

Escalabilidade binária de rastros Pajé

Vinícius A. Herbstrith, Lucas Mello Schnorr
{vaherbstrith,schnorr}@inf.ufrgs.br



Contexto

- Com o desenvolvimento de aplicações paralelas e distribuídas surge o interesse da análise do seu desempenho.
- O formato Pajé é largamente utilizado para a análise e visualização de rastros de aplicações paralelas.
- Prós:
 - Extensibilidade: possibilita seu uso nos mais diversos contextos de análise de desempenho e rastreamento
- Contras:
 - Estrutura textual: leitura lenta do arquivo de rastro no momento da análise

Objetivos

- Criação de um formato binário para o arquivo de rastro Pajé
 - Otimização da leitura
 - Maior escalabilidade
- Três tarefas:
 1. Definição da versão binária correspondente ao formato de rastro
 2. A geração de arquivos de rastro binário
 3. Análise de desempenho da leitura do arquivo binário

Conclusões

- A estrutura do arquivo binário promoveu uma melhora de 32% no desempenho da leitura do arquivo de rastro com o método standard.
- Ganho de 27% no desempenho da leitura do arquivo usando o método reference
- Redução do arquivo de rastro usando o método reference:
 - Rastro LU com 0.77 GB: redução de 48% do tamanho
 - Rastro cálculo do gradiente conjugado de 2.1 GB: a redução foi de 42%
- Trabalhos futuros:
 - Técnicas eficientes de compactação e descompactação
 - Análise da intrusão de rastro do formato binário

Referências

- [1] Benhur Stein. Depuração de programas paralelos. In *I Escola Regional de Alto Desempenho do Estado do Rio Grande do Sul*, pages 151 – 176, Gramado, RS, 2001. Sociedade Brasileira de Computação (SBC).
- [2] L. M. Schnorr. Poti. <http://github.com/schnorr/poti>, 2014.
- [3] L. M. Schnorr. Pajé trace file format. <http://paje.sf.net>, 2014.
- [4] L. M. Schnorr. Pajeng. <http://github.com/schnorr/pajeng>, 2014.
- [5] L. M. Schnorr. Librastro tracing library. <http://github.com/schnorr/akypuera/tree/master/librastro>, 2014.
- [6] The GNU Project Robert Corbett. Bison. <http://www.gnu.org/software/bison>, 2014.
- [7] Vern Paxson. Flex. <http://flex.sourceforge.net>, 2014.

Implementação

- Criação e leitura do arquivo:
 1. As informações do arquivo binário são salvas usando a biblioteca libRastro, uma biblioteca de geração de arquivos de rastros binários
 2. A biblioteca Poti – responsável por implementar o formato de rastro Pajé – foi utilizada para a criação do rastro no formato binário, usando como suporte a libRastro.
 - Dois conversores, Paje2Rastro e Rastro2Paje, para converter arquivos binários e textuais.
 3. A leitura e análise do novo formato é feita dentro da PajeNG, uma ferramenta de visualização de rastros no formato Pajé.
 - Componente PajeRastroReader foi criado para realizar a decodificação dos arquivos no formato binário
 - Componente PajeRastroTraceEvent armazena a informação dos eventos
- Opções de arquivo:
 - Método standard: Tradução direta do arquivo textual Pajé para o arquivo binário da libRastro.
 - Método reference: Busca obter uma otimização quanto ao tamanho do arquivo gerado através do uso de referências do tipo int short para dados do tipo string que costumam se repetir durante o arquivo de rastro, gerando um arquivo binário de tamanho reduzido.

