Applied Operations Scheduling and Routing

S. E. Butt\textsuperscript{1} and A. Houshyar\textsuperscript{*}

With the changes in the workplace and the added element of competition, it is apparent that the efficient scheduling of resources such as machines, labor, and material is extremely important. Scheduling covers a broad array of topics, such as machine scheduling and sequencing, vehicle and fleet scheduling, crew scheduling, employee scheduling, project scheduling, lot-size scheduling, interval (or reservation) scheduling and timetabling. This article reviews the literature to identify the state-of-the-art research on scheduling and routing problems.

INTRODUCTION

Changes in the nature of work have significantly contributed to the modification of work time patterns. In the past few decades, the makeup of the employee workforce, modes of transportation and communication, increase of employment in the service industries and consumer expectations for products and services have significantly altered the standard work week. Technological advances, employees and consumers desire for greater freedom from time constraints and the need for flexibility in modes of operation all interact simultaneously to create today's workplace.

With changes in the workplace and the added element of competition, it is apparent that the efficient scheduling of resources such as machines, labor, and material is extremely important. Scheduling covers a broad array of topics, such as machine scheduling and sequencing, vehicle and fleet scheduling, crew scheduling, employee scheduling, project scheduling, lot-size scheduling, interval (or reservation) scheduling and timetabling. In general, many of the problems encountered in scheduling are NP-hard; however, with the increased speed of computers and the efficient use of memory by software, many of these problems can now be solved. Furthermore, the latest developments in heuristic search methods have also opened up successful solution strategies to problems that were once intractable.

Many scheduling problems can be formulated as instances of special zero-one integer programs known as set partitioning or set covering problems. Basically, a set covering problem involves a given 0–1 matrix with costs attached to all columns. The objective is to choose a minimum-cost solution set of columns such that the number of 1’s appearing in each row of the selected columns is at least one. If this number is required to be exactly one, the set-partitioning problem results. In this case, the rows are partitioned into a number of subsets covered by exactly one selected column. Set covering and set partitioning problems have been studied extensively over the last three decades. In the past, their use was limited due to the large size of the resulting integer program. Fortunately, techniques such as constraint branching and column generation have been developed and refined, so that realistic problems can now be solved efficiently and in many cases optimally [1-3]. A recent application of using the combination of these techniques to solve a constrained vehicle routing problem can be found in [4].

Local search techniques have become very popular and useful tools in scheduling. The most prominent are tabu search, simulated annealing and genetic algorithms. (Each of these techniques will be discussed in the next section.) The application of these general solution strategies to scheduling can be found in [5-9].

NOVEL APPROACHES TO SCHEDULING

Among the more recent approaches to scheduling problems are: 1) Simulated annealing, 2) Tabu search, 3) Genetic algorithms and 4) branch-and-price techniques. In what follows, each approach is briefly described and sources for additional readings are identified.

---

\textsuperscript{1} Department of Industrial & Manufacturing Engineering, Western Michigan University, Kalamazoo, Michigan, 49008, USA.

\textsuperscript{*} Corresponding Author Department of Industrial & Manufacturing Engineering, Western Michigan University, Kalamazoo, Michigan, 49008, USA.
Simulated Annealing

Simulated annealing is analogous to the annealing of materials to produce sound low energy states. Simulated annealing as an approach to optimization was developed by Kirkpatrick and colleagues in the early 1980's [16]. Simulated annealing addresses the problem of escaping from local optima by adopting a process controlled by a parameter known as the temperature. A heuristic designed for the problem in hand defines the neighborhood of an initial solution and a possible move within that neighborhood is generated at random. The move is accepted if it produces a better solution; a worsening move is accepted with a probability governed by the cost of the move and the value of the temperature parameter. At a high temperature a given worsening move is more likely to be accepted, leading possibly to an escape beyond the contours of a local optimum. As the process continues, the temperature is reduced until at a very low temperature only non-worsening moves are accepted. A simulated annealing bibliography was published in 1988 by Collins, Eglese and Golden [11]. Recent scheduling research in this area can be found in [12–14].

Tabu Search

Tabu search, first developed by Glover [15,16], is in many ways similar to simulated annealing, i.e., they both move from one schedule to another with the next schedule being possibly worse than the one before. The basic difference between tabu search and simulated annealing lies in the mechanism used for approving a candidate schedule. In tabu search, the mechanism is not probabilistic, but rather deterministic in nature. At any stage of the process, a tabu-list of mutations, which the procedure is not allowed to perform, is kept. Mutations on the tabu-list may be, for example pairs of jobs that may be interchanged. Every time a move is made by a mutation in the current schedule, the reverse mutation is entered at the top of the tabu-list, all other entries are pushed down one position and the bottom entry is deleted. The reverse mutation is put on the list to avoid returning to a local minimum that has been visited before. Note that if the number of entries in the tabu-list is too small, cycling may occur and if the number is too large, the search may be unduly constrained. Tabu search has been applied to many scheduling problems. Recent applications include [17–20].

