Hartmann, Sönke; Briskorn, Dirk

Working Paper
A survey of variants and extensions of the resource-constrained project scheduling problem

Working Paper, No. 02/2008

Provided in Cooperation with:
Hamburg School of Business Administration (HSBA)

Suggested Citation: Hartmann, Sönke; Briskorn, Dirk (2008) : A survey of variants and extensions of the resource-constrained project scheduling problem, Working Paper, No. 02/2008, Hamburg School of Business Administration (HSBA), Hamburg

This Version is available at:
http://hdl.handle.net/10419/45542

Terms of use:
Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.
A Survey of Variants and Extensions of the Resource-Constrained Project Scheduling Problem

Sönke Hartmann\textsuperscript{a,∗}, Dirk Briskorn\textsuperscript{b,c,†}

\textsuperscript{a} HSBA Hamburg School of Business Administration
Alter Wall 38, D-20457 Hamburg, Germany
e-mail: soenke.hartmann@hsba.de
www.hsba.de

\textsuperscript{b} Lehrstuhl für Produktion und Logistik
Christian-Albrechts-Universität zu Kiel, D-24098 Kiel, Germany
e-mail: briskorn@bwl.uni-kiel.de
www.bwl.uni-kiel.de/Prod/team/briskorn/

\textsuperscript{c} Leonard N. Stern School of Business
New York University, 44 West 4th Street, New York, NY 10012, USA

∗ supported by the HSBA Foundation

† supported by a fellowship within the Postdoc-Program of the German Academic Exchange Service (DAAD)

Abstract

The resource-constrained project scheduling problem (RCPSP) consists of activities that must be scheduled subject to precedence and resource constraints such that the makespan is minimized. It has become a well-known standard problem in the context of project scheduling which has attracted numerous researchers who developed both exact and heuristic scheduling procedures. However, it is a rather basic model with assumptions that are too restrictive for many practical applications. Consequently, various extensions of the basic RCPSP have been developed. This paper gives an overview over these extensions. The extensions are classified according to the structure of the RCPSP. We summarize generalizations of the activity concept, of the precedence relations and of the resource constraints. Alternative objectives and approaches for scheduling multiple projects are discussed as well. In addition to popular variants and extensions such as multiple modes, minimal and maximal time lags, and net present value-based objectives, the paper also provides a survey of many less known concepts.

Keywords: Project scheduling, modeling, resource constraints, temporal constraints, networks.
1 Introduction

Project scheduling is an important task in project management. The presence of scarce resources as well as precedence relations between activities makes project scheduling a difficult task. In practice, specific software is used to support the scheduling process. The foundation of such software is a formal model that allows to describe the actual project by a set of scheduling constraints and an objective function.

During the last decades, the resource-constrained project scheduling problem (RCPSP) has become a standard problem for project scheduling in the literature. It can be summarized as follows. The RCPSP considers a project with \( J \) activities which are labeled \( j = 1, \ldots, J \). The processing time (or duration) of an activity \( j \) is denoted as \( p_j \). Once started, an activity may not be interrupted, i.e., preemption is not allowed. Due to technological requirements, there are precedence relations between some of the activities. They are given by sets of immediate predecessors \( P_j \) indicating that an activity \( j \) may not be started before each of its predecessors \( i \in P_j \) is completed. The precedence relations can be represented by an activity-on-node network which is assumed to be acyclic. Each activity requires certain amounts of resources to be performed. The resources are called renewable because their full capacity is available in every period. We have \( K \) renewable resources labeled \( k = 1, \ldots, K \). For each resource \( k \) the per-period availability is assumed to be constant over time, it is given by \( R_k \). Activity \( j \) requires \( r_{jk} \) units of resource \( k \) in each period it is in process. We consider two additional activities \( j = 0 \) and \( j = J + 1 \) representing the start and the completion of the project, respectively. Both are “dummy” activities with durations of 0 and no resource requests.

All information is assumed to be deterministic and known in advance. The parameters are assumed to be nonnegative and integer valued. A schedule is an assignment of start times \( S_j \) to the activities \( j = 0, 1, \ldots, J + 1 \). The objective is to find a schedule which leads to the earliest possible end of the project, i.e., the minimal makespan. Blazewicz et al. [21] have shown that the RCPSP belongs to the class of the strongly NP-hard problems. A mathematical model for the RCPSP has been developed by Pritsker et al. [150]. Brucker et al. [31] have provided a notation to classify resource-constrained project scheduling problems. This notation follows the famous three field notation \( \alpha | \beta | \gamma \) for machine scheduling problems introduced in Graham et al. [75]. In the context of project scheduling \( \alpha \) specifies characteristics of the resources, \( \beta \) describes activities (and can be assigned more than one value), and \( \gamma \) denotes the problem’s objective. The standard RCPSP is denoted by \( PS|\text{prec}|C_{\text{max}} \).

While the RCPSP as given above is already a powerful model, it cannot cover all situations that occur in practice. Therefore, many researchers have developed more general project scheduling problems, often using the standard RCPSP as a starting point. Since the 1990s, several survey papers on project scheduling have been published. Those papers focus on methods for the RCPSP (see Hartmann and Kolisch [81], Kolisch and Hartmann [108, 109]) and a few common variants (see Brucker [29], Brucker et al. [31], Herroelen et al. [88], Herroelen [85], Kolisch and Padman [111], Özdamar and Ulusoy [146] and Tavares [172]).

The purpose of this paper is to provide a broad overview over variants and extensions of the RCPSP that have been proposed in the literature. We restrict ourselves to deterministic approaches (for an introduction to stochastic problems, the reader is referred to Neumann [133] and Herroelen and Leus [86]). Since a huge number of papers are available, we cannot give a fully exhaustive reference list, but we work out the diversity of problem settings currently available for project scheduling. If several papers are available for a concept, we focus on references from the last ten years. For those concepts presented here that are covered by the classification scheme of Brucker
et al. [31] we point out to the related notation. Moreover, we propose extensions for some common concepts that are not yet covered by the existing classification scheme.

The outline is as follows: Section 2 lists generalizations of the activity concept. Alternative precedence constraints and network characteristics are summarized in Section 3. Section 4 gives extensions of the resource concept. Different objectives are outlined in Section 5. Finally, Section 6 deals with the simultaneous consideration of multiple projects, and Section 7 briefly summarizes generators that have been developed to produce test instances for various project scheduling problems.

2 Generalized Activity Concepts

2.1 Preemptive Scheduling

The basic RCPSP assumes that an activity cannot be interrupted once it has been started. Bianco et al. [20], Brucker and Knust [30], Debels and Vanhoucke [48], Demeulemeester and Herroelen [51] and Nudtasomboon and Randhawa [144] allow activity preemption at discrete points in time, that is, an activity can be interrupted after each integer unit of its processing time. In the $\alpha|\beta|\gamma$ notation of Brucker et al. [31], preemption is represented by adding the value $prmt$ to $\beta$.

Ballestin et al. [11] consider a variant in which an activity may be interrupted at most $m$ times; their focus is on the case $m = 1$.

Debels and Vanhoucke [48] extend the concept of preemption by a so-called fast tracking option: The parts of an activity that result from preemption need not to be processed in sequence, i.e. they can also be carried out in parallel.

Franck et al. [73] propose a calendar concept for project scheduling which includes preemptive scheduling. A calendar is defined as a binary function that determines for each period whether activity execution is possible or a break occurs during which an activity may not be started or continued. (Note that there is an analogy between the calendar concept and resource capacities varying with time, cf. Section 4.6.) Activities are allowed to be interrupted but only due to calendar breaks (a similar concept is considered in Schwindt and Trautmann [162]). Moreover, each activity $j$ has a minimum processing time $\varepsilon_j$ during which it may not be interrupted.

