

TITLE: Communication-aware scheduling of precedence-constrained jobs

TIME: 5:15 PM, 4th August (Friday), 2023.

VENUE: KD102

ABSTRACT:

The problem of scheduling precedence-constrained jobs has a rich history, dating back to the classic list scheduling algorithm of Graham from the 1960s. The problem has applications in diverse scenarios and has been studied extensively in both the theory and systems communities. One such setting is where the underlying machines are heterogeneous and distributed, with communication among them incurring delays, which impact the completion times of the jobs. This setting arises in the scheduling of neural network training jobs over a network of GPUs, CPUs, and TPUs, where it has been referred to as the device placement problem.

A survey by Bansal in 2017 lists communication-aware scheduling of precedence-constrained jobs as one of the top ten open problems in scheduling theory. In this talk, we will present a framework for designing approximation algorithms for the problem, which addresses machine speeds, job sizes, and general models of machine and job delays. Our framework yields the first polylogarithmic-approximation algorithms for the problem.

The talk will be self-contained. We will motivate and present the problem, explore examples, review a simple combinatorial algorithm for a special case, present the linear programming framework with an overview of the results, and close with the many open problems remaining in the space.

(Based on a FOCS 2020 paper with Biswaroop Maiti, David Stalfa, Zoya Svitkina, and Aravindan Vijayaraghavan, and an ICALP 2023 paper with David Stalfa and Sheng Yang.)

BIO:

Rajmohan Rajaraman is a Professor and Associate Dean in the Khoury College of Computer Sciences at Northeastern University. He is a member of the Algorithms and Theory group at Northeastern. Rajmohan's contributions to the field have been well cited, specifically his research on distributed hash tables, which has been incorporated in peer-to-peer systems. He has obtained significant results for fundamental combinatorial optimization problems, including facility location, edge-disjoint paths, multiprocessor scheduling, and dominating sets.

Previously, Rajmohan held positions as a research scientist at Cambridge Mobile Telematics, Google Research, and Akamai Technologies, and as a postdoctoral fellow at DIMACS. He was the recipient of an NSF Career Award and was a co-author of award-winning papers at ACM SPAA, ACM PODC, IEEE ICDCS, and ACM WiSec. He has served as the program chair and track chair for leading conferences in parallel computing and distributed systems. He is an Associate Editor for the journals Theory of Computing Systems and International Journal on Foundations of Computer Science. He earned his PhD in Computer Science from the University of Texas at Austin and his B. Tech in Computer Science from IIT Kanpur in India. He is thrilled to return to his alma mater!