Genetic Algorithms

Genetic algorithms were introduced in the United States in the 1970s. To use a genetic algorithm, a solution must be represented to a given problem as a genome (or chromosome). The genetic algorithm then creates a population of solutions and applies genetic operators, such as mutation and crossover, to evolve the solutions in an attempt to find the best alternatives. The three most important aspects of using genetic algorithms are (1) Definition of the objective function, (2) Definition and implementation of the genetic representation and (3) Definition and implementation of the genetic operators. Once these three have been defined, the generic genetic algorithm typically works fairly well. Recent scheduling research involving genetic algorithms includes [21–24].

Branch-and-Price Techniques

The branch-and-price solution strategy has been shown to be the state-of-the-art for several hard combinatorial optimization problems, such as the traveling salesman problem, the linear ordering problem and the maximum cut problem. Barnhart et al. [3] present a comprehensive review of recent developments and future directions of this approach with respect to scheduling problems, including advances in column generation and constraint branching. Additional information on the efficient use of column generation and constraint branching in a branch-and-bound framework can be found in [1,25].

CONTEMPORARY APPROACHES TO SCHEDULING

To classify different scheduling topics into a few major groups, the authors decided to characterize them into: 1) Machine scheduling, 2) Employee scheduling, 3) Vehicle scheduling and routing and 4) Reservation scheduling and timetabling. In what follows, the literature on each of these classes of scheduling problems is surveyed and additional reading references are offered.

Machine Scheduling

The problem of scheduling a set of jobs on one or several machines with the objective to optimize a certain performance measure has been the subject of extensive research for the past four decades. Machine scheduling and sequencing are traditional topics found in most graduate industrial engineering programs. Topics in this area are often categorized as single machine, parallel machines, flow-shop, job-shop and open shop models. In scheduling terminology, a distinction is made between a sequence and a schedule. A sequence usually corresponds to a permutation of the job set or the order in which jobs are to be processed on a given machine. A schedule refers to an allocation of jobs within a more complicated setting of machines, which could allow for preemption of jobs by other jobs that
are released at later points in time. Most textbooks on "scheduling" cover machine scheduling and sequencing topics in depth. The most recent are [26–31].

For the most part, research in machine scheduling has focused on problems involving objective functions that are non-decreasing in job completion times. Examples of these objectives can be found in recent textbooks [26–31]. Conway, Maxwell and Miller [32] refer to such objective functions as regular performance measures. For example, mean completion time, maximum completion time, mean tardiness and number of tardy jobs are all usual regular performance measures that have been extensively studied in the past [33,34]. The mean tardiness criterion, in particular, has been the standard way to measure conformance to due dates, although it ignores the consequences of jobs completing early.

Single machine problems arise when there is only one machine available and arriving jobs require service from this machine. Jobs are processed by the machine one at a time. The most common objective is to sequence jobs on the machine so as to minimize the penalty of finishing jobs after their due date. Single machine problems can often be solved using dispatching rules that identify the sequence with which the jobs will be processed on a machine. Examples include the Shortest Processing Time rule (SPT), the Earliest Due Date rule (EDD) and the Cost Over Time rule (COVERT).

Parallel machine problems involve a number of identical or non-identical machines on which jobs can be processed on any one of them. The typical objective is to minimize makespan. Flow shop problems stem from jobs which must be processed on multiple machines in an identical sequence. However, the processing time of each job on each machine may be different. A job shop includes a number of different machines and a job may require some or all of these machines in a specific sequence, the only restriction is that a job cannot use the same machine more than once. An open shop is similar to a job shop except that a job may be processed on the machines in any sequence the job needs. Two other machine scheduling problems that are often encountered include batch processing with or without sequence dependent set-up times (see [35,36]) and minimizing tool changeover times on NC machines (see [37,38]). Finally, before leaving the discussion on the traditional machine scheduling models, it should be noted that a general solution approach to machine scheduling was recently proposed by Lee, Bhaskaran and Pinedo [39], in which a three-step algorithm combines dispatching rules with simulated annealing or tabu search.

The emphasis on regular performance measures in the field of scheduling has changed with the current interest in Just-In-Time (JIT) or Zero-Inventory (ZI) production philosophy, which espouses the notion that earliness, as well as tardiness, should be discouraged (see [40–43]). In a JIT production environment, jobs that are completed earlier than their due dates tie up capital, increase inventory levels, take up scarce floor space, cause losses owing to deterioration, while jobs that are completed after their due dates may cause a customer to shut down operations. Kanet [44] mentioned the example of a job shop that produces components for subsequent assembly into finished products. Efficient scheduling, in general, implies that the needed parts on the production line arrive just in time. Merten and Muller [45] pointed out the appropriateness of minimizing the variance of retrieval times of records in a filing system, especially in online systems, since it is important to provide uniform response times to users. The production of perishable goods is yet another example where earliness as well as tardiness penalties are realistic. Inman and Bullfin [46] reformulated a JIT assembly system problem as a single machine earliness-tardiness problem. In the JIT assembly system, there are demands for $n$ products that are evenly spaced over a given planning horizon and the objective is to match production as closely as possible to demand, that is, to minimize the deviation between cumulative production and cumulative demand over the given planning horizon. Therefore, an ideal schedule in a JIT production system is one in which all jobs finish exactly at their assigned due dates. Alternatively, if jobs do not finish exactly at their due dates, then both earliness and tardiness have to be penalized. This can be translated to a scheduling objective in several ways. The most obvious is to minimize the deviation of job completion times from their due dates.

