PRODUCTION AND OPERATIONS MANAGEMENT
POMS
Vol. 13, No. 1, Spring 2004, pp. 77–92
ISSN 1059-1478 | 04 | 1301 | 077$1.25

Planning and Scheduling in Supply Chains: An Overview of Issues in Practice

Stephan Kreipl • Michael Pinedo
SAP Germany AG & Co.KG, Neurottstrasse 15a, 69190 Walldorf, Germany
Stern School of Business, New York University, 40 West Fourth Street, New York, New York 10012

This paper gives an overview of the theory and practice of planning and scheduling in supply chains. It first gives an overview of the various planning and scheduling models that have been studied in the literature, including lot sizing models and machine scheduling models. It subsequently categorizes the various industrial sectors in which planning and scheduling in the supply chains are important; these industries include continuous manufacturing as well as discrete manufacturing. We then describe how planning and scheduling models can be used in the design and the development of decision support systems for planning and scheduling in supply chains and discuss in detail the implementation of such a system at the Carlsberg A/S beerbrewer in Denmark. We conclude with a discussion on the current trends in the design and the implementation of planning and scheduling systems in practice.

Key words: planning; scheduling; supply chain management; enterprise resource planning (ERP) systems; multi-echelon inventory control
Submissions and Acceptance: Received October 2002; revisions received April 2003; accepted July 2003.

1. Introduction

This paper focuses on models and solution approaches for planning and scheduling in supply chains. It describes several classes of planning and scheduling models that are currently being used in systems that optimize supply chains. It discusses the architecture of decision support systems that have been implemented in industry and the problems that have come up in the implementation and integration of systems in supply chains. In the implementations considered, the total cost in the supply chain has to be minimized, i.e., the stages in the supply chain do not compete in any form with one another, but collaborate in order to minimize total costs. This paper focuses primarily on how to integrate medium term planning models (e.g., lot sizing models) and detailed scheduling models (e.g., job shop scheduling models) into a single framework.

A medium term production planning model typically optimizes several consecutive stages in a supply chain (i.e., a multi-echelon model), with each stage having one or more facilities. Such a model is designed to allocate the production of the different products to the various facilities in each time period, while taking into account inventory holding costs and transportation costs. A planning model may make a distinction between different product families, but usually does not make a distinction between different products within a family. It may determine the optimal run length (or, equivalently, batch size or lot size) of a given product family when a decision has been made to produce such a family at a given facility. If there are multiple families produced at the same facility, then there may be setup costs and setup times. The optimal run length of a product family is a function of the trade-off between the setup cost and/or setup time and the inventory carrying cost. The main objectives in medium term planning involve inventory carrying costs, transportation costs, tardiness costs, and the major setup costs. However, in a medium term planning model, it is typically not customary to take the sequence dependency of setup times and setup costs into account. The sequence dependency of setups is difficult to incorporate in an integer programming formulation and can increase the complexity of the formulation significantly.

A short term detailed scheduling model is typically only concerned with a single facility, or, at most, with a single stage. Such a model usually takes more detailed information into account than a planning model. It is typically assumed that there are a given number of jobs and each one has its own parameters (including sequence-dependent setup times and sequence-dependent setup costs). The jobs have to be scheduled in such a way that one or more objectives are minimized, e.g., the number of jobs that are shipped late, the total setup time, and so on.

Clearly, planning models differ from scheduling models in a number of ways. First, planning models often cover multiple stages and optimize over a medium term horizon, whereas scheduling models are usually designed for a single stage (or facility) and optimize over a short term horizon. Second, planning models use more aggregate information, whereas scheduling models use more detailed information. Third, the objective to be minimized in a planning model is typically a total cost objective and the unit in which this is measured is a monetary unit; the objective to be minimized in a scheduling model is typically a function of the completion times of the jobs and the unit in which this is measured is often a time unit. Nevertheless, even though there are fundamental differences between these two types of models, they often have to be incorporated into a single framework, share information, and interact extensively with one another.

Planning and scheduling models may also interact with other types of models, such as long term strategic models, facility location models, demand management models, and forecasting models; these models are not discussed in this paper. The interactions with these other types of models tend to be less intensive and less interactive. In what follows, we assume that the physical settings in the supply chain have already been established; the configuration of the chain is given, and the number of facilities at each stage is known.

Supply chains in the various industries are often not very similar and may actually give rise to different sets of issues and problems. This paper considers applications of planning and scheduling models in supply chains in various industry sectors. A distinction is made between two types of industries, namely the continuous manufacturing industries (which include the process industries) and the discrete manufacturing industries (which include, for example, automotive and consumer electronics). Each one of these two main categories is subdivided into several subcategories. This categorization is used because of the fact that the planning and scheduling procedures in the two main categories tend to be different. We focus on the frameworks in which the planning and scheduling models have to be embedded; we describe the type of information that has to be transferred back and forth between the modules and the kinds of optimization that is done within the modules.

There is an extensive literature on supply chain management. Many papers and books focus on supply chain coordination; a significant amount of this work has an emphasis on inventory control, pricing issues, and the value of information; see Simchi-Levi, Kaminsky, and Simchi-Levi (2000), Chopra and Meindl (2001), and Stadtler and Kilger (2000). There is also an extensive literature on production planning and scheduling theory. A significant amount of research has been done on the solution methods applicable to planning and scheduling models; see Shapiro (2001). Planning models and scheduling models have often been studied independently from one another in order to obtain elegant theoretical results. Planning models are often based on (multi-echelon) inventory theory and lot sizing; see Zipkin (2000), Kimms (1997), Drexl and Kimms (1997), Muckstadt and Roundy (1993), and Dobson (1987, 1992). Scheduling models typically focus on how to schedule a number of jobs in a given machine environment in order to minimize some objective. For general treatises on scheduling, see Bhaskaran and Pinedo (1992), Brucker (1998), Pinedo (2002), and Pinedo and Chao (1999). For applications of scheduling to supply chain management, see Hall and Potts (2000) and Lourenco (2001). Some research has been done on more integrated models in the form of hierarchical planning systems; this research has resulted in frameworks that incorporate planning and scheduling; see Bowersox and Closs (1996), Barbarosoglu and Ozgur (1999), Dhaenens-Flipo and Finke (2001), Shapiro (2001), and Miller (2002). For examples of descriptions of successful industrial implementations, see Haq (1991), Arntzen, Brown, Harrison, and Trafton (1995), Hadavi (1998), and Shepherd and Lapide (1998).

This paper is organized as follows. The second section describes and categorizes some of the typical industrial settings. The third section discusses the overall frameworks in which planning models as well as scheduling models have to be embedded. The fourth section describes a standard mixed integer programming formulation of a planning model for a supply chain. The fifth section covers a typical formulation of a scheduling problem in a facility in a supply chain. The sixth section describes an actual implementation of a planning and scheduling software system at the Danish beerbrewer Carlsberg A/S. The last section presents the conclusions and discusses the impact of the Internet on decision support systems in supply chains.

2. Supply Chain Settings and Configurations

This section gives a concise overview of the various types of supply chains. It describes the differences in the characteristics and the parameters of the various categories. It first describes the various different industry groups and their supply chain characteristics and then discusses how the different planning and scheduling models analyzed in the literature can be used in the management of these chains. One can make a distinction between two types of manufacturing industries, namely:

(I) Continuous manufacturing industries (e.g., the process industries),

(II) Discrete manufacturing industries (e.g., cars, semiconductors).

These two industry sectors are not all-encompassing; the borderlines are somewhat blurry and may overlap. However, planning and scheduling in continuous manufacturing (the process industries) often have to deal with issues that are quite different from those in discrete manufacturing.

Continuous Manufacturing. Continuous manufacturing (process) industries often have various types of different operations. The most common types of operations can be categorized as follows:

(I-a) Main processing operations,

(I-b) Finishing or converting operations.

Main Processing Operations in Continuous Manufacturing (I-a). The main production facilities in the process industries are, for example, paper mills, steel mills, aluminum mills, chemical plants, and refineries. In paper, steel, and aluminum mills, the machines take in the raw material (e.g., wood, iron ore, alumina) and produce rolls of paper, steel, or aluminum, which afterwards are handled and transported with specialized material-handling equipment. Machines that do the main processing operations typically have very high startup and shutdown costs and usually work around the clock. A machine in the process industries also incurs a high changeover cost when it has to switch over from one product to another. Various methodologies can be used for analyzing and solving the models for such operations, including cyclic scheduling procedures and Mixed Integer Programming approaches.