Buddhakulsomsiria and Kim [32, 33] follow a similar approach. Their approach permits activity splitting due to resource vacations (i.e., resource capacities varying with time). This concept is embedded into a multi-mode RCPSP with renewable resources only (cf. Section 2.4).

Damay et al. [43] consider two types of activities. An activity of the first type may not be interrupted. An activity of the second type may be interrupted at arbitrary (not necessarily integer) points in time.

2.2 Resource Requests Varying with Time

The activities in the standard RCPSP require constant amounts of renewable resources, that is, the per-period request for a resource does not change during the duration of an activity. This can be generalized by resource requests varying with time. To incorporate this formally, we denote with $r_{jkt}$ the request of activity $j$ for renewable resource $k$ in the $t$-th period of its processing time.

Hartmann [80] describes a real-world medical research project with time-dependent resource requests. There, each activity requires certain laboratory equipment, but only in the last period of its duration.
Cavalcante et al. [35] follow a similar problem setting. They define activities with time-dependent resource requests for one renewable resource which is interpreted as labor.

Drezet and Billaut [65] deal with time-dependent requests for resources which represent software developers. A minimum and a maximum resource request per period is given. Additional constraints reflect legal requirements such as the maximal working time per day.

Bartusch et al. [16] point out that resource requests varying with time can be transformed into constant requests if maximal time lags (see Section 3.2) are available. An activity is split into parts with constant requests. Between two sub-activities $i$ and $i+1$ a precedence relation is introduced as well as a maximal time lag to avoid preemption of the original activity.

### 2.3 Setup Times

In some cases, it may be required that a resource (e.g., a machine) must be prepared before a certain activity can be started. The time needed for such a preparation is called setup time.

Mika et al. [127] consider three types of setup times. Sequence-independent setup times depend only on the activity and the resource on which the activity will be performed. Sequence-dependent setup times additionally depend on the sequence of the activities, i.e., on the previous activity executed on the same resource. Finally, schedule-dependent setup times depend on the occupation of resources by activities over time, that is, the setup time related to the processing of activity $j$ depends on the resources the direct predecessors of $j$ are processed on. These concepts are embedded into an RCPSP with multiple modes (see Section 2.4). The authors note that schedule-dependent setup times could be captured by adding activities with specific modes, although this would lead to a huge number of additional activities and modes. In the notation of Brucker et al. [31], $s_j$ and $s_{ij}$ as values of $\beta$ can be used to represent sequence-independent and sequence-dependent setup times, respectively.

Mika et al. [126] discuss many further aspects of setup times. They consider setup times between single activities as well as between families of activities. Moreover, they distinguish between inseparable setups which must be executed immediately before an activity starts and separable activities which may also start earlier. A setup may be executed either after the predecessors are finished or independently from the predecessors, or it may overlap with the predecessor for a given time. The setups of several resources may be required to be synchronous or not. Finally, so-called removal times are analogous to setup times but occur after an activity is finished.

Vanhoucke [178] incorporates sequence-independent setup times into an RCPSP with preemption and fast tracking (see Section 2.1). Each time an activity is resumed after having been interrupted, the setup time occurs.

Schwindt and Trautmann [162] incorporate sequence-dependent setup times into an RCPSP with various extensions. Their goal is to capture a scheduling problem that arises in batch production.

Drexl et al. [64] employ sequence-dependent setup times (which are called changeover times here) into a multi-mode RCPSP (see Section 2.4). The setup time of activity $j$ performed in mode $m_j$ depends on the previous activity $i$ and its mode $m_i$.

Nonobe and Ibaraki [141] model setup times as additional activities that have to be completed before the activity which requires the setup is started.
2.4 Multiple Modes

The standard RCPSP assumes that an activity can only be executed in a single way which is determined by a fixed duration and fixed resource requirements. Starting with the work of Elmaghraby [66], the activity concept as given in the standard RCPSP has been extended by allowing several alternatives or modes in which an activity can be performed. Each mode reflects a feasible way to combine a duration and resource requests that allow to accomplish the underlying activity.

The assumptions of the multi-mode RCPSP can be summarized as follows. An activity \( j \) must be performed in one of its modes which are labelled \( 1, \ldots, M_j \) with \( M_j \) being the number of modes. Once started in one of its modes, an activity must be completed in that mode; mode changes and preemption are not permitted. The processing time of activity \( j \) being executed in mode \( m \) is given by \( p_{jm} \). The request of activity \( j \) being executed in mode \( m \) for resource \( k \) is \( r_{jmk} \). In addition to renewable resources, often also nonrenewable resources are considered in multi-mode models, see Section 4.1. A schedule for the multi-mode RCPSP assigns a start time \( S_j \) and a mode \( m_j \) to each activity \( j \). The most popular objective in the literature is the minimization of the makespan.

The multi-mode RCPSP is often referred to as MRCPSP. Clearly, if only one mode per activity and no nonrenewable resources are given, we obtain the standard RCPSP. Note that there might not be a feasible schedule if a nonrenewable resource is given, and as shown by Kolisch and Drexl [107] the related feasibility problem (i.e., the problem to determine whether there is a feasible schedule) is NP-complete if at least two nonrenewable resources are present. To capture multiple modes for activities in the notation of Brucker et al. [31], \( \alpha \) is set to \( MPS \).

Recent papers on this problem class include Alcaraz et al. [6], Bouleimen and Lecocq [25], Hartmann [79], Jarboui et al. [94], Józefowska et al. [97], Özdamar [145] and Pesch [149] while Varma et al. [187] discuss a multi-mode problems without nonrenewable resources. Multi-mode problems with generalized precedence constraints have been considered by Barrios et al. [14], Brucker and Knust [30], Calhoun et al. [34], de Reyck and Herroelen [47], Drexl et al. [64], Heilmann [82, 83], Nonobe and Ibaraki [141], and Sabzehparvar and Seyed-Hosseini [156] (see also Section 3). Zhu et al. [198] employ a multi-mode problem with generalized resource constraints (see also Section 4).

Erenguc et al. [68] introduce so-called crashable modes. The duration of a mode can be shortened at the expense of additional costs, which is essentially a special case of the mode concept described above.

Bellenguez and Néron [18] consider activities which require staff members with certain skill levels. This can be seen as a special case of the mode concept where each mode corresponds to a feasible subset of staff members that can carry out the activity with regard to the required skill levels.

Li and Womer [121] employ the mode concept to take quality considerations into account. Each mode \( m \) of an activity \( j \) is associated with a quality \( \gamma_{jm} \). The sum of the qualities related to the actual modes in the schedule must be at least \( \Gamma \). The authors use the resulting multi-mode RCPSP to capture the configuration of a supply chain, where the quality represents the reliability of a node in the supply chain.

Tareghian and Taheri [171] follow a similar approach. Their objective is to maximize the quality while the project’s deadline and budget (nonrenewable resource capacity) must be observed.

Tiwari et al. [173] also consider quality. In their problem setting, an activity may be started in a mode that does not allow to complete the activity at a required quality level. Processing in such a mode then has to be followed by a rework mode that completes the activity. Their approach was motivated by customer training projects in a telecommunication firm. There, an activity may be
started by a less skilled employee who is available and then completed by an employee with the required skills.

Nudtasomboon and Randhawa [144] allow preemption of activities in a multi-mode RCPSP. When an activity is resumed, the mode may not change. Various alternative objectives are included (see Section 5).

Salewski et al. [158] and Drexl et al. [64] extend the multi-mode RCPSP by introducing so-called mode identity constraints. The motivation for this is that there may be several activities that should be performed in the same way, e.g., by allocating the same resources to them. To cover this, the set of all activities is partitioned into sets of activities \( H_u, u = 1, \ldots, U \). The activities of each set \( H_u \) must be performed in the same mode. That is, \( m_i = m_j \) is must hold for all activities \( i, j \in H_u \) (note that this requires \( M_i = M_j \)).

Schultmann and Rentz [160] present a case study that demonstrates how the multi-mode RCPSP can be applied to projects which consist of the dismantling of buildings.

