



Automatic Tuning of Collectives

By Rajesh Nishtala, Kushal Chakrabarti,
Kaushal Sanghavi, and Neil Patel

Advisors: Katherine A. Yelick and
James W. Demmel
University of California at Berkeley,
Computer Science Division



Outline

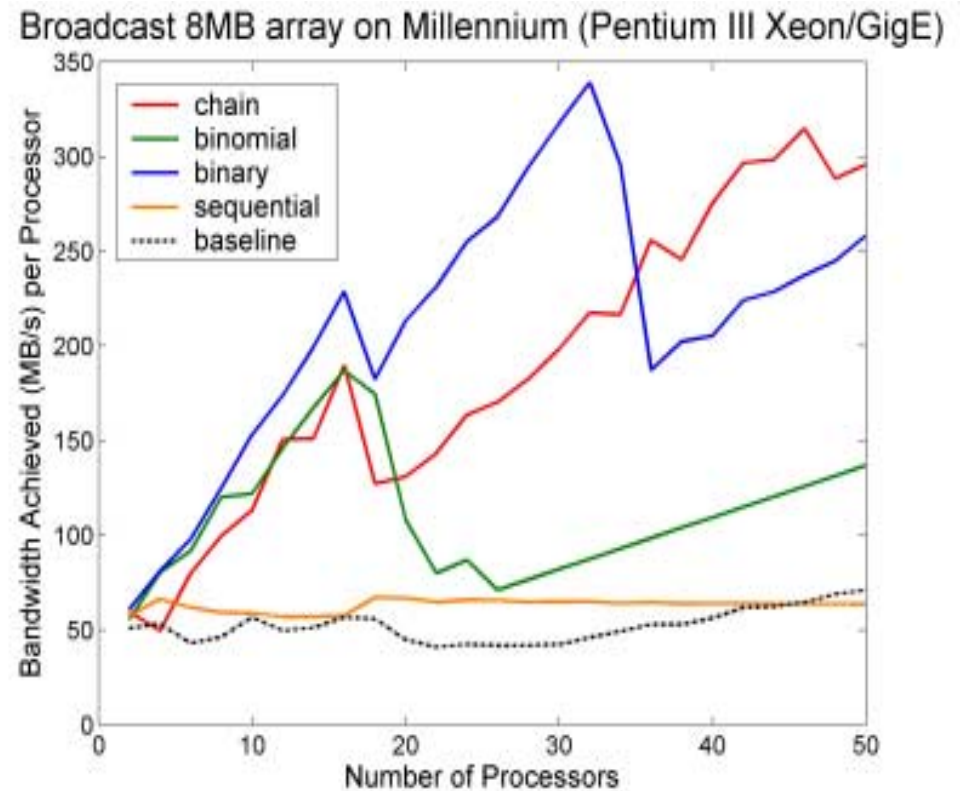
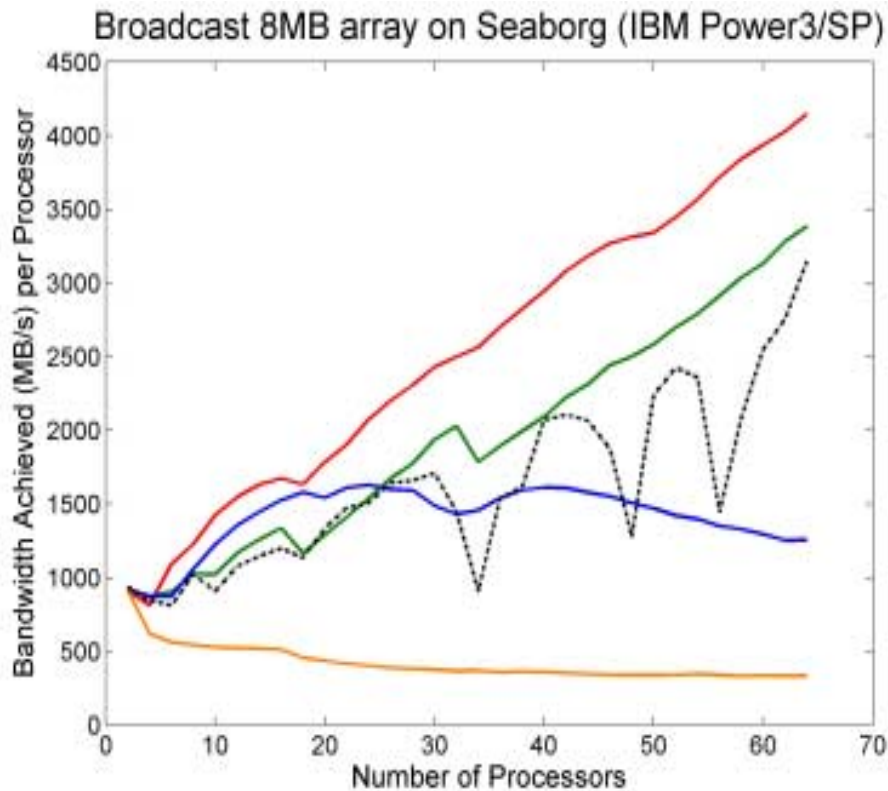