Finishing Operations in Continuous Manufacturing (I-b). Many process industries have some form of finishing operations that do some converting of the output of the main production facilities. This converting usually involves cutting of the material, bending, folding, and possibly painting or printing. These operations often (but not always) produce commodity-type items, for which the producer has many clients. For example, a finishing operation in the paper industry may produce cut size paper out of the rolls that come from the paper mill. The paper finishing business is often a mixture of Make-To-Stock (MTS) and Make-To-Order (MTO). If it operates according to MTO, then the scheduling is based on customer due dates and sequence-dependent setup times. This leads often to single machine and parallel machine scheduling models. If it operates according to MTS, then it may follow a so-called *s-S* or *Q-R* inventory control policy. If it is a mixture of MTO and MTS, then the scheduling policies become a mixture of inventory control and detailed scheduling rules.

Discrete Manufacturing. The discrete manufacturing industry sector is quite diverse and includes the automotive industry, the appliances industry, and the PC industry. From the perspective of planning and scheduling, a distinction can be made between three different types of operations in this sector. The reason for making such a distinction is based on the fact that planning and scheduling in these three segments are quite different.

(II-a) Primary converting operations (e.g., cutting and shaping of sheet metal),

(II-b) Main production operations (e.g., production of engines, PCBs, wafers), and

(II-c) Assembly operations (e.g., cars, PCs).

Primary Converting Operations in Discrete Manufacturing (II-a). Primary converting operations are somewhat similar to the finishing operations in the process industries. These operations may typically include stamping, cutting, or bending. The output of this operation is often a particular part that is cut and bent into a given shape. There are usually few operations done on such an item, and the routing in such a facility is relatively simple. The final product of a primary converting facility is usually not a finished good, but basically a part or piece made of a single material (boxes, containers, frames, stamped body parts of cars, and so on). Examples of the types of operations in this category are stamping plants that produce body parts for cars, and plants that produce epoxy boards of various sizes for the facilities that produce Printed Circuit Boards. The planning and scheduling procedures under *II-a* may be similar to those under *I-b*. However, they may be here more integrated with the operations downstream.

Main Production Operations in Discrete Manufacturing (II-b). The main production operations are those operations that require multiple different operations by different machine tools, and the product (as well as its parts) may have to follow a certain route through the facility going through various work centers. Capital investments have to be made in various types of machine tools (lathes, mills, chip fabrication equipment). For example, in the semiconductor industry, wafers typically have to undergo hundreds of steps. These operations include oxidation, deposition, and metallization, lithography, etching, ion implantation, photoresist stripping, and inspection and measurements. It is often the case that certain operations have to be performed repeatedly and that certain orders have to visit certain workcenters in the facility several times, i.e., they have to recirculate through the facility. In semiconductor and Printed Circuit Board manufacturing, the operations are often organized in a job shop fashion. Each order has its own route through the system, its own quantity (and processing times), and its own committed shipping date. An order typically represents a batch of identical items that requires sequence-dependent setup times at many operations.

Assembly Operations in Discrete Manufacturing (II-c). The main purpose of an assembly facility is to put different parts together. An assembly facility typically does not alter the shape or form of any one of the individual parts (with the possible exception of the painting of the parts). Assembly operations usually do not require major investments in machine tools, but do require investments in material handling systems (and possibly robotic assembly equipment). An assembly operation may be organized in workcells, in assembly lines, or according to a mixture of workcells and assembly lines. For example, PCs are assembled in workcells, whereas cars and TVs are typically put together in assembly lines. Workcells typically do not require any sequencing, but they may be subject to learning curves. In assembly operations that are set up in a line, the sequencing is based on grouping and spacing heuristics combined with committed shipping dates. The schedules that are generated by the grouping and spacing heuristics typically affect not only the throughput of the line, but also the quality of the items produced.

Supply chains in both continuous and discrete manufacturing may have, besides the stages described above, additional stages. In a supply chain in a process industry, there may be a stage preceding Stage I-a in which the raw material is being gathered at its point of origination (which may be a forest or a mine) and taken to the main processing operations. There may also be a distribution stage following stage I-b. A company may have its own distribution centers in different geographical locations, where it keeps certain SKUs in stock for immediate delivery. The company may also ship directly from its manufacturing operations to customers. A supply chain in a discrete manufacturing industry also may have other types of stages. There may be a stage preceding stage II-a in which raw material is being collected at a supplier (which may be an operation of the type I-b) and brought to a primary converting operation. There may also be a stage following stage II-c which would consist of distribution operations (e.g., dealerships).

Supply chains in both continuous and discrete manufacturing may have several facilities at each one of the stages, each one feeding into several facilities at stages downstream. The configuration of an entire chain may be quite complicated: For example, there may be assembly operations that produce subassemblies that have to be fed into a production operation.

Table 1

Sector	Processes	Time horizon	Clock-speed	Product differentiation
(I-a)	planning	long-medium	low	very low
(I-b)	planning/scheduling	medium/short	medium/high	medium/low
(II-a)	planning/scheduling	medium/short	medium	very low
(II-b)	planning/scheduling	medium/short	medium	medium/low
(II-c)	scheduling	short	high	high

There are some basic differences between the parameters and operating characteristics of the facilities in the two main categories described above. Several of these differences have an impact on the planning and scheduling processes, including the differences in (i) the planning horizon, (ii) the clock-speed, and (iii) the level of product differentiation.

(i) The planning horizon in continuous manufacturing facilities tends to be longer than the planning horizon in the discrete manufacturing facilities. In continuous as well as in discrete manufacturing the planning horizons tend to be shorter more downstream in the supply chain.

(ii) The so-called "clock-speed" tends to be higher in a discrete manufacturing facility than in a continuous manufacturing facility. A high clock-speed implies that existing plans and schedules often have to be changed or adjusted; that is, planning and scheduling is more reactive. In continuous as well as in discrete manufacturing, the clock-speed increases the more downstream in the supply chain.

(iii) In discrete manufacturing, there may be a significant amount of mass customization and product differentiation. In continuous manufacturing, mass-customization does not play a very important role. The number of SKUs in discrete manufacturing tends to be significantly larger than the number of SKUs in continuous manufacturing. The number of SKUs tends to increase more downstream in the supply chain.

These operating characteristics are summarized in Table 1. Because of these differences, the planning and scheduling issues in each one of the sectors can be very different. Table 2 presents a summary of the model types that can be used in the different categories as well as the corresponding solution techniques.

Note that problems that have continuous variables may lead to Mixed Integer Programming (MIP) formulations, whereas problems that have only discrete variables may lead to pure Integer Programming (IP) formulations (or Disjunctive Programming formulations). However, a discrete problem in which certain variables assume large values (i.e., the number of units to be produced) may be replaced by a continuous problem, resulting in a Mixed Integer Programming formulation rather than a pure Integer Programming

Table 2

Sector	Models	Solution techniques
(I-a)	Lot sizing models (multi-stage); cyclic scheduling models	Mixed Integer Programming formulations
(I-b)	Single machine scheduling models; parallel machine scheduling models	Batch scheduling; mixtures of inventory control rules and dispatching rules
(II-a)	Single machine scheduling models; Parallel machine scheduling models	Batch scheduling and dispatching rules
(II-b)	Flow Shop and Job Shop Scheduling Models with specific routing patterns	Integer Programming formulations; shifting bottleneck heuristics; dispatching rules
(II-c)	Assembly Line Models; Workcell Models	Grouping and Spacing Heuristics; Make-to-Order/Just-In-Time

formulation. Planning models typically result in Mixed Integer Programming formulations with a mix of continuous and discrete variables. Scheduling models usually do not have any continuous variables; they may have continuous variables when preemptions and job splitting are allowed. When there are few discrete variables, it makes a lot of sense to solve the Linear Programming relaxation of the MIP. The solution may provide a useful lower bound and may give indications regarding the structure of the optimal solutions of the MIP. If the formulation of the problem is a pure Integer Program (which is often the case with a scheduling problem), then solving the linear relaxation typically does not provide a significant amount of benefit.