Experimentos

- Análise de desempenho:
 - Dois arquivos de rastro como entrada:
 - * Fatoração LU: arquivo de rastro de 0.77 GB
 - * Cálculo de gradiente conjugado: arquivo de rastro de 2.1 GB
 - Três versões (Obtidas com a ferramenta paje2rst)
 - * Textual (a versão original)
 - * Binária standard
 - * Binária reference
 - Utilizamos gráficos de caixa a partir de 30 execuções para cada versão e arquivo
- Plataforma experimental:
 - Processador Intel i7-4770 CPU (3.40GHz)
 - Disco rígido HDD 1TB SATA II
 - Memória de 8GiB DIMM DDR3(1.6GHz)

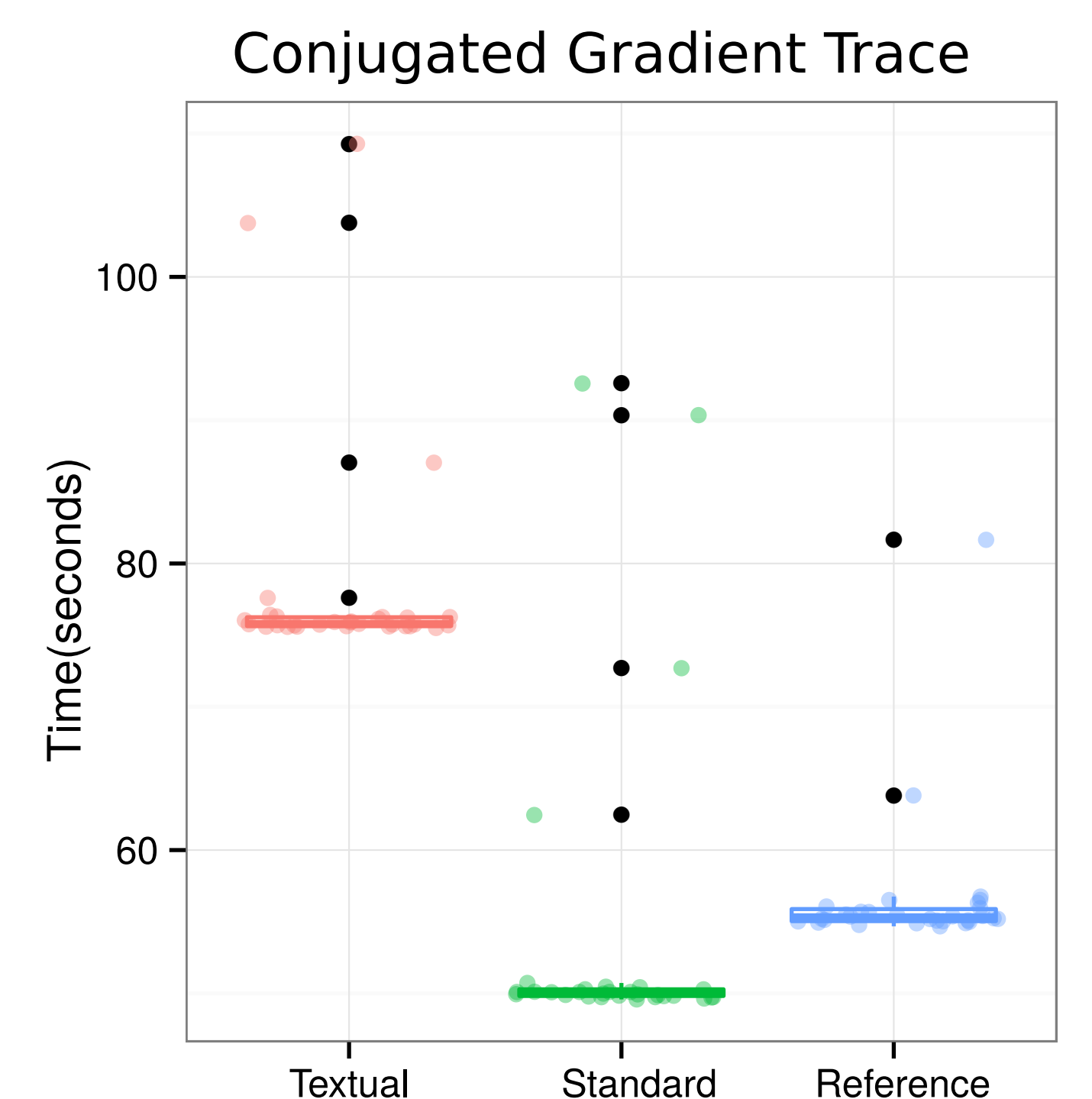


Figura 1: Tempo de execução dos formatos Pajé com arquivo de rastro cálculo de gradiente

