Enabling Optimization under Uncertainty using High Performance Computing

Parallel Algorithms for Stochastic Integer Optimizations

Akhil Langer (4th year PhD Student), Udatta Palekar* (Co-Advisor), Laxmikant (Sanjay) Kale (Advisor)
Department of Computer Science, *Department of Business, University of Illinois at Urbana-Champaign

PARALLEL PROGRAMMING LABORATORY

COLLEGE of BUSINESS at ILLINOIS

MOTIVATION

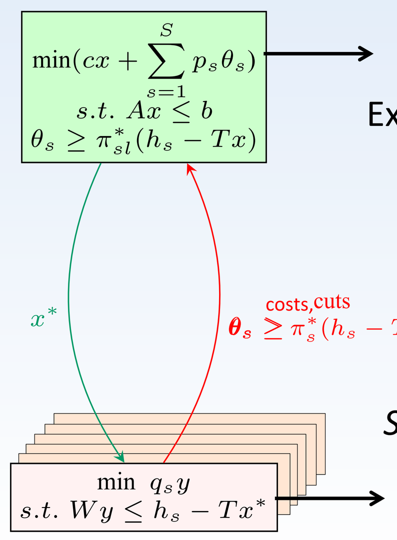
US Air Mobility Command Mission:
Optimally schedule 1300 aircraft for

- Cargo shipment
- Personnel movement
- Distinguished visitor support
- Air refueling

CHALLENGE
Mission requirements subject to enormous uncertainty, even in peacetime

- Demand delays
- Demand changes
- Aircraft breakdown
- Weather events
- Natural disaster
- Conflicts

STOCHASTIC OPTIMIZATION OVERVIEW



Stage 1 (strategic decision)
Cost of resource allocation + Expected cost of stage 2 decisions
(Integer) Linear Program

Planning Stage: Makes decision about known parameters

Stage 2 (operational decisions)
Execution cost
(Integer) Linear Program

Execution stage: Assumes probabilistic distribution of uncertain parameters is known and evaluates them for the stage 1 decision

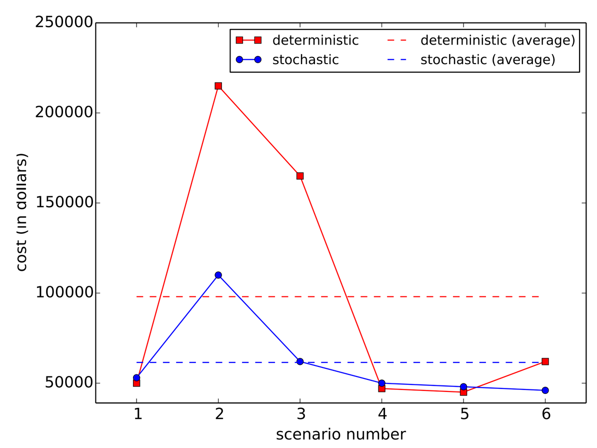
APPLICATIONS

RELATED WORK

- Extensive Formulation
  Infeasible for toy problems, commercial IP solvers have poor parallel efficiency
- PySP solver (Sandia National Lab)
  Heuristic method, inefficient IP solves
- Lubin et al (Argonne National Lab)
  Poor load balancing, scaling up to 32 cores

STATE-OF-THE-ART CANNOT SOLVE LARGE PROBLEMS TO OPTIMALITY

CASE STUDY: US MILITARY AICRAFT ALLOCATION PROBLEM

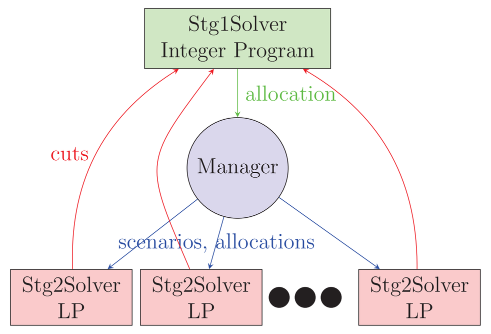


US AMC yearly expenses: ~USD 4 Billion
- Average cost benefit: 10%
- Robustness: 66% reduction in variance

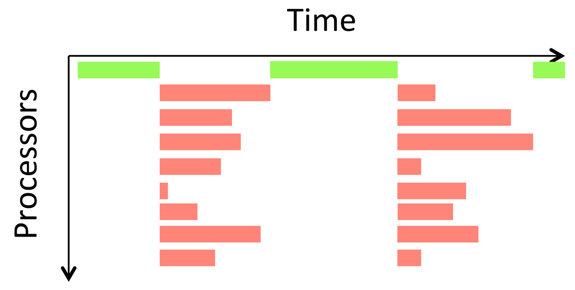
What is different/ new that we are doing?

