MATHEMATISCHES FORSCHUNGSINSTITUT OBERWOLFACH

Report No. 19/2004

Mathematics in the Supply Chain

Organised by
Robert E. Bixby (Houston, TX, USA)
Alexander Martin (Darmstadt)
David Simchi-Levi (Cambridge, MA, USA)
Uwe T. Zimmermann (Braunschweig)

April 11th – April 17th, 2004

Introduction by the Organisers

Mathematical models and methods have been established in various fields of applications in recent years. These include transport and traffic (see for instance the Oberwolfach workshop on "Traffic and Transport Optimization" in 1999), production planning (see Handbooks in Operations Research and Management Science, the volume on "Logistics of Production and Inventory" Graves, Rinnooy Kan and Zipkin (eds.), North-Holland, 1993), communication networks (see "Handbook of Discrete and Combinatorial Mathematics" Rosen (ed.), or "Handbook in Operations Research and Management Science" the volume on "Network Models" Ball, Magnanti, Monma and Nemhauser (eds.)) and financial engineering ("Options, Futures, and Other Derivatives" by J. Hull). The success of mathematical methods does not only rely on progress in computer technology, but in a fundamental way it relies on the improvement of the underlying mathematical models, methods and associated algorithmic developments. In recent years new challenges from industry and business have arisen. Modern problems contain more and more structures whose origins lie in various disciplines of mathematics, including graphs and networks, optimization, control and stochastic processes. A prominent field in which such problems arise is supply chain management.

Supply chain management is typically defined as:

> A set of approaches utilized to efficiently integrate suppliers, manufacturers, warehouses, and stores so that merchandise is produced and distributed in the right quantities, to the right locations, and

at the right time, in order to minimize systemwide costs while satisfying service level requirements.

The idea of the workshop was to bring people from mathematics who have successfully applied mathematical methods for the solution of practical problems together with people who are in touch with the real-world problems that constitute supply chain management. To motivate and initiate interactions between the two groups, the mathematicians presented their applied methods and explained the situations in which they are effective as well as the underlying theory. These methods and concepts cover a wide spectrum, ranging from integer programming and dynamic programming methods through approximation algorithms to stochastic optimization methods. Similarly, supply chain management experts explained their problems, pointed out where current methods will help and where further developments are necessary. The scope of these problems is quite impressive. It includes distribution problems that integrate warehousing and transportation, inventory management models, the integration of procurement and manufacturing activities with demand planning processes and pricing and auction models that are used to improve business-to-consumer and business-to-business interactions. In this way we hope that the workshop provided a forum for open discussions about interesting new mathematical problems that are of central importance to supply chain management.

The workshop itself was organized as follows. We had a series of plenary talks from world leading experts in mathematics and in supply chain management about the current state-of-the-art in these fields. We completed the program by having a series of shorter talks where people reported on their chosen models and solution methods for these models.

In particular we invited and encouraged young people to participate in the workshop. They used the chance to gain as well as to present insights into new developments in applied mathematics and supply chain management. The interconnection of the two disciplines showed to provide a great source for new, interesting research projects in the future.

The Oberwolfach workshop on "Mathematics in the Supply Chain" turned out to be of great interest for both experts in mathematics and experts in supply chain management. The meeting offered a platform for mathematicians to become acquainted with an important future-oriented field of applications. Vice versa, it offered a platform for supply chain management people to learn more about current mathematical models and methods potentially useful in solving further problems. We are very happy that we had the opportunity to organize such an interdisciplinary workshop at Oberwolfach.

Workshop on Mathematics in the Supply Chain

Table of Contents

Abstracts

Beyond Supply Chain Optimization to Enterprise Optimization
Jeremy Shapiro

The number and scope of mixed integer programming applications to supply chain management are large and increasing. Models for strategic planning include those for analyzing: Worldwide sourcing of chemical products; consolidation of the supply chains of two companies after a merger; the design of the supply chain for a new product; and, the classic problem of re-locating distribution centers. In most instances, the objective function driving analysis is minimizing the total supply chain cost of meeting fixed and given product demand. Many scenarios reflecting variations in demand and other planning parameters may be optimized before senior managers, using a combination of model results and judgement, commit to important strategic decisions.