- Motivation
- Description of Optimized Collectives
 - Preliminary results are for MPI collectives using MPI point-to-point operation
 - Extension of work by Dongarra et al.
- Description of Optimizations
- Automatic Tuning
- Conclusions
- Future Work



Motivation

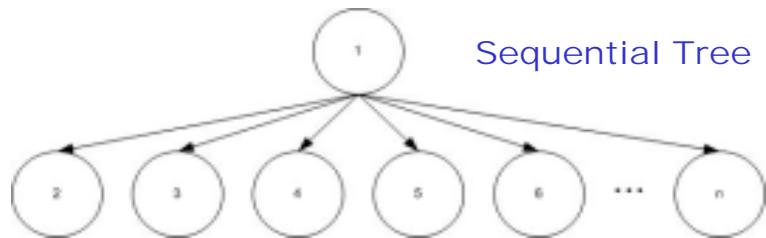
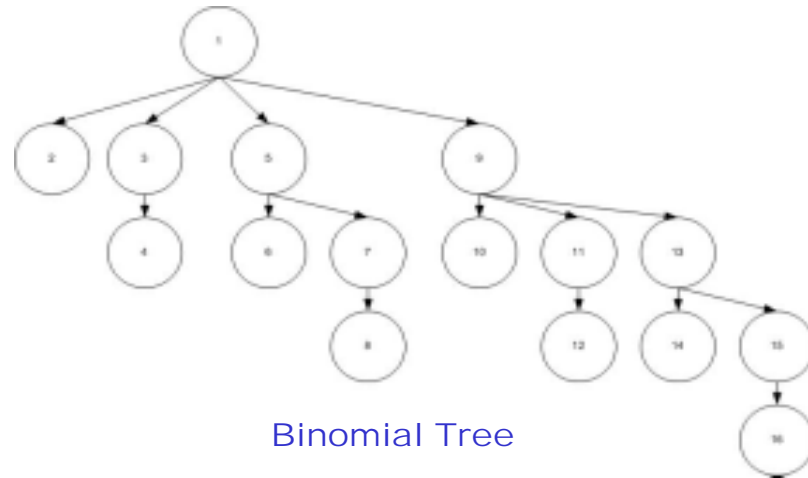
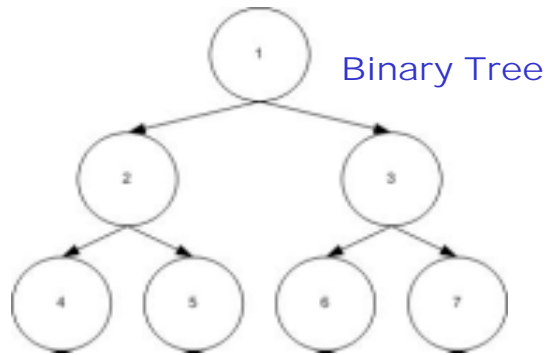
- No single optimal implementation for collectives
- Best algorithm varies across number of nodes, cluster architecture, and message size.