- Solving large scale stochastic optimization problems
  - Not looked before because stochastic-optimization is a hard problem
- Combine stochastic programming with high performance computing
  - Facilitates reconciliation of myriad possible outcomes in a timely manner
  - Makes it feasible to solve integer programs

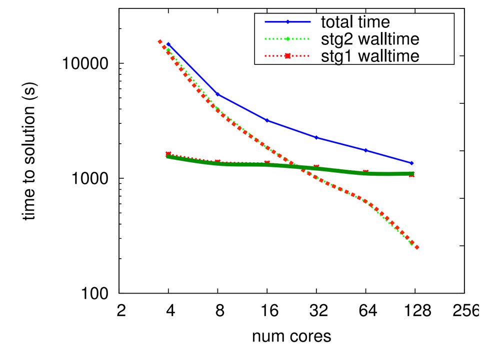
LIMITATIONS OF NAÏVE PARALLELIZATION



Parallel Design



Timeline



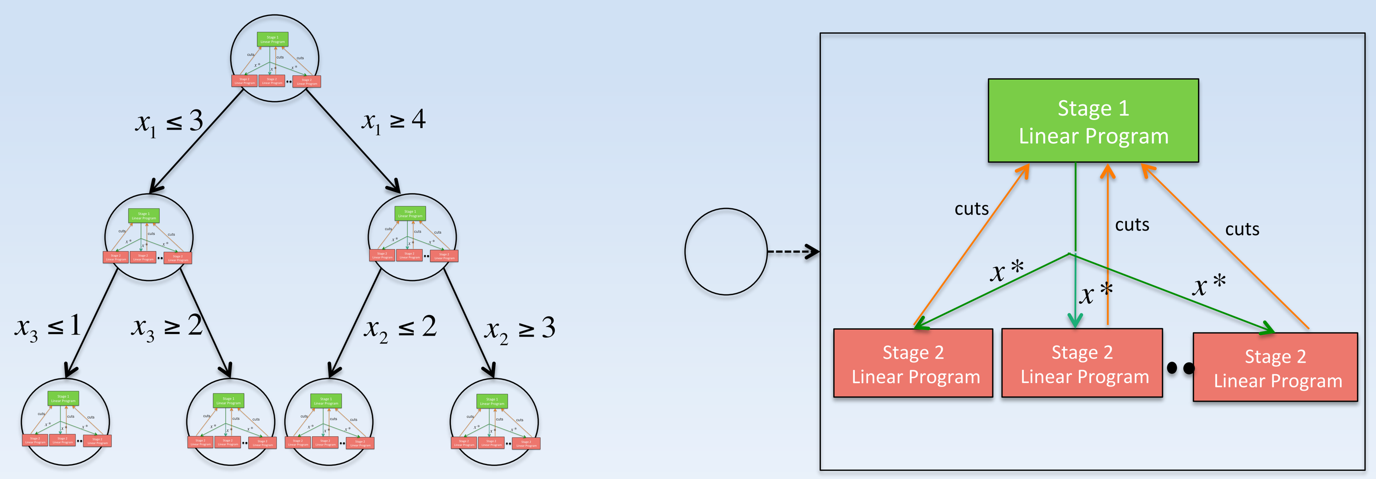
Scalability: 120 scenarios

Limitations of the design
- Scalability limited by #scenarios
- Stage 1 bottleneck
  - Parallel efficiency decreases with increase in stage 1 size