2.5 Tradeoff Problems

In the discrete time-resource tradeoff problem, the workload \( \psi_j \) for each activity \( j \) with regard to a single renewable resource \( k = 1 \) is given. Activity \( j \) can be performed in each discrete combination of processing time \( p_j \) and resource request \( r_{jk} \) that allows to reach the workload, that is, \( p_j \cdot r_{jk} \geq \psi_j \). Note that a combination of \( p_j \) and \( r_{jk} \) must be considered only if it is efficient, that is, \((p_j - 1) \cdot r_{jk} < \psi_j \) and \( p_j \cdot (r_{jk} - 1) < \psi_j \). This problem setting has been discussed by Demeulemeester et al. [53] and Ranjbar and Kianfar [151]. Ranjbar et al. [153] consider the case of multiple renewable resources.

The discrete time-cost tradeoff problem includes one nonrenewable resource (which is interpreted as the budget of the project) and no renewable resource; for a description of nonrenewable resources refer to Section 4.1. Demeulemeester et al. [52] consider this problem setting with makespan minimization as objective, and they also discuss a “dual” version with a deadline and a resource-based objective (see Section 5.4). A further approach to the discrete time-cost tradeoff problem has been presented by Akkan et al. [4].

Note that these two discrete tradeoff problems are special cases of the multi-mode RCPSP and, thus, are covered by \( MPS_{\text{prec}}|C_{\text{max}} \) employing the notation of Brucker et al. [31]. A mode can be defined for each efficient combination of processing time and resource request that covers the workload. The difference is that tradeoff problems usually specify the workload whereas multi-mode problems list every possible mode explicitly. Due to this difference, we propose to use \( \alpha = T_{tr}PS \) and \( \alpha = T_{tc}PS \) to specify the discrete time-resource and the discrete time-cost tradeoff problem, respectively, considering the notation of Brucker et al. [31].

Deckro et al. [49] consider the continuous version of the time-cost tradeoff problem where activity durations are not limited to discrete values. A deadline is imposed, and the cost function is quadratic and increases with an increasing deviation of the actual duration from the given normal duration. Extensions include a budget constraint and bonus or penalty payments in case of early or late project completion, respectively.

2.6 Further Activity Concepts

Drexl et al. [64] introduce the concept of forbidden periods. Each activity is associated with a set of periods in which the activity may not be carried out.
In the classical RCPSP, the activity durations are assumed to be integer-valued. Icmeli and Rom [92] and Rom et al. [154] allow the durations to be continuous. They present a mathematical model with a continuous timeline that is not partitioned into periods of equal length.

3 Generalized Temporal Constraints

3.1 Minimal Time Lags

In the classical RCPSP, an activity must have finished before any of its successors can be started. This basic precedence concept can be extended by so-called minimal time lags \( d_{ij}^{FS} \) between the completion time \( C_i \) of an activity \( i \) and the start time \( S_j \) of successor activity \( j \), which leads to constraints \( C_i + d_{ij}^{FS} \leq S_j \) (note that we always have \( d_{ij}^{FS} = 0 \) in the standard RCPSP). Allowing negative minimal time lags implies that the corresponding activities may overlap. Minimal time lags are captured by the value \( temp \) for \( \beta \) in the \( \alpha|\beta|\gamma \) notation of Brucker et al. [31].

Minimal time-lags have recently been considered by, e.g., Chassiakos and Sakellaropoulos [37], Klein [101, 102], Klein and Scholl [103, 104], Kolisch [106], and Vanhoucke [180]. De-meulemeester and Herroelen [50] show how minimal time-lags can be used to capture sequence-independent setup times as well as batches in production projects. Drexl et al. [64] consider a multi-mode problem with minimum time-lags that depend on the modes.

In addition to a minimal time lag between the completion time of activity \( i \) and the start time of activity \( j \), we may also consider minimal time lags between the start time of \( i \) and the start time of \( j \), between the start time of \( i \) and the completion time of \( j \), and between the finish time of \( i \) and the completion time of \( j \). Note that these four types of minimal time-lags can all be transformed into each other such that one type is sufficient to obtain the full expressive power of minimal time lags. We omit further details here and refer to the transformation rules provided in Bartusch et al. [16]. It should be mentioned, however, that these transformations are not applicable if a multi-mode problem is considered.

3.2 Maximal Time Lags

Similarly to minimal time lags, we denote a maximal time lag between the completion time of activity \( i \) and the start time of successor activity \( j \) as \( \bar{d}_{ij}^{FS} \), which means that \( C_i + \bar{d}_{ij}^{FS} \geq S_j \) must hold. In other words, activity \( j \) may not start later than \( \bar{d}_{ij}^{FS} \) periods after the completion time of activity \( i \). Analogously, maximal start-start, start-finish, and finish-finish time lags can be given. The RCPSP with minimal and maximal time lags is often referred to as RCPSP/max. Note that maximal time lags might lead to infeasible project instances, and the associated feasibility problem is NP-complete. Like minimal time lags, maximal time lags are covered by \( temp \) for \( \beta \) in the classification scheme of Brucker et al. [31].

Problems with minimal and maximal time lags have been discussed by, e.g., Bartusch et al. [16], Cesta et al. [36], Dorndorf et al. [58], Neumann and Zimmermann [135, 137], Neumann et al. [138], Schwindt [161], and Schwindt and Trautmann [162]). The concepts of minimal and maximal time lags has also been considered within the multi-mode RCPSP, see Barrios et al. [14], Brucker and Knust [30], Heilmann [82, 83], de Reyck and Herroelen [47] and Calhoun et al. [34]. Sabzehparvar and Seyed-Hosseini [156] consider minimal and maximal mode-dependent time lags, i.e. any time lag between two activities \( i \) and \( j \) depends on the modes \( m_i \) and \( m_j \) (thus, a maximal finish-start time lag can be written as \( d_{im_i,jm_j}^{FS} \)).
It should be noted that maximal time lags typically lead to cyclic network structures, see Franck and Neumann [70] for a discussion of this issue. A survey of exact and heuristic algorithms for problems with maximal time lags has been given by Neumann and Zimmermann [136].

3.3 Release Dates and Deadlines

A release date \( r_j \) is the earliest time at which activity \( j \) may be started. Likewise, a deadline \( d_j \) denotes the latest time instant at which activity \( j \) must be finished. In the notation of Brucker et al. [31], release dates and due dates are represented in the \( \beta \) field by \( r_j \) and \( d_j \), respectively. Recent problems with release dates and deadlines include Drezet and Billaut [65], Kis [99, 100], and Klein and Scholl [103, 104]. Baptiste et al. [13] consider what they call the cumulative scheduling problem which is an RCPSP with a single renewable resource and without precedence relations, but with release dates and deadlines for the activities (the objective is to find a feasible schedule).

Note that a release date \( r_j \) is equivalent to a minimal finish-start time lag between the dummy source activity and activity \( j \). Analogously, a deadline \( d_j \) can be imposed by a maximal finish-finish time lag between the dummy source activity and activity \( j \).

Whereas a deadline as defined above may not be violated, Ballestín et al. [10], Brânzei et al. [26], Chiu and Tsai [40], Neumann et al. [138], and Özdamar et al. [147] consider due dates which can be exceeded at some penalty cost. Objective functions considering due dates are discussed in Section 5.1. Due dates are also important when multiple projects are planned simultaneously, see Section 6.2.

3.4 Time-Switch Constraints

Time-switch constraints have been introduced by Yang and Chen [196]. The planning horizon is divided into cycles of work and rest time windows. An activity can start only in a work window. This approach allows to capture working times (e.g., working is allowed from Monday till Friday but not on weekends). Note that there is a certain relationship to the concept of forbidden periods (see Section 2.6), but a main difference is that the forbidden periods depend on the activities.

