

Um algoritmo paralelo para priorização de testes baseada em similaridade usando OpenMPI

ERADSP 2017 @ ICMC-USP

Carlos Diego N. Damasceno, Paulo S. L. Souza, Adenilso Simao

damascenodiego@usp.br, {pssouza,adenilso}@icmc.usp.br
Instituto de Ciências Matemáticas e de Computação – ICMC
Universidade de São Paulo – USP (ICMC-USP)

Abril de 2017



Resumo

Priorização de casos de teste baseada em similaridade:

- ▶ Funções de similaridade auxiliam execução de testes¹
- ▶ Maximizar diversidade

Calculo de *matriz de similaridade (SM)*:

- ▶ Comparação entre *todos os pares* de teste (i.e. $O(n^2)$)
- ▶ Paralelismo \implies Melhoria de performance²

Neste estudo propomos e avaliamos:

- ▶ Geração paralela de matriz de similaridade (**PGSM**)
- ▶ LMDP paralelo (**PLMDP**) (i.e. pares mais distintos)

¹Emanuela G. Cartaxo et al. "On the use of a similarity function for test case selection in the context of model-based testing". Em: *Software Testing, Verif. and Reliab.* (2011).

²Felicia Gîză-Belciug e Stefan-Gheorghe Pentiuc. "Parallelization of similarity matrix calculus in ontology mapping systems". Em: *2015 14th RoEduNet NER*. IEEE, set. de 2015, pp. 50–55, Tobias Rawald et al. "Massively Parallel Analysis of Similarity Matrices on Heterogeneous Hardware". Em: *Proceedings of the Workshops of the EDBT/ICDT 2015 Joint Conference, Brussels, Belgium*. 2015.

Geração paralela de matriz de similaridade (PGSM)

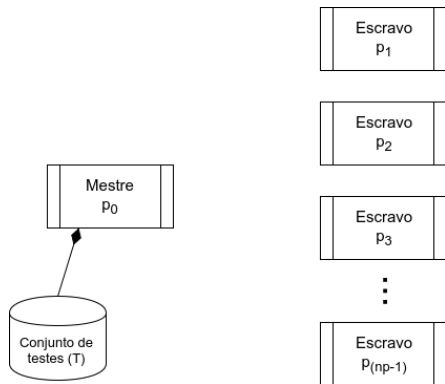


Figura 1: Considere np processos, tal que p_0 é mestre e $p_1, p_2, \dots, p_{(np-1)}$ são escravos, e um conjunto de testes T

Geração paralela de matriz de similaridade (PGSM)

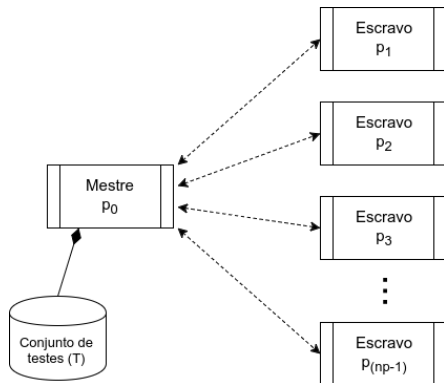


Figura 2: Mestre p_0 transmite $T = \{t_1, t_2, \dots, t_{nr}\}$ para todos os escravos

Geração paralela de matriz de similaridade (PGSM)

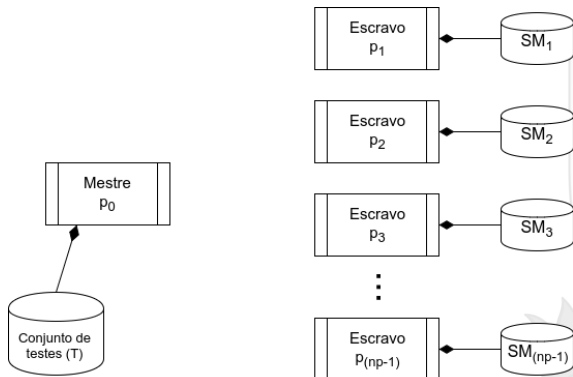


Figura 3: Escravos calculam seus respectivos graus de similaridade (SM_p)

(i.e. balanceamento de tarefas)

LMDP paralelo (PLMDP)

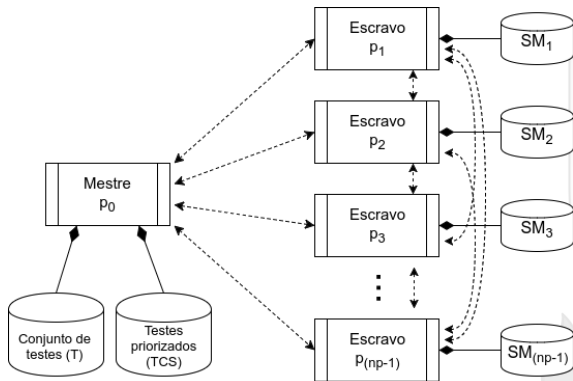
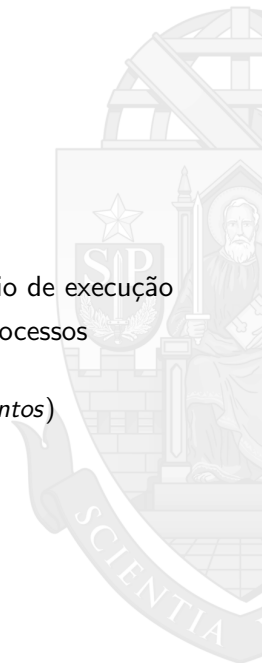


Figura 4: **Broadcast** e **Allreduce** são usados para buscar os *pares de teste mais distintos* e **Send/Recv** são usados para *consultar os graus de similaridade e identificadores dos testes*

Considerações finais

- ▶ Experimento em um cluster (320 ensaios)
- ▶ Número de processos influenciou o tempo médio de execução
- ▶ Speedup crescente em função do número de processos
- ▶ PGSM \implies Baixa eficiência (i.e. *overhead*)
 - ▶ Escravos ociosos (*remoção de pares mais distintos*)



Obrigado!

