HL	LL	LH
1	$LI_1 - LS_1$	$\%HH_1$	$\%HL_1$	$\%LL_1$	$\%LH_1$
2	$LI_2 - LS_2$	$\%HH_2$	$\%HL_2$	$\%LL_2$	$\%LH_2$
...	...	...	...	...	...
23	$LI_{23} - LS_{23}$	$\%HH_{23}$	$\%HL_{23}$	$\%LL_{23}$	$\%LH_{23}$
24	$LI_{24} - LS_{24}$	$\%HH_{24}$	$\%HL_{24}$	$\%LL_{24}$	$\%LH_{24}$

Neste trabalho, desenvolvemos um algoritmo de agrupamento de jobs de uma carga de trabalho, como parte da técnica de caracterização. O algoritmo agrupa os grupos consecutivos (*i.e.*, que possuem  $LS_i = LI_{i-1}$ )

em que a diferença entre os valores de %HH, %HL, %LL e %LH é menor ou igual a 10. Nesta caracterização, utilizamos o valor 10 para limitar o desvio máximo permitido dentro do grupo, testamos também com os valores 3, 5, 7, mas por limitação de espaço mostramos apenas com o valor 10.

Toda vez que é realizado um agrupamento do grupo  $i$  com o grupo  $i+1$  forma-se um novo grupo que assume a posição de  $i$  originário e recebe os seguintes valores:

- i) Intervalo:  $LI_i = LI_i$
- ii)  $LS_i = LS_i + 1$
- iii)  $\%HH_i = (\%HH_i + \%HH_{i+1})/2$
- iv)  $\%HL_i = (\%HL_i + \%HL_{i+1})/2$
- v)  $\%LL_i = (\%LL_i + \%LL_{i+1})/2$
- vi)  $\%LH_i = (\%LH_i + \%LH_{i+1})/2$

O algoritmo executa diversas iterações até que não sejam realizados novos agrupamentos. A técnica foi implementada em linguagem Perl utilizando-se o ambiente ActiveState ActivePerl 5.6. A implementação recebe como entrada um *log* no formato definido em [5] e o período que se deseja caracterizar (início e fim) e como saída gera um arquivo contendo as tabelas de agrupamento como a apresentada na Tabela 2.

#### 4. Resultados Experimentais

Para avaliar a técnica de caracterização de carga neste trabalho realizamos um estudo de caso com o *log* LANL-CM5-1994-3 do supercomputador *Thinking Machines CM-5* [5].

Esse *log* foi escolhido por ser de um supercomputador real e por armazenar dados de período superior a um ano (abrange de 4 de outubro de 1994 a 24 de setembro de 1996), possibilitando gerar a caracterização com dados de todo o ano.

Desse *log*, para o escopo de testes, são utilizados os *jobs* submetidos entre as 0:00 horas do dia 01 de janeiro de 1995 e 23:00 horas do dia 31 de dezembro de 1995 e que possuem tempo de execução e número de processos maior ou igual a 1.

**Tabela 3 – Agrupamento dos domingos**

Domingo					
i	Intervalo	HH	HL	LL	LH
1	[0 - 5]	46,29	11,02	18,3	24,31
2	6	27,41	41,48	17,04	14,08
3	[7-23]	39,22	19,61	23,73	17,44

Do *log* LANL-CM5-1994-3, utilizado no estudo de caso, selecionou-se 66560 *jobs* sob os critérios

apresentados na seção 3. Dessa amostra obteve-se a mediana 32 para o número de processos e 320 segundos para o tempo de execução. A caracterização gerada pela técnica é apresentada nas Tabelas de 3 a 9 de acordo com o dia da semana.

**Tabela 4 – Agrupamento das segundas-feiras**

Segunda-Feira					
i	Intervalo	HH	HL	LL	LH
1	[0 - 5]	43,3	21,83	9	26,15
2	6	32,07	32,61	19,56	15,76
3	[7 - 10]	38,45	20,27	22,83	18,46
4	[11 - 22]	27,6	17,9	34,08	20,41
5	23	38,15	15,6	27,75	18,5

**Tabela 5 – Agrupamento das terças-feiras**

Terça-Feira					
i	Intervalo	HH	HL	LL	LH
1	0	37,34	17,72	27,85	17,09
2	[1-2]	50,13	9,16	15,16	25,54
3	[3-5]	40,48	23,69	12,26	23,57
4	6	24,32	34,95	26,75	13,98
5	[7-22]	32,05	16,78	30,43	20,74
6	23	28,83	28,83	26,58	15,76

**Tabela 6 – Agrupamento das quartas-feiras**

Quarta-feira					
i	Intervalo	HH	HL	LL	LH
1	(0-23)	27,85	20,91	32,68	18,56

**Tabela 7 – Agrupamento das quintas-feiras**

Quintas-feiras					
i	Intervalo	HH	HL	LL	LH
1	0	47,97	21,62	11,49	18,92
2	1	25,42	55,37	7,34	11,87
3	[2-4]	40,75	19,56	23,22	16,46
4	5	32,69	11,54	35,58	20,19
5	6	24,08	29,32	32,46	14,14
6	[7-8]	39,44	21,91	20,68	17,97
7	[9-10]	28,35	23,61	29,02	19
8	[11-12]	17,77	31,02	35,76	15,45
9	[13-17]	27,63	15,46	38,28	18,23
10	18	32,39	16,97	27,25	23,39
11	[19-20]	30,37	12,4	37,56	19,67
12	[21-23]	32,71	19,93	26,6	20,76