Brucker and Knust [30] mention such time windows as well and point out that they can be taken into account by adding a renewable resource with a capacity of one during each work window and 0 otherwise. Vanhoucke et al. [184] employ time-switch constraints within a discrete time-cost tradeoff problem (see 2.5).

3.5 Further Temporal Constraints

Bartels and Zimmermann [15] introduce so-called partially ordered destructive relations for projects that consist of engineering and testing activities in the automotive industry. Such a relation between activities \( i \) and \( j \) implies that either \( i \) must be completed before \( j \) starts, or these activities must be executed in different modes (i.e., use different resources). This is relevant for the case that activity \( j \) destroys a resource (e.g., a test vehicle which would be destroyed in a crash test) that could also be used by activity \( i \) which would not destroy the resource. See also Section 4.3 on the resource types used in this problem setting.

In a case study on a real-world project, Hartmann [80] notes that many medical research projects require that certain activities may not finish in the same period. In order to avoid systematic errors, replications of the same experiment (which correspond to activities) are not allowed to end at the same time. Thereby, it is not prescribed which activity shall finish first. It is
shown that this type of temporal constraint is, from a mathematical point of view, a special case of
time-dependent requests for renewable resources (see Section 2.2).

Brucker and Knust [30] introduce parallelity constraints which force two activities to be pro-
cessed in parallel for at least one period. They also discuss the opposite case in which two activ-
ities may not be processed in parallel at all. The latter is extended by minimal separating times
between two activities which may not be carried out in parallel. Note that a minimal separating
time is different from a minimal time lag since the former is not associated with a precedence
relation between the activities. Finally, Brucker and Knust [30] introduce so-called bounding con-
straints which require that a set of activities is carried out for at least \( l \) and at most \( u \) time units
within a given set of periods.

Kis [100] describes so-called feeding precedence constraints. Each precedence relation be-
tween two activities \( i \) and \( j \) is associated with a percentage \( g_{ij} \) which indicates that \( j \) may not start
before \( g_{ij} \) percent of \( i \) are completed. Moreover, at any point in time the percentage of \( j \) that is
completed may not be greater than the percentage of \( i \) that is completed.

Bomsdorf and Derigs [23] consider a class of precedence relations where no activity \( k \) may be
scheduled between two precedence-related activities \( i \) and \( j \).

Nonobe and Ibaraki [141] use a similar concept which adds a dependency on renewable re-
sources. An immediate precedence constraint \( i \prec \prec r j \) between activities \( i \) and \( j \) implies that \( i \) must
be finished before \( j \) is started and that no other activity \( k \) that requires resource \( r \) is allowed to be
started between the completion time of \( i \) and the start time of \( j \). Nonobe and Ibaraki [141] use
this to ensure that no other activity is started between a setup activity and the activity that requires
the setup. They also mention that these immediate precedence constraints can be used to enforce
equal start times for activities or to enforce that two activities overlap.

Krüger and Scholl [115] introduce temporal constraints caused by resource transfers. Trans-
ferring resource \( r \) from the location where activity \( i \) is carried out to the location of activity \( j \)
requires \( \Delta_{ijr} \) time units. In an extended formulation, executing a resource transfer may require
another resource.

### 3.6 Logical Dependencies

Several researchers, e.g., Elmaghraby [67] and Belhe and Kusiak [17], have extended the prece-
dence network concept by adding logical nodes. In the RCPSP, an activity node which has several
successor nodes implies that this activity must be finished before any of the successors may be
started. This controls the temporal arrangement of the activities. As all successor activities must
be executed, the activity nodes can be viewed as logical “AND” nodes. Extending this concept,
some activity nodes might be declared as logical “OR” nodes, implying that at least one successor
activity of an “OR” node has to be performed. Analogously, “XOR” (exclusive-or) nodes induce
that exactly one successor activity must be executed. These network extensions are, for example,
useful to model research and development projects: After an “XOR”-activity “test prototype” has
been performed, either activity “modify prototype” or activity “start production” is executed, de-
pending on the outcome of the test. In the project management literature, networks with logical
nodes are mostly treated in connection with stochastic concepts, see Neumann [133].

Another type of logical dependencies has been proposed by Kuster and Jannach [119]. A so-
called mutual exclusion relationship between activities \( i \) and \( j \) implies that if \( i \) is executed then
\( j \) shall not be executed. Analogously, a mutual inclusion relationship between activities \( i \) and \( j \)
means that if \( i \) is executed then \( j \) shall be executed as well. The resulting problem is used for
scheduling ground processes on airports such as fueling, cleaning and catering.
3.7 Network Representations

Finally, we briefly address the way the activities are represented within the network. In the RCPSP and many of its extensions, the so-called *activity-on-node* representation is used, that is, each activity corresponds to a node in the network, while the precedence relations are given by arcs between the nodes. In the literature, also an alternative representation can be found, namely the *activity-on-arc* network. There, each activity corresponds to an arc in the network, while the nodes represent events. This representation has been used by, e.g., Achuthan and Hardjawidjaja [2], Brânzei et al. [26], Elmaghraby [66], Smith-Daniels et al. [167], and Tareghian and Taheri [171]. For a brief discussion of the similarities and differences between both representations, we refer to Kolisch and Padman [111].

4 Generalized Resource Constraints

4.1 Nonrenewable and Doubly Constrained Resources

The basic RCPSP features only one type of resources which is called renewable because it is available in each period with its full capacity. In project scheduling problems with multiple modes (MPS|prec|C_{max}, see Section 2.4), often three different kinds of resources are considered, namely renewable, nonrenewable, and doubly constrained resources. This resource categorization has been developed by Słowiński [165] and Węglarz [191].

Renewable resources such as manpower and machines are limited on a per-period basis. In contrast, nonrenewable resources have a limited capacity for the entire project, an example would be a budget for the project. While renewable resources are already part of the basic (single-mode) RCPSP, nonrenewable ones need only be considered in a multi-mode environment. In the classification scheme of Brucker et al. [31], we propose to reflect a multi-mode problem with both renewable and nonrenewable resource with \( \alpha = MPS; R; N \). This slightly modifies the original notation of Brucker et al. [31], but it enables us to integrate further resource categories throughout the following sections.

Doubly constrained resources are limited both for each period and for the whole project (money is an example if both the budget and the per-period cashflow of the project are limited). As a doubly constrained resource can be incorporated by a renewable and a nonrenewable resource, doubly constrained resources do not enhance the expressive power of the problem setting and need not be considered explicitly.

4.2 Partially Renewable Resources

Partially renewable resources (which generalize both renewable and nonrenewable resources) have been introduced by Böttcher et al. [24]; see also Alvarez-Valdes et al. [7, 8], Drexl et al. [63, 64] and Schirmer and Drexl [159]. For each partially renewable resource \( k \), we have a set \( \Pi_k \) of so-called period subsets. Resource \( k \) has a total availability of \( R_k^j(Q_{ki}) \) during the time given by period subset \( Q_{ki} \subseteq \{1, \ldots, T\} \), where \( Q_{ki} \in \Pi_k \) and \( T \) is the planning horizon. The request of activity \( j \) for partially renewable resource \( k \) is again denoted as \( r_{jk} \). In an equivalent normalized formulation, a separate partially renewable resource is defined for each period subset. Considering the notation of Brucker et al. [31], we propose to use \( \alpha = PS; PR \) to reflect project scheduling with partially renewable resources.

For illustration of this concept, consider an employee who may work on every day from Monday through Friday and either on Saturday or on Sunday but not both. Obviously, each day (Mon-
day through Sunday) would be one period. A partially renewable resource would then be defined as follows: Each day from Monday through Friday makes up one separate period subset with an associated availability of 1. The weekend restriction is imposed by defining another period subset containing both Saturday and Sunday with a capacity of 1.

Zhu et al. [198] integrate partially renewable resources into the multi-mode RCPSP. Nudtassomboon and Randhawa [144] employ a special case of partially renewable resources where each period subset is an interval $[t_1, t_2]$.