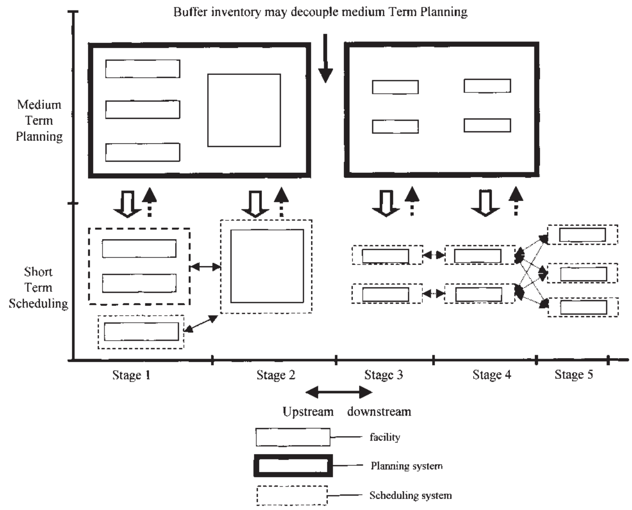
Examples of applications of planning and scheduling in continuous manufacturing can be found in Haq (1991), Murthy et al. (2001), and Rachlin et al. (2001). Examples of planning and scheduling in discrete manufacturing are described in Arntzen, Brown, Harrison, and Trafton (1995), De Bontridder (2001), and Vandaele and Lambrecht (2001).

In the following four sections, we discuss frameworks for planning and scheduling in supply chains, we present some examples of planning and scheduling models that have formed a basis for several systems that have been implemented in practice, and we describe an actual implementation. These four sections have been inspired primarily by the design of systems developed and implemented by SAP Germany AG; see Braun (2001), Braun and Groenewald (2000), and Strobel (2001).

3. Frameworks for Planning and Scheduling in Supply Chains

The main objective in a supply chain or production distribution network is to produce and deliver finished products to end consumers in the most cost-

Figure 1 Planning and Scheduling in Supply Chains



effective and timely manner. Of course, this overall objective forces each one of the individual stages to formulate its own objectives.

Since planning and scheduling in a global supply chain requires the coordination of operations in all stages of the chain, the models and solution techniques described in the previous section have to be integrated within a single framework. Different models that represent successive stages have to exchange information and interact with one another in various ways. A continuous model for one stage may have to interact with a discrete model for the next stage.

Planning and scheduling procedures in a supply chain are typically used in various phases: a first phase involves a multi-stage medium term planning process (using aggregate data), and a subsequent phase performs a short term detailed scheduling at each one of those stages separately. Typically, whenever a planning procedure has been applied and the results have become available, each facility can apply its scheduling procedures. However, scheduling procedures are usually applied more frequently than planning procedures. Each facility in every one of these stages has its own detailed scheduling issues to deal with; see Figure 1.

If successive stages in a supply chain belong to the same company, then it is usually the case that these stages are incorporated into a single planning model. The medium term planning process attempts to minimize the total cost over all the stages. The costs that have to be minimized in this optimization process include production costs, storage costs, transportation costs, tardiness costs, non-delivery costs, handling costs, costs for increases in resource capacities (e.g., scheduling third shifts), and costs for increases in storage capacities.

In this medium term optimization process, many input data are only considered in an aggregate form. For example, time is often measured in weeks or months rather than days. Distinctions are usually only made between major product families, and no distinctions are made between different products within one family. A setup cost may be taken into account, but it may only be considered as a function of the product itself and not as a function of the sequence.

The results of this optimization process are daily or weekly production quantities for all product families at each location or facility as well as the amounts scheduled for transport every week between the locations. The production of the orders requires a certain amount of the capacities of the resources at the various facilities, but no detailed scheduling takes place in the medium term optimization. The output consists of the allocations of resources to the various product families, the assignment of products to the various facilities in each time period, and the inventory levels of the finished goods at the various locations. As stated before, in this phase of the optimization process, a distinction may be made between different product families, but not between different products within the same family. The model is typically formulated as a Mixed Integer Program. Variables that represent quantities that have to be produced are often continuous variables. The integer (discrete) variables are often 0-1 variables; they are, for example, needed in the formulation when a decision has to be made whether or not a particular product family will be produced at a certain facility during a given time period.

The output of the medium term planning process is an input to the detailed (short term) scheduling process. The detailed scheduling problems typically attempt to optimize each stage and each facility separately. So, in the scheduling phase of the optimization process, the process is partitioned according to:

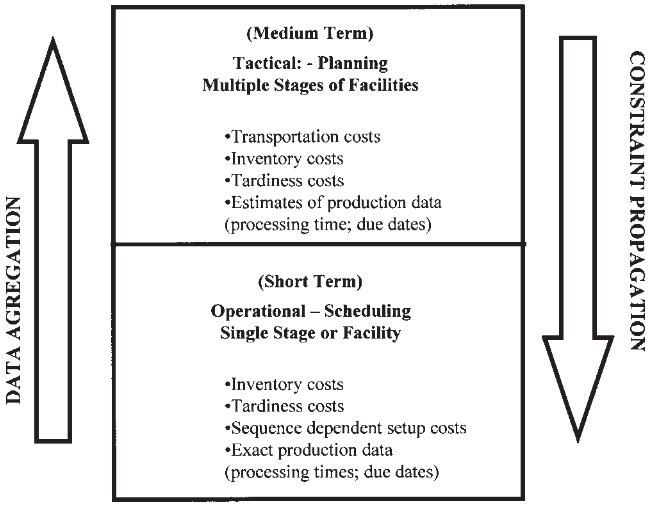
(i) the different stages and facilities, and
(ii) the different time periods.

So, in each detailed scheduling problem the scope is considerably narrower (with regard to time as well as space), but the level of detail taken into consideration is considerably higher; see Figure 2. This level of detail is increased in the following dimensions:

(i) the time is measured in a smaller unit (e.g., days or hours); the process may be even time continuous,
(ii) the horizon is shorter,
(iii) the product demand is more precisely defined, and
(iv) the facility is not a single entity, but a collection of resources or machines.

The product demand now does not consist, as in the medium term planning process, of aggregate demands for entire product families. In the detailed scheduling process, the demand for each individual product within a family is taken into account. The minor setup times and setup costs in between different products from the same family are taken into account as well as the sequence dependency.

Figure 2 Data Aggregation and Constraint Propagation



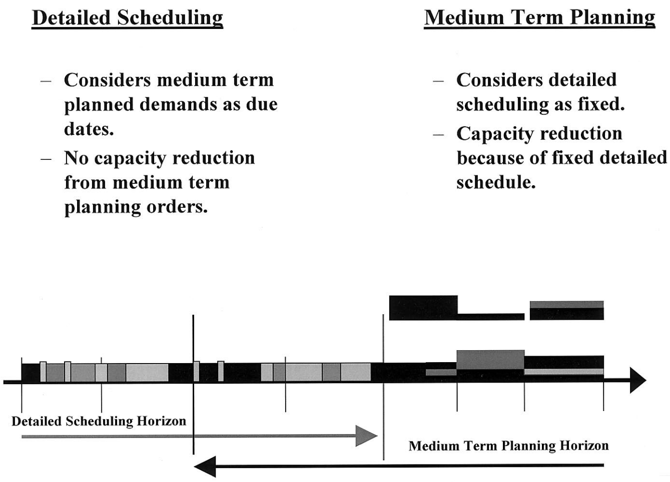
The factory is now not a single entity; each product has to undergo a number of operations on different machines. Each product has a given route and given processing requirements on the various machines. The detailed scheduling problem can be analyzed as a job shop problem and various techniques can be used, including:

(i) dispatching rules,
(ii) shifting bottleneck techniques,
(iii) local search procedures (e.g., genetic algorithms), or
(iv) integer programming techniques.

The objective takes into account the individual due dates of the orders, sequence-dependent setup times, sequence-dependent setup costs, lead times, as well as the costs of the resources. However, if two successive facilities (or stages) are tightly coupled with one another (i.e., the two facilities operate according to the JIT principle), then the short term scheduling process may optimize the two facilities jointly. It actually may consider them as a single facility with the transportation in between the two facilities as another operation.

The interaction between a planning module and a scheduling module may be intricate. A scheduling module may cover only a relatively short horizon (e.g., one month), whereas the planning module may cover a longer horizon (e.g., six months). After the schedule has been fixed for the first month (fixing the schedule for this month required some input from the planning module), the planning module does not consider this first month any more; it assumes the schedule for the first month to be fixed.

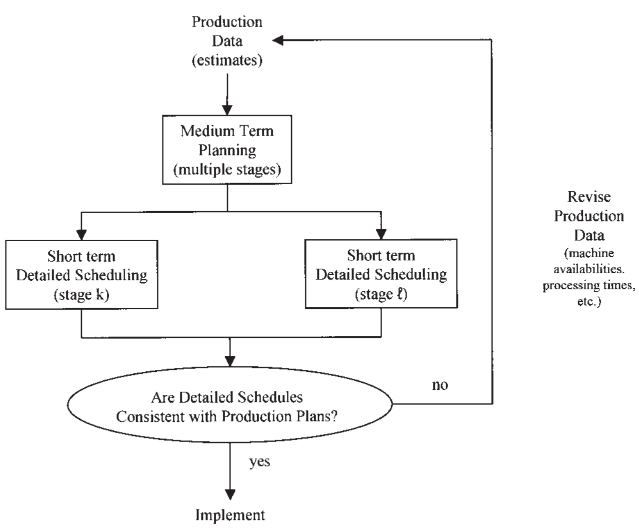
Figure 3 Scheduling and Planning Horizons



However, the planning module still tries to optimize the second up to the sixth month. Doing so, it considers the output of the scheduling module as a boundary condition. However, it also may be the case that the time periods covered by the detailed scheduling process and the medium term planning process overlap; see Figure 3.

