

Resumo

A comparação de rastros de execução de aplicações paralelas permite uma análise detalhada das diferenças entre duas versões ou execuções do programa. A comparação automática de rastros enormes apresenta desafios em automação, escalabilidade, e visualização de dados. Neste poster é apresentada uma estratégia para superar tais dificuldades baseada em um algoritmo de alinhamento de seqüências usado em bioinformática. Concluiu-se que o desenvolvimento de uma aplicação escalável utilizando essa estratégia é plausível. Porém, experimentos preliminares indicam a incapacidade da técnica, por si só, de providenciar uma visualização significativa ao analista para os casos testados.

1. Objetivos

Desenvolvimento de uma ferramenta de comparação de rastros que atenda aos critérios de:

- Automação
- Escalabilidade
 - Visualização compreensível
 - Utilização de recursos moderada
- Software livre

2. Desafios

- Identificação automática de regiões semelhantes
- Método escalável para comparação
- Visualização escalável dos resultados

3. Método

O método consiste em alinhar regiões semelhantes das seqüências de eventos registradas nos arquivos de rastro, afim de ressaltá-las ao usuário, facilitando a identificação das diferenças.

- Alinhamento de seqüências → algoritmo de Hirschberg
 - Tempo quadrático → paralelização
 - Espaço linear
- Templates agnósticas ao formato do arquivo de entrada
- Paralelização baseada na estrutura do rastro
- Formato de rastro Pajé para escrita dos resultados → Resultados em ferramenta de visualização qualquer

4. Visualização

- Casos de teste simples
 - Loop paralelizado com OpenMP
 - Escalonador dinâmico vs. guiado
- Comparação manual
 - Fácil identificação da distribuição das threads
 - Clareza no tempo de execução de cada caso
- Comparação automática (via alinhamento)
 - Desloca eventos no tempo para alinhar a semelhantes
 - Difícil identificação da distribuição das threads
 - Perda da informação cronológica da execução

As figuras 1 e 2 mostram os rastros visualizados pela ferramenta Pajé, que gera um diagrama de Gantt. Além do deslocamento de eventos no tempo, sempre que um evento está presente em apenas uma das seqüências são inseridos “eventos de lacuna” durante o alinhamento.