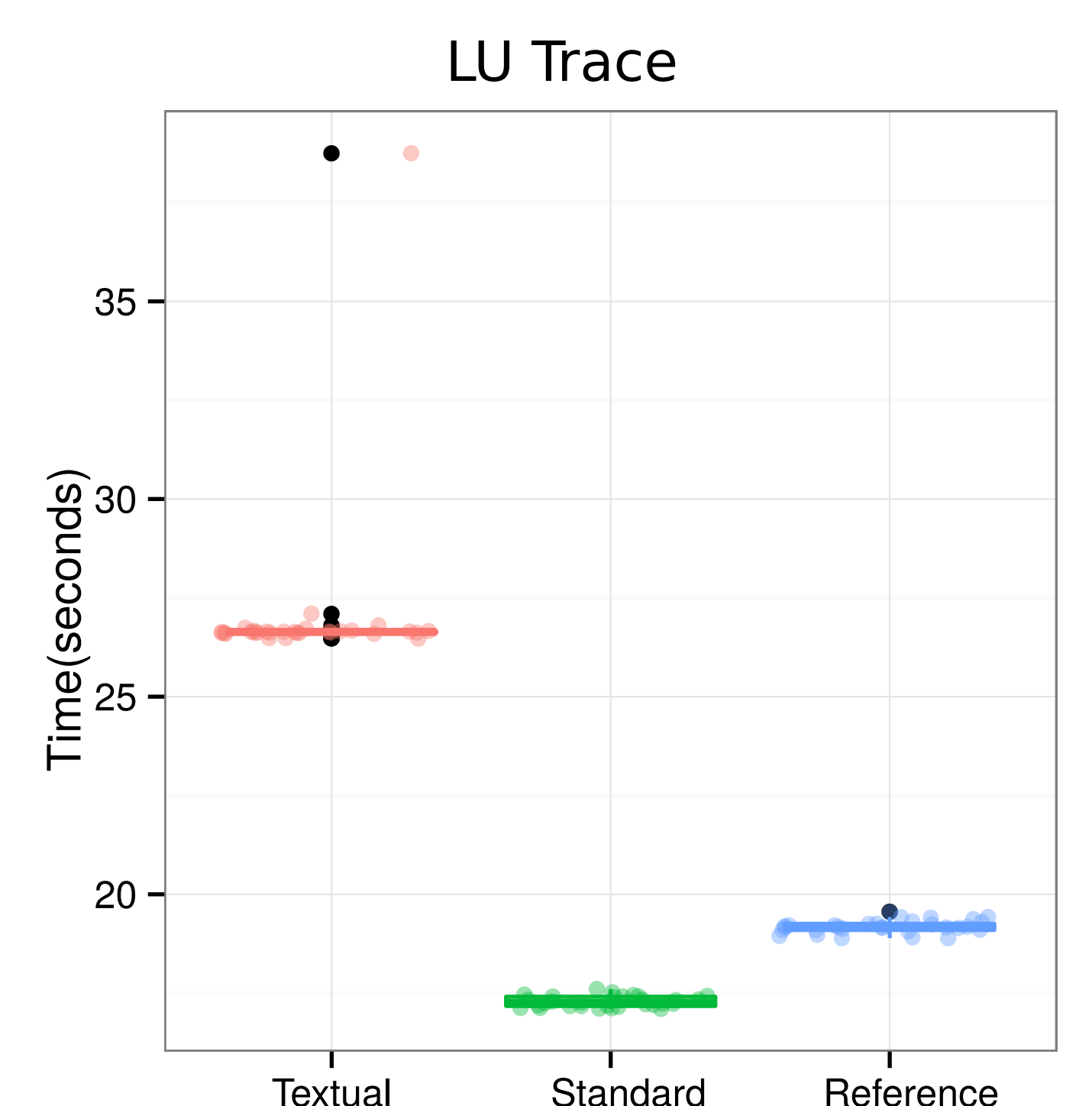


Figura 2: Tempo de execução dos formatos Pajé com arquivo