Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

J.K. Lenstra, A.H.G. Rinnooy Kan

# New directions in scheduling theory

Department of Operations Research and System Theory     Report OS-R8401 January

The Centre for Mathematics and Computer Science is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

NEW DIRECTIONS IN SCHEDULING THEORY

J.K. LENSTRA

*Centre for Mathematics and Computer Science, Amsterdam*

A.H.G. RINNOOY KAN

*Erasmus University, Rotterdam*

This is an assessment of new developments in the theory of sequencing and scheduling. After a review of recent results and open questions within the traditional class of scheduling problems, we focus on the probabilistic analysis of scheduling algorithms and next discuss some extensions of the traditional problem class that seem to be of particular interest.

69F21, 69F13

1980 MATHEMATICS SUBJECT CLASSIFICATION: 90B35, 68C15, 68C25.

KEY WORDS & PHRASES: scheduling, jobs, machines, algorithm, computational complexity, probabilistic analysis.

NOTE: This paper will appear in *Operations Research Letters*.

Report OS-R8401

Centre for Mathematics and Computer Science

P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

## 1. INTRODUCTION

This paper is the second in a series to be published in *Operations Research Letters*, the purpose of which is to assess new developments in areas that fall under the general heading of Operations Research. Rather than providing an exhaustive survey of such areas, each paper in the series will try to identify new ideas and techniques and important recent results. The series may be compared to D.S. Johnson's quarterly column in the *Journal of Algorithms*, which keeps track of new developments in the theory of NP-completeness.

To the extent that the NP-completeness column is a continuing update of Garey & Johnson's standard text on complexity theory [15], our report on new directions in scheduling theory follows in the wake of several books and surveys, in which more complete and detailed information on many of the topics to be dealt with (and on the many topics *not* dealt with) can be found. In particular, we mention the books by Baker [1], Coffman [5] and French [11], our survey with Graham and Lawler [18], its update [27], Lawler's tutorial [26], and a recent annotated bibliography [28].

Although the scope of each of these predecessors is somewhat different, they all concern themselves with what is commonly accepted as the principal domain of scheduling theory: the optimal allocation over time of scarce resources in the form of *machines* to activities known as *jobs*, subject to the basic constraints that, at any point in time, no machine processes more than one job and no job is processed by more than one machine.

We start, in Section 2, by reviewing the status of a few problems in this class that have been the object of recent research efforts. Next, in Section 3, we focus on the probabilistic analysis of the behavior and performance of scheduling algorithms, an approach that is receiving an increasing amount of attention. Finally, in Section 4, we discuss some extensions of the traditional problem class that seem to represent interesting new directions in scheduling research.

## 2. THE TRADITIONAL CLASS OF SCHEDULING PROBLEMS

The traditional class of scheduling problems was first systematically investigated by Conway, Maxwell and Miller [7] and is best summarized by referring to the problem classification introduced in [18]. Here, scheduling problems are defined by specifying three components, namely *machine environment* (single machine, parallel machines of various types, open shop, flow shop, job shop), *job characteristics* (preemption allowed or not, precedence constraints of various types, arbitrary or equal release dates, arbitrary or unit processing times, etc.), and *optimality criterion* as a function of the completion times $C_j$ of the jobs j (maximum completion time $C_{max} = max_j\{C_j\}$, total completion time $C_{sum} = \sum_j C_j$, etc.).

The class so constructed contains thousands of problem types. Indeed, a computer program is being used to record the known results on their computational complexity and to deduce the consequences of new complexity results [21,22]. NP-completeness theory has been strikingly successful in identifying the borderline between *well-solvable* scheduling problems (that can be solved in time bounded by a polynomial function of problem size) and *NP-hard* ones. Of the 4,536 problems in the class administered by our program, 416 problems (9%) are well solvable, 3730 (82%) are NP-hard, and the remaining 390 (9%) are still open. In reviewing the most important recent results and the most prominent open questions, it is appropriate to consider the current status of the four open problems mentioned in [30].

