High Performance Computing 2023 - Exercise 1

Marco Zampar - SM3800032

September 25, 2024

Problem Statement

A profiling study reports that **80%** of the total execution time of MPI applications is consumed by **MPI collective operations**.

This project aims to:

- Estimate the latency of the default OpenMPI implementation.
- Vary the number of processes and message sizes.
- Compare the results with other algorithms and process mappings.

Computational Resources

- 2 THIN nodes from the ORFEO cluster.
- Each node has 2 CPUs with 12 cores in a single NUMA region.

More information: ORFEO cluster documentation.

Broadcast Algorithms

Flat Tree Algorithm

- Single-level tree topology.
- Root node transmits to P-1 child nodes.

Chain Tree Algorithm

- Internal nodes have one child.
- Messages are split into segments, transmitted in a pipeline.

Binary Tree Algorithm

- Internal process has two children.
- Segmentation is used to improve communication parallelism.

Latency Analysis

Latency vs Message Size (Fixed Processes)

- Scaling is linear across different message sizes.
- Observed consistent results across different process mappings.

Latency vs Number of Processes (Fixed Message Size)

- Different behaviors for small/medium and large message sizes.
- Mapping by core, socket, and node influences latency trends.

Performance Models

Linear regression models were used to estimate the latency surface, varying the number of processes and message sizes.

Results:

- Core mapping shows better results for the linear algorithm.
- Binary Tree and Chain Tree algorithms perform similarly with node and socket mappings.

Barrier Algorithms

Algorithms Analyzed:

- **Linear**: All nodes report to a root.
- **Tree**: Hierarchical synchronization in a tree-like structure.
- **Recursive Doubling**: Logarithmic communication steps, optimal for powers of 2.

Conclusion

- OpenMPI's default algorithm doesn't always choose the optimal one in terms of latency.
- However, differences in performance are often minor, making the default a reasonable choice in many cases.
- Custom mapping and algorithm selection can lead to performance improvements for specific configurations.

Appendix

Refer to the 3D plots and latency vs process mappings in the appendix for detailed visual insights.