Our research begins with the premise that, from the perspective of fact-based strategic planning, cost minimizing objectives that focus only on supply chain activities are timid and shortsighted. To expand the scope, we first examine extensions of supply chain optimization models that incorporate marketing science models for demand management. The combined models address decisions for maximizing net revenues by allowing product mix to vary. We show how simple extensions can be constructed if product sales volumes by market and time period are primarily functions of product prices.

Relationships involving location sensitive revenue functions, which are applicable to supply chains involving commodities and retail products, are also presented. Specifically, we discuss how logit models may be applied to consumer surveys to determine market share and sales as a function of the distance to the nearest supply point (DC or retail store). These relationships may be used to determine a supply chain network design that maximizes net revenues.

We also examine a more complex supply chain/demand management model for the case when demand is a function not only of price but advertising, promotions, and sales force effort. This model is applicable to the study of consumer package goods. For such applications, we discuss a generalized programming (column generation) method that integrates a marketing science model relating demand to values for these factors with a supply chain mixed integer programming model for minimizing the cost of meeting demand. In effect, the integrated model maximizes net revenues by harmonizing decisions about money spent to create demand with decisions about money spent to supply products to meet this demand.

Second, we examine extensions of strategic supply chain/demand management models to models that include strategic corporate financial management, which involves decisions regarding capital investment, long-term debt, and, in general, the financial performance of the firm. Although not yet widely applied, mixed integer programming models for examining these decisions based on funds flow equations for each planning period are available and provide a powerful, holistic

view of the firm's corporate financial decisions. We demonstrate how these models can be seamlessly combined with supply chain/demand management models that compute optimal earnings before interest and taxes in each time period, key variables affecting the firm's financial performance.

Corporate financial decision-making also requires multi-objective methods to measure tradeoffs of after-tax profits, returns on fixed assets, the speed of payback on capital investments, and possibly other criteria. We examine a Lagrange multiplier method for exploring the efficient frontier defined by these criteria. We also discuss the use of statistical methods to map points on the frontier into shareholder value.

The Dance of the 30 Ton Trucks: Demand Dispatching in a Dynamic Environment

KARLA HOFFMANN

(joint work with Martin T. Durbin)

The planning, scheduling, dispatching, and delivery of perishable items in a time-constrained environment are recognized as some of the most challenging problems in manufacturing. In the concrete industry, the challenge is dramatically increased due to overbooking and the requirement to always complete multi-truck, time-synchronized orders once they are started. Additionally, weather and traffic conditions can adversely affect expected travel-time in an environment where more than 90

Virginia Concrete (VC) is a very forward-thinking, innovative company that was in the process of installing GPS equipment and sensors on board every truck in their delivery fleet. This equipment allowed operators to track the position of the truck and the status of the delivery. Additionally, each truck was equipped with an on-board computer and mobile digital transmitter to support communication between the truck and the central dispatch centre. VC wanted to use this information to increase the efficiency of their delivery process. The company noticed that their trucks were consistently queueing up at both their plants and customer sites and wondered whether a decision support system might alleviate some of these problems.

Virginia Concrete delivers 400 - 600 loads per day with approximately 125 trucks, which results in 100,000 - 150,000 loads per year. They estimate that saving a modest five minutes on every load would generate between $500,000 and 750,000$ in annual savings. Additionally, dispatching is a high-pressure job where efficiency is highly dependent upon the skills of individual dispatchers. When a highly effective dispatcher is absent, the operational efficiency of the company is negatively impacted. Similarly, plant scheduling, next-day scheduling, and the determination of driver arrival times are costly and time-consuming processes. VC believed that the development of a decision-support tool would not only generate savings, it would also i) decrease stress for dispatchers, schedulers, and plant managers, ii)

create consistency in these functions across the company, iii) reduce the time required to train new dispatchers, and iv) enable easier substitution and relocation of dispatchers.