A planning and scheduling framework for a supply chain typically must have a mechanism that allows feedback from a scheduling module to the planning module; see Figure 4. This feedback mechanism enables the optimization process to go through several iterations. It may be used under various circumstances: First, the results of the detailed short term optimization process may indicate that the estimates used as input data for the medium term planning process were not accurate. (The average production times in the planning processes do not take the sequence dependency of the setup times into account; setup times are estimated and embedded in the total production times. The total setup times in the detailed schedule may actually be higher than the setup times anticipated in the planning procedure.) If the results of the detailed scheduling process indicate that the input to the planning process has to be modified, then new input data for the planning process have to be generated and the planning process have to be redone.

Figure 4 Information Flows Between Planning and Scheduling Systems



Second, there may be an exogenous reason necessitating a feedback from the detailed scheduling process to the medium term planning process. A major disruption may occur on the factory floor level, e.g., an important machine goes down for an extended period of time. A disruption may be of such a magnitude that its effects cannot be contained within the facility where it occurs. The entire planning process may be affected and therefore the scheduling processes at other facilities as well. So a framework with a feedback mechanism may allow the overall optimization process to iterate (see Figure 4).

The individual modules within the planning and scheduling framework for a given chain may have other interesting features. Two types of features that are often incorporated are decomposition features and so-called discretization features. Each feature can be activated and deactivated by the user of the system.

Decomposition is often used when the optimization problem is simply too large to be dealt with effectively by the routines available. A decomposition process partitions the overall problem in a number of subproblems and solves the (smaller) subproblems separately. At the end of the process, the partial solutions are put together in a single overall solution. Decomposition can be done according to:

(i) time;
(ii) available resources (facilities or machines);
(iii) product families; and
(iv) geographical areas.

Some of the decompositions may be designed in such a way that they are activated automatically by the system itself, and other decompositions may be designed in such a way that they have to be activated by the user of the system. Decomposition is used in medium term modules as well as in detailed scheduling modules. In medium term planning, the decomposition is often based on time and/or on product family (these may be internal decompositions activated by the system itself). The user may specify in a medium term planning process a geographical decomposition. In the detailed scheduling process, the decomposition is often machine-based (such a decomposition may be done internally by the system or imposed by the user).

One type of discretization feature may be used when the continuous version of a problem (for example, a linear programming relaxation of a more realistic integer programming formulation) does not yield sufficiently accurate results. To obtain more accurate results, certain constraints may have to be imposed on given variables. For example, production quantities are often not allowed to assume just any values, but only values that are multiples of given fixed amounts or lot sizes (e.g., the capacity of a tank in the brewing of beer). The quantities that have to be transported between two facilities also have to be multiples of a fixed amount (e.g., the size of a container). This type of discretization may transform the problem from a continuous problem (i.e., a linear program) to a discrete optimization problem.

Another type of discretization can be done with respect to time. It allows the user of the system to determine the size of the time unit. If the user is only interested in a rough plan, he may set the time unit to be equal to a week. That is, the results of the optimization then only specify what is going to be produced that week, but will not specify what is going to be produced within each day of that week. If the user sets the time unit equal to one day, the result will be significantly more precise. Besides specifying the sizes of the time units, a system may use time units of different sizes for different periods. The discretization feature is often implemented in the medium term planning modules. The first week of a three-month planning period may be specified on a daily basis, the next three weeks may be determined on a weekly basis, and all activities beyond the first month are planned based on a continuous model. Discretization with respect to time does *not* change the nature of the problem; if the problem is a linear program, then it will remain a linear program.

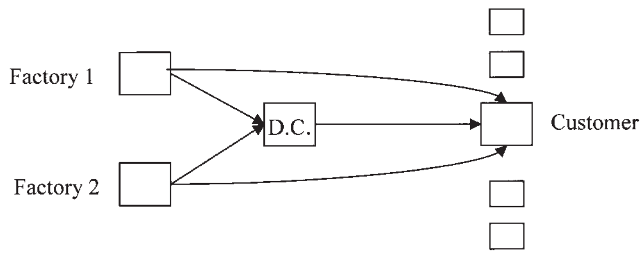
The APO system of SAP Germany enables the modeler to activate and deactivate the discretization of various types of constraints in order to improve the performance of the optimization process. For example, discretization may be used for daily and weekly time buckets, but not for monthly time buckets in which Linear Programming is used without discretization.

4. Medium Term Planning Models for Supply Chains

This section considers a standard medium term planning model for a supply chain. It does not present the model in its full generality; the notation needed for a more general model is simply too cumbersome. A description is given of a model with many of the relevant parameters having fixed values (in order to simplify the notation). It also does not incorporate all of the features described in the previous section (e.g., all the time units are of the same size).

Consider three stages in series. The first and most upstream stage (Stage 1) has two factories in parallel. They both feed Stage 2, which is a distribution center (DC). Both Stages 1 and 2 can deliver to a customer, which is a part of Stage 3; see Figure 5. The factories have no room for finished goods storage and the customer does not want to receive any early deliveries.

Figure 5 **A System with Three Stages**



The problem has the following parameters and input data. The two factories work around the clock; so their available weekly production capacity is $24 \times 7 = 168$ hours. There are two major product families, F_1 and F_2. As stated before, in the medium term planning process, all the products within a family are considered identical. The demand forecasts for the next four weeks are known (the unit of time being one week). In this section, the subscripts and superscripts have the following meaning.

The subscript i ($i = 1, \ldots, 4$), refers to time period i.
The subscript j ($j = 1, 2$), refers to product family j.
The subscript k ($k = 1, 2$), refers to factory k.
The subscript l ($l = 1, 2, 3$) refers to stage l;
- $l = 1$ refers to the two factories,
- $l = 2$ refers to the distribution center, and
- $l = 3$ refers to the customer.

The superscript p refers to a production parameter.
The superscript s refers to a storage parameter.
The superscript τ refers to a transportation parameter.

The demand for product family j, $j = 1, 2$, at the DC level (stage 2) by the end of week i, $i = 1, \ldots, 4$, is denoted by D_{ij2}. The demand for product family j, $j = 1, 2$, at the customer level (stage 3) by the end of week i, $i = 1, \ldots, 4$, is denoted by D_{ij3}. Production times and costs are given:

c^p_{jk} = the cost to produce 1 unit of family j in factory k.
t^p_{jk} = the time (in hours) to produce 1000 units of family j in factory k.

The t^p_{jk} is the reciprocal of the rate of production. Storage costs and transportation data include:

c^s_2 = the storage cost for one unit of any type in the DC per week.
$c^{\tau}_{k2\circ}$ = the transportation cost for a unit of any type from factory k to the DC.
$c^{\tau}_{k\circ 3}$ = the transportation cost for a unit of any type from factory k to the customer.

$c^{\tau}_{\circ 23}$ = the transportation cost for a unit of any type from the DC to the customer.

t^{τ} = the transportation time from any one of the two factories to the DC, from any one of the two factories to the customer, and from the DC to the customer; all transportation times are assumed to be identical and equal to one week.

The following weights and penalty costs are given:

w''_j = the tardiness cost per unit per week for an order of family j products that arrive late at the DC.

w'''_j = the tardiness cost per unit per week for an order of family j products that arrive late at the customer.

π = the penalty for never delivering one unit of product.

The objective is to minimize the total of the production costs, storage costs, transportation costs, tardiness costs, and penalty costs for non-delivery over a horizon of four weeks. In order to formulate this problem as a Mixed Integer Program, the following decision variables have to be defined:

x_{ijk} = number of units of family j produced at plant k during period i.

y_{ijk2} = number of units of family j transported from plant k to the DC in week i.

y_{ijk3} = number of units of family j transported from plant k to customer in week i.

z_{ij} = number of units of family j transported from the DC to the customer in week i.

q_{0j2} = number of units of family j in storage at the DC at time 0.

q_{ij2} = number of units of family j in storage at the DC in week i.

v_{ij2} = number of units of family j that are tardy (have not yet arrived) at the DC in week i.

v_{4j2} = number of units of family j that have not been delivered to the DC by the end of the planning horizon (the end of week 4).

v_{0j3} = the number of units of family j that are tardy (have not yet arrived) at the customer at time 0.

v_{ij3} = number of units of family j that are tardy (have not yet arrived) at the customer in week i.

v_{4j3} = the number of units of family j that have not been delivered to the customer by the end of the planning horizon (the end of week 4).

