Chapter 2
Literature Review on Personnel Scheduling

In this chapter we review the relevant literature for our work. In Sect. 2.1 we start with reviewing the general personnel scheduling literature and some classification schemes. In Sect. 2.2 we summarize the relevant work on physician scheduling. Afterwards we review in Sect. 2.3 work on implicit modeling techniques which is required for the new implicit shift modeling approach to schedule physicians in a hospital given in Chap. 3. Finally, we conclude in Sect. 2.4 with a focus on research that considers column generation approaches and B&P methodologies for personnel scheduling problems in the service industry. This is the relevant literature for Chap. 4 where we present a new column generation and B&P approach for the flexible shift scheduling of physicians in a hospital.

2.1 General Personnel Scheduling

There is a vast amount of literature on personnel scheduling so our review will be limited to the most relevant work. An introductory tutorial to staff scheduling is given in Blöchliger [34] that presents the basic concepts of the scheduling problem and discusses some facets of staff scheduling. To avoid any ambiguity, we introduce the following three definitions. For the first and the second definition we give an example that shows a binary representation of the respective verbal definition.

- **Shift**: A set of consecutive periods within a day. Its length is the total amount of time it covers (see [14]).
  
  Example: 1-day planning horizon with 24 periods/hours
  
  0 0 0 0 0 0 0 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0

- **Roster**: A combination of shifts and days off assignments that covers a fixed period of time; a line-of-work (see [99]).
  
  Example: 3-day planning horizon with 72 periods/hours
  
  0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 (day 1 shift)
  
  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 (day 2 off)
  
  0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 (day 3 shift)
Flexibility: For each employee, the ability to assign arbitrary shift lengths, shift starting times, and break periods, and to accommodate individual preferences, requests, and constraints.

Note, that in the second example we would get one binary sequence of length 72, however we used three lines to facilitate presentation. In the service industry, some of the most studied applications of personnel scheduling include call center staffing, airline crew pairing, nurse rostering, and postal worker tour construction (e.g., see [5, 14, 41, 45, 49, 54, 65, 83, 87, 107, 118]). Ernst et al. [65] classify the related problems into six planning modules that address demand forecasting, days off scheduling, shift scheduling, line-of-work construction, task assignment, and staff assignment. The first module, demand forecasting, determines the requirements per period or per shift. It distinguishes between a task-based view and a flexible view. Days-off scheduling (second module) deals with how rest days are interspersed between work days. Shift scheduling selects from a large pool of candidate shifts to cover the demand over some planning horizon. Candidates are constructed in association with general and individual work regulations. The line-of-work construction module combines days off scheduling with shift scheduling and uses the results to build feasible work patterns (tours) over the planning horizon. The task assignment module then assigns tasks to shifts. The last module, staff assignment, involves the assignment of staff to lines-of-work. Normally, these modules are considered sequentially by planners. In an annotated bibliography Ernst et al. [64] focus mainly on algorithms for personnel scheduling. A collection of some 700 references is classified according to the type of problem type, application area, and method. An earlier classification scheme of general manpower problems is introduced by Tien and Kamiyama [112]. The review classifies the formulations of the problems into five stages, namely, the determination of temporal manpower requirements, total manpower requirement, recreation blocks, recreation/work schedule, and shift schedule.

With respect to the classification scheme by Ernst et al. [65], the majority of the research on personnel scheduling has focused on modules two and three, shift scheduling and days off scheduling, with the objective of minimizing total cost. In what follows we will review the most relevant literature in this area.

Alfares [4] provides an efficient algorithm for the rostering or tour scheduling problem that assigns shifts and two consecutive days off to employees whereas Baily [8] integrates the shift and days off scheduling in a linear program (LP). The main model objective is to minimize the cost of premium pay. Compared to a heuristic approach savings of almost 14% are reported. Following the approach of Burns and Carter [43] one has first to develop lower bounds on the workforce size and then introduce them as an additional constraint in a linear programming model. This is sufficient to ensure integer solutions. Research that considers especially the staffing decision which is not the focus of this research is given in Baker Magazine [12], Burns and Narasimhan [42], and Mundschank and Drexel [94].