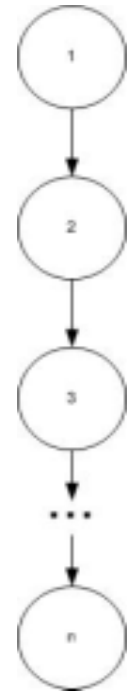


Optimizations (Tree Structure)

- Used 4 Tree Structures



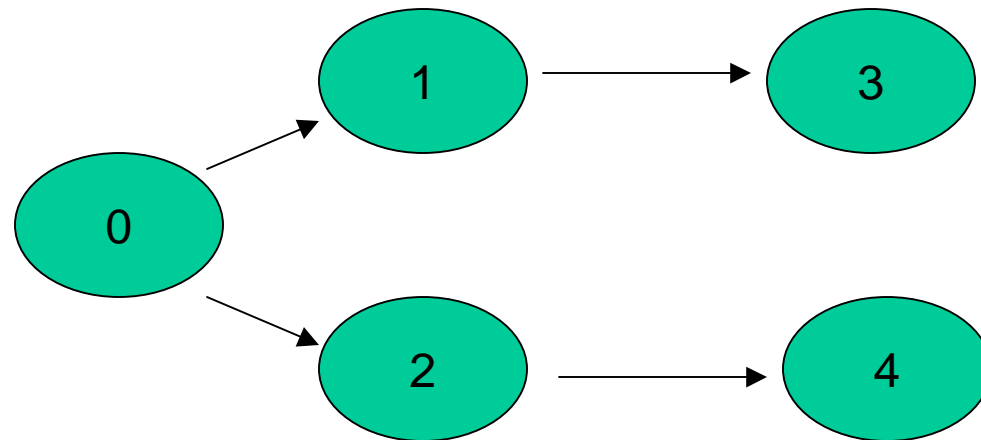
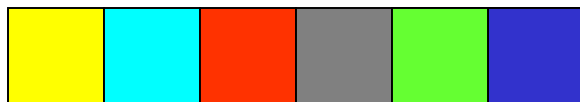
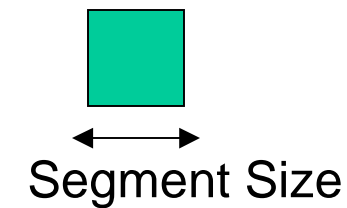
Chain Tree





Optimizations (Pipelining)