There are various constraints in the form of upper bounds UB_{jkl} and lower bounds LB_{jkl} on the quantities of family j to be shipped from plant k to stage l. The integer program can now be formulated as follows:

minimize

$$\sum_{i=1}^{4}\sum_{j=1}^{2}\sum_{k=1}^{2} c^{p}_{jk}x_{ijk} + \sum_{i=1}^{4}\sum_{j=1}^{2}\sum_{k=1}^{2} c^{\tau}_{k2\circ}y_{ijk2} + \sum_{i=1}^{4}\sum_{j=1}^{2}\sum_{k=1}^{2} c^{\tau}_{k\circ 3}y_{ijk3}$$

$$+ \sum_{i=1}^{4}\sum_{j=1}^{2} c^{\tau}_{\circ 23}z_{ij} + \sum_{i=1}^{4}\sum_{j=1}^{2} c^{s}_{2}q_{ij2} + \sum_{i=1}^{3}\sum_{j=1}^{2} w''_j v_{ij2}$$

$$+ \sum_{i=1}^{3}\sum_{j=1}^{2} w'''_j v_{ij3} + \sum_{j=1}^{2} \pi v_{4j2} + \sum_{j=1}^{2} \pi v_{4j3}$$

subject to the following weekly production capacity constraints:

$$\sum_{j=1}^{2} t^{p}_{j1}x_{ij1} \le 168 \quad i = 1, \ldots, 4;$$

$$\sum_{j=1}^{2} t^{p}_{j2}x_{ij2} \le 168 \quad i = 1, \ldots, 4;$$

subject to the following transportation constraints:

$$y_{ij1l} \le UB_{j1l} \quad i = 1, \ldots, 4;$$

$$y_{ij1l} \ge LB_{j1l} \text{ or } y_{ij1l} = 0 \quad i = 1, \ldots, 4;$$

$$y_{ij2l} \le UB_{j2l} \quad i = 1, \ldots, 4;$$

$$y_{ij2l} \ge LB_{j2l} \text{ or } y_{ij2l} = 0 \quad i = 1, \ldots, 4;$$

$$\sum_{l=2}^{3} y_{ijkl} = x_{ijk} \quad i = 1, \ldots, 4; \quad j = 1, 2, \quad k = 1, 2;$$

$$\sum_{k=1}^{2} y_{ijk3} + z_{ij} \le D_{i+1,j,3} + v_{ij3} \quad i = 1, \ldots, 3; \quad j = 1, 2;$$

$$z_{1j} \le \max(0, q_{0j2}) \quad j = 1, 2;$$

$$z_{ij} \le q_{i-1,j,2} + y_{i-1,j,1,2} + y_{i-1,j,2,2} \quad i = 2, 3, 4; \quad j = 1, 2;$$

subject to the following storage constraints:

$$q_{1j2} = \max(0, q_{0j2} - D_{1j2} - z_{1j}) \quad j = 1, 2;$$

$$q_{ij2} = \max(0, q_{i-1,j,2} + y_{i-1,j,1,2} + y_{i-1,j,2,2} - D_{ij2} - z_{ij} - v_{i-1,j,2}) \quad j=1,2 \quad i=2,3,4;$$

subject to the following constraints regarding number of jobs tardy and number of jobs not delivered:

$$v_{1j2} = \max(0, D_{1j2} - q_{0j2}) \quad j = 1, 2;$$

$$v_{ij2} = \max(0, D_{ij2} + v_{i-1,j,2} + z_{ij} - q_{ij2} - y_{i-1,j,1,2} - y_{i-1,j,2,2}) \quad j = 1, 2; \quad i = 2, 3, 4;$$

$$v_{1j3} = \max(0, D_{1j3}) \quad j = 1, 2;$$

$$v_{ij3} = \max(0, D_{ij3} + v_{i-1,j,3} - z_{i-1,j} - y_{i-1,j,1,3} - y_{i-1,j,2,3}) \quad j = 1, 2; \quad i = 2, 3, 4.$$

Table 3

	Week 1	Week 2	Week 3	Week 4
D_{f12}	20,000	30,000	15,000	40,000
D_{f22}	0	50,000	30,000	50,000
D_{f13}	10,000	5,000	15,000	40,000
D_{f23}	0	10,000	0	5,000

It is clear that most variables in this Mixed Integer Programming formulation are continuous variables. However, the transportation variables y_{ijkl} are subject to disjunctive constraints. In an alternative formulation of the problem, some integer (0–1) variables are needed to ensure that the continuous (transportation) variables y_{ij1l} are either 0 or larger than the lower bound LB_{j1l}. Note that the Linear Programming relaxation of the formulation above (i.e., the formulation without the disjunctive constraints) provides a valid lower bound on the total cost.

The following numerical example illustrates an application of the model described above.

Example 4.1. Consider the following instance of the problem described above. The production times and costs concerning factory 1 are: $t^p_{11} = 1$ hour and $t^p_{21} = 2$ hours; $c^p_{11} = 1$ and $c^p_{21} = 0.50$. The production times and costs concerning factory 2 are: $t^p_{12} = 2$ hours and $t^p_{22} = 3$ hours; $c^p_{12} = 0.50$ and $c^p_{22} = 0.25$.

The storage cost for a unit of any type of product at the DC (c^s_2) is 0.10 per unit per week. The transportation costs are $c^\tau_{12\circ} = 0.10$ per unit; $c^\tau_{22\circ} = 0.30$ per unit; $c^\tau_{k\circ 3} = 0.05$ for $k = 1, 2$, $c^\tau_{\circ 23} = 0.50$ per unit. The forecast demand at the DC and from the customer for the two different product families are presented in Table 3.

From Factory 1 to the DC, there has to be each week at least a shipment of 10,000 units of product family 1 or otherwise nothing, i.e., $LB_{112} = 10{,}000$. From Factory 2 to the DC, there has to be each week at most a shipment of 10,000 units of product family 2, i.e., $UB_{222} = 10{,}000$. The transportation time t^τ is 1 week.

The tardiness cost w''_1 (w''_2) is \$10.00 (\$5.00) per unit per week. The tardiness cost w'''_1 (w'''_2) is \$20.00 (\$15.00) per unit per week. The penalty cost π for not delivering at all is \$1000.00 per unit. The boundary condition v_{0j3} is 0.

Running these data through a Mixed Integer Programming solver yields the production and transportation decisions shown in Tables 4 and 5. The total cost of this solution is \$3,004,950.20.

Table 4

	Week 1	Week 2	Week 3	Week 4
x_{f11}	0	0	0	0
x_{f21}	47,333	20,000	52,333	0
x_{f12}	65,000	30,000	80,000	0
x_{f22}	12,667	10,000	2,667	0

Table 5

	Week 1	Week 2	Week 3
y_{f112}	0	0	0
y_{f113}	0	0	0
y_{f122}	50,000	15,000	40,000
y_{f123}	15,000	15,000	40,000
y_{f212}	47,333	20,000	50,000
y_{f213}	0	0	2,333
y_{f222}	2,667	10,000	0
y_{f223}	10,000	0	2,667

If an additional constraint is added to this problem requiring the production lot sizes to be multiples of 10,000 (such a constraint is fairly easy to formulate), then we obtain the solution in Tables 6 and 7. The total cost is in this case \$3,029,672.00, which is indeed higher than the total cost without the production constraint that items have to be produced in lots of 10,000. However, the increase is less than 1%. The increased costs are mainly due to excess production (the total production quantities now exceed the total demand quantities) and, consequently, additional transportation and storage costs.

It is clear that this formulation of this medium term planning problem can be extended very easily to more time periods, more factories at the first stage, and more product families. An extension to more stages may be a little bit more involved if there is an increase in the complexity of the routing patterns.