Beaumont [23] takes an expansive view of shift scheduling and includes worker availability, the maximum number of workers who can start at the same time, the relative efficiency of a worker, the cost of making a customer wait, annual leave
factors, the expected number of jobs an employee can complete in each period, the number of contractors, and the maximum number of jobs that can be done in each period, as parameters in his model. He also permits a limited amount of queuing of customers, but ultimately, real-world problems turn out to be too large to solve. Similar problems, for instance, are studied by Azmat et al. [7] and Ingolfsson et al. [79]. In another research Bailyn et al. [10] consider self-scheduling as an option. The study finds out that self-scheduling might have positive effects for both nurses and nurse managers. But everyone loses when the objective is not to balance between individual and unit benefits.

When demand is constant over some period cyclic scheduling might be the right choice. Baker [11] presents a simple staffing model that assigns days off to full-time staff given a cyclic seven-day demand pattern. A vast amount on literature that tackles the cyclic scheduling problem can be found, among others, in Bard and Purnomo [18], Bartholdi et al. [21], Beliën and Demeulemeester [29], Hochbaum and Levin [77], Millar and Kiragu [92], and Purnomo and Bard [99].

Looking at advanced computational approaches, Brusco [37] evaluates the performance of dual all-integer cutting planes for solving the tour scheduling problem. He shows that a cutting plane enhanced by a linear programming objective cut and a sophisticated row selection rule improves solution times when compared to a commercial branch-and-bound (B&B) code. Bard and Purnomo [16] use a column generation algorithm to investigate the problem of individual preference considerations in the midterm scheduling process. A similar problem is investigated by Azaiez and Al Sharif [6]. A 0-1 goal programming model is solved for a 6-month period. The model takes into account nurse preferences like ratio of night shifts and weekends off as well as hospital objectives as ensuring a continuous service with appropriate nursing skills. Jaumard et al. [81] use an exact branch-and-price algorithm to solve a similar problem and were successful for small size problems. Dowling et al. [59] develop a software package based on simulated annealing to schedule ground staff at a large international airport. They start with a feasible solution and try to improve on it until some termination criteria is fulfilled. Tabu search has also been used frequently for midterm personnel scheduling problems, especially for nurses (e.g., see [32, 39, 40, 60]). Other computational methods as various heuristics, stochastic programming or decomposition approaches have been in the focus of the research community too (e.g., see [15, 27, 36, 44, 50, 68]).

Planning for overtime, our chief concern, dates back to at least the work of Mc Manus [89] who investigates how best to allocate overtime in the British Post Office. His main goal was to identify the optimal level of staffing for given fluctuations in daily demand. By making assumptions about how the workload changes from day to day, he is able to derive a series of optimal scheduling rules. Easton and Rossin [63] argue that the increasing per capita labor expenses have forced service sector employers to increase the use of overtime and decrease the use of part-time labor. They evaluate the effects of alternative overtime staffing and scheduling policies on critical performance measures, such as total labor costs, labor utilization, and workforce size and find out that using overtime with premium pay provides significant savings. An important conclusion was that the ideal workforce size and proportion
of overtime allocated for a given scheduling policy seems to be relatively insensitive to changes in per capita labor costs.

2.2 Physician Scheduling

While much has been written on nurse scheduling, only a handful of published research exists on the more complex problem of physician scheduling. There are several reasons for this imbalance. To begin, the demand for nurses is relatively constant over a shift, but for physicians it can fluctuate widely. This makes it more difficult to match supply with demand. In addition, more complex labor agreements and individual contract clauses that physicians are able to negotiate make their scheduling problem less general. Thirdly, shifts for physicians can start at different times depending to some extent on operating room bookings, but nurses work standard 8-h or 12-h shifts with little or no variability in their starting times. Fourthly, the need to account for on-call service requirements adds a complicated series of constraints to the physician scheduling problem. For the most part, nurse scheduling is uniform from one hospital to the next, whereas physician scheduling is much more hospital-centric.

Beaulieu et al. [22] are one of the first to develop a mathematical model for scheduling emergency room physicians. They formulate the problem as a mixed integer program with three distinct 8-h shifts and allow requests for vacations, days off, and particular shifts. Attempts to solve the model with a commercial code for a 6-month planning horizon are not successful so they divide the planning horizon into six 4-week periods. Their decomposition strategy is employed to generate schedules for up to 20 physicians including five with part-time contracts. In related studies, Franz and Miller [70] develop a MIP model for scheduling residents at a large teaching hospital for training purposes. They use a rounding heuristic to get good schedules and provide an adjustment procedure to resolve infeasibilities inherent in the problem, while Cohn et al. [52] present results derived from mathematical programming techniques used to schedule the teaching phase of special training programs for medical residents at Boston university school of medicine. Their method provides one-year schedules and includes the consideration of on-call service as well as a possibility to accommodate vacation requests by each resident. White and White [119] address the problem of scheduling hospital rounds by speciality for teams comprising senior, junior, and resident physicians as well as interns. Tabu search coupled with a logic constraint formalism is used to provide monthly schedules. Ovchinnikov and Milner [97] describe an implementation of a spreadsheet model for assigning medical residents over a 1-year horizon in radiology at the university of Vermont’s college of medicine. They argue that spreadsheets are preferable to free standing codes for small size problems, especially when practitioners are the ultimate user.