**Tabela 8 – Agrupamento das sextas-feiras**

Sexta-Feira					
i	Intervalo	HH	HL	LL	LH
1	[0-1]	25,01	21,77	36,18	17,04
2	2	29,51	8,2	47,54	14,75
3	[3-5]	35,63	16,52	33,69	14,16
4	6	24,64	29,29	29,64	16,43
5	[7-20]	33,18	15,6	31,47	19,75
6	[21-23]	38,59	17,1	19,42	24,89

**Tabela 9 – Agrupamento dos sábados**

Sábado					
i	Intervalo	HH	HL	LL	LH
1	0	31,82	23,48	25,76	18,94
2	[1-4]	41,3	23,94	13,52	21,23
3	[5-7]	31,04	35,1	17,43	16,42
4	[8-20]	40,13	20,43	21,62	17,81
5	[21-22]	31,42	13,25	33	22,33

As tabelas de agrupamentos obtidos mostram a sua eficácia em extrair de um *log*, grande e complexo, padrões, agrupamentos e informações importantes para a entrada de escalonador reconfigurável de tarefas.

O resultado obtido no estudo de caso consiste em um modelo do *log* [5] com os dados necessários para a entrada do escalonador reconfigurável de tarefas.

Para analisar esse modelo pode-se tomar como métrica a quantidade de horas agrupadas em um mesmo grupo, assim, pode-se constatar que nas quartas-feiras, Tabela 6, obteve-se o melhor caso onde todas as horas do dia foram agrupadas e um único grupo.

Nesse dia o escalonador poderá utilizar, por exemplo, apenas um algoritmo de escalonamento uma vez que a maioria dos *jobs* que executam nesse dia possui estrutura semelhante.

De segunda à sexta, da Tabela 3 à 8, às seis horas tem-se um grupo de *jobs* que não se agrupam com nenhum dos grupos vizinhos. Além disso, no horário comercial (8 às 18), existe pouca variação na carga de trabalho na maioria dos dias.

Por último, observamos que nas madrugadas (0 às 6 horas) dos finais de semana as cargas de trabalho são basicamente compostas por *jobs* que demandam muito tempo de processamento e alto número de processamentos.

## 5. Conclusão e Trabalhos Futuros

Concluímos que a técnica de caracterização de carga de trabalho desenvolvida atende ao objetivo de fornecer dados confiáveis para o escalonador RGSA.

Neste trabalho também identificamos formas de modelagem de cargas de trabalho para escalonadores paralelos de tarefas. Isso, através do agrupamento dos *jobs* pelos seus tempos de submissão e execução, e pelo número de processos.

Através da utilização de dados reais e de fonte confiável, e pela implementação de um algoritmo de agrupamento extraímos informações importantes ao processo de escalonamento de tarefas.

Como trabalhos futuros nós destacamos o teste e avaliação do escalonamento em um ambiente real.

Destacamos também o desenvolvimento de um algoritmo de agrupamento adaptativo para um escalonador reconfigurável dinâmico. Nesse processo o algoritmo deve ser capaz de gerar tabelas de agrupamento dinâmicas que serão alteradas à medida que se alterar o padrão dos *jobs* executados naquele determinado tempo e espaço.

## 6. Referências

- [1] Feitelson, D., “Workload Modeling for Performance Evaluation”, Performance Evaluation of Complex Systems: Techniques and Tools, pp. 114-141, 2002.
- [2] Feitelson, D. and Nitzberg, B. “Job characteristics of a production parallel scientific workload on the NASA Ames iPSC/860”, Workshop on Job Scheduling Strategies for Parallel Processing, pp.337-360, 1995.
- [3] Góes, L.F. W., Martins, C. A. P. S., “Escalonamento Paralelo de Tarefas: Conceitos Simulação e Análise de Desempenho”, Workshop de Computação de Alto Desempenho, 2004.
- [4] Góes, L. F. W., C. A. P. S., “Proposta e Desenvolvimento de um Algoritmo Reconfigurável de Escalonamento Paralelo de Tarefas”, Dissertação de Mestrado, PPGEE, PUC Minas, 2004.
- [5] LANL-CM5-1994-3  
<http://www.cs.huji.ac.il/labs/parallel/workload/logs.html>
- [6] Subhlok, J., Venkataramaiah, S., Singh, A., “Characterizing NAS Benchmark Performance on Shared Heterogeneous Networks”, IEEE International Parallel and Distributed Processing Symposium, 2002.
- [7] Lee, C.B., et al. “Are user runtime estimates inherently inaccurate?”, 10th Workshop on Job Scheduling Strategies for Parallel Processing, 2004.
- [8] Cirne, W., Brasileiro, F., Paranhos, D., Góes, L. F. W., Voorsluys, W., “On the efficacy, efficiency and emergent behavior of task replication in large distributed systems”, Parallel Computing, 2007.
- [9] Góes, L. F. W., Martins, C. A. P. S., “Reconfigurable Gang Scheduling Algorithm”, 10th Workshop on Job Scheduling Strategies for Parallel Processing, Lecture Notes in Computer Science, New York, 2004.