

Distributed Order Recording Techniques for Efficient Record-and-Replay of Multi-threaded Programs



(IEEE Cluster 2024 [49])

Xiang Fu^{†1}, Shiman Meng^{†1}, Weiping Zhang^{†1}, Luanzheng Guo^{†2}, Kento Sato^{†3},
Dong H. Ahn^{†4}, Ignacio Laguna^{†5}, Gregory L. Lee^{†5}, Martin Schulz^{†6}



^{†1}: Nanchang Hongkong University, ^{†2}: PNNL, ^{†3}: RIKEN, ^{†4}: NVIDIA, ^{†5}: LLNL, ^{†6}: Technical University of Munich

Best Paper Finalist

Background

- Applications can include non-deterministic (ND) bugs due to use of asynchronous algorithms by multi-threading, asynchronous communications
- ND bugs require significant amount of time to reproduced, identify and fix them
- We developed ReMPI for MPI record-and-replay (R&R) [1] and for R&R hybrid parallelism, we extended ReMPI support R&R for OpenMP

Approach:

- We developed an OpenMP recode-and-replay tool (ReOMP) to reproduce ND behaviors in OpenMP by recording order of threads executing: Critical section, Reduction, Atomic operations, data racy accesses
- However, recording thread scheduling behavior easily lead I/O performance bottleneck
- To reduce I/O performance bottleneck, we proposed efficient distributed order recording which can be used for other order-record problem

Results

- We could minimize I/O performance overhead for recording thread scheduling behaviors
- We successfully R&R non-determinism in MPI+OpenMP hybrid proxy applications (e.g., HACC, HPCCG, miniFE, QuickSilver and AMG)

- Github: <https://github.com/PRUNERS/ReMPI>

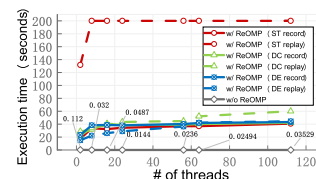
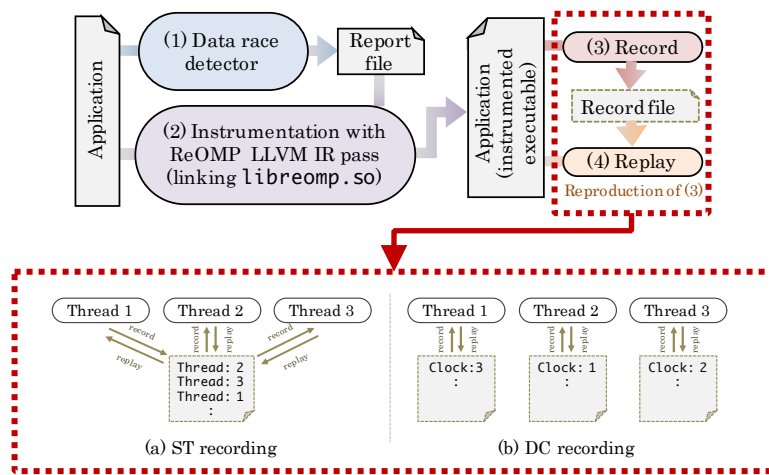


Fig. 13: Execution time of OpenMP AMG

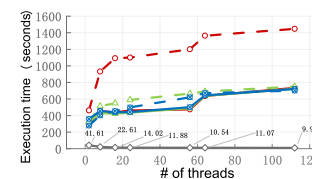


Fig. 14: Execution time of OpenMP QuickSilver

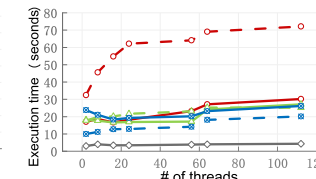


Fig. 15: Execution time of OpenMP miniFE

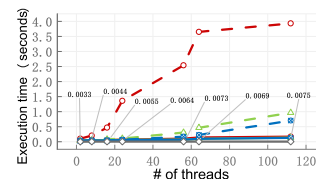


Fig. 16: Execution time of OpenMP HACC

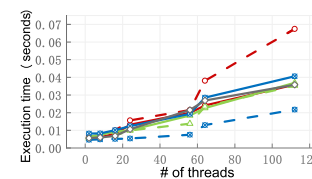


Fig. 17: Execution time of OpenMP HPCCG

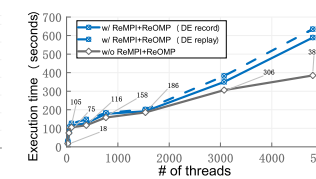


Fig. 18: Execution time of OpenMP+MPI HACC

[49] Xiang Fu, Shiman Meng, Weiping Zhang, Luanzheng Guo, Kento Sato, Dong H. Ahn, Ignacio Laguna, Gregory L. Lee, Martin Schulz, “Distributed Order Recording Techniques for Efficient Record-and-Replay of Multi-threaded Programs”, IEEE Cluster 2024, Kobe, Hyogo, Japan. (Best Paper Nomination)

[1] Kento Sato, Dong H. Ahn, Ignacio Laguna, Gregory L. Lee and Martin Schulz, “Clock Delta Compression for Scalable Order-Replay of Non-Deterministic Parallel Applications”, In Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis 2015 (SC15), Austin, USA, Nov, 2015