We now list some of the complexities of the scheduling problem:

- Concrete is a perishable product. There is a maximum amount of time that concrete can remain in a truck before it hardens. Unexpected delays in traffic or at customer sites can result in concrete that is unacceptable for delivery and must be unloaded from the truck before it begins to harden (as little as 2 hours).
- Most concrete orders require more than one load of concrete, and the trucks are required to arrive sequentially, with the customer specifying the inter-arrival rate.
- Concrete must be poured in a continuous fashion. To pour continuously, the trucks should be spaced sufficiently to allow the customer time to unload the concrete from one truck before the next arrives, but at a rate brisk enough to allow the concrete to be poured fluidly.
- The customer rarely knows the exact size of the order. Typically, while unloading the last of the ordered amount of concrete the customer determines if any additional concrete is needed. At this time, the customer will determine the additional yardage required, if any, and inform the driver.
- Most trucks hold nine cubic yards of concrete and weigh approximately thirty tons when loaded but newer trucks hold ten cubic yards.
- Some orders are restricted to a subset of the trucks, some orders to a subset of the drivers, and some to a subset of the plants.
- Plants can load a specific number of trucks per hour. Therefore, it may be more cost effective to load a truck from a plant other than the closest.
- All concrete is not the same. Customers vary the concrete strength and viscosity requirements depending upon the desired usage. Furthermore, depending upon the weather, different chemical additives may be used to ensure that the concrete meets application requirements. Some orders require that all material must come from the same plant to maintain consistency.
- When a customer orders concrete to a site that has a pump, there is the requirement to send a small initial load that contains a watery concrete, called grout. The grout load must be delivered as the first load.
- Occasionally the customer will request a gap in the delivery to account for a planned delay during the pouring of the concrete.
- Drivers have restrictions on their workday imposed by unions and government regulations.

The environment in which concrete is delivered is very dynamic. Obviously, the time it takes to travel from a plant to a customer site (or vice versa) can vary significantly within a day and from day-to-day. Traffic tie-ups in the Northern Virginia area are likely, but unpredictable. In addition, customer behaviour

can impact the efficiency of the delivery process. Some very common customer problems are:

- Customers may not be ready when the driver arrives with the concrete. This results in the delayed usage of the truck, the backup of additional trucks at the site, wasted time, and the delayed availability of trucks for later deliveries.
- When placing an order, customers request a specific unload time for a truckload, but customers are consistently optimistic regarding their ability to unload a truck.
- The customer seldom knows the exact amount of concrete they need. Underestimating by as little as one yard could require an additional truck load. In addition, the customer often delays or cancels an order right at the last moment.

Weather can impact the delivery process in a variety of ways. The most obvious is that inclement weather impacts the expected travel-time and causes delays in job-site arrival. Additionally, concrete cannot be delivered during storms, or extreme cold.

Another factor making the environment more dynamic is the breakdown of either trucks or plants. The breakdown of an empty truck can result in delivery delays. The breakdown of a loaded truck has additional costs associated with material replacement; late delivery and removal of hardened product from the drum of the truck. The breakdown of a plant can create a significant disruption in the delivery plan for the entire company. When this occurs, the company is mainly interested in system recovery and places priority on continuing any customer delivery that is currently in progress.

In the ready-mix concrete industry, much of the concrete that is delivered requires the time-synchronized, staggered deployment of several trucks to a customer site. The standard dispatching method utilized in the industry is truck-based. In *truck-based dispatching*, each customer is assigned a specific number of trucks and a specific plant as the source of the product. These trucks then make trip after trip to the same customer until all of the concrete requested by the customer has been delivered. The static allocation of trucks and plants can be inefficient in a dynamic environment. Consequently, a *demand dispatching* capability has been developed. In demand dispatching, trucks are assigned to a specific delivery (as opposed to an entire job) when they enter the yard of a plant. When the driver completes the delivery, they are then directed to a plant, which may or may not be the same plant from which they received their previous load of concrete. Dispatching trucks in this intelligent, responsive fashion results in a complex, intertwined movement of trucks throughout the day. This intricate behaviour is called *the Dance of the Thirty-ton Trucks*.