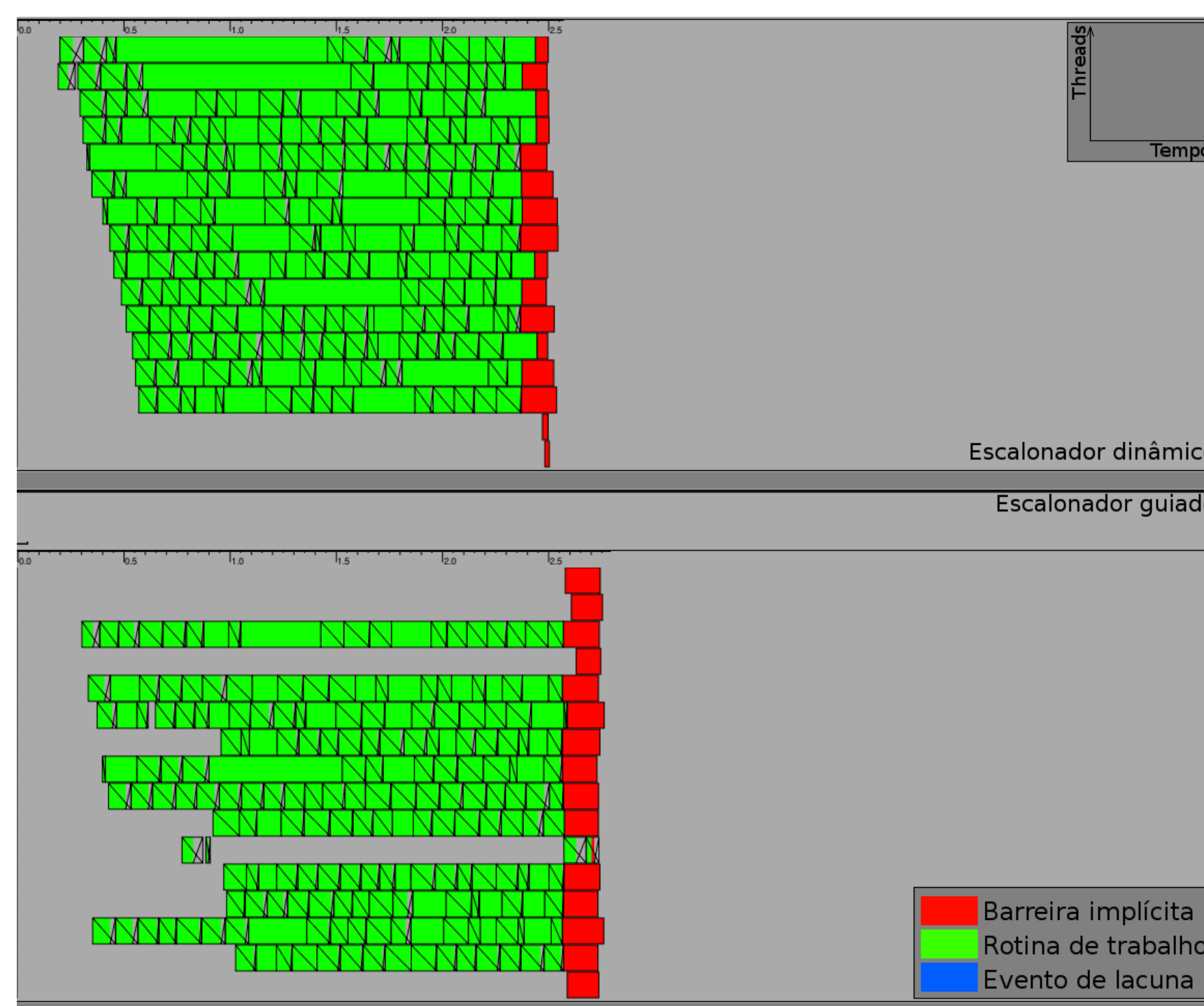


Figura 1: Rastros de execução não alinhados.