- Pipeline the messages too improve the throughput

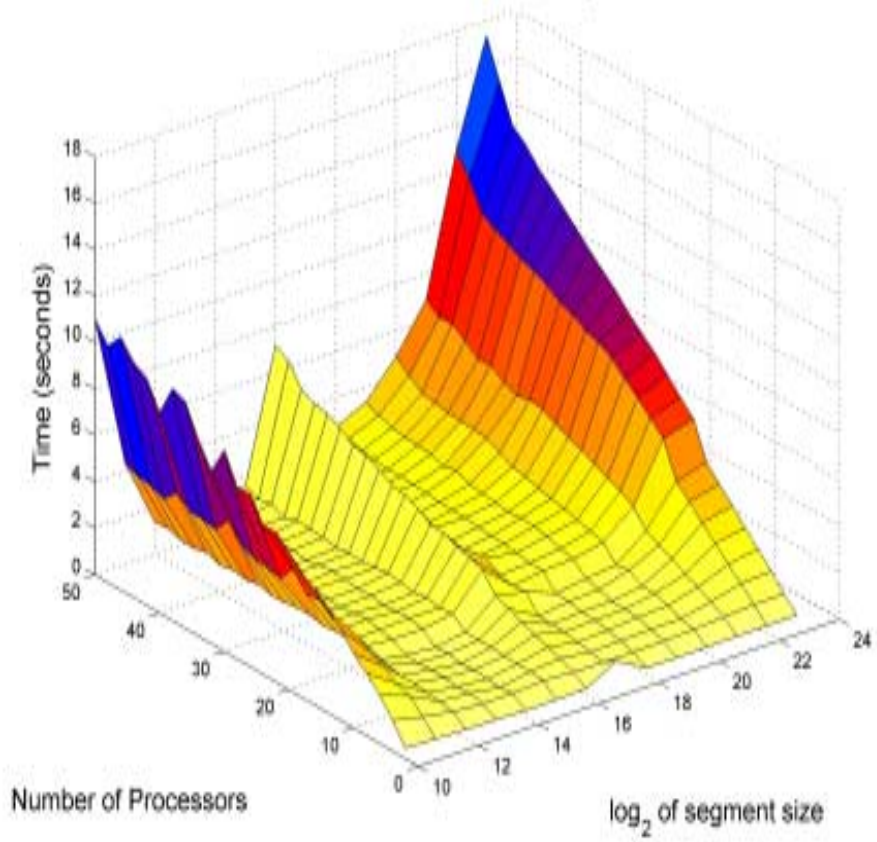




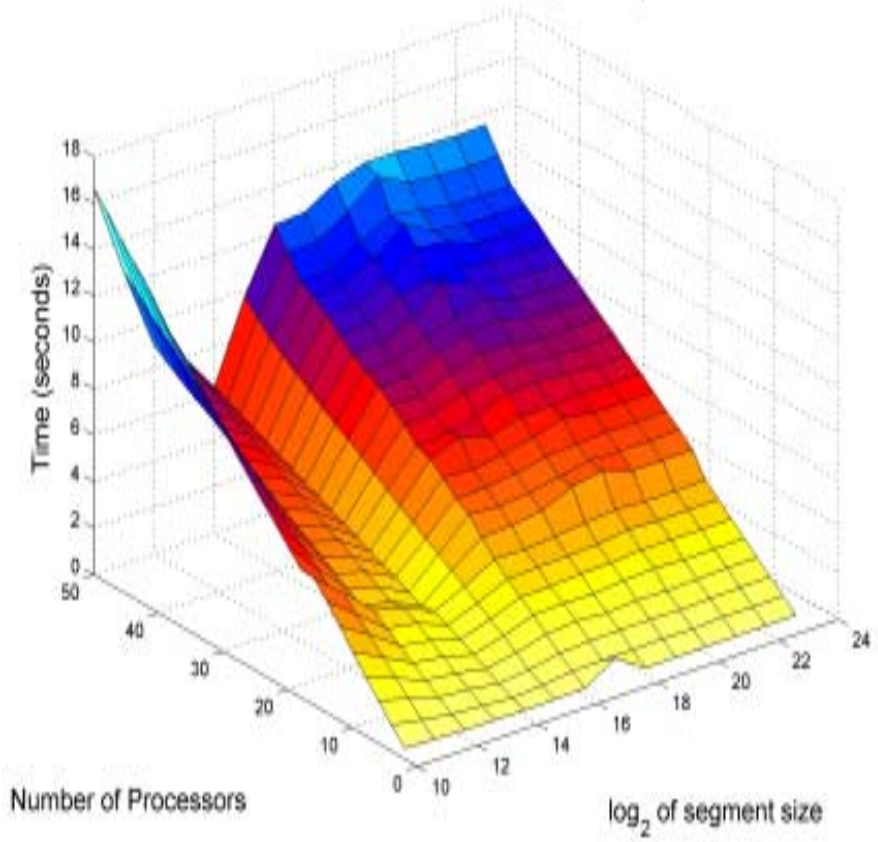
Selection of Segment Choice

- Performance is sensitive to segment size choice

Millennium 8MB broadcast using a chain tree



Millennium 8MB broadcast using a sequential tree

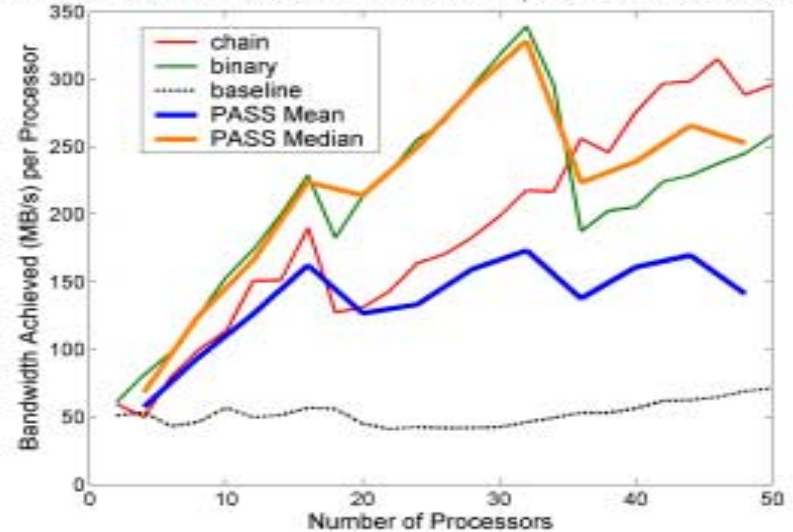




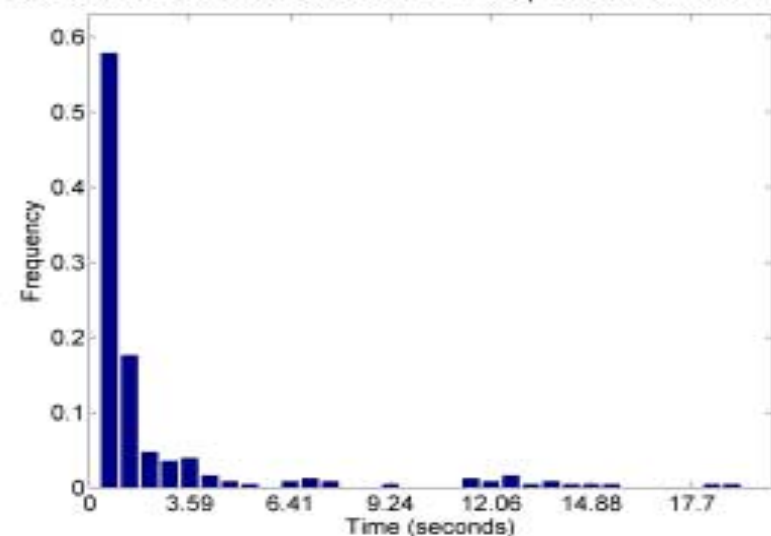
Automatic Tuning

- Use a probabilistic algorithm selection system (PASS) to choose the algorithms we run
- Based on history, faster implementations are chosen more frequently with the ability to search

Broadcast with PASS on Millennium (Pentium III Xeon/GigE)



Distribution of PASS execution times (Pentium III Xeon/GigE)





Conclusions



- Offline tuning clearly pays off
 - Also important to be able to tune applications during runtime
- Search space for automatic tuning increases with GASNet and UPC



Future Work and References



- Analyze more collectives such as all-to-all
- Refine automatic tuning system
 - reduce the penalty of search
- Experiment on more interconnects and novel cluster architectures

- References and detailed summary of work:
 - http://www.cs.berkeley.edu/~rajeshn/mpi_opt.pdf
- Questions?
 - email: rajeshn@eecs.berkeley.edu