The first one is the *single machine total tardiness* problem, in which n jobs j with processing times $p_j$ and due dates $d_j$ (j = 1,...,n) have to be scheduled so as to minimize $\sum_j max\{C_j-d_j,0\}$. Lawler [24] has developed a pseudopolynomial dynamic programming recursion to solve the problem in $O(n^4\sum p_j)$ time, but this has neither been complemented by an NP-hardness proof nor been improved by a strictly polynomial algorithm. In other words, the problem is well solvable if a *unary* encoding of the problem data is allowed and open with respect to a *binary* encoding. It is one of the few open single machine problems, and the fact that it is so simple to state makes this all the more annoying.

The second problem has been high on several lists of open problems for the past years. It is the *precedence constrained three-processor scheduling* problem, in which unit-time jobs have to be scheduled on three identical parallel machines subject to precedence constraints so as to minimize $C_{max}$. Two immediate simplifications can be solved in linear time: the case of tree-like constraints (even for an arbitrary number of machines) [19] and the case of two machines (and arbitrary precendence constraints) [14]; the generalization to an arbitrary number of machines is NP-hard [15]. There is a wealth of complexity results for other variants (see [20,29] for brief summaries), but the problem itself has stayed out of reach.

The third and fourth problem are of the *no wait flow shop* type: each job has to spend given amounts of time on machines 1,...,m in that order without waiting time between its completion time on one machine and its starting time on the next. While the case of two machines and the $C_{max}$ criterion is solvable in O(n log n) time, the cases (m = 3, $C_{max}$) and (m = 2, $C_{sum}$) have remained open for a long time. Röck [36,37] has settled these questions by providing two ingenious NP-hardness proofs.

The most impressive recent algorithmic progress for scheduling problems has occurred in the area of parallel machine problems. Polynomial algorithms have been obtained for the problems of finding nonpreemptive schedules for unit-time jobs on identical machines subject to (possibly nonintegral) release dates and deadlines [16,38], preemptive schedules subject to precedence constraints [25], and preemptive schedules on uniform machines (that may differ in speed) subject to release dates and deadlines [32].

We cannot resist mentioning one more open problem - not a general problem *type*, but a specific problem *instance* that no one has been able to solve since it was generated in 1963. It is a *job shop* problem with only ten jobs and ten machines, and all we know is that the optimal $C_{max}$ value is at least 874 and at most 935 [10]. Simple combinatorial arguments and surrogate duality relaxations appear to be of no help here, and this may be the point where polyhedral combinatorics may successfully enter scheduling theory. Initial work in this direction is being carried out [2].

In concluding this section, it should be said that we have presented our personal choice of recent results and open questions. A quite different selection is given in [20]. Many more results are reviewed in [28], e.g. on the computational complexity of *open shop* problems and on the worst case performance of *flow shop* heuristics. In the next section, we turn to other interesting questions that can be asked about familiar problems and algorithms.

## 3. PROBABILISTIC ANALYSIS OF SCHEDULING ALGORITHMS

As in other parts of combinatorial optimization, a probabilistic approach has recently become fashionable in the analysis of the behavior and performance of scheduling algorithms. Here, the *behavior* (or efficiency) of an algorithm refers to its running time, and the *performance* (or effectiveness) of an algorithm provides information about the value produced by it. Probabilistic analyses require the specification of a *probability distribution* over the class of problem instances. In scheduling theory, it is typically assumed that the job processing times $p_j$ are independent random variables following identical distributions or distributions from the same class that differ only in some parameters.

One of the great challenges in the analysis of algorithmic *behavior* is the development of powerful tools to determine whether an enumerative method, with superpolynomial behavior in the worst case, requires at most polynomial running time on the average. Such an analysis is usually very difficult due to the dependencies that arise when the enumeration proceeds; a typical example is the stochastic conditioning effect of the branching rule in a branch-and-bound method on the subproblems created. Yet, the amazing efficiency of certain enumerative methods in, for instance, minimizing maximum lateness $\max_j\{C_j-d_j\}$ on a single machine subject to release dates [4,23] asks for a theoretical explanation that has not been put forward so far.