### 4.3 Cumulative Resources

Cumulative resources have been introduced by Neumann and Schwindt [134] who incorporate them into an RCPSP with minimal and maximal time lags. If a project represents a production process, then an activity may require some intermediate product which is withdrawn from the storage area, or it may manufacture a product which is put into the storage area. A cumulative resource $k$ is given by the capacity $R_k$ and the minimum inventory level (or safety stock) $R_k$ of the storage area. An activity $j$ may produce $r_{jk}$ items (then we have $r_{jk} > 0$) or withdraw $r_{jk}$ items (then we have $r_{jk} < 0$); otherwise we have $r_{jk} = 0$ (note that events rather than activities are used to describe production and withdrawal formally). Neumann and Schwindt [134] point out that cumulative resources subsume both renewable and nonrenewable resources. We propose to use $\alpha = PS; Cu$ to specify the RCPSP with cumulative resources in the notation of Brucker et al. [31].

Bartels and Zimmermann [15] employ cumulative resources within a multi-mode RCPSP with minimal and maximal time lags. They consider a project of engineering and testing activities in the automotive industry. A test vehicle is modeled as a cumulative resource since it can be built, used (which temporarily withdraws its capacity) and destroyed in a crash test (which withdraws its capacity until the project end).

Neumann et al. [139] use cumulative resources in an RCPSP incorporating minimal and maximal time lags to capture a scheduling problem related to batch production in the process industry. Schwindt and Trautmann [162] further extend this setting.

### 4.4 Continuous Resources

In the problem settings discussed up to now, the considered resources were available in discrete quantities only, as in the cases of, e.g., manpower and machines. Węglarz et al. [194] generalized the concept of renewable resources by allowing continuously divisible resources. This is useful if the resources correspond to, e.g., energy or raw material like liquids. An overview is provided in Blazewicz et al. [22]. Considering the notation of Brucker et al. [31], we suggest to use $\alpha = PS; Co$ to reflect the RCPSP with continuous resources.

Further problems with continuous resources have been considered by Józefowska et al. [96], Kis [99, 100] and Waligóra [190]. Węglarz [192] presents an extension which is essentially a continuous variant of the doubly constrained resource concept.

### 4.5 Dedicated Resources

Dedicated resources are resources that can be assigned to only one activity at a time. They can be expressed by a renewable resources with an availability of one unit per period, $R_k = 1$. Consequently, they are included in the RCPSP as a special case. Project scheduling problems with dedicated resources are studied in Bianco et al. [19, 20]. Dorndorf et al. [57] refer to the RCPSP with dedicated resources as disjunctive scheduling problem, see also Dorndorf et al. [59]. In the
notation of Brucker et al. [31], dedicated resources can be covered by setting \( \alpha = PS; Rm, 1, 1 \) which means that we have \( m \) renewable resources with a capacity of one unit and a request of at most one unit.

4.6 Resource Capacities Varying with Time

Up to this point, the availability of the renewable resources has been assumed to be constant over time, i.e., the capacity is the same in each period. This assumption might be too strict in some cases such as changing availability of workers due to vacation or varying equipment capacities due to maintenance. To capture resource availabilities varying with time, we denote the capacity of renewable resource \( k \) in period \( t \) as \( R_{kt} \). Time-dependent resource capacities have been discussed by, e.g., Bomsdorf and Derigs [23], Klein [101, 102], Klein and Scholl [103], Nonobe and Ibaraki [141], Pesch [149] and Schwindt and Trautmann [162]. Hartmann [80] uses time-dependent capacities to capture the availability of researchers and laboratory equipment in a medical research project. Resource capacities varying with time are sometimes discussed together with preemptive scheduling, see Section 2.1 and the references cited therein. In the notation of Brucker et al. [31], time-varying capacities of renewable resources can be captured using \( \alpha = PS; Rm, \cdot, \cdot \) which indicates that we have \( m \) renewable resources with a time-dependent capacity and an arbitrary request.

Brucker and Knust [30] consider so-called disjunctive resources which have time-dependent capacities of up to 1, that is, \( R_{kt} \in \{0, 1\} \). Observe that these resources are a time-dependent version of the dedicated resources described in Section 4.5.

Icmeli and Rom [92] deal with time-dependent capacities in a problem with a continuous timeline where changes in the resource availability occur at certain points in time called milestones.

Bartusch et al. [16] show that resource capacities varying with time can be transformed into constant capacities if minimal and maximal time lags are available. The constant capacity would be defined as the maximum of the time-dependent capacity over time, and for each time interval with a smaller capacity, an artificial activity is defined to reduce the capacity appropriately. Each artificial activity is fixed to the desired time interval using a minimal and a maximal time lag.

Also note that the RCPSP with time-varying resource capacities is a special case of the RCPSP with partially renewable resources, since we can define a period subset for each period with an individual capacity.

5 Alternative Objectives

5.1 Time-Based Objectives

While the minimization of the makespan is among the most popular objectives (see, e.g., the references in Kolisch and Hartmann [108]), there are various other time-based objectives. Objectives based on lateness, tardiness, and earliness are of particular importance. The lateness \( L_j \) of an activity \( j \) is the deviation of the completion time \( C_j \) from a given due date \( d_j \), hence \( L_j = C_j - d_j \). The tardiness \( T_j \) is similar but it cannot be negative; it is defined as \( T_j = \max \{0, C_j - d_j\} \). The earliness \( E_j \) is defined analogously as \( E_j = \max \{0, d_j - C_j\} \).

Ballestín et al. [10], Kolisch [106], Nudtasomboon and Randhawa [144], and Viana and de Sousa [188] consider the minimization of the weighted tardiness. Note that this generalizes the makespan objective. Neumann et al. [138] describe the minimization of the maximum lateness and of the weighted total tardiness. Vanhoucke et al. [182] discuss a just-in-time objective which
is achieved by minimizing the weighted sum of all earliness and tardiness values. Franck and Schwindt [71] mention a multi-mode RCPSP with the objective to minimize the sum of the earliness and tardiness values. An alternative objective minimizes the maximum value of all earliness and tardiness values. Lorenzoni et al. [122] propose a variant where earliness and tardiness are measured w.r.t. a given time window in which an activity should be carried out.

Vanhoucke [179] defines a set of time windows for each activity. Carrying out an activity within one of its time windows is desired due to quality considerations. The objective function minimizes penalties that are caused by executing activities outside their time windows. The approach is motivated by a bio-technology project. Vanhoucke [179] notes that this concept is similar to the time-switch constraints (see Section 3.4) where time windows are reflected by hard constraints.

Nudtasomboon and Randhawa [144] propose to minimize the sum of all activity completion times, while Rom et al. [154] minimize the weighted sum of the completion times. Similarly, Nazareth et al. [132] suggest to minimize the mean flow time, which is the average of all activity completion times. Note that minimization of total completion time and minimization of average completion time is equivalent.

In the $\alpha|\beta|\gamma$ notation of Brucker et al. [31], these objectives can be captured in the $\gamma$ field by symbols that are well known from machine scheduling: We can use $\gamma = C_{\text{max}}$ for the makespan, $\gamma = L_{\text{max}}$ for the maximal lateness, $\gamma = \sum w_j C_j$ for total weighted completion time, $\gamma = \sum w_j T_j$ for total weighted tardiness, $\gamma = \sum w_j U_j$ for weighted number of tardy jobs, or $\gamma = \sum (w_e E_j + w_t T_j)$ for sum of total weighted earliness and total weighted tardiness ($w_j$ is the weight of activity $j$ in the objective and $U_j \in \{0, 1\}$ indicates whether activity $j$ is late or not).

