Laboratory for Parallel Programming

Bachelor's / Master's Thesis

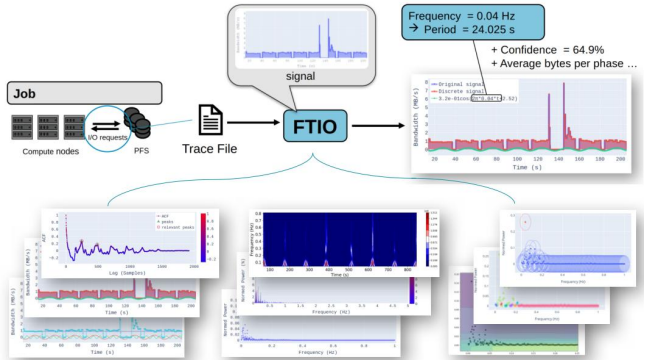
TECHNISCHE UNIVERSITÄT DARMSTADT

Parallel Programming

Modeling and Predicting I/O Phases in HPC

Motivation

Characterizing the temporal I/O behavior of an application is a challenging task. Yet, many HPC applications perform their I/O in bursts that follow a periodic pattern. System providers can leverage the knowledge about the periodicity of I/O phases to reduce file-system contention by actively scheduling I/O bandwidth or for burst buffer management. To find the period of the I/O phases, we developed the tool FTIO[a], which uses a well-known technique from the signal processing domain, namely the discrete Fourier transformation, combined with machine learning methods (e.g., outlier detection). Together with our tool TMIO[b], which traces MPI-IO online, we have demonstrated the applicability of our approach to I/O scheduling [2], boosting system utilization by 26% and reducing I/O slowdown by 56%. Both tools can operate independently, and several extensions with other popular HPC tools already exist [1].



FTIO in a Nutshell. The I/O requests are captured using TMIO and analyzed using FTIO to obtain the frequency of the I/O phases alongside the confidence in the results.

[a] https://github.com/tuda-parallel/FTIO/
[b] https://github.com/tuda-parallel/TMIO/

Task

As this is a rolling call, there are no fixed tasks to report here. You can approach us, and we can figure out together what best suits you and your interests. While topics on TMIO are available, FTIO is currently the focus. Possible directions to extend FTIO include: Reducing noise using filters, applying other signal processing techniques to enhance prediction accuracy, using AI methods to model and predict the I/O phases, or improving the online time window adaption (see [1]).

Requirements

- Depending on the topic: Python or C++
- Knowledge in parallel programming and MPI

Contact

Ahmad Tarraf <ahmad.tarraf@tu-darmstadt.de>

References

1. Ahmad Tarraf et al. "Capturing Periodic I/O Using Frequency Techniques". In: 2024 IEEE International Parallel and Distributed Processing Symposium (IPDPS). San Francisco, CA, USA, 2024-05-27/2024-05-30, pp. 1–14.
2. Jean-Baptiste Besnard et al. "Introducing the Metric Proxy for Holistic I/O Measurements". In: The 10th HPC I/O in the Data Center Workshop (HPC-IODC'24), held in conjunction with the ISC High Performance Conference. Hamburg, Germany, May 2024, pp. 1–15.

Issue date: **July 2024**