In order to support demand dispatching, a decision-support tool was created that consists of both planning and execution modules. The solution of the model formulation assists customer service representatives and dispatchers in determining 1) the feasibility of accepting additional orders, 2) the arrival times for drivers

reporting to work, 3) the scheduling of all orders, 4) the real-time assignment of drivers to delivery loads, 5) the dispatching of these drivers to customers and back to plants, and 6) the scheduling of plants. This dissertation describes the series of *optimization models* required to implement a decision-support tool, the implications of imperfect data, and implementation issues associated with real-time requirements. The foundation for the solution is a time-space network representation of the problem incorporating a multitude of alternatives for delivering an order to a customer. Choosing a single delivery alternative for each customer adds restrictive integrality constraints to the network model. In addition to the time-space network formulation, a minimum-cost network flow model and a Tabu Search heuristic are utilized as other modules in the planning of orders and the dispatching of trucks.

Recommendations are made to dispatchers and customer-service representatives in real-time, responding to delivery events and customer order modifications as they occur. To make effective recommendations in such a dynamic environment, the real-time dispatcher and order planning tools are run on a cyclic basis. Currently, every **five minutes**.

The majority of scheduling solutions currently in use today focus on planning models. Solving a real-time scheduling problem with perishable products in a time-dependent, dynamic environment has proven to be an extraordinarily challenging task. On a good day, 90

Due to the success of this project, development of the decision-support tool described in this document is being expanded to include all of the sister companies in Northern Virginia. Virginia Concrete's parent company, Florida Rock, is interested in deploying this application throughout the corporation, thereby expanding the number of trucks dispatched from 125 to 1400, and the number of plants from 10 to 150. More impressively, Virginia Concrete is sufficiently convinced of the importance of this research and its application that they have begun to promote it as a "best practice" throughout the concrete industry.

Multi-Period Strategic Supply Chain Models

Stefan Nickel

(joint work with M. T. Melo and F. Saldanha da Gama)

We propose a mathematical modelling framework for supply chain network design which captures many aspects of practical problems that have not received adequate attention in the literature. The aspects considered include: dynamic planning horizon, generic supply chain structure, inventory and distribution opportunities for goods, facility configuration, budget constraints, and storage limitations. Moreover, the gradual relocation of facilities over the planning horizon is considered. A generic mathematical programming model is described in detail.

1. Problem Formulation

Index sets

$$
\begin{array}{rcl}
L & : & \text{set of facilities} \\
S & : & \text{set of } \textit{selectable} \text{ facilities}, \ S \subset L \\
S^c & : & \text{set of } \textit{selectable existing} \text{ facilities}, \ S^c \subset S \\
S^o & : & \text{set of potential sites for establishing } \textit{new} \text{ facilities}, \ S^o \subset S \\
P & : & \text{set of product types} \\
T & : & \text{set of periods with } |\,T\,| = n
\end{array}
$$

The set L contains all types of facilities. These are categorized in so-called *selectable* and *non-selectable* facilities. Selectable facilities form the set S and include existing facilities (S^c) as well as potential sites for establishing new facilities (S^o). At the beginning of the planning horizon, all the facilities in the set S^c are operating. Afterwards, capacity can be shifted from these facilities to new facilities located at the sites in S^o. Note that $S^c \cap S^o = \emptyset$ and $S^c \cup S^o = S$. The second category of facilities, the so-called non-selectable group, forms the set $L \setminus S$ and includes all facilities that exist at the beginning of the planning project and which will remain in operation. Examples of such facilities include plants and warehouses that should continue supporting supply chain activities, that is, are not subject to relocation decisions. Non-selectable facilities may also have demand requirements, that is, they may correspond to customers.

Costs

$$
\begin{array}{rcl}
PC^t_{\ell,p} & : & \text{variable cost of purchasing one unit of product } p \in P \text{ from an} \\
& & \text{external supplier by facility } \ell \in L \text{ in period } t \in T \\
TC^t_{\ell,\ell',p} & : & \text{variable cost of shipping one unit of product } p \in P \text{ from facility} \\
& & \ell \in L \text{ to facility } \ell' \in L \ (\ell \neq \ell') \text{ in period } t \in T \\
IC^t_{\ell,p} & : & \text{variable inventory carrying cost per unit on hand of product} \\
& & p \in P \text{ in facility } \ell \in L \text{ at the end of period } t \in T \\
MC^t_{i,j} & : & \text{unit variable cost of moving capacity from the existing facility} \\
& & i \in S^c \text{ to a new facility established at site } j \in S^o \text{ at the beginning} \\
& & \text{of period } t \in T \setminus \{1\} \\
OC^t_{\ell} & : & \text{fixed cost of operating facility } \ell \in L \text{ in period } t \in T \\
SC^t_i & : & \text{fixed cost charged in period } t \in T \setminus \{1\} \text{ for having shut down} \\
& & \text{the existing facility } i \in S^c \text{ at the end of period } t-1 \\
FC^t_j & : & \text{fixed setup cost charged in period } t \in T \setminus \{n\} \text{ when a new facility} \\
& & \text{established at site } j \in S^o \text{ starts its operation at the beginning of} \\
& & \text{period } t+1
\end{array}
$$