### 5.2 Robustness-Based Objectives

During the execution of a project, delays may occur that could not be foreseen when the schedule was determined. Therefore, a project manager might be interested in a robust schedule, i.e. a schedules in which a delay has only a limited effect. This approach is often referred to as proactive scheduling. (While there are several robustness-oriented approaches that make use of stochastic concepts to capture the uncertainty explicitly, we restrict ourselves to deterministic problems.)

Al-Fawzan and Haouari [5] define the free slack $s_l_j$ of activity $j$ as the amount of time the completion of $j$ can be delayed without affecting any other activity. They propose a bi-objective problem incorporating makespan minimization and maximization of total free slack which is used as a measure of robustness. This problem setting is also treated by Abbasi et al. [1]. Kobylianski and Kuchta [105] provide a discussion of this approach and propose to use maximization of the minimum free slack $\min_i s_l_i$ as sole objective. Quality according to the makespan is assured using an additional project deadline constraint.

Chtourou and Haouari [41] follow the approach of Al-Fawzan and Haouari [5] and add further measures of robustness. They suggest to weight the free slack of an activity with the number of its immediate successors and/or the sum of its resource requests. Moreover, they outline alternative measures in which the free slack $s_l_j$ of activity $j$ is replaced with a binary value $\alpha_j = 0$ if $s_l_j = 0$ and $\alpha_j = 1$ if $s_l_j > 0$. The idea behind the latter approach is to avoid a bias caused by a very large free slack values. Such large slacks might not be particularly important, given that a smaller slack is already a sufficient buffer for typical delays.

Icmeli-Tukel and Rom [93] assume that activities need to be reworked. $R_{T_{jt}}$ is the expected rework time (i.e., delay) of activity $j$ when completed at time $t$, while $R_{C_{jt}}$ is the expected rework cost. It is assumed that $R_{T_{jt}}$ and $R_{C_{jt}}$ increase with time. The objective is to find a schedule with
5.3 Objectives for Rescheduling

Rescheduling is necessary if the project is already in progress, but due to unexpected events (e.g., delays) the schedule that has been calculated before the start of the project is no longer valid. In such a situation, the problem’s characteristics may have changed: Some activities might already be finished and can be ignored, some activities might be in progress and must be considered fixed (as long as no preemption is allowed), and the resource availability may have changed and might even have switched from time-independent to time-dependent. In contrast to proactive scheduling which anticipates disruptions by building robust schedules, we now consider the case that some disruption has occurred and a new schedule has to be determined. This case is often referred to as reactive scheduling.

Calhoun et al. [34] propose to minimize the perturbation of the original schedule by minimizing the number of activities that receive a different start time in the new schedule. Van de Vonder et al. [177] propose to reschedule the remaining activities such that the sum of deviations of the new finishing times from the original ones is minimized. They refer to a just in time problem developed by Vanhoucke et al. [182] where the weighted sum of earliness and tardiness of each activity is minimized. They suggest to use the original finishing times as due dates in order to obtain a just-in-time problem. Sakkout and Wallace [157] measure the perturbation to be minimized as the sum of deviations of starting and finishing times of all activities.

Zhu et al. [197] propose to penalize changes in resource utilization and in the selected modes in a multi-mode RCPSP. Moreover, it is considered to fix some activities by additional constraints. This can be useful if the activities have been scheduled that close to the current point of time that no change is possible, or if the schedule should catch up with the original one at a specific time in the future.

5.4 Objectives Based on Renewable Resources

Various objectives related to renewable resources have been discussed in the literature. A comprehensive overview is provided by Neumann et al. [140].

In the basic RCPSP, the makespan is to be minimized while a given capacity level of each renewable resource has to be observed. A “dual” version of this is the resource investment problem where the costs for providing a certain capacity level are to be minimized while a deadline for the project has to be observed. The objective is to minimize the sum of availability costs of all resources, i.e., \( \sum_k C_k(R_k) \), where \( C_k \) is a discrete non-decreasing cost function of resource \( k \) and the capacity levels \( R_k \) are variables. A special case is obtained from the linear cost function \( C_k(R_k) = c_k \cdot R_k \), where \( c_k \) is the per-unit cost of resource \( k \). The resource investment problem has recently been tackled by Drexl and Kimms [61], Neumann and Zimmermann [137], Neumann et al. [138], Ranjbar et al. [152], and Yamashita et al. [195].

Shadrokh and Kianfar [163] present an extension of the resource investment problem. Instead of a deadline, a due date is used, i.e., delayed completion of the project is allowed. The objective is to minimize the sum of resource availability costs and tardiness penalty. The latter results from a constant cost factor for each period the project is delayed.

Nübel [143] suggests the resource renting problem in which the renewable resources have to be rented. Renting resource \( k \) is associated with fixed costs \( c^f_k \) per unit and variable costs \( c^v_k \) per unit and period. Thus renting \( a \) units of resource \( k \) over \( t \) periods leads to costs of \( a \cdot (c^f_k + t \cdot c^v_k) \). The
fixed costs could be the delivery costs of the resource to the project’s location while the variable costs represent the actual rent. The objective is to minimize the total renting costs. Observe that \( c^k_0 = 0 \) leads to the resource investment problem. The resource renting problem has been picked up by Ballestín [9].

Another resource-based objective is to achieve a smooth resource profile, which leads to the resource leveling problem. The objective there is to minimize the changes in the level of resource usage from period to period in the schedule while a deadline has to be observed. This might be measured as the maximum change or the sum of all changes, see Neumann and Zimmermann [137], Neumann et al. [138] and Nudtasomboon and Randhawa [144]. Bandelloni et al. [12] propose to minimize the sum of all squared changes.

Davis et al. [44], Neumann and Zimmermann [137] and Viana and de Sousa [188] propose to minimize the utilized renewable resource units of each resource that exceed a given level. Nudtasomboon and Randhawa [144] minimize the cumulated deviations of the resource utilization from a given level, i.e., they consider both over- and underutilization. Bomsdorf and Derigs [23] discuss the minimization of both the number and the length of gaps in the resource profile. Kis [99] distinguishes between internal and external capacities of the renewable resources. The objective is to minimize the expenses for using external capacities.

In the notation of Brucker et al. [31], these objectives are represented by \( \sum c^k f(r^k(S,t)) \) in the \( \gamma \) field. Here \( c_k \) represent the cost per unit of resource \( k \) and \( f \) is a function of resource utilization profile \( r^k(S,t) \). Note that the resource investment problem is obtained by setting \( f = \max \).

### 5.5 Objectives Based on Nonrenewable Resources

The multi-mode RCPSP (see Section 2.4) requires that the resource capacities are not exceeded while the makespan is minimized. Analogously to the resource investment problem of Section 5.4, an alternative setting would be to impose a deadline on the project while minimizing the consumption of the nonrenewable resources.

Such a nonrenewable resource-based objective has been considered by Akkan et al. [4] and Demeulemeester et al. [52] within the discrete time-cost tradeoff problem, where the only non-renewable resource is interpreted as money. Similarly, Nudtasomboon and Randhawa [144] and Tareghian and Taheri [171] minimize the consumption of a nonrenewable resource which again represents money. Nudtasomboon and Randhawa [144] and Viana and de Sousa [188] suggest to minimize the consumed nonrenewable resource units that exceed the capacities. These objectives can be specified by \( \gamma = \sum c^k f(r^k(S)) \) in the notation of Brucker et al. [31], where \( f \) is a function of the consumption \( r^k(S) \) of nonrenewable resource \( k \) in schedule \( S \).

### 5.6 Objectives Based on Costs

Maniezzo and Mingozzi [124] and Möhring et al. [128, 129] consider costs \( c_j \) for each activity \( j \) that depend on the start time \( t \) of \( j \). The objective is to minimize the sum of these costs (the underlying problem is an RCPSP without resources). As pointed out by Möhring et al. [129], this objective includes various other well-known objectives as special cases, such as makespan minimization, earliness-tardiness-based objectives and maximization of the net present value (for the latter see Section 5.7).

