

Automatic Tuning of Collective Communications in MPI

Rajesh Nishtala, Kushal Chakrabarti,
Neil Patel, Kaushal Sanghavi
Computer Science Division
University of California at Berkeley

- No single implementation of MPI collectives is optimal across all environmental variables (eg. node architecture and network load).
- Our Probabilistic Algorithm Selection System (PASS) accounts for this and optimizes the collectives by learning from the implementations' past performance
- PASS optimizes operations above the level of underlying MPI point to point operations, making it cluster and implementation independent and thus adaptive and extensible.
- Our new tuned implementations yield up to 10x speedups through the use of pipelining.

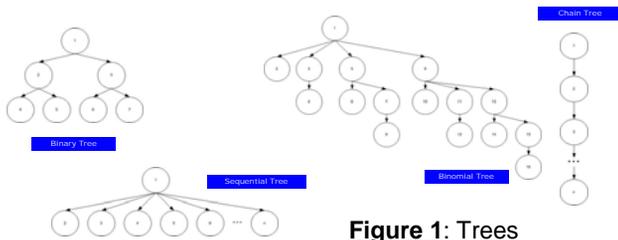


Figure 1: Trees

Performance results for four different implementations of MPI collectives were gathered varying the following parameters:

- Cluster interconnect
- Node architecture
- Number of nodes
- Pipeline segment size

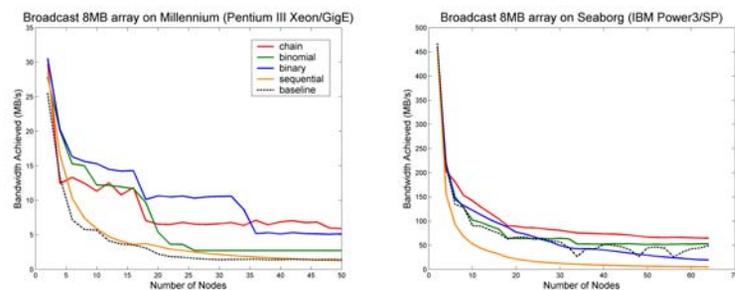


Figure 2: The best implementation varies across clusters and the number of nodes. For instance, the chain tree is always the best implementation on Seaborg, while it is only best for large numbers of nodes on Millennium. Because the binary tree is better for smaller numbers of nodes and the base implementation is suboptimal, space exists for performance tuning.

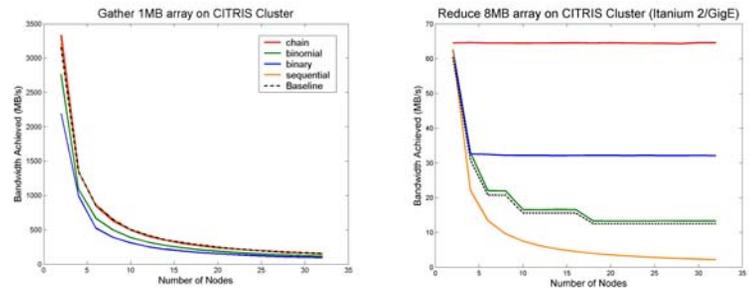


Figure 3: Because there is no definitive winning implementation of gather on CITRIS, transient cluster conditions could cause a superior implementation to emerge. PASS dynamically accounts for these changes and selects the optimal implementation.

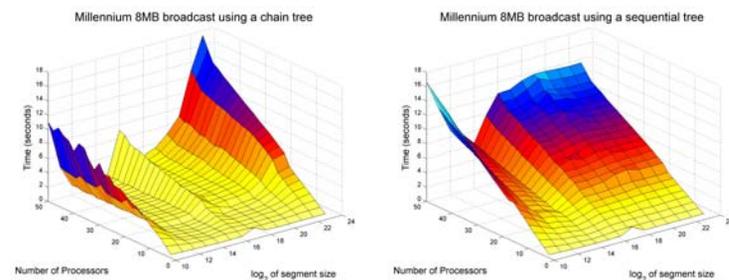


Figure 4: Pipelined transmission of messages yields significant improvements. Although performance is very sensitive to the unit of transmission (segment size), PASS will discover the optimal size.

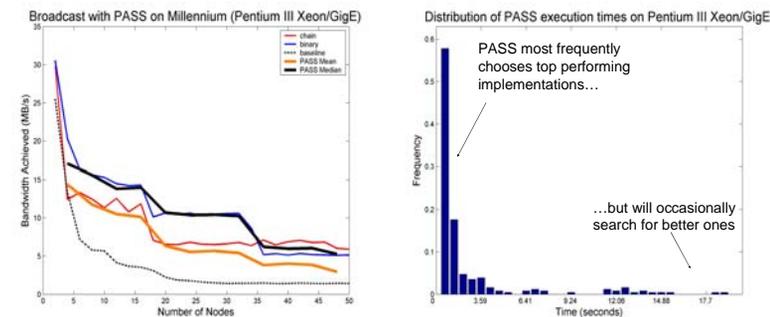


Figure 5: PASS accounts for varying performance times (shown above) by usually choosing the optimal ones. Averaging over multiple trials, the difference between the PASS mean and median lines illustrate the effect of exploration. Future work will focus on fine-tuning PASS so that the negative effects of exploration are negligible.