



Evento	Salão UFRGS 2015: SIC - XXVII SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS
Ano	2015
Local	Porto Alegre - RS
Título	Escalabilidade binária de rastros Pajé
Autor	VINÍCIUS ALVES HERBSTRITH
Orientador	LUCAS MELLO SCHNORR

Escalabilidade binária de rastros Pajé

Vinícius Herbstrith (bolsista), Lucas Mello Schnorr (orientador)
Instituto de Informática – Universidade Federal do Rio Grande do Sul

Com o desenvolvimento de aplicações paralelas e distribuídas surge o interesse da análise do seu desempenho. Esta análise é baseada em eventos selecionados e representativos registrados durante a execução da aplicação em arquivos de rastro de execução. Com o atual aumento da capacidade das máquinas paralelas, podemos ter bilhões de fluxos concorrentes gerando milhares de eventos, tornando a escalabilidade um fator de grande importância na geração do rastro.

O formato Pajé é largamente utilizado para a análise e visualização de rastros de aplicações paralelas. O formato se destaca no atributo extensibilidade, possibilitando seu uso nos mais diversos contextos de análise de desempenho e rastreamento. No entanto, a maior desvantagem do formato é a sua estrutura textual, que leva a uma leitura lenta do arquivo de rastro no momento da análise.

O objetivo deste trabalho é a criação de um formato binário para o arquivo de rastro Pajé, almejando a otimização da leitura e trazendo uma maior escalabilidade para o formato Pajé. Isto nos leva a três objetivos: a definição da versão binária correspondente ao formato de rastro Pajé, a geração de arquivos de rastro binário e a análise de desempenho da leitura deste arquivo.

As informações do arquivo binário são salvas usando a biblioteca LibRastro, uma biblioteca de geração de arquivos de rastro. A LibRastro tem como principais características uma baixa intrusão na coleta do rastro e flexibilidade nas definições de eventos, características estas que coincidem com o nosso objetivo. Temos então, um arquivo de rastro gerado na LibRastro que contém informações equivalentes ao rastro original no formato Pajé.

A biblioteca Poti – responsável por implementar o formato de rastro Pajé – foi utilizada para a implementação da criação do rastro no formato binário, usando como suporte a libRastro. Foram feitas modificações na biblioteca com o objetivo de gerar a mesma saída Pajé da Poti no novo formato binário proposto. Como a execução de aplicações em sistemas paralelos e distribuídos é uma tarefa custosa, é de grande interesse promover ao usuário a possibilidade de converter rastros gerados anteriormente. Para isto, foram criadas duas ferramentas dentro da Poti: `rst2paje` e `poti2rst`. A ferramenta `poti2rst` faz a conversão de um arquivo no formato de rastro Paje para o formato binário proposto, usando Flex e Bison para fazer a análise da entrada textual. Já a ferramenta `rst2paje` faz a conversão no sentido contrário.

A leitura e análise do novo formato é feita dentro da PajeNG, uma ferramenta de visualização de rastros no formato Pajé. O componente PajeRastroReader foi criado para realizar a decodificação dos arquivos no formato binário proposto.

A análise de desempenho da leitura do formato binário foi feita sobre dois arquivos de rastro de duas aplicações distintas, cada um em duas versões, textual – a versão original – e binário – obtido pela `paje2rst`. Uma das aplicações é uma fatoração LU que gerou um arquivo de rastro de 0.77 GB, a outra é um cálculo de gradiente conjugado com um rastro de tamanho 2.1 GB. A máquina usada para os testes possui processador Intel i7-4770 CPU (3.40GHz), HD 1TB SATA II (3.0Gb/s) e 8GiB DIMM DDR3(1.6GHz). Foram feitas 30 execuções para cada versão, binária e textual, dos dois arquivos, de forma a se obter uma amostra significativa para comparação sobre o desempenho do formato proposto. Comparando os tempos e usando metodologias estatísticas de teste sobre os tempos de execução, obtivemos uma redução de 26% do tempo total de leitura usando o formato binário em ambos os arquivos de rastro.

A partir destes resultados positivos, podemos buscar maiores otimizações para o formato binário em próximos trabalhos. Como possibilidades, temos o uso de técnicas de compactação e o uso de referências para informações repetidas que levariam à um arquivo de rastro menor e conseqüentemente um melhor tempo de leitura.