The probabilistic *performance* analysis of scheduling algorithms has been much more successful. Here, one may be interested in an asymptotic expression for the optimal solution value or in an estimate of the error produced by a heuristic. The former result frequently comes as a byproduct of the latter. We distinguish between the *absolute* error of a heuristic, which is the difference between the approximate and optimal solution values, and the *relative* error, which is the ratio of the two.

Results of such analyses confirm that, for certain problem types, the combinatorial difficulties tend to average out when the problem size becomes sufficiently large. A typical example is provided by the minimization of maximum completion time for n independent jobs on m identical parallel machines. Suppose that the $p_j$ are independent and identically distributed with a finite expectation and a positive derivative of the density function at 0.

Then the optimal value for the nonpreemptive problem is asymptotic to the optimal value for the preemptive problem, $\sum p_j/m$; i.e., their ratio converges to 1 almost surely. Moreover, the absolute error of the *longest processing time rule*, which assigns jobs to the first available machine in order of decreasing $p_j$, converges to 0 almost surely [12], so that this heuristic is asymptotically optimal in a very strong sense. Results of this type are quite rare. Asymptotic optimality in the weaker, relative, sense is often easier to establish. Indeed, for the model under consideration it is not hard to see that the relative error of the *list scheduling rule*, which assigns jobs to the first available machine in any order, converges to 0 almost surely.

Usually, probabilistic performance results are of an asymptotic nature and give no indication for the speed of convergence; for an exception, see [12]. While nonasymptotic results would be preferable, asymptotic ones often represent all that is known. They should serve not so much as a guarantee, as worst case performance results do, but as an explanation of why certain NP-hard problems appear to be so amenable to solution by simple rules, whose worst case performance in no way reflects the average case.

Probabilistic analyses have also been performed for the parallel machine model with the total completion time criterion [13] and for a single machine model with release dates [17]. The results have found application in the design and analysis of multistage heuristics for hierarchical scheduling problems [8,13,31]. Altogether, this is still very much virgin territory, where a lot of work remains to be done.

## 4. EXTENSIONS OF THE TRADITIONAL PROBLEM CLASS

Some of the most interesting new directions in scheduling research may be found in the extensions of the traditional problem class that have been proposed over the past years.

Those extensions sometimes take the form of a more careful analysis of the original class itself, as in the case of *multicriteria problems* or, more importantly, of *robustness and sensitivity analysis*, virtually neglected topics of obvious practical relevance. They may also take the form of further refinements, as in the case of more general *resource constraints*. This generalization produces a host of NP-hard problems, but a more detailed problem classification could help in deciding which type of heuristic would work well on which type of problem. We refer to [3] for some typical results in this problem class and to [6] for a survey of the subclass of *bin packing* problems; see also [20,28]. The most fruitful extensions, however, seem to be those that link scheduling theory as a whole to other areas in which similar problems are studied.

This is exemplified by the increasing use of techniques from *queueing theory* in *stochastic scheduling*. An example of a stochastic scheduling problem is the minimization of expected maximum completion time for independent jobs, with independent and identically distributed processing times, on identical parallel machines. (Note that, in contrast to the model dealt with in the previous section, realizations of the processing times are not given prior to the actual scheduling and that the optimality criterion is deterministic.) Optimality of the *longest expected processing time rule* for the above problem has been established, but results of this type demand a great deal of technical expertise and the surface has only been scratched so far. We refer to [9], a collection of papers emphasizing these results, in particular [39] on parallel machine models and [35] on shop scheduling, and to [33,34] which present an impressive survey of the area.

Equally importantly, there appear to be good opportunities for research on the interface between scheduling and *inventory theory*. Both are investigating aspects of the general *production planning* problem but have been developed in complete mutual isolation. So many theoretical tools are now available that interaction seems to be feasible and practical.

An integration of scheduling, queueing and inventory theory should give rise to more realistic models that will provide challenges for fundamental research in all three areas. Embedded in this way, the problems that scheduling theory was intended to cope with will be as important as ever.