Table 6

	Week 1	Week 2	Week 3	Week 4
x_{f11}	0	0	0	0
x_{f21}	60,000	20,000	60,000	0
x_{f12}	70,000	30,000	80,000	0
x_{f22}	0	10,000	0	0

Table 7

	Week 1	Week 2	Week 3
y_{f112}	0	0	0
y_{f113}	0	0	0
y_{f122}	55,000	15,000	40,000
y_{f123}	15,000	15,000	40,000
y_{f212}	50,000	20,000	55,000
y_{f213}	10,000	0	5,000
y_{f222}	0	10,000	0
y_{f223}	0	0	0

5. Short Term Scheduling Models for Supply Chains

The short term scheduling problem for a facility in a supply chain can be described as follows: The output of the medium term planning problem specifies that over the short term n_j items of family j have to be produced. The scheduling problem can either be modeled as a job shop (or flexible flow shop) that takes all the production steps in the facility into account, or as a somewhat simpler single (or parallel) machine scheduling problem that focuses only on the bottleneck operation. If the operations in a facility are well balanced and the location of the bottleneck depends on the types of orders that are in the system, then the entire facility may have to be modeled as a job shop. If the bottleneck in the facility is a permanent bottleneck (that never moves), then a focus on the bottleneck is justified. If the bottleneck stage is modeled as a parallel machine scheduling model, then the parallel machines may not be identical. They may also be subject to different maintenance and repair schedules.

There is, of course, a close relationship between the time t^p_{jk} in the medium term planning process and the processing time of an order in the short term detailed scheduling problem. The t^p_{jk} in the medium term planning process has to be estimated and may be a value anywhere in between the average processing time of an order at the bottleneck operation and the total (estimated) throughput time of an order through the facility. The t^p_{jk} is a function of the processing times p_{ij} as well as of the sequence-dependent setup times s_{ijk}.

Any given order cannot be released before all the required raw material has arrived (these dates are typically stored in a Material Requirements Planning (MRP) system). That is, order j has an earliest possible starting time that is typically referred to as a release date r_j, a committed shipping date d_j, and a priority factor or weight w_j. Dependent upon the manufacturing environment, preemptions may or may not be allowed. Every time a machine switches over from one type of item to another type of item, a setup cost may be incurred and a setup time may be required. If a schedule calls for a large number of preemptions, a large number of setups may be incurred.

The objective to be minimized may include the minimization of the total setup times on the machines at the bottleneck as well as the total weighted tardiness, which is denoted by $\sum w_j T_j$. So the objective may be formulated as

$$\alpha_1 \sum w_j T_j + \alpha_2 \sum I_{ijk} s_{ijk},$$

where the α_1 and the α_2 denote the weights of the two parts of the objective function. The first part of the objective function is the total weighted tardiness, and the second part of the objective represents a total of all setups; the indicator variable I_{ijk} is 1 if job j is followed by job k on machine i, the indicator variable is 0 otherwise.

This scheduling problem may be tackled via a number of different techniques, including a combination of dispatching rules, such as the Shortest Setup Time (SST) first rule, the Earliest Due Date first (EDD) rule, and the Weighted Shortest Processing Time first (WSPT) rule. Other techniques may include genetic algorithms or integer programming approaches. In this phase, however, integer programming approaches are not often used because they are computationally quite intensive.

Example 5.1. Consider the two factories described in the medium term planning process in the previous section. In the detailed scheduling process, the two factories may be scheduled independently from one another and the scheduling is done one week at a time. Consider Factory 1 with the two product families. The production process in this factory consists of various steps, but one of these steps is the clear bottleneck. This bottleneck consists of a number of resources in parallel. Consider the operations of Factory 2 in the example in the previous section and only on the first week of operations. The solution of the integer program yields $x_{i12} = 65{,}000$ and $x_{i22} = 12{,}667$. Of the 65,000 of product family 1, a total of 50,000 has to be shipped to the DC and the remainder has to go to the customer. Of the 12,667 of product family 2, a total of 2667 has to be shipped to the DC and the remaining 10,000 has to go to the customer.

Assume now that the following more detailed information is available (which was not taken into account in the medium term planning process). The time unit in the scheduling process is 1 hour in contrast to the 1 week in the medium term planning process (in actual implementations the time unit in the scheduling process can be made arbitrarily small). The scheduling horizon is 1 week.

Recall that 2 hours of the bottleneck resource are required to produce 1000 units of Family 1 in Factory 2, whereas 3 hours of the bottleneck resource are required for 1000 units of Family 2. This implies that, based on these estimated production times, the planned production takes the full capacity of the bottleneck resource (in hours):

$$65 \times 2 + 12.667 \times 3 = 168.$$

However, the 2 and 3 hours requirement of the bottleneck resource are only estimates. They are estimates that are being used in the medium term planning process in order not to have to make a distinction between sequence-dependent setup times and run times. The actual run times (or processing times), excluding any setup times are as follows: To produce in

Table 8

Job	1	2	3	4
p_j	87.5	26.25	6.67	25
r_j	0	0	36	36
d_j	168	120	168	120

Factory 2 1000 units of Family 1, 1.75 hours of the bottleneck resource is required, whereas 1000 units of Family 2 requires 2.5 hours of the bottleneck resource. To start producing units of Family 1, a setup of 16 hours is required. To start producing units of Family 2, a setup of 6 hours is required. If each one of the products was to be produced in a single run in that week, then the entire production could be done within 168 hours, since

$$16 + 65 \times 1.75 + 6 + 12.66 \times 2.5 = 167.4.$$

So, if there are not too many setup times, the original assumptions for the medium term planning model are appropriate.

However, the shipment to the customer is supposed to go on a truck at time 120 (after 5 days), whereas the shipment to the DC takes place at the end of the week at time 168. All the raw material required to produce Family 1 products are available at time 0, whereas the material necessary to produce family 2 products are only available after 2 days, i.e., at time 48.

This problem can be modeled as a single machine scheduling problem with the jobs having different release dates and sequence-dependent setup times and as objective

$$\alpha_1 C_{\max} + \alpha_2 \sum w_j T_j.$$

There are 4 different jobs (Table 8), with the following processing times, release dates, and sequence-dependent setup times. Each job is characterized by its family type and its destination. Jobs 1 and 2 correspond to jobs from the same family, so there is a 0 setup time if one job is followed by the other. If either job 1 or job 2 follows job 3 or 4, then a setup of 16 hours is required. If job 3 or 4 follows job 1 or 2, a setup of 6 hours is required.

Two rules can be applied. One rule would follow the Shortest Setup Time first rule (with ties being broken according to the Earliest Due Date rule). This rule would generate the schedule 1, 2, 4, 3, since after job 1 has been completed, job 2 has to be started since it has a 0 setup time. All jobs are then completed by time 168. However, job 4 is completed tardy. It had to be shipped by time 120 and it is shipped by time 161.

The second rule follows the Earliest Due Date rule (with ties being broken according to the Shortest Setup Time first rule). It results in the schedule 2, 4, 3, 1. Since there is an additional setup time, the makespan is 167.4 + 16 = 183.4. The shipment to the customer leaves on time, but the shipment to the DC leaves late.

Which one of these two schedules is preferable depends on the weights α_1 and α_2 in the objective function.

The results coming out of the detailed scheduling problem may be, for various reasons, not acceptable. When trying to minimize the makespan (in order to ensure the production of the required quantities in the one week), it may turn out that there does not exist a schedule that would complete the requested production within one week. The reason may be the following: the production times t_{jk}^p that were entered in the medium term planning problem were estimated based on plant data, that include the average processing times on the bottleneck machines, the expected throughput times, the expected setup times, and so on. However, the value t_{jk}^p did not represent an accurate cycle time, since the average production time may depend on the run length of the batches at the bottleneck. It may be that the schedule generated in the detailed scheduling process has batch sizes that are very short and therefore an average production time that is larger than the estimates used in the medium term planning process. If there is a major discrepancy (i.e., the frequency of the setups is considerably higher than usual), then a new estimate may have to be developed for the t_{jk}^p in the medium term planning process and the integer programming problem has to be solved again.

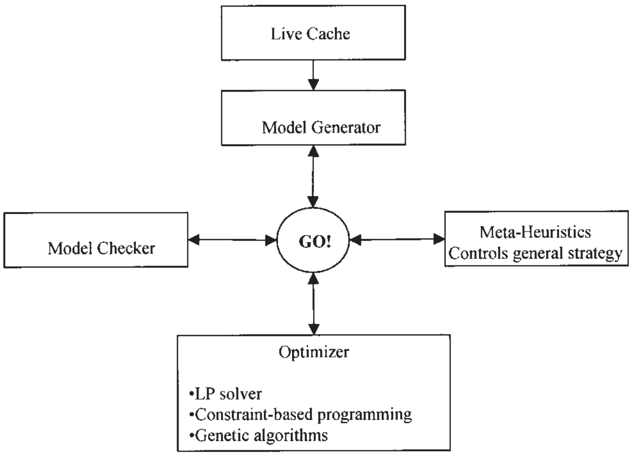
6. Carlsberg Denmark: An Example of a System Implementation

There are many software vendors that sell custom-made solutions for supply chain planning and scheduling. One of the largest companies in this field is SAP, based in Walldorf (Germany). SAP has a division that develops its Advanced Planner and Optimizer (APO) system. This supply chain planning and scheduling system has functionalities on various levels, including (i) the tactical level and (ii) the operational level.