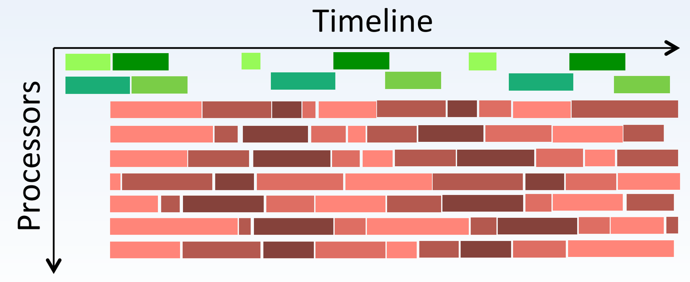
PROPOSED NESTED BRANCH-AND-BOUND PARALLELISM

Relax integer program to linear program → Each vertex is a stochastic linear program

Unlike traditional applications, a tree vertex is non-atomic execution unit



Nested Parallelism



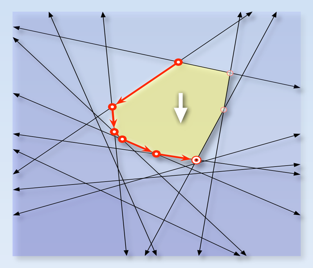
- Parallel exploration of branch-and-bound tree vertices
- Simultaneous evaluation of scenarios

DESIGN CHALLENGES AND THEIR PROPOSED SOLUTIONS

Challenges different than those in typical, scientific, iterative applications

Challenge 1: Fundamental unit of computation is linear program optimization which are *coarse-grained and have unpredictable sizes*

IBM ILOG CPLEX · Gurobi Optimization · COIN-OR CLP · GLPK



Challenge 2: Variable amount of available parallelism in branch-and-bound tree
Solution: Use pull-based load balancing

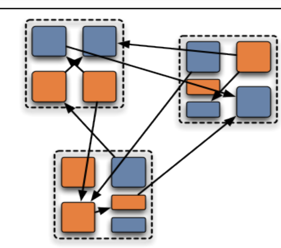
Challenge 3: Solver libraries maintain internal state for warm starts
- limits concurrency due to memory overheads
Solution: Dedicate processors to solvers

Challenge 4: Better processor utilization ≠ Better performance
Solution: Prioritized exploration of tree vertices

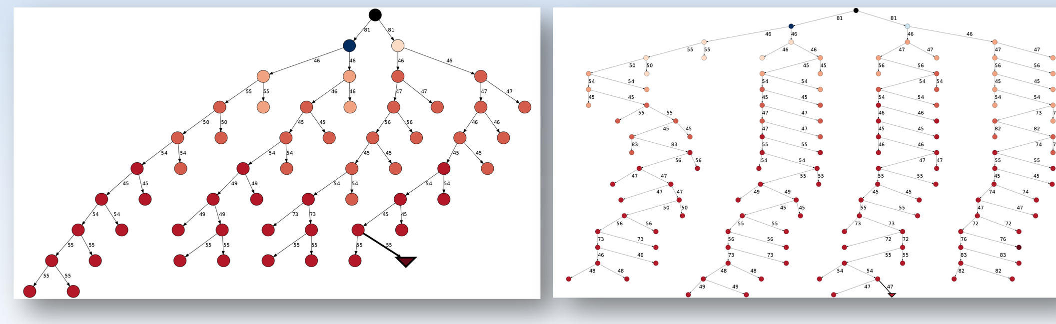
PARALLEL PROGRAMMING MODEL

Charm++ parallel programming model fits well into B&B parallelism

- Object-based expression
- One-sided messaging
- Prioritized execution



PERFORMANCE VARIATION ACROSS IDENTICAL TRIALS

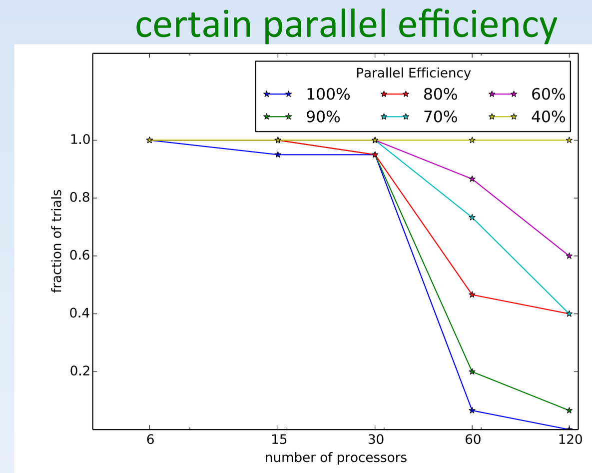


Cause of Variation
- Incumbents found in different order
- Sequence of vertex optimization determines solution, when LPs are degenerate