de rastro LU

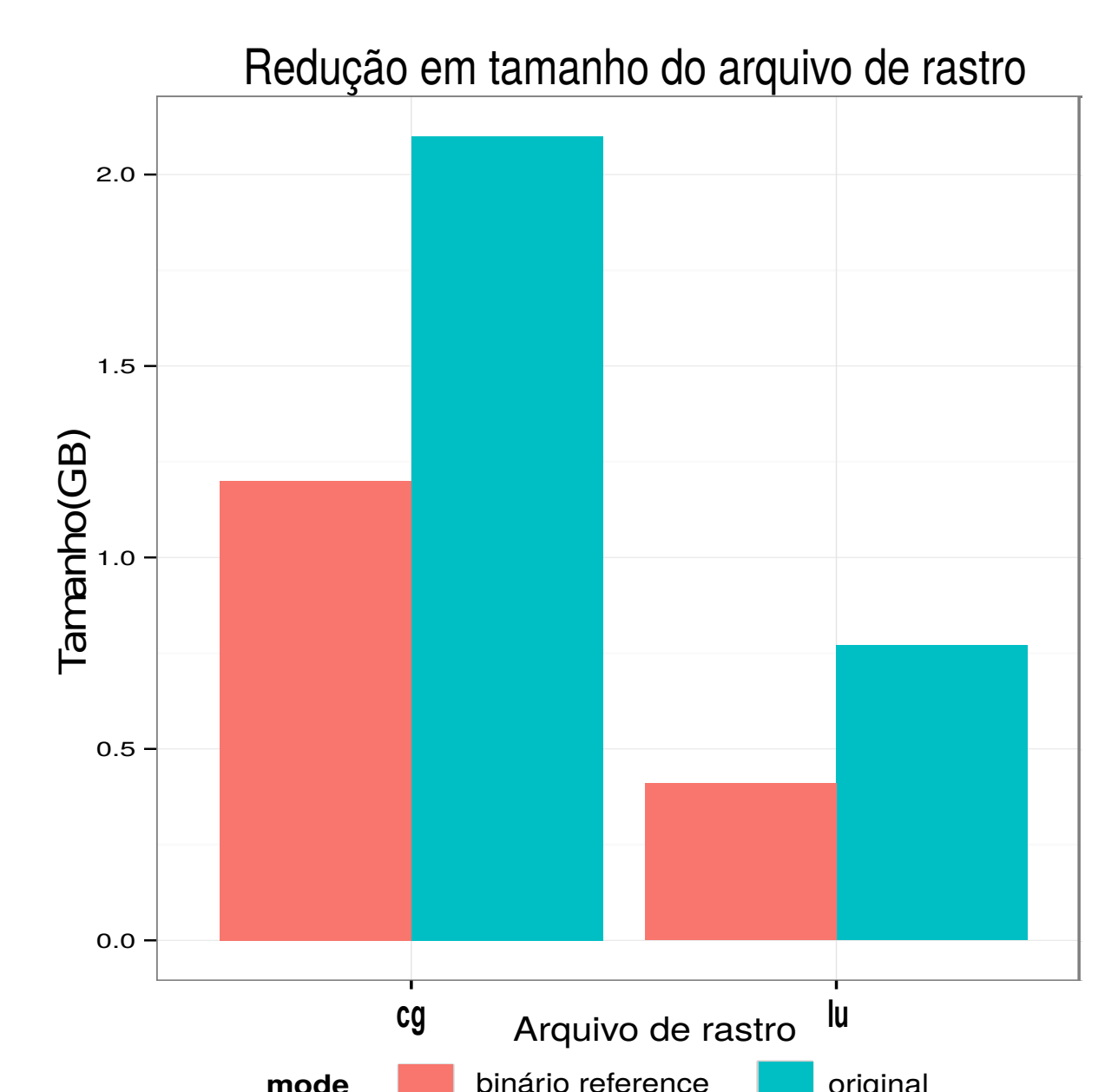


Figura 3: Redução do tamanho de arquivo