SCHEDULING WITH OUTLIERS TO MINIMIZE LOAD AND FLOW-TIME

SYAMANTAK DAS

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY DELHI
OCTOBER 2016

SCHEDULING WITH OUTLIERS TO MINIMIZE LOAD AND FLOW-TIME

by

SYAMANTAK DAS

Department of Computer Science and Engineering

Submitted

in Fulfillment of the Requirements of the Degree of Doctor of Philosophy

to the

Indian Institute of Technology Delhi

OCTOBER 2016

Certificate

This is to certify that the thesis titled **"Scheduling with Outliers to Minimize Load and Flow-time"** being submitted by **Syamantak Das** to the **Indian Institute of Technology Delhi**, for the award of the degree of **Doctor of Philosophy**, is a record of bonafide research carried out by him under our supervision. The results obtained in this thesis have not been submitted to any other university or institute for the award of any other degree or diploma.

Naveen Garg
Professor
Department of Computer Science and Engineering
Indian Institute of Technology Delhi
Delhi 110016

Amit Kumar
Professor
Department of Computer Science and Engineering
Indian Institute of Technology Delhi
Delhi 110016

Acknowledgments

During the last five and half years, I have often wondered about the day when I would be writing the Acknowledgment Section of my PhD thesis. Now that the day has finally arrived, I sincerely hope I do not miss out on the name of any person who has been a part, in some way or the other, of this wonderful and exciting journey.

I have been advised for my thesis work by two of the most beautiful minds I have ever come across - Prof. Naveen Garg and Prof. Amit Kumar. Prof. Garg is the reason I decided to pursue a PhD in the field of theoretical computer science in the first place. A two hours lecture delivered by him on Approximation Algorithms at Cadence Designs System, Noida, where I had been an employee at that time, was so inspirational that I left my job and decided to pursue an academic career. During the next five years, we have interacted umpteen number of times through his class lectures, group meetings and one-to-one discussions. Among several important things, I sincerely hope that I have imbibed, albeit in a small percentage, his striking ability to strip down some of the most complicated mathematical ideas to its bare-bone simplicity. His penchant for pursuing deep problems as opposed to producing large volume of results will remain an inspiration for a long time.

Prof. Amit Kumar has been the person with whom I have spent the longest hours working together on several problems. The space is too short to write about all the things I have

learnt from those sessions. But the one thing for which I shall be ever indebted to him is teaching me how to discipline my thoughts and convert vague ideas into organized mathematical statements. His oft-repeated simple advise "try to work out all calculations yourself, whether your own or from somebody else's paper" would be always be my mantra.

I extend my thanks to one of my oldest friends since school days and now a co-author Anamitra Roy Choudhury. He played a pivotal role in most of the work that goes into this thesis. His ability to keep his cool even in the most dire situations was something to learn from. Work was never dull with his witty interruptions providing welcome breaks. My thanks goes to Suman Kalyan Bera, another good friend and co-author for one of my papers.

I extend my gratitude to Tata Consultancy Services for supporting my entire PhD through the TCS Research Scholar Program Fellowship. A special thanks goes to Mr. Sachin Parkhi of TCS who has been instrumental in making sure that all logistic issues, be it monthly stipends or international travel funds, are managed seamlessly. I am grateful to IMPECS (Indo-German Max Planck Center for Computer Science) for supporting my visits to Max Planck Institut für Informatik, Saarbruecken, Germany (MPII) twice.

At MPII, I spent some extremely exciting days brainstorming with Antonios Antoniadis and Sebastian Ott. Tobias Moemke and Andreas Weise were two other great people there to spend time with, discussing theory and other stuff.

There are several people in the Theory community whom I have not met too many times, but whose work have taught me to search for elegance in algorithm design and proofs. I cannot thank enough Nikhil Bansal, Debmalya Panigrahi, Deeparnab Chakrabarti and Janardhan Kulkarni, among many others, for the wonderful papers they have written.

I am thankful to Prof. S.N. Maheshwari for his scintillating lectures on Network Flows and Submodular Functions as well as the invaluable advise on research he has been kind enough to share. My gratitude goes to Prof. Sandeep Sen and Dr. Ragesh Jaiswal for being part of my SRC and for their constant encouragement. A heartfelt thanks goes to Dr. Sambuddha Roy, Dr. Yogish Sabharwal and Dr. Amitabha Bagchi for being kind enough to be on my SRC. Dr. Shweta Agrawal, who recently joined the CSE Department as an Assistant Professor, has been more of a friend - thanks Shweta for the CCD sessions, the 'gyans' about life and work and for giving a much needed impetus to the Theory Reading Group. I convey my gratefulness to Prof. Preeti Ranjan Panda and Prof. Subodh Kumar who have been PhD coordinators during my tenure and helped me with several issues at various points of time.