Parameters

$\overline{K}_\ell^t$: maximum capacity of facility $\ell \in L$ in period $t \in T$

$\underline{K}_\ell^t$: minimum required throughput at the selectable facility $\ell \in S$ in period $t \in T$

$\mu_{\ell,p}$: unit capacity consumption factor of product $p \in P$ at facility $\ell \in L$

$H_{\ell,p}$: stock of product $p \in P$ at facility $\ell \in L$ at the beginning of the planning horizon (observe that $H_{\ell,p} = 0$ for every $\ell \in S^o$)

$D_{\ell,p}^t$: external demand of product $p \in P$ at facility $\ell \in L$ in period $t \in T$

α^t : unit return factor on capital not invested in period $t \in T \setminus \{n\}$, that is, $\alpha^t = 1 + \beta^t/100$ with β^t denoting the interest rate in period t

B^t : available budget in period $t \in T$

Since each existing selectable facility may have its capacity transferred to one or more new facilities, it is assumed that its maximum capacity is non-increasing during the planning horizon, that is, $\overline{K}_i^t \geq \overline{K}_i^{t+1}$ for every $i \in S^c$ and $t \in T \setminus \{n\}$. Without loss of generality, it is assumed that $\overline{K}_i^1$ denotes the actual size of facility $i \in S^c$ at the beginning of the planning horizon. The above condition permits to impose capacity transfers in specific periods or even complete shutdowns. Similarly, potential new facilities have non-decreasing capacities, that is, $\overline{K}_j^t \leq \overline{K}_j^{t+1}$ for every $j \in S^o$ and $t \in T \setminus \{n\}$. Clearly, at the beginning of the planning project we have $\overline{K}_j^1 = 0$ for every new site $j \in S^o$.

Decision variables

$b_{\ell,p}^t$ = amount of product $p \in P$ purchased from an outside supplier by facility $\ell \in L$ in period $t \in T$

$x_{\ell,\ell',p}^t$ = amount of product $p \in P$ shipped from facility $\ell \in L$ to facility $\ell' \in L$ ($\ell \neq \ell'$) in period $t \in T$

$y_{\ell,p}^t$ = amount of product $p \in P$ held in stock in facility $\ell \in L$ at the end of period $t \in T \cup \{0\}$ (observe that $y_{\ell,p}^0 = H_{\ell,p}$)

$z_{i,j}^t$ = amount of capacity shifted from the existing facility $i \in S^c$ to a newly established facility at site $j \in S^o$, at the beginning of period $t \in T$

ξ^t = capital not invested in period $t \in T$

δ_ℓ^t = $\begin{cases} 1 & \text{if the selectable facility } \ell \in S \text{ is operated during period } t \in T \\ 0 & \text{otherwise} \end{cases}$

In view of the assumptions made on the time points for paying fixed facility costs, it follows that a new facility can never operate in the first period since that would force the company to invest in its setup before the beginning of the planning horizon. Similarly, an existing facility cannot be closed at the end of the last period since the corresponding fixed shutdown costs would be charged in a period beyond the planning horizon. Hence, $z_{i,j}^1 = 0$ for $i \in S^c$ and $j \in S^o$, $\delta_i^1 = 1$ for $i \in S^c$, and $\delta_j^1 = 0$ for $j \in S^o$.