Achuthan and Hardjawidjaja [2] minimize total project costs which consist of earliness and tardiness costs with regard to due dates as well as costs related to the durations of the activities (the durations can be shortened at additional costs). The project must be finished not later than a
given deadline, and renewable resources are not considered.

Dodin and Elimam [55] aim at minimizing costs which includes costs for activity crashing (i.e., shortening the duration), material costs and inventory holding costs. Based on a given due date, a bonus for early project completion or a penalty for late completion is added.

Nonobe and Ibaraki [142] consider activity durations which have to be between a lower and an upper limit. This leads to an objective that minimizes the costs related to the durations, where the cost function for each activity is convex.

Rummel et al. [155] consider a cost-based objective that consists of two components. The project duration costs are proportional to the makespan. Activities can be combined (consolidated) in order to reduce the makespan, but this causes consolidation costs. The objective is used in a problem without resource constraints.

5.7 Objectives Based on the Net Present Value

Another important type of objective emerges if cash flows occur while the project is carried out. Cash outflows are induced by the execution of activities and the usage of resources. On the other hand, cash inflows result from payments due to the completion of specified parts of the project. Typically, discount rates are also included. Note that cash flows related to activity $j$ might occur at several points in time during execution of $j$. However, they can easily be compounded to a single cash flow at the beginning or the end of $j$. These considerations result in problems with the objective to maximize the net present value (NPV) of the project subject to the standard RCPSP constraints. Recent papers include Kimms [98], Mika et al. [125], Padman and Zhu [148], and Vanhoucke et al. [181]. In the notation of Brucker et al. [31], the objective to optimize the NPV is represented by $\sum c^F_j \beta^j$ for $\gamma$, where $c^F_j$ is the cash flow of activity $j$ and $\beta$ is the discount factor.

Furthermore, NPV objectives have been investigated for the multi-mode RCPSP (see Ulusoy et al. [176], Varma et al. [187], and Waligora [190]), the resource investment problem (see Najafi and Niaki [131]), and the RCPSP with minimal and maximal time lags (see Neumann and Zimmermann [135, 137]). Icmeli and Rom [92] employ the NPV objective in a problem with continuous activity durations and time-dependent resource capacities. An overview of problem settings with NPV objectives is given by Herroelen et al. [87]. A comparison of problems with cash flow based objectives is provided by Dayanand and Padman [46].

Several variants of the NPV objective mentioned above arise in literature. Instead of fixed cash flows, Etgar et al. [69] consider individual cash flow functions of the completion time for each activity. In Vanhoucke et al. [185], at specific points of time a cash inflow related to activity $j$ occurs which amounts to the fraction of $j$ being finished at this moment. Vanhoucke et al. [183] assume the cash flow of an activity to be a linear non-increasing function of the activity’s completion time. Najafi and Niaki [131] consider a cash flow related to a subset of activities, that is, a cash flow is initiated when the last activity of the subset is finished. Furthermore, the cash flow associated with activity $j$ might depend on the mode chosen for $j$ as lined out in Icmeli and Erenguc [91].

Some authors have extended the NPV objective by additional payments upon project completion. Icmeli and Erenguc [90] as well as Özdamar et al. [147] consider a penalty being charged for each period the project is finished after a given due date. Chiu and Tsai [40], Doersch and Patterson [56], and Sung and Lim [170] extend this by including also bonus payments for early completion.

Smith-Daniels et al. [167] and Sung and Lim [170] propose the objective to maximize the discounted amount of cash available in each period. This amount is influenced by cash flows
associated with each activity. In Smith-Daniels et al. [167] this is extended by a constraint: An activity $j$ may not be started at a specific point in time if the cash outflow at the beginning of $j$ exceeds the available cash. Cash can be seen as a renewable resource if the outflow equals the inflow for each activity. However, if this does not hold, the capacity of this resource depends on the point in time and the activities being already started (note that there is a similarity to the concept of cumulative resources, see Section 4.3). In Sung and Lim [170] there is no limit for the amount of cash available. In particular, the amount of available cash might be negative which can be interpreted as being indebted.

Ulusoy and Cebelli [175] investigate the negotiation process to find the timing of payments and the amount of each specific payment between a client and a contractor. Obviously, the client seeks to minimize the NPV while the contractor aims at maximizing it. The objective in Ulusoy and Cebelli [175] is to find the payment structure which minimizes each party’s loss in comparison to the respective ideal payment structure. Dayanand and Padman [45] treat a similar problem but restrict themselves to the client’s point of view. The client might associate a specific value with each event (start or completion of a job). Cash outflows can be assigned to each event having a positive value. The problem is to find a project schedule and decide cash outflows to happen at a given number of events. The total outflow might exceed the total value of finished activities (minus a percentual retention) at no point of time. The objective is to find a solution such that discounted cash inflow (associated with finishing the project) minus total discounted cash outflow is maximized.

5.8 Multiple Objectives

The problems discussed above have a single objective function (e.g., makespan minimization) while all other properties of a schedule are controlled by means of constraints (e.g., resource usages and costs). However, several authors have employed multiple performance measures into their project scheduling problems.

A widely used approach to cope with multiple objectives is to define one overall objective as the weighted sum of all performance measures considered. This is done by Nudtasomboon and Randhawa [144] who include various objectives into the multi-mode RCPSP such as makespan, weighted tardiness, resource leveling and nonrenewable resource consumption. Vöß and Witt [189] employ the multi-mode RCPSP with an objective that contains makespan, weighted tardiness and setup costs. The inclusion of setup costs supports batching of activities. The problem setting is motivated by a production planning problem at a steel manufacturer. Bomsdorf and Derigs [23] employ an objective for movie shooting projects that consists of several components which are allowed to be squared. The components include specific criteria such as the minimization of location changes over time (each activity is associated with a location). Al-Fawzan and Haouari [5] combine makespan minimization and maximization of total free slack into one objective.

Another way to deal with multiple objectives is the generation of Pareto-optimal schedules. This approach is followed by several authors. Davis et al. [44] minimize the makespan as well as the overutilization of each renewable resource. Viana and de Sousa [188] add to these the minimization of overutilization of each nonrenewable resource and the mean weighted tardiness. Hapke et al. [77] propose a multi-criteria approach which allows to simultaneously consider several objectives, namely time based, resource based, and financial ones. Słowiński et al. [166] consider the multi-mode RCPSP with various objectives including makespan, mean weighted lateness, total number of tardy activities, smoothness of the resource profile, total and weighted resource consumption and net present value. Dörner et al. [60] employ three objectives within a variant of
the time-cost tradeoff problem. The first objective is a function of the project makespan, while the second and the third are functions of the monetary and nonmonetary costs for crashing the activities, respectively. The case of Pareto-optimal schedules can be represented in the classification scheme of Brucker et al. [31] by setting \( \gamma \) to a vector containing all symbols of the objectives under consideration.

Nabrzynski and Węglarz [130] present a knowledge-based approach to a project scheduling problem with multiple modes and a set of time and cost criteria. The latter includes the makespan, the smoothness of the resource profile, mean weighted lateness, mean weighted flow time, total and weighted resource consumption as well as the net present value.

6 Multiple Projects

6.1 Networks for Multiple Projects

In practice, often not only one but several dependent projects have to be scheduled simultaneously. This is important if two or more projects which may be processed in parallel share at least one resource. Herroelen [85] stresses the importance of frameworks for scheduling multiple projects.

The most common way to deal with multiple projects is to comprise their networks into a “super-network” by adding a “super-source” and a “super-sink,” while a common pool of resources is considered. This approach has been proposed by Pritsker et al. [150]. Confessore et al. [42] consider a set of projects in which each project has its own resources while one additional resource is shared by the projects. The advantage of integrating multiple projects in a single network is that this provides a formal basis for the application of scheduling methods for single projects also to the case of multiple projects.

