Review of "Approximation and Online Algorithms" by RNDr. T. Tichý

Rob van Stee

June 26, 2008

This thesis deals with several tough problems in the area of online scheduling. In online problems, information about the problem is revealed bit by bit, and an online algorithm needs to make decisions without knowing the as yet unrevealed information. Its performance is then compared to the optimal offline solution, which potentially requires exponential time to be computed. In competitive analysis, we are interested in the highest possible ratio between the cost of an online algorithm and the optimal cost, the so-called competitive ratio.

The thesis deals with single-machine problems, and in particular with randomized algorithms for such problems and algorithms that use various forms of resource augmentation. A randomized algorithm is allowed to use random bits in its decision-making, and its performance is measured by comparing its expected cost or profit for a particular problem to the optimal cost or profit. While it is clear that randomization can help to get better algorithms, a property which is also known for approximation algorithms, it is often very hard to calculate the competitive ratio of a given randomized online algorithm, and it is also more difficult to provide good lower bounds. Resource augmentation is a way to give the online algorithm more power, for instance in scheduling problems it can be given a faster machine. It is then still compared to the optimal solution on an original machine. This sometimes allows us to identify algorithms which may work well in practice despite having a poor competitive ratio.

In this thesis, jobs typically have a size of 1, with varying release times, deadlines, and weights. The goal is always to maximize the total weight of the jobs that are fully processed (i.e., that complete before their deadline). The model where profit is also gained for jobs that are only partially completed is also considered. All results are for one machine only, though in one case, a lower bound is derived for the case where the online algorithm has m machines and is compared to the optimal solution on a single machine. In the following, I will highlight some of the results.

For the problem as formulated above, without additional constraints, the author presents a randomized algorithm which beats the best possible deterministic algorithm and has a competitive ratio of 1.58. The algorithm uses a clever probability distribution to select the next job to run from among the set of relatively heavy and urgent jobs.

The case where all jobs have a difference of 2 between their release time and their deadline is solved completely for deterministic algorithms: a tight bound for the competitive ratio is derived. The algorithm is quite complicated, but this appears to be unavoidable. The analysis is quite impressive. The author derives lower bounds for randomized algorithms, also for the more general case where the difference between release time and deadline is a constant. For the case where this difference is only upper bounded by a constant, upper bounds are derived. In particular, if the difference is 2, an algorithm with competitive ratio 1.25 is presented, which is optimal.

The attention then shifts to algorithms that use resource augmentation and restarts. In particular, for algorithms that are allowed to restart a running job (meaning that the work done on the job is lost, and the job will have to be run again from scratch), an upper bound of 3/2 is presented. The idea of this algorithm is quite simple: only restart if a tight job arrives that would otherwise be lost, and all currently pending jobs can still be completed even after the restart. The analysis of this algorithm uses a nontrivial charging scheme. It is also shown that this algorithm is optimal among deterministic algorithms.

Most of the results in this thesis have appeared in prestigious journals and conferences, which vouches for their newness and correctness. The only unpublished results are in Section 5.3; to the best of my knowledge, these results are also correct and new, though the result of Theorem 5.3.4 appears to subsume the result of Theorem 5.3.2. The results presented in this thesis have significantly increased our understanding of the problem of maximizing the weight of the completed jobs on a single machine in an online setting. The techniques used to prove the several upper and lower bounds may also be applicable in other online problems.

I feel obligated to point out that the presentation of this thesis could have been improved. For instance, the overview of paradigms on page 13 is a bit confused: item 3 should be a subitem of item 2, and item 4 is basically orthogonal to items 1 and 2. Also, it might have made more sense to put the preliminaries (Chapter 3) before the results overview (Chapter 2). The titles of Chapters 4 and 5 are quite general and vague. Finally, the thesis would have benefited from additional proofreading and spellchecking.

In summary, in my opinion, the results presented here clearly prove that the author can do creative scientific work, and I look forward to discussing this thesis with him at the Ph.D. defense.

Saarbrücken, June 26, 2008

Priv.-Doz. Dr. Rob van Stee