Sherali et al. [106] investigate the problem of allocating night shifts to residents taking into consideration departmental staffing and skill requirements as well as
2.3 Implicit Shift Modeling

An observation from the above given literature is that virtually all models are based on an explicit representation of a shift. In particular, pre-defined shifts are used as an input to the problem and then used as the basic element of constraint construction. Three 8-h shifts, two 12-h shifts and various combinations are commonly used in practice to cover the demand during an arbitrary day. A few authors, though, have tackled the shift scheduling problem, at least in part, with implicit modeling techniques. Furthermore, Bechtold and Jacobs [26] show that implicit optimal modeling for shift scheduling has size and runtime benefits over the general applied set covering method for shift scheduling. For a single-day problem, Thompson [110] developed a linear model built on the premise that if a shift starts in a particular period then it must end in one of several successor periods. In this approach, solutions provide starting and ending periods for the minimum number of required shifts but not the shifts themselves. In a post-processing step, he applies a first-in-first-out rule to construct the actual shifts. The approach derives from the work of Moondra [93] who defined variable sets for the number of shifts starting and ending in each period and formulated constraints to impose limits on their allowable durations. Bechtold and Jacobs [24] extended this idea in the development of an implicit modeling scheme for including lunch breaks in shifts. In a subsequent work Bechtold and Jacobs [25] investigate shift length and break placement flexibility on labor utilization. Addou and Soumis [1] generalize the ideas introduced by Bechtold and Jacobs [24] and confute the hypothesis that the implicit modeling is only valid when no extraordinary overlap exists. To generate weekly rosters with two rest days, Cezik et al. [47] combined Thompson’s modeling technique with days off scheduling requirements in a network flow framework. Outputs were weekly rosters with two days off. Burke et al. [40] worked with time interval-based demand and allowed for the possibility of splitting and then combining different shifts to generate new
schedules within a tabu search framework. Greater flexibility was achieved as the time interval was reduced.

Bailey and Field [9] present a linear program that uses 6-, 8-, and 10-h shifts rather than a standard 8-h shift to cover demand periods. Tests of the Flexshift model on a data set of 12-h days report an average gain in labor of 24.2%. Some other research studies consider also implicit modeling techniques (e.g., see [38, 80, 100]).

2.4 Column Generation and B&P

The success story of column generation (CG) to solve large-scale problems dates back to middle of the 20th century. General reviews that discuss column generation are given in, for example, Wilhelm [120], Soumis [109], and Desrosiers [58]. The idea of implicit dealing with the decision variables in a multicommodity flow problem is introduced by the work of Ford and Fulkerson in the 1950s (see [69]). Dantzig and Wolfe [53] use this idea to extend linear programs columnwise on an as needed basis in the solution process. Then, the cutting stock problem is the first real application where this fundamental work has been implemented (see [73, 74]).

Savelsbergh [104] concentrates on solving the general assignment problem using column generation embedded in a B&B framework to achieve integer solutions. Compared with this Vanderbeck and Wolsey [117] present a general framework that can simultaneously deal with columns and right-hand side integer vectors. Additional they give advice to use lower bounds and to reduce the tailing off effect. Other applications dealing with column generation solving integer programs (IPs) are described in Al-Yakoob and Sherali [3], Bard and Purnomo [16], Sarin and Aggarwal [103], Purnomo and Bard [99], as well as in Gamache et al. [71].

Desaulniers et al. [56] present various accelerating strategies to speed up a column generation algorithm. However, the application is restricted to a class of vehicle routing and crew scheduling problems. Desaulniers et al. [56] concentrate on each step of the solution process, namely preprocessing, subproblem, master problem, branch-and-bound and postprocessing. Nevertheless, the strategies and suggestions can easily be adapted to other research areas like physician scheduling, the focus of our work.