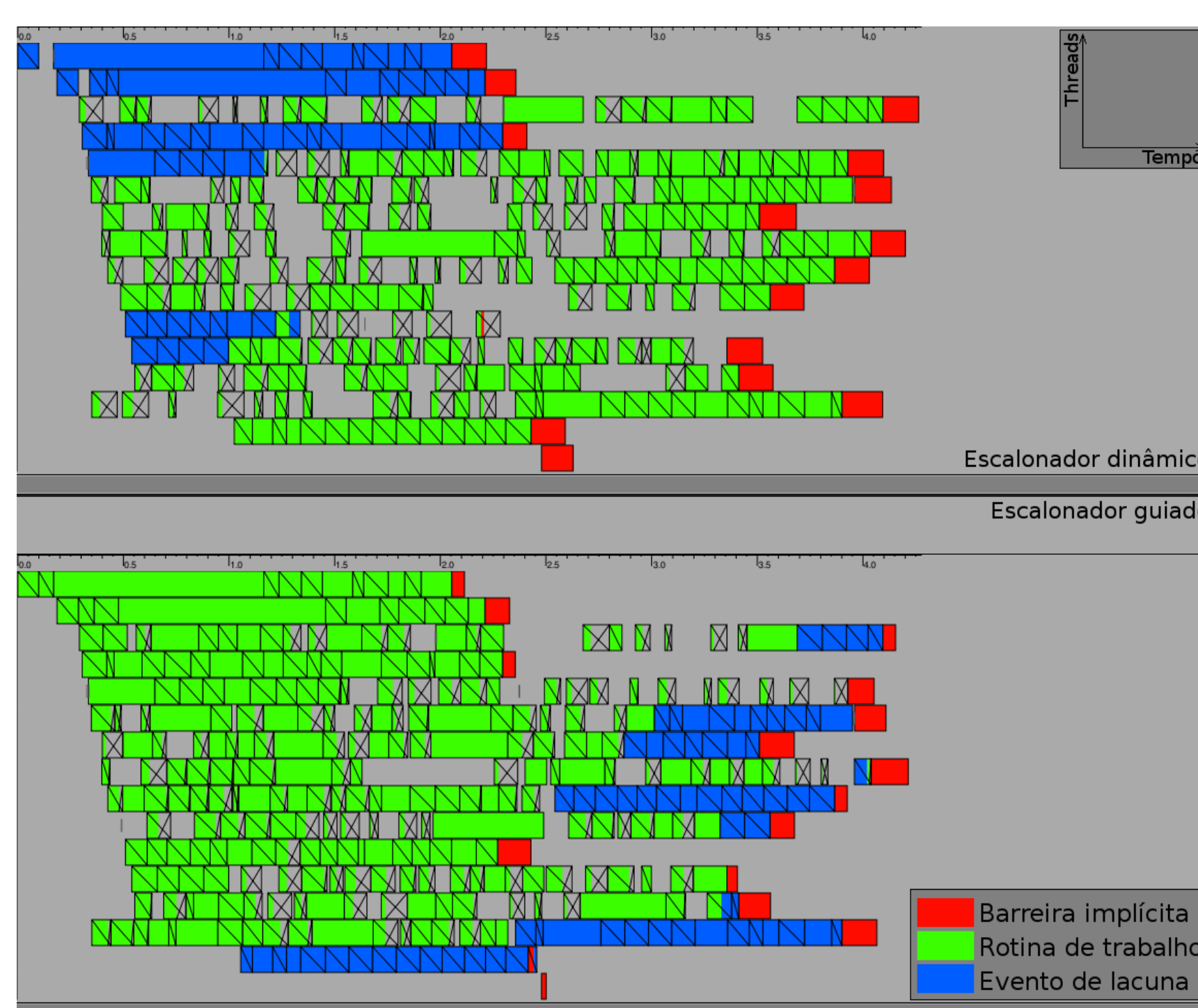


Figura 2: Rastros de execução alinhados.

5. Desempenho

- Testes com versão inicial da ferramenta
 - Rastros artificiais para controlar o número de eventos
 - Tempos crescem com taxa menor que quadrática [3]
- Testes preliminares com versão posterior
 - Rastros de aplicações reais
 - Tempos próximos a trabalhos anteriores [4]

Número de eventos	Pior	Médio	Melhor
961, 621	2.57s	2.54s	2.44s
34, 611, 196	3.31m	3.30m	3.29m

Todos testes foram realizados em um computador pessoal moderno com quatro processadores; trinta execuções foram realizadas para cada caso de teste.

6. Agregação de dados

Avaliou-se a possibilidade de comparar rastros de execução agregando regiões homogêneas entre eles, ao invés de agregar regiões homogêneas dentro de um mesmo arquivo, como as ferramentas de agregação geralmente fazem. Para isso, usou-se a ferramenta Ocelotl [2] e uniu-se dois arquivos de rastro em um, deixando a ferramenta agregá-los como bem entendesse. Foram realizados testes com os mesmos rastros da seção 4 e o resultado visual foi insatisfatório, como exibido nas figuras 3 e 4. Sugere-se unir a técnica à de alinhamento para que apenas regiões semelhantes sejam agregadas.