On the tactical level, the medium term planning scenarios for the global chain are monitored (from distribution centers to plants and suppliers). The optimizer automatically processes bills of materials while taking capacities into account, and optimizes transportation costs, production costs, storage costs, and revenues for demand. The sheer complexity of this global view is handled by a rough-cut model that aggregates the time in buckets (e.g., day or week) and products and resources in families.

On the operational level, APO has a detailed scheduling model. At this level, the short term, day-to-day operations are monitored, focusing especially on exceptions in supply chain operations. The optimizer

Figure 6 SAP-APO Optimizer Architecture



schedules orders according to manufacturing constraints that handle complex manufacturing environments with alternative routings and resources, secondary resources, and multi-stage production.

Figure 6 shows the architecture of the optimizer in APO. For long and medium term planning, APO uses its LP solvers (CPLEX). For detailed short term planning and scheduling, APO has various approaches, including Constraint Programming, Genetic Algorithms, and Repair Algorithms.

This section describes an implementation of the SAP-APO system at the beerbrewer Carlsberg A/S in Denmark. The modeling that forms the basis for this case is somewhat similar to the models described in Sections 5 and 6 of this paper. Carlsberg Denmark A/S, the largest beerbrewer in Scandinavia, started in 2001 a supply chain project with the objective to decrease inventory costs, to optimize sourcing decisions, to increase customer service level, and in general to change the business to a more demand-driven process. Carlsberg selected APO. The scope of the project gives an example of how a real-life implementation is considering the planning and scheduling issues over several stages. The system is operational since the end of 2002.

The supply chain considered in the project consists of three stages. The first stage is the production process of the beer at two breweries with two and four filling lines, respectively. Each filling line has a different capacity. The second stage consists of the centralized warehouses, and the third stage consists of the local warehouses; see Figure 5. In the first stage, there are three production steps, namely brewing (and fermentation), filtering, and filling of the beer. All three steps have a limited capacity, but the bottleneck is usually the filling step. The resources for the filling operations at the two plants have different costs and processing times. When creating the production orders for brewing and filling, different lot size constraints have to be taken into account. Production orders for the brewing have always a fixed lot size because the brewing tank has to be filled. If the demand quantity is higher than the fixed lot size, then additional production orders have to be created for the brewing process (each with the fixed lot size as the production quantity). Orders below the minimal lot size are increased to the minimal lot size and orders above the minimal lot size are either rounded up or down to the closest integer value. The filling resources have to be filled up to 100%. There is further a split in the business processes according to the sales volumes of the various products. There are three categories: A, B, and C. Category A are the fast movers and include the well-known brands Carlsberg Pils and Tuborg Green. Category C are the (more expensive) slow movers.

Once the beer is bottled, it has to be transported either to a central depot or to a local depot. Depending on the different products and the quantities to be transported, either a direct delivery from the factory to a local depot or a transport via the centralized depot is better. Again, lot size constraints have to be taken into consideration when creating transport orders. The transport durations depend, of course, on the origin and the destination.

One of the main objectives of Carlsberg is to provide a given level of service to its customers. A typical way to achieve a given service level is to keep safety stocks at the depots. The higher the safety stocks, the higher the service level, but also the higher the inventory costs. One function of a supply chain management system is the computation of the lowest levels of safety stocks that achieve the desired service levels. Carlsberg uses advanced safety stock methods to compute safety stock values for all its products at its central as well as at its local depots. These safety stock levels depend on the given service level, the demand forecast, the uncertainty in the forecast, the replenishment lead time, and the typical lot sizes.

The medium term planning module plans ahead for 12 weeks, with the first 4 weeks in days and the remaining 8 weeks in weekly periods. Assuming a given demand pattern (sales orders and forecasts), APO creates a Mixed Integer Program, along the lines described in Section 5, and tries to find a solution with minimum cost. The total costs are the sum of the production costs, the storage costs, the transportation costs, the late delivery (tardiness) costs, the non-delivery costs, and the violation of the safety stock levels computed in the first step. Some of the costs mentioned above can be specified in an exact way, such as the production and transportation costs. Other costs, such as storage, violation of safety stock, and late and non-delivery costs, merely represent the priorities of

Carlsberg. If, for example, Carlsberg considers the safety stock in the local warehouses more critical than the safety stock in the central warehouse, then the cost assigned to the violation of safety stock for a product at the central warehouse is less than the cost of violating the safety stock of the same products at the local warehouses. If neither safety stock can be maintained, then the system will create a transport from the DC to the local warehouse (provided the difference between the costs of safety stock violations at the DC and at the local warehouse is higher than the transportation cost from the DC to the local warehouse). Clearly, all cost types are strongly related with one another, and modifying one type of cost can have many unforeseen consequences in the solution generated. Carlsberg developed its own cost model for storage costs; this model, for example, takes into account the location occupied by a pallet, the maximum number of levels pallets can be stacked, the number of products per pallet, and the warehouse itself. Based on these parameters for each product at each location, storage costs can be computed.

The following constraints have to be taken into consideration, namely the production times in the three production steps, the capacity of the bottling resources on a daily or weekly level, the transportation times between locations, the lot size constraints, the existing stock, and the resource consumptions.

The medium term plan is the result of various costs trade-offs and material consumption. The system generates for the next 12 weeks the planned production quantities for the three production steps in detail (including the quantity of each product to be bottled on each filling resource as well as the quantities to be transported between the locations).

The short term scheduling starts its computations using the results obtained from the medium term plan. The planned production orders for the first week that come out of the medium term planning system are transformed into short term production orders on which a detailed scheduling procedure has to be applied. These production orders are then scheduled on the filling resources by applying a genetic algorithm with as objective the minimization of the sum of the sequence-dependent setup times and the sum of the tardinesses. The due dates are specified by the medium term planning problem and are equal to the starting times of the transportation orders. It is possible that the results of the medium term plan are changed by the short term scheduling procedure (i.e., a different filling resource may be selected in the same plant). After the detailed scheduling has been completed, the transportation planning and scheduling is done. In this step, the trucks between the locations are filled based on the results of the planned transportation orders that come out of the medium term plan and the results that come out of the detailed scheduling procedure. An attempt is made to fill up the trucks with several products in order to ship mainly full trucks.

A new medium term plan is generated every day. The daily run takes into account the most up-to-date capacity situation of all the available resources, the results of the previous day detailed schedule, and the most current demand forecast. Afterwards, another detailed schedule and transportation plan are generated.

The generation of the medium term plan was split into three Mixed Integer Programs, which were solved in consecutive runs. Each MIP had between 100,000 and 500,000 variables and between 50,000 and 150,000 constraints. Total running time was about 10–12 hours. Each MIP used product decomposition methods, which created 5 to 10 subproblems each. The generation of the subproblems has to take into account different priorities of the finished beer products and the fact that the same brewed and filtered beer type may end up in different end products. The quality of the solution is measured in business terms as well as in technical terms. Only by considering all dimensions one can speak of a "good" or a "bad" solution. The technical data and measures tend to be easy to collect and understand; the more important business measures are harder to understand and verify. The two most important technical measures are (i) the difference between the costs of the MIP solution and the LP relaxation solution, and (ii) the difference between the overall delivery percentages of the MIP solution and the LP relaxation solution. The difference between the costs was on average between 0.2% and 10%, but sometimes shot up to 400%. A huge cost difference between the MIP and the LP relaxation could occur in a case where the LP could fulfill all demands, while the MIP could not fulfill all demands because of lot size constraints. As unfilled demand brought about very