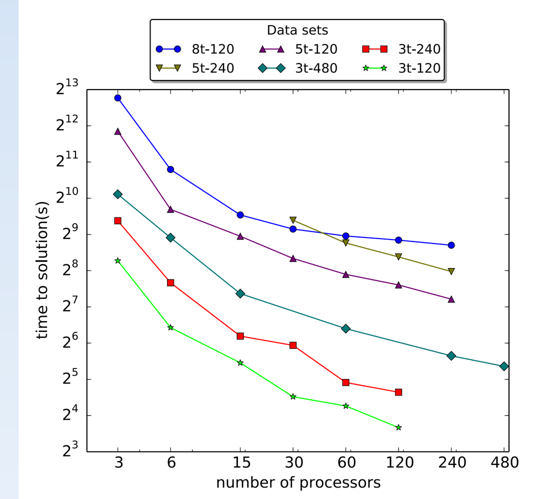
How bad is performance variation?
Makes it difficult to guarantee better performance with increase in the number of processors

PERFORMANCE COMPARISON WITH COMMERCIAL SOLVERS

Performance measurement in presence of variation
Solution: Fraction of trials achieving certain parallel efficiency



Scaling



SIPS Performance
1. 96x speedup from 3 to 480 cores
2. Parallel efficiency > 0.4 at 120 cores in ~100% of the cases

VS

Gurobi IP Solver Performance
Typical parallel efficiency of just 0.1 at 32 cores

ABLE TO SOLVE LARGE PROBLEMS IN A TIMELY MANNER WITH SIGNIFICANT PARALLEL SPEEDUPS

ONGOING WORK

- Decomposition Schemes
  - Scenario based decomposition
  - Lagrangean based temporal decomposition
- Automated distribution of processes to stage 1 and stage 2

IMPACT

Robust planning of **Real-time US Air Mobility Command operations** leading to
- **reduced transportation costs,**
- **improved personnel satisfaction, and**
- **timely delivery of mission-critical cargo**

while doing **dynamic mission re-planning** as disruptions occur

POTENTIAL IMPACT

Energy grid optimizations for *renewable energy resources* that are unpredictable by nature
- Unit commitment problem
- Energy dispatch problem

TAKEAWAYS

- Enabled large-scale stochastic optimizations leading to robust planning of US AMC operations
- Reduced memory overhead and time to solution using state sharing amongst tree vertices
- Asynchronous parallel programming model for maximum productivity and performance

REFERENCES

- Akhil Langer, Ramprasad Venkataraman, Udatta Palekar, and Laxmikant V. Kale. "Parallel branch-and-bound for two-stage stochastic integer optimization." In *High Performance Computing (HiPC), 2013 20th International Conference on*, pp. 266-275. IEEE, 2013.
- Langer, Akhil, Ramprasad Venkataraman, Udatta Palekar, Laxmikant Kale, and Steven Baker. "Performance Optimization of a Parallel, Two Stage Stochastic Linear Program." In *2012 IEEE 18th International Conference on Parallel and Distributed Systems*, pp. 676-683. IEEE, 2012.
- Laxmikant Kale, Anshu Arya, Nikhil Jain, Akhil Langer, Jonathan Liander, Harshitha Menon, Xiang Ni, Yanhua Sun, Ehsan Totoni, Ramprasad Venkataraman, and Lukasz Wesolowski. Migratable Objects + Active Messages + Adaptive Runtime = Productivity + Performance A Submission to 2012 HPC Class II Challenge. Technical Report 12-47, Parallel Programming Laboratory, November 2012.
- Langer, Akhil, Ramprasad Venkataraman, Gagan Gupta, Laxmikant Kale, Udatta Palekar, Steven Baker, and Mark Surina. "Poster: enabling massive parallelism for stochastic optimization." In *Proceedings of the 2011 companion on High Performance Computing Networking, Storage and Analysis Companion*, pp. 89-90. ACM, 2011.