ACKNOWLEDGEMENT

This paper is partly based on discussions held at a workshop devoted to research directions in scheduling and routing, held at Cornell University on June 23 and 24, 1983.

REFERENCES

1. K.R. BAKER (1974). *Introduction to Sequencing and Scheduling*, Wiley, New York.

2. E. BALAS (1983). On the facial structure of scheduling polyhedra. Management Science Research Report MSRR-496, Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburgh, PA.

3. J. BLAZEWICZ, J.K. LENSTRA, A.H.G. RINNOOY KAN (1983). Scheduling subject to resource constraints: classification and complexity. *Discrete Appl. Math. 5*, 11-24.

4. J. CARLIER (1982). The one-machine sequencing problem. *European J. Oper. Res. 11*, 42-47.

5. E.G. COFFMAN, JR. (ed.) (1976). *Computer & Job/Shop Scheduling Theory*, Wiley, New York.

6. E.G. COFFMAN, JR., M.R. GAREY, D.S. JOHNSON (1983). Approximation algorithms for bin-packing - an updated survey. Unpublished manuscript, Bell Laboratories, Murray Hill, NJ.

7. R.W. CONWAY, W.L. MAXWELL, L.W. MILLER (1967). *Theory of Scheduling*, Addison-Wesley, Reading, MA.

8. M.A.H. DEMPSTER, M.L. FISHER, L. JANSEN, B.J. LAGEWEG, J.K. LENSTRA, A.H.G. RINNOOY KAN (1983). Analysis of heuristics for stochastic programming: results for hierarchical scheduling problems. *Math. Oper. Res. 8*, 525-537.

9. M.A.H. DEMPSTER, J.K. LENSTRA, A.H.G. RINNOOY KAN (eds.) (1982). *Deterministic and Stochastic Scheduling*, Reidel, Dordrecht.

10. M.L. FISHER, B.J. LAGEWEG, J.K. LENSTRA, A.H.G. RINNOOY KAN (1983). Surrogate duality relaxation for job shop scheduling. *Discrete Appl. Math. 5*, 65-75.

11. S. FRENCH (1982). *Sequencing and Scheduling: an Introduction to the Mathematics of the Job-Shop*, Horwood, Chichester.

12. J.B.G. FRENK, A.H.G. RINNOOY KAN (1984). The asymptotic optimality of the LPT scheduling heuristic. Report, Erasmus University, Rotterdam, to appear.

13. J.B.G. FRENK, A.H.G. RINNOOY KAN, L. STOUGIE (1984). A hierarchical scheduling problem with a well-solvable second stage. *Ann. Oper. Res. 1*, to appear.

14. H.N. GABOW, R.E. TARJAN (1983). A linear-time algorithm for a special case of disjoint set union. *Proc. 15th Annual ACM Symp. Theory of Computing*, 246-251.

15. M.R. GAREY, D.S. JOHNSON (1979). *Computers and Intractability: a Guide to the Theory of NP-Completeness*, Freeman, San Francisco.

16. M.R. GAREY, D.S. JOHNSON, B.B. SIMONS, R.E. TARJAN (1981). Scheduling unit-time tasks with arbitrary release times and deadlines. *SIAM J. Comput. 10*, 256-269.

17. P.G. GAZMURI (1981). Probabilistic analysis of a machine scheduling problem. Unpublished manuscript, Computer Science Division, University of California, Berkeley.

18. R.L. GRAHAM, E.L. LAWLER, J.K. LENSTRA, A.H.G. RINNOOY KAN (1979). Optimization and approximation in deterministic sequencing and scheduling: a survey. *Ann. Discrete Math. 5*, 287-326.

19. T.C. HU (1961). Parallel sequencing and assembly line problems. *Oper. Res. 9*, 841-848.

20. D.S. JOHNSON (1983). The NP-completeness column: an ongoing guide. *J. Algorithms 4*, 189-203.

21. B.J. LAGEWEG, E.L. LAWLER, J.K. LENSTRA, A.H.G. RINNOOY KAN (1981). Computer-aided complexity classification of deterministic scheduling problems. Report BW 138, Centre for Mathematics and Computer Science, Amsterdam.