Barnhart et al. [20] review the B&P literature and attempt to derive a general methodology. They present several MIPs whose structure is amenable to B&P and discuss major difficulties with respect to the modeling and branching. They also provide valuable insights on implementation and computational issues. Jaumard et al. [81] are the first to propose an exact B&P algorithm to solve the midterm scheduling problem for nurses and where successful for units with up to a dozen nurses. Their main contribution is in the formulation of the subproblem as a resource constrained shortest path problem and in the efficiency of the solution methodology. Rather than modeling resources on the vertices as in many routing applications they model them on the arcs. Taking the same approach, Beliën and Demeulemeester [30] present an integrated operating room and nurse scheduling procedure based on
2.4 Column Generation and B&P

Column generation. In the decomposition, one subproblem generates new roster for the nurses while a second produces new workload patterns for the surgery unit. The former is formulated as a shortest path problem with side constraints and is solved with dynamic programming techniques. It is called more often than the latter, which has no special structure, and hence the solutions to the IP formulation are obtained with standard software (CPLEX). Integrality is only forced on the workload patterns when creating the search tree. To find feasible solutions, the master problem is solved as a MIP at each node of the tree after no more column(s) price out favorably. Final results provide lower and upper bounds on the number of nurses required to staff the operating theaters. A broad study is carried out using data from the authors’ earlier work (see [28]).

Bard and Purnomo [16] use a column generation scheme to solve the midterm scheduling problem for nurses with an emphasis on preference considerations. Their set covering formulation includes five different shift types and employs a double swapping heuristic to generate promising columns. The cost of a new column is calculated based on the degree to which a roster violates individual preferences. Computational results using data provided by a US hospital show that high-quality rosters can be obtained within a few minutes. In subsequent work, Bard and Purnomo [17] extend their approach to allow for downgrading in the shift scheduling process.

In a study using techniques similar to those presented in Sect. 4.3, Purnomo and Bard [99] propose a new model for the cyclic preference scheduling problem and are able to derive rosters for up to 200 nurses at a time. Their B&P algorithm exploits different branching strategies to limit the size of the search tree and applies an extremely effective rounding heuristic to find integer solutions. The efficiency of the proposed approach is enhanced by applying double aggregation to the rotational profiles that define the subproblems. Runtimes of no more than 10 min are reported for almost all instances investigated. The problem associated with the daily adjustments of midterm schedules is discussed in Bard and Purnomo [19]. A 24-h planning horizon is considered and is solved in a rolling way each 8 h but just the first 8 h are implemented. To solve the problem a B&P algorithm is presented that takes advantage of two branching strategies. The first tackles fractional assignment variables in master problem but is implemented on subproblem variables. The second is associated with slack variables in master problem. The two basic components of the algorithm that contribute most to its effectiveness are a feasibility heuristic to find high-quality integer solutions quickly and a mixed-integer rounding cut procedure to improve on the lower bounds in the search tree. Results are shown for up to 200 nurses. Various other interesting studies that deal with column generation are given in Ceselli and Righini [48], Dumas et al. [61], Easton and Rossin [62], Eveborn and Ronqvist [66], Hoffmann and Padberg [78], Lavoie et al. [85], Lübbecke and Desrosiers [86], Mehrotra et al. [90], and Ni and Abeledo [95].

Our research falls in the combined categories of shift scheduling, days off scheduling, and line-of-work construction, as described by Ernst et al. [64]. We take an integrated approach that allows for flexible start times, variable shift lengths, break inclusion, overtime and on-call service, shift spillover from one day to
the next, and various individual preferences and constraints. One of the primary contributions of the proposed implicit modeling approach is that it allows for flexible starting times and variable shift lengths. In addition, it treats breaks implicitly rather than explicitly, which is a more efficient way of including them in the formulation. In the second part of the research we focus on solution methodologies. First, an efficient decomposition heuristic is presented and then we show a B&P algorithm to solve the underlying MIP. Several branching strategies and methods to find high quality integer solutions are reported to reduce the size of the search tree and to terminate with an optimal solution. To the best knowledge there is no other work in the personnel scheduling literature that combines implicit shift modeling ideas with B&P techniques to solve shift scheduling problems, especially when physicians are the main scheduling object.
Flexible Shift Planning in the Service Industry
The Case of Physicians in Hospitals
Brunner, J.O.
2010, XI, 117 p. 25 illus., Softcover
ISBN: 978-3-642-10516-6