CODES FOR STORAGE WITH QUEUES FOR ACCESS

Mentors: Dr. Emina Soljanin and Graduate student Amir Behrouzi-Far

Dalton Burke and Elise Catania
DIMACS 2018

Background

- In industry, there is an emphasis on big data and machine learning
- Projects can involve a great deal of information for computers to process and compute
- Work is divided between different servers, “workers”
- This division of work is called *parallelism*

Helpful Analogy-Time for the Grocery Store! How 3 People Buy 30 Pints of Ice Cream

- Step 1: Have three people acquire 30 pints of ice cream.
- Step 2: Walk with 30 pints to the checkout lines.
- Step 3: Realize scanning 30 pints of ice cream in one line would take forever and you really want to eat the ice cream.
- Step 4: Parallelize-give each person 10 pints of ice cream to check out
- Step 5: Each person gets into a random (*for now*) line
- Step 6: Enjoy the ice cream in 1/3 the time!

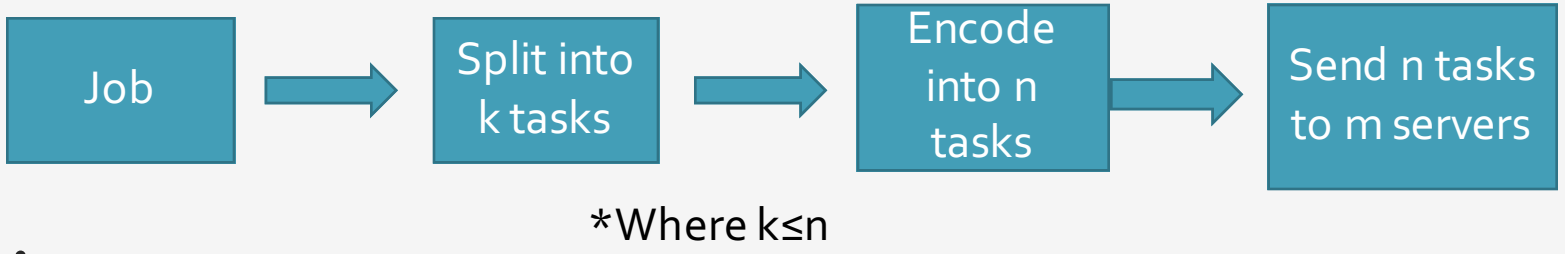
However, what if one person had a really long wait to check out?

That could have significantly slowed down the job.

New and Improved Method

- Step 1: Have three people acquire 45 pints of ice cream.
- Step 2: Walk with 45 pints to the checkout lines.
- Step 3: Realize scanning 45 pints of ice cream in one line would take forever and you really want to eat the ice cream.
- Step 4: Parallelize- give each person 15 pints of ice cream to check out
- Step 5: Each person gets into a random (*for now*) line
- Step 6: When the first 2 people check out, the job is done!
- Step 7: Call the last friend to get out of their line and go enjoy your 30 pints of ice cream!

Project described a little more mathematically



Any k sized subset of tasks recovers all of the information we need to finish our job!

Simplifying Assumptions (These will change)

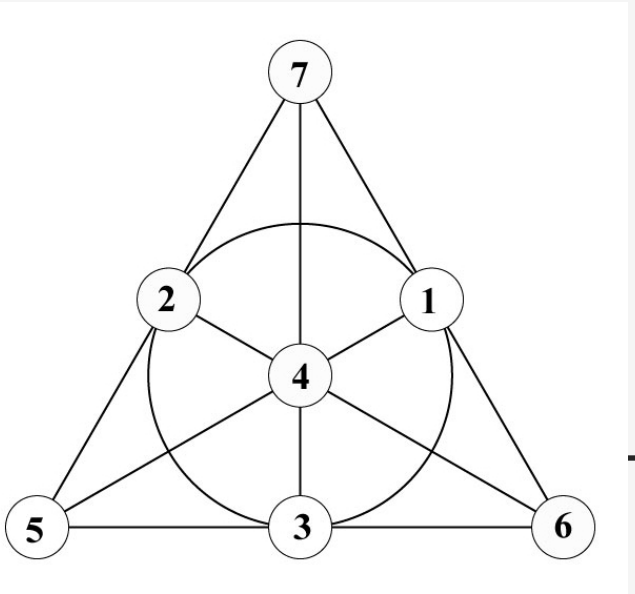
1) Each task size is the same
2) All of the computers have the same resources

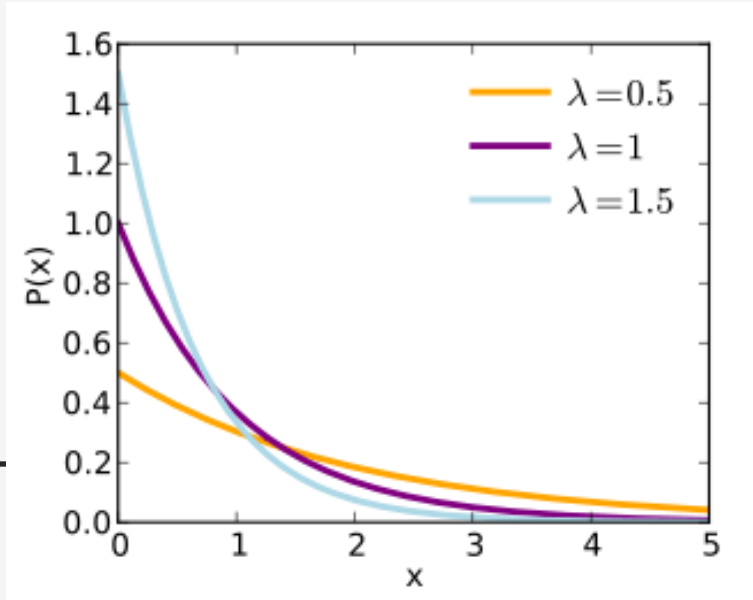
Goal of this Project

We want to find the optimal expected time for a job to be completed with random parameters. In our examples in the grocery store, we had people get into random queues. However, we are hoping to do better than random which is the major focus of our research.

What We Will Be Using in This Project

- Combinatorial designs
- Order statistics of random variables





Acknowledgements

Thank you NSF for supporting our work through grant CCF-1559855 and CCF-1717314