Visualização 3D para Análise de Aplicações Paralelas

Lucas Mello Schnorr[1,2], Guillaume Huard[2], Philippe O. A. Navaux[1]

[1] Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brazil

[2]Equipe MOAIS – Laboratoire d'Informatique de Grenoble
ZIRST 51, avenue Jean Kuntzmann - 38330 - Montbonnot Saint Martin - France

{lmschnorr,navaux}@inf.ufrgs.br, Guillaume.Huard@imag.fr

1. Resumo

Aplicações voltadas para sistemas distribuídos como grids necessitam de uma análise de desempenho. Essa análise permite ao desenvolvedor melhorar a sua aplicação levando em conta características do ambiente e correlacionando-as com o comportamento da aplicação. Sendo uma operação complexa, a visualização de dados é uma solução comumentemente utilizada para realizar a análise de aplicações paralelas. Grande parte das ferramentas utilizam gráficos em duas dimensões, onde uma dimensao é o tempo e a outra dimensão é usada para se listar os componentes da aplicação, como processos e fluxos de execução. Alguns exemplos dessas ferramentas são Pajé [de Kergommeaux et al. 2000] e Vampir [Nagel et al. 1996].

O uso de duas dimensões para a análise de aplicações paralelas em ambientes como grids pode ser limitante. O primeiro problema que surge é que com ferramentas 2D, não é possível analisar as aplicações levando-se em conta a topologia da rede. Em ambientes como grid, as características de sua interconexão podem ser decisivas na obtenção de alto desempenho. O impacto da topologia de rede na aplicação grid pode ser ainda mais relevante quando a interconexão é fortemente hierárquica, com diferentes tecnologias de rede em diferentes níveis dessa hierarquia. Um exemplo disso é a interconexão do Grid'5000 [Capello et al. 2008], onde os recursos do grid são agrupados por clusters homogêneos internamente organizados com Gigabit Ethernet e globalmente conectados por fibra óptica. O segundo problema é que com ferramentas 2D não é possível analisar o padrão de comunicação entre os processos, uma vez que são observadas sempre as comunicações propriamente ditas entre eles e não o padrão global da comunicação levando-se em conta todas as comunicações.

Este trabalho consiste em propor novas técnicas de visualização para a análise de aplicações paralelas. Um primeiro esforço nesse sentido é a visualização em três dimensões dos dados de rastreamento de uma aplicação [Schnorr et al. 2008]. Nessa técnica, duas dimensões são utilizadas para representar a topologia de um grid e os componentes da aplicação paralela; e a terceira dimensão é usada como tempo. Assim, é possível ver a evolução dos processos de uma aplicação paralela e relacioná-los aos recursos utilizados. As comunicações são então dispostas nesse ambiente 3D sendo possível a verificação do padrão de comunicação da aplicação assim como quais foram os canais de comunicação utilizados.

Um protótipo chamado Triva - *Three Dimensional and Interactive Visualization Analysis* [Schnorr 2008] - foi criado para validar a técnica. O protótipo utiliza, como en-

trada, dados gerados pelo DIMVisual [Schnorr et al. 2006], adaptado para a leitura de rastros de aplicações KAAPI [Gautier et al. 2007]. Internamente, o protótipo contém adaptadores para o simulador Pajé, a biblioteca Graphviz e Ogre3D. Para alimentar o protótipo com dados, foi realizado o rastreamento de diferentes execuções de aplicações paralelas KAAPI no Grid'5000. As aplicações eram compostas por uma quantidade de processos que varia de 30 à 2900. A figura 1 mostra alguns exemplos de visualizações 3D possíveis com o protótipo. A primeira imagem mostra o padrão de comunicação de uma aplicação e os outros dois, duas aplicações diferentes com a topologia de rede e as comunicações da aplicação.

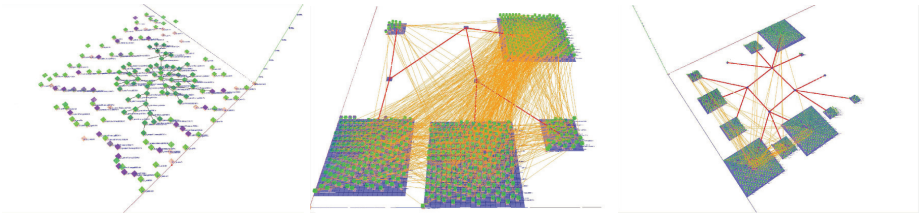

Figura 1. Visualizações 3D dos rastros de aplicações paralelas.

Como trabalhos futuros, pretende-se realizar uma análise mais profunda dos resultados já obtidos e trabalhar com intervalos de tempo, agregação e integração de dados para permitir uma sumarização dos rastros.

Referências

Capello, F., Plateau, B., Priol, T., Vandromme, D., Cappello, F., Dayde, M., Desprez, F., Jeannot, E., Jegou, Y., Lanteri, S., Melab, N., Namyst, R., Quinson, M., Richard, O., and Primet, P. V.-B. (2008). Grid5000. http://www.grid5000.fr. Accessed: 17.Dec.2008.

de Kergommeaux, J., Stein, B., and Bernard, P. (2000). Paje, an interactive visualization tool for tuning multi-threaded parallel applications. *Parallel Computing*, 26(10):1253–1274.

Gautier, T., Besseron, X., and Pigeon, L. (2007). Kaapi: A thread scheduling runtime system for data flow computations on cluster of multi-processors. In *Proceedings of the international workshop on Parallel symbolic computation*, pages 15–23, New York, NY, USA. ACM.

Nagel, W., Arnold, A., Weber, M., Hoppe, H., and Solchenbach, K. (1996). VAMPIR: Visualization and Analysis of MPI Resources. *Supercomputer*, 12(1):69–80.

Schnorr, L. M. (2008). Triva. http://triva.gforge.inria.fr. Last access: December, 17, 2008.

Schnorr, L. M., Huard, G., and Navaux, P. O. A. (2008). 3d approach to the visualization of parallel applications and grid monitoring information. In *Proceedings of the 9th IEEE/ACM International Conference on Grid Computing*, Washington, DC, USA. IEEE Computer Society.

Schnorr, L. M., Navaux, P. O. A., and de Oliveira Stein, B. (2006). Dimvisual: Data integration model for visualization of parallel programs behavior. In *Proceedings of the Sixth IEEE International Symposium on Cluster Computing and the Grid*, pages 473–480, Washington, DC, USA. IEEE Computer Society.