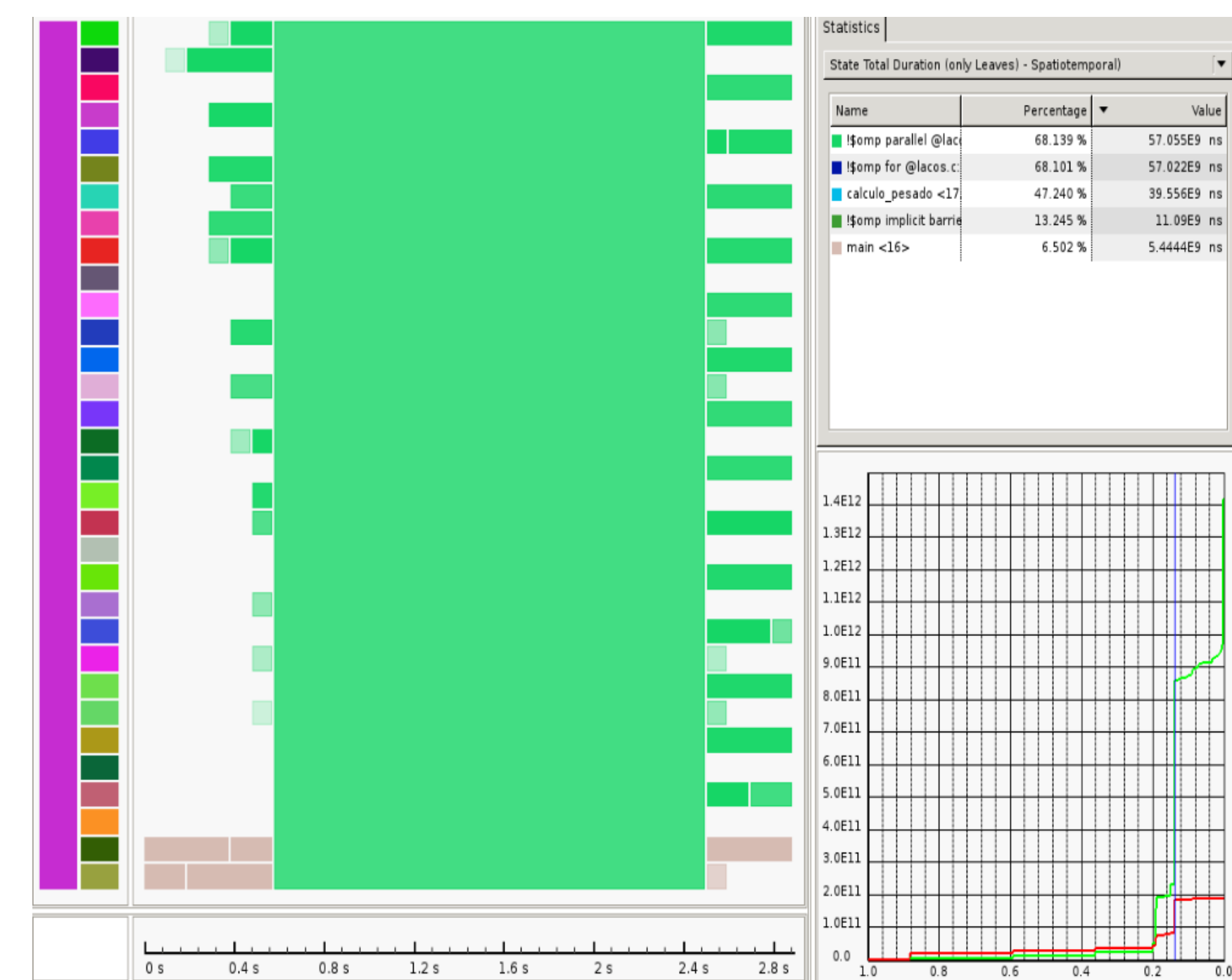


Figura 3: Rastros de execução agregados.