22. B.J. LAGEWEG, J.K. LENSTRA, E.L. LAWLER, A.H.G. RINNOOY KAN (1982). Computer-aided complexity classification of combinatorial problems. *Comm. ACM 25*, 817-822.

23. B.J. LAGEWEG, J.K. LENSTRA, A.H.G. RINNOOY KAN (1976). Minimizing maximum lateness on one machine: computational experience and some applications. *Statist. Neerlandica 30*, 25-41.

24. E.L. LAWLER (1977). A "pseudopolynomial" algorithm for sequencing jobs to minimize total tardiness. *Ann. Discrete Math. 1*, 331-342.

25. E.L. LAWLER (1982). Preemptive scheduling of precedence-constrained jobs on parallel machines. M.A.H. DEMPSTER, J.K. LENSTRA, A.H.G. RINNOOY KAN (eds.). *Deterministic and Stochastic Scheduling*, Reidel, Dordrecht, 101-123.

26. E.L. LAWLER (1983). Recent results in the theory of machine scheduling.

A. BACHEM, M. GRÖTSCHEL, B. KORTE (eds.). *Mathematical Programming: the State of the Art - Bonn 1982*, Springer, Berlin, 202-234.

27. E.L. LAWLER, J.K. LENSTRA, A.H.G. RINNOOY KAN (1982). Recent developments in deterministic sequencing and scheduling: a survey. M.A.H. DEMPSTER, J.K. LENSTRA, A.H.G. RINNOOY KAN (eds.). *Deterministic and Stochastic Scheduling*, Reidel, Dordrecht, 35-73.
28. J.K. LENSTRA, A.H.G. RINNOOY KAN (1983). Scheduling theory since 1981: an annotated bibliography. Report BW 188, Centre for Mathematics and Computer Science, Amsterdam.
29. J.K. LENSTRA, A.H.G. RINNOOY KAN (1984). Two open problems in precedence constrained scheduling. *Ann. Discrete Math.*, to appear.
30. J.K. LENSTRA, A.H.G. RINNOOY KAN, P. BRUCKER (1977). Complexity of machine scheduling problems. *Ann. Discrete Math. 1*, 343-362.
31. J.K. LENSTRA, A.H.G. RINNOOY KAN, L. STOUGIE (1984). A framework for the probabilistic analysis of hierarchical planning systems. *Ann. Oper. Res. 1*, to appear.
32. C. MARTEL (1982). Preemptive scheduling with release times, deadlines and due times. *J. Assoc. Comput. Mach. 29*, 812-829.
33. R.H. MÖHRING, F.J. RADERMACHER, G. WEISS (1983A). Stochastic scheduling problems I: general strategies. Unpublished manuscript, RWTH Aachen.
34. R.H. MÖHRING, F.J. RADERMACHER, G. WEISS (1983B). Stochastic scheduling problems II: set strategies. Unpublished manuscript, RWTH Aachen.
35. M. PINEDO, L. SCHRAGE (1982). Stochastic shop scheduling: a survey. M.A.H. DEMPSTER, J.K. LENSTRA, A.H.G. RINNOOY KAN (eds.). *Deterministic and Stochastic Scheduling*, Reidel, Dordrecht, 181-196.
36. H. RÖCK (1984A). The three-machine no-wait flow shop problem is NP-complete. *J. Assoc. Comput. Mach.*, to appear.
37. H. RÖCK (1984B). Some new results in no-wait flow shop scheduling. *Z. Oper. Res.*, to appear.
38. B. SIMONS (1983). Multiprocessor scheduling of unit-time jobs with arbitrary release times and deadlines. *SIAM J. Comput. 12*, 294-299.
39. G. WEISS (1982). Multiserver stochastic scheduling. M.A.H. DEMPSTER, J.K. LENSTRA, A.H.G. RINNOOY KAN (eds.). *Deterministic and Stochastic Scheduling*, Reidel, Dordrecht, 157-179.