As for the single-project case, further constraints can be added. Pritsker et al. [150] add due dates and deadlines for the sink activities of the single projects. Franck et al. [72] consider a network of multiple projects with minimal and maximal time lags. Krüger and Scholl [115] discuss a multi-project problem that includes resource transfers. Kumanan et al. [116] consider multiple projects the activities of which can be performed in one of several modes.

6.2 Objectives for Multi-Project Scheduling

Several different objectives for scheduling multiple projects have been discussed. In many multi-project problems, each project is associated with a due date, and the tardiness appears to be among the most widely used performance measures for multi-project scheduling.

Chen [39] employs the minimization of weighted tardiness as well as the minimization of costs that exceed the budget of each individual project as well as the total budget. Chiu and Tsai [40] deal with an NPV objective that includes earliness and tardiness payments with regard to the project due dates. Franck et al. [72] consider resource leveling as well as resource investment objectives. Goncalves et al. [74] aim at minimizing the weighted sum of earliness and tardiness as well as the flow time of each project. Homberger [89] discusses the minimization of the average makespan of the projects, where the makespan of each individual project is defined as its completion time minus its release date. Lawrence and Morton [120] deal with the minimization of weighted tardiness. Lova et al. [123] consider various objectives such as mean tardiness, makespan of the “super-project,” resource leveling and project splitting. The latter is defined as the number of periods a project is interrupted, i.e., no activity of this project is carried out after the project has been started and before it is finished (but probably activities related to other projects).
Kurtulus and Davis [118] consider the total delay, where the delay of each project is measured as the difference between completion time in the actual schedule and completion time in the resource-unconstrained case. Kurtulus [117] extend this approach by defining several functions that assign different delay penalties to the projects. Browning and Yassine [27] follow this definition of delays and discuss objectives including average delay per project, average percentage delay per project (measured as percentage of the critical path length) as well as percentage delay of the “super-network.”

6.3 Project Selection and Scheduling

Chen and Askin [38] develop a multi-project problem which contains two types of decisions: First, a set of candidate projects is given, and the projects to be carried out are selected (this allows to decide, e.g., which product development projects should be executed). Second, the selected projects are scheduled subject to the usual precedence and renewable resource constraints. The project selection and scheduling decisions are made simultaneously with the objective to maximize the net present value.

Kolisch and Meyer [110] integrate project selection and scheduling into a problem for pharmaceutical research projects. The selection of projects is captured by modes reflecting that an activity is not executed, while the mode identity concept ensures that either all activities of a project are carried out or none. The problem further includes time-dependent resource requests and an objective that is based on the net present value.

6.4 Further Specific Multi-Project Problems

Heimerl and Kolisch [84] propose a multi-project problem where the schedule of each project is fixed. Consequently, each project can be viewed as an activity with time-dependent resource requirements. The resources are the internal and external workforce where different skills and efficiencies are taken into account. The objective is to minimize costs for regular and overtime work as well as for external resources.

Kolisch [106] presents a problem for scheduling multiple assembly projects. It includes due dates for the individual projects and tardiness minimization. Only a limited number of individual projects can be in progress at the same time because space in the assembly area is limited. Moreover, an activity can only start when all parts required for the assembly are available (note that this is similar to the concept of cumulative resources, see Section 4.3).

Shtub et al. [164] develop multi-project problems which assume that a project is repeated several times, as it may occur in small batch production. Their problem settings consider learning effects that can be exploited if a (human) resource performs several identical projects, that is, the time needed for one project may decrease.

Tukel and Wasti [174] discuss problems for product development projects which include outsourcing of sub-projects to suppliers. The “contractual approach” assumes that the suppliers schedule their sub-projects in the first step. The second step is the scheduling of the main project, where each sub-project of a supplier is treated as a single activity with the duration corresponding to the sub-project’s makespan. The “partnership approach” is a multi-project problem setting that integrates the sub-project’s activities into the overall project network.

Vanhoucke [180] considers a project which is repeated several times. The resulting projects are represented in a super-network as described above, and the super-sink is associated with a deadline. An activity requires the same resource in each project (which is referred to as work
continuity), and the objective is to minimize the resource idle time.

7 Test Instance Generators

When testing exact or heuristic methods for project scheduling, test instances are necessary. In recent years, several parameter-driven instance generators for the RCPSP and many of its extensions have been developed. This section gives a brief overview of the problem variants that have been covered (a discussion of the control parameters is beyond the scope of this work).

Kolisch et al. [113] develop a generator called ProGen for the classical RCPSP as well as the multi-mode extension. It has been applied to generate several sets with test instances for these problem classes. These sets which are comprised in the internet-based library PSPLIB (see Kolisch and Sprecher [112] and Kolisch et al. [114]) have been used in a huge number of studies. ProGen has been extended to cover further extensions of the RCPSP. Schwindt [161] develops a version called ProGen/max in order to include minimal and maximal time lags. This generator can also produce activities with multiple modes as well as instances for the resource leveling and the resource investment problem. DrexI et al. [64] introduce a version called ProGen/πx which incorporates partially renewable resources, mode identity, mode-dependent and sequence-dependent setup times (called changeover times here), mode-dependent minimum time lags, and forbidden periods for activities.

Agrawal et al. [3] suggest a generator called DAGEN for activity-on-arc project networks. Demeulemeester et al. [54] develop RanGen, a generator for single-mode and multi-mode project instances which is based on different control parameters than ProGen. Vanhoucke et al. [186] provide RanGen2, which extends RanGen by incorporating further topological network measures. Browning and Yassine [28] propose a generator for problems consisting of multiple projects where the activities are associated with due dates.

8 Conclusions

We have summarized and classified publications on various variants and generalizations of the well-known resource-constrained project scheduling problem (RCPSP). Among the most popular extensions are multiple modes, generalized time lags, and objectives based on the net present value. Beyond these well-researched areas, many other concepts have been developed in recent years. This paper can serve as a guide through these developments.

Several recent problem settings have been motivated by specific industries. This includes production and engineering where Bartels and Zimmermann [15] consider the automotive industry, Schwindt and Trautmann [162] deal with the process industry, Kolisch [106] discusses the production of palletizing systems, and Voß and Witt [189] consider a steel manufacturer. Specific requirements of research projects have been discussed for medical research (Hartmann [80]), pharmaceutical research (Kolisch and Meyer [110]), and bio-technology (Vanhoucke [179]). Further examples include training in the telecommunication industry (Tiwari et al. [173]) and aviation industry (Haase et al. [76]), arrivals of ships in ports (Lorenzoni et al. [122]), IT projects (Heimerl and Kolisch [84]) and movie projects (Bomsdorf and Derigs [23]). Also note that some of the above papers consider ongoing processes rather than projects in the strict sense (i.e., a set of activities which are carried out to achieve a predefined goal). Apparently, project scheduling concepts are powerful enough to capture such problems as well.
The field of project scheduling has also attracted researchers who examined the underlying mathematical structures and their relationships to other optimization problems. For example, Sprecher [168] shows that the flow shop, open shop, job shop, and assembly line balancing problems are special cases of project scheduling problems. Hartmann [78] outlines how bin packing, strip packing, and knapsack problems can be modeled as project scheduling problems. Stadtler [169] compares the multi-level capacitated lotsizing problem and the RCPSP and presents an integrated model. Brucker and Knust [30] and Drexl and Salewski [62] show how school and university timetabling problems can be captured as project scheduling problems.

These studies show that the use of the RCPSP with all its extensions is not limited to applications in its original field, the scheduling of projects. The RCPSP and its generalizations are also perceived as a collection of powerful tools that allow to describe many highly complex optimization problems from other areas. If some problem is captured as a resource-constrained project scheduling problem, it is possible to employ the solution methods that have been developed for the project scheduling problem. Hence, algorithms originally designed for project scheduling can be transferred to problems from other fields. We believe that this underscores the importance of research on models and methods for project scheduling.

References