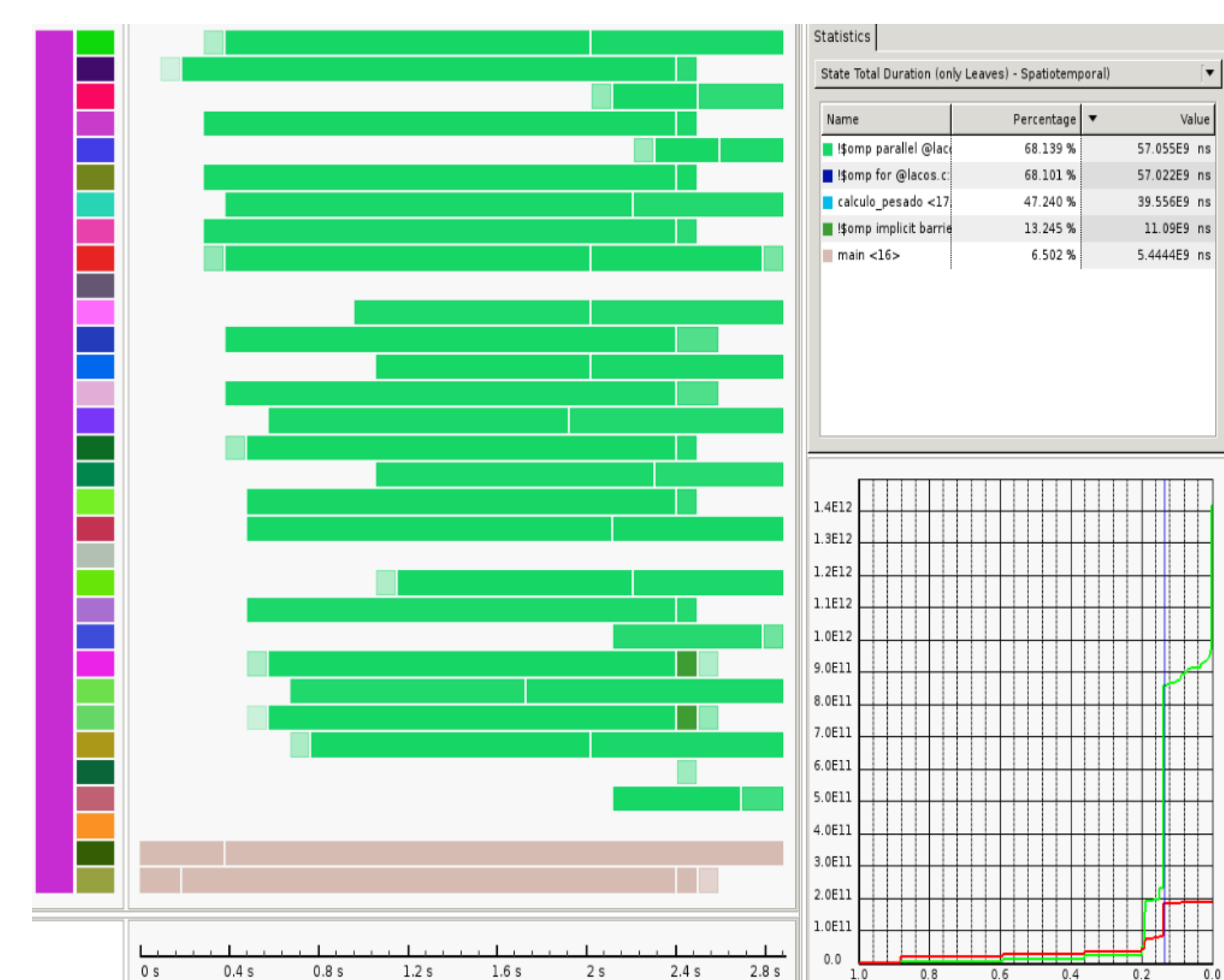


Figura 4: Nível reduzido de agregação.

7. Conclusão

Diferente dos autores do estudo original [4], concluiu-se que a técnica de alinhamento pode apresentar resultados insatisfatórios em certos casos. Métricas para identificar os deslocamentos no tempo podem mitigar o problema.

- Alinhamento de seqüências
 - Confuso em certos casos → métricas?
 - Desempenho e complexidade aceitáveis
- Agregação de dados
 - Confuso em certos casos → alinhamento?

Julgamos que mais testes são necessários para nosso trabalho, especialmente com rastros mais complexos, e apresentamos algumas sugestões abaixo.

- Solução híbrida utilizando MPI e OpenMP
- Alinhamento semi-global [1]

Referências

- [1] Sean R Eddy. A probabilistic model of local sequence alignment that simplifies statistical significance estimation. *PLoS computational biology*, 4(5):e1000069, 2008.
- [2] Damien Dosimont et. al. A spatiotemporal data aggregation technique for performance analysis of large-scale execution traces. In *Cluster Computing (CLUSTER), 2014 IEEE International Conference on*, pages 149–157, Sept 2014.
- [3] A. Farah and L.M. Schnorr. Comparação de rastros para análise de desempenho de aplicações paralelas. *Anais da XV Escola Regional de Alto Desempenho*, pp. 201–204, 2015.
- [4] M. Weber, R. Brendel, and H. Brunst. Trace file comparison with a hierarchical sequence alignment algorithm. In *Parallel and Distributed Processing with Applications (ISPA), IEEE 10th International Symposium on*, pages 247–254, 2012.