Surviving through a PhD tenure requires immense support from strong and motivated peers. Folks of the theory group - Manoj, Jatin, Anup and Nikhil were always ready to brainstorm on problems and ideas. Theory won't have been this much fun without them. I met some of the most interesting and warm individuals in these five and half years. Thanks Shibashis, Chinmay, Nisha, Suvam, Brojeswar, Sayak, Shashank and Gayathri for the endless chats on research and life over coffee, lunch or dinner. It was only because of you guys that I could keep my head straight a midst all the work load. A special thanks to Yamuna for helping with the thesis submission procedure.

Outside IIT, there are two persons whose contribution to this thesis is more than I can express. The first one is Souvik, whose house in C.R. Park was a regular weekend haven. The second person is Niladrish, one of my oldest friends, who was always patient enough to listen to my problems and offer simple and comforting words of encouragement.

Perhaps the most difficult task is to find words of gratitude for the three most important persons in my life - my parents and my wife Rijurekha. They are the reason for everything I could achieve. My parent's have made tireless efforts and immense sacrifices to make sure that I always receive the best education. Without their constant endeavor and encouragement, my PhD would have never happened. My wife Rijurekha has been like a pillar in my life. Whenever I was about to slip, I knew she would be there to hold on to. This thesis is dedicated to Baba, Maa and Riju.

Syamantak Das

Abstract

In this thesis, we consider problems of scheduling jobs in a multiprocessor environment in order to minimize some standard metrics. Central to all the problems considered is a model where one is allowed to reject *or not process some jobs, but only upto a bounded fraction of the entire input. We consider such problems in both online and offline settings.*

In the online setting*, we work in both restricted assignments and unrelated machines models and consider the objectives of load balancing, minimizing maximum weighted flow-time and more generally, minimizing ℓ_p-norm of weighted flow-time. Online algorithms are usually analyzed using the notion of competitive ratio which compares the solution obtained by the algorithm to that obtained by an online adversary for the worst possible input sequence. Often this measure turns out to be too pessimistic, and one popular approach especially for scheduling problems has been that of "resource augmentation" which was first proposed by Kalyanasundaram and Pruhs [27]. Although resource augmentation has been very successful in dealing with a variety of objective functions, for the problems stated above, even a (arbitrary) constant speedup cannot lead to a constant competitive algorithm. In this work, we propose a new "rejection model" which requires no resource augmentation but which permits the online algorithm to not serve an epsilon-fraction of the*

requests, whereas the offline optimal has to serve all of them. Under this model, we give the following results:

1. Load Balancing

   For the load balancing problem where the objective is to minimize the maximum load on any machine, we give $O(\log^2 1/\varepsilon)$-competitive algorithm which rejects at most an ε-fraction of the jobs.

2. Maximum Weighted Flow-time

   - *For minimizing maximum weighted flow-time in the restricted assignments model, we give an an immediate dispatch non-migratory $O(1/\varepsilon^4)$-competitive algorithm which can reject at most an ε-fraction of the jobs by weight. We extend this result to a more general setting where the weights of a job for measuring its weighted flow-time and its contribution towards total allowed rejection weight are different. This is useful, for instance, when we consider the objective of minimizing the maximum stretch. We obtain an $O(1/\varepsilon^6)$-competitive algorithm in this case.*
   - *In the unrelated machines model, we give a $1/\varepsilon^{O(1)}$-competitive algorithm for minimizing maximum unweighted flow-time and load balancing problems under ε-rejection.*

   These results were obtained jointly with Anamitra Roy Choudhury, Naveen Garg and Amit Kumar [17].

3. ℓ_p-norm of Weighted Flow-time

   For minimizing ℓ_p-norm of weighted flow-time in the restricted assignments setting, our main result is an immediate dispatch non-migratory $1/\varepsilon^{O(1)}$-competitive algorithm when one is allowed to reject at most ε-fraction of the total weight of jobs arriving. This is in contrast with the speed augmentation model under which no online algorithm for this problem can achieve a competitive ratio independent of p.

   This result was obtained jointly with Anamitra Roy Choudhury and Amit Kumar [18].

In the offline setting*, we give the first logarithmic approximation for minimizing average flow-time of jobs in the restricted assignments setting under a single knapsack constraint. In a knapsack constraint setting, each job has a profit. A feasible schedule is one which processes a set of jobs with total profit of at least a quantity Π. This result is an extension of the work of Gupta, Krishnaswamy, Kumar and Segev [24] who considered the special case where the profit of each job is unity, in the identical machines model. Our algorithm is based on iteratively rounding a natural LP relaxation for this problem. This result was obtained through joint work with Suman K. Bera and Amit Kumar [13].*

Contents

List of Figures