Figure 7 Carlsberg-Denmark Planning System User Interface

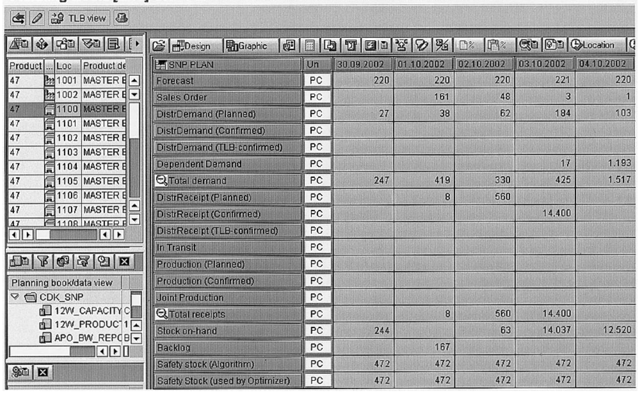
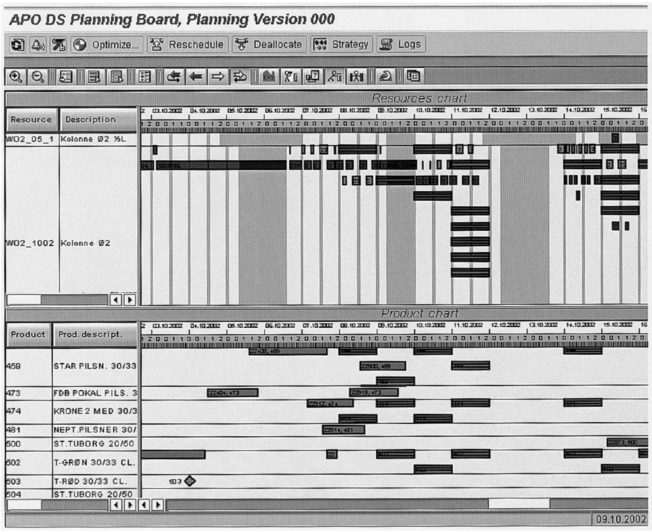


Figure 8 **Carlsberg-Denmark Scheduling System User Interface**



high penalty costs, the cost difference between the MIP and the LP relaxation turned out to be very high.

The user interfaces of the systems are, of course, quite elaborate and do include the typical Gantt charts; see Figures 7 and 8.

7. Discussion

The purpose of this paper is to provide insights into the use of planning and scheduling models in supply chain management, as well as into the information sharing and interactions that occur between the different types of models that are embedded in one system. In the literature, planning models have often been analyzed in detail; scheduling models, on the other hand, have been studied less often within a supply chain management framework. The interactions and information sharing between the planning models and the scheduling models also deserve more attention.

There are several reasons why it is not that easy to incorporate planning systems and scheduling systems in one framework. One reason is that, in the planning stage, the objective is measured in dollar terms (e.g., minimization of total cost), whereas in the schedule stage, the objective is typically measured in time units (e.g., minimization of total tardiness). A second reason is that the time periods over which the planning module and the scheduling module optimize may overlap only partially (see Figure 3). The horizon over which the scheduling module optimizes typically starts at the current time and covers a relatively short term. The planning module optimizes over a period that starts at some time in the future (since it may assume that all schedules before this point in time already have been fixed) and covers a long term. The units of time used in the two modules may be different as well. In the scheduling module, the unit may be an hour or a day; in the planning module, it may be a week or a month.

Comparing the modeling that is done in practice for medium term planning processes with models that have been studied in the research literature, it becomes clear that there are differences in emphasis. When multi-stage models are considered in the planning and scheduling research literature, there is more of an emphasis on setup costs (typically sequence-independent) and less of an emphasis on transportation costs; in the modeling that is done in practice, there is a very strong emphasis on transportation costs and less of an emphasis on setup costs. Incorporating setup costs as well as transportation costs in a multi-stage planning model may cause the number of variables to become prohibitively large.

References

Arntzen, B. C., G. G. Brown, T. P. Harrison, L. L. Trafton. 1995. Global Supply Chain Management at Digital Equipment Corporation. *Interfaces* **25** 69–93.

Barbarosoglu, G., D. Ozgur. 1999. Hierarchical Design of an Integrated Production and 2-Echelon Distribution System. *European Journal of Operational Research* **118** 464–484.

Bhaskaran, K., M. Pinedo. 1992. Dispatching in *Handbook of Industrial Engineering*, G. Salvendy (ed.), Wiley, New York, 2184–2198.

Bowersox, D. J., D. J. Closs. 1996. *Logistical Management, the Integrated Supply Chain Process*. McGraw-Hill, New York.

Braun, H. 2001. Optimization in Supply Chain Management. *SAP Insider*. SAP AG., Vol. 2.

Braun, H., C. Groenewald. 2000. New Advances in SAP APO to Optimize the Supply Chain. *SAP Insider*. SAP AG. Vol. 1.

Brucker, P. 1998. *Scheduling Algorithms* (Second Edition). Springer Verlag, Berlin.

Chen, Z-L., N. Hall. 2000. Supply Chain Scheduling. Working Paper, Fisher College of Business, The Ohio State University, Columbus, Ohio.

Chopra, S., P. Meindl. 2001. *Supply Chain Management—Strategy, Planning, and Operation*. Prentice-Hall, Upper Saddle River, New Jersey.

De Bontridder, K. 2001. Integrating Purchase and Production Planning—Using Local Search in Supply Chain Optimization. PhD Thesis, Technische Universiteit Eindhoven, the Netherlands.

Dhaenens-Flipo, C., G. Finke. 2001. An Integrated Model for an Industrial Production-Distribution Problem. *IIE Transactions* **33** 705–715.

Dobson, G. 1987. The Economic Lot Scheduling Problem: Achieving Feasibility Using Time-Varying Lot Sizes. *Operations Research* **35** 764–771.

Dobson, G. 1992. The Cyclic Lot Scheduling Problem with Sequence-Dependent Setups. *Operations Research* **40** 736–749.

Drexl, A., A. Kimms. 1997. Lot Sizing and Scheduling—Survey and Extensions. *European Journal of Operational Research* **99** 221–235.

Hadavi, K. C. 1998. Supply Chain Planning Integration: Design Consideration or Afterthought? *APS Magazine* **1** 2, MFG Publishing, Inc., Beverly, Massachusetts.

Hall, N. G., C. N. Potts. 2000. Supply Chain Scheduling: Batching and Delivery. Working Paper, Fisher College of Business, The Ohio State University, Columbus, Ohio.

Haq, A. N. 1991. An Integrated Production-Inventory-Distribution Model for Manufacturing of Urea: A Case. *International Journal of Production Economics* **39** 39–49.

Kimms, A. 1997. *Multi-Level Lot Sizing and Scheduling*. Physica Verlag, Heidelberg.

Lourenco, H. R. 2001. Supply Chain Management: An Opportunity for Metaheuristics. Technical Report, Universitat Pompeu Fabra, Barcelona, Spain.

Miller, T. C. 2002. *Hierarchical Operations and Supply Chain Planning*. Springer Verlag, London.

Muckstadt, J. A., R. Roundy. 1993. Analysis of Multi-Stage Production Systems. Chapter 2 in *Handbooks in OR and MS* **4** 59–132, S. C. Graves et al. (eds.), Elsevier Science Publishers.

Murthy, S., R. Akkiraju, R. Goodwin, P. Keskinocak, J. Rachlin, F. Wu. 2001. Cooperative Multi-Objective Decision-Support for the Paper Industry. Working Paper, IBM T.J. Watson Research Center, Yorktown Heights, New York.

Pinedo, M. 2002. *Scheduling—Theory, Algorithms, and Systems* (Second Edition). Prentice-Hall, Upper Saddle River, New Jersey.

Pinedo, M., X. Chao. 1999. *Operations Scheduling with Applications in Manufacturing and Services*. Irwin/McGraw-Hill.

Rachlin, J., R. Goodwin, S. Murthy, R. Akkiraju, F. Wu, S. Kumaran, R. Das. 2001. A-Teams: An Agent Architecture for Optimization and Decision Support. Working Paper, IBM Supply Chain Optimization Solutions, Atlanta, Georgia.

Shapiro, J. F. 2001. *Modeling the Supply Chain*. Duxbury, Pacific Grove, California.

Shepherd, J., L. Lapide. 1998. Supply Chain Optimization: Just the Facts. AMR Research, Inc.

Simchi-Levi, D., P. Kaminsky, E. Simchi-Levi. 2000. *Designing and Managing the Supply Chain*. McGraw-Hill/Irwin, Burr Ridge, Illinois.

Stadtler, H., C. Kilger (eds.). 2000. *Supply Chain Management and Advanced Planning*. Springer, Berlin.

Strobel, K. 2001. mySAP Supply Chain Management: Managing the Supply Chain beyond Corporate Borders. *SAP Insider*. SAP AG., Vol. 2.

Vandaele, N. J., M. Lambrecht. 2001. Planning and Scheduling in an Assemble-to-Order Environment: Spicer-Off-Highway Products Division. Technical Report, Antwerp University, Belgium.

Zipkin, P. H. 2000. *Foundations of Inventory Management*. McGraw-Hill/Irwin, Burr Ridge, Illinois.