Employee Scheduling

Employee scheduling is concerned with the issues, problems and methods of assigning work hours to the personnel of an organization. It is an important component of human resource planning and management activity, particularly in organizations where demands for services vary with time. It is this challenge of meeting varying levels of demand while maintaining good service levels and a stable work force that most service operations managers now face. Quality of life and the structure of daily and weekly activity of people is directly affected by their work schedules. A stable, knowledgeable and motivated staff is very important, since service personnel interact directly with the customer, determining the quality of the service and the reputation of the provider. The only textbook devoted entirely to the topic of employee scheduling was written by Nanda and Browne [47]. The textbooks by Parker [27], Sule [29] and Pinedo and
Chao [31] do, however, each contain a chapter in this area.

The main objective of employee scheduling is to ensure that employees are working at the times when they are needed. To meet customer demand for goods or services, it is important to assess the anticipated demands and convert them into projected staffing requirements. Once the requirements are known, employee scheduling is typically broken down into 2 categories of problems, shift scheduling and tour scheduling. Shift work implies that a set of employees replaces another set of employees in the same work place at fixed times over a repeating time period. Tour scheduling involves the determination of a set of starting and ending times for employees, over a given time horizon, to meet staffing requirements over the same time period. Tour scheduling has many applications in the transportation industry (e.g., developing airline crew rosters and locomotive engineer schedules) and the service industry (e.g., scheduling hotel staff and fast-food restaurant staff). Examples can be found in [48–50]. Tour scheduling also has important implications in the development of the alternative work schedules found in the health services community, as well as, the current trends used to schedule part-time employees. For an overview of recent developments see [25]. In this paper, the authors describe several heuristic and optimization techniques in staff scheduling.

Vehicle Scheduling and Routing

The sequencing of vehicle activities in both space and time is at the heart of vehicle scheduling and routing problems. The type of problem that evolves is a function of the constraints imposed upon the formation of the schedules, the tasks being serviced and the locations where these tasks must be carried out. The cost associated with operating vehicles (and crews) for delivery purposes forms an important component of the total distribution costs. Consequently, small percentage savings in these expenses could result in substantial total savings over a number of years. The basic output of all vehicle routing and scheduling systems is essentially the same. For each vehicle or driver, a route and a schedule are provided. Generally the route specifies the sequence of locations to be visited and the schedule identifies the times at which the activities at these locations are to be carried out. If the entities to be serviced have no temporal restrictions and there are no precedence relations among these entities, then a routing problem is encountered. If each entity has a definitive service time, then a scheduling problem results. Otherwise, one is dealing with a combined routing and scheduling problem. Generally routing and scheduling problems involve both precedence relations and time windows. Bodin et al. [51] and Desroches et al. [52] present detailed classification schemes for vehicle routing and scheduling problems. An overview of exact and approximate algorithms is given in [4,53–58].

Reservation Scheduling and Timetabling

Reservation systems can be identified in both the manufacturing and service industries. Reservation scheduling typically refers to a parallel machine environment with n jobs available for processing. The processing time of each has to fit within a given time window and there may or may not be any slack. A release date and a due date specify the time window of a job. It may be the case that it is not possible to process all n jobs in the time allocated and thus the scheduler must decide which jobs to process based on an objective, such as maximizing the total number of jobs processed. A typical manufacturing example is when "customers" try to reserve one or more machines in a factory for a specific time period and the factory has to decide which orders to accept and which ones to reject. While the problems are computationally similar, instances of reservation scheduling are more visible in the service industry. Take for example hotel room reservations and car rental reservations, where the scheduler has the similar problem of deciding which reservations to accept and which ones to reject within a given time interval. For an introduction to work in this area see [59–61].

Timetabling is defined as the allocation of a set of resources to a set of objects, placed in space and time, such that a set of desirable objectives is satisfied as nearly as possible. Examples found in academia are course and examination timetabling. The basic university timetabling problem is to assign times to meetings, consisting of classes, teachers and possibly rooms, so that no class, teacher or room is scheduled more than once in any time period. Some forms of personnel allocation, such as the staffing of toll booths subject to a given number of personnel and the scheduling of sporting events are also considered timetabling issues (e.g. [62,63]). Recent developments in timetabling are discussed in [64,65]. Textbook chapters on timetabling can be found in [27,31]. In addition to heuristic search algorithms and set-covering and set-partitioning models, graph coloring is a very popular solution approach to timetabling problems (e.g. [66–69]).

REFERENCES


68. Weng, M.X. and Ventura, J.A. “Scheduling about a large common due date with tolerance to minimize mean absolute deviation of completion times”, *Naval Research Logistics*, 14, pp 843-851 (1994a).