Assuming that all parameters are non-negative our MIP formulation is as follows.

$$\text{MIN} \quad \sum_{t \in T} \sum_{\ell \in L} \sum_{p \in P} PC^t_{\ell,p}\, b^t_{\ell,p} + \sum_{t \in T} \sum_{\ell \in L} \sum_{\ell' \in L \setminus \{\ell\}} \sum_{p \in P} TC^t_{\ell,\ell',p}\, x^t_{\ell,\ell',p}$$

$$(1) \qquad + \sum_{t \in T} \sum_{\ell \in L} \sum_{p \in P} IC^t_{\ell,p}\, y^t_{\ell,p} + \sum_{t \in T} \sum_{\ell \in S} OC^t_\ell\, \delta^t_\ell + \sum_{t \in T} \sum_{\ell \in L \setminus S} OC^t_\ell$$

s. t.

$$b^t_{\ell,p} + \sum_{\ell' \in L \setminus \{\ell\}} x^t_{\ell',\ell,p} + y^{t-1}_{\ell,p} =$$

$$(2) \qquad D^t_{\ell,p} + \sum_{\ell' \in L \setminus \{\ell\}} x^t_{\ell,\ell',p} + y^t_{\ell,p}, \quad \ell \in L,\, p \in P,\, t \in T$$

$$(3) \qquad \overline{K}^1_i - \sum_{\tau=1}^{t} \sum_{j \in S^o} z^\tau_{i,j} \le \overline{K}^t_i\, \delta^t_i, \quad i \in S^c,\, t \in T$$

$$(4) \qquad \sum_{\tau=1}^{t} \sum_{i \in S^c} z^\tau_{i,j} \le \overline{K}^t_j\, \delta^t_j, \quad j \in S^o,\, t \in T$$

$$(5) \qquad \sum_{\tau=1}^{t} \sum_{j \in S^o} z^\tau_{i,j} + \delta^t_i\, \epsilon \le \overline{K}^1_i, \quad i \in S^c,\, t \in T$$

$$\sum_{p \in P} \mu_{i,p} \left(b^t_{i,p} + \sum_{\ell \in L \setminus \{i\}} x^t_{\ell,i,p} + y^{t-1}_{i,p} \right)$$

$$(6) \qquad \le \overline{K}^1_i - \sum_{\tau=1}^{t} \sum_{j \in S^o} z^\tau_{i,j}, \quad i \in S^c,\, t \in T$$

$$\sum_{p \in P} \mu_{j,p} \left(b^t_{j,p} + \sum_{\ell \in L \setminus \{j\}} x^t_{\ell,j,p} + y^{t-1}_{j,p} \right)$$

$$(7) \qquad \le \sum_{\tau=1}^{t} \sum_{i \in S^c} z^\tau_{i,j}, \quad j \in S^o,\, t \in T$$

$$(8) \qquad \sum_{p \in P} \mu_{\ell,p} \left(b^t_{\ell,p} + \sum_{\ell' \in L \setminus \{\ell\}} x^t_{\ell',\ell,p} + y^{t-1}_{\ell,p} \right) \le \overline{K}^t_\ell, \quad \ell \in L \setminus S,\, t \in T$$

$$(9) \qquad \sum_{p \in P} \mu_{\ell,p} \left(b^t_{\ell,p} + \sum_{\ell' \in L \setminus \{\ell\}} x^t_{\ell',\ell,p} + y^{t-1}_{\ell,p} \right) \ge \underline{K}^t_\ell\, \delta^t_\ell, \quad \ell \in S,\, t \in T$$

$$(10) \qquad \delta^t_i \ge \delta^{t+1}_i, \quad i \in S^c,\, t \in T \setminus \{n\}$$

$$(11) \qquad \delta^t_j \le \delta^{t+1}_j, \quad j \in S^o,\, t \in T \setminus \{n\}$$

$$(12) \qquad \sum_{j \in S^o} FC_j^1 \, \delta_j^2 + \xi^1 = B^1$$

$$\sum_{i \in S^c} \sum_{j \in S^o} MC_{i,j}^t \, z_{i,j}^t + \sum_{i \in S^c} SC_i^t \left(\delta_i^{t-1} - \delta_i^t \right) + \sum_{j \in S^o} FC_j^t \left(\delta_j^{t+1} - \delta_j^t \right)$$

$$(13) \qquad + \quad \xi^t = B^t + \alpha^{t-1} \, \xi^{t-1}, \quad t \in T \setminus \{1, n\}$$

$$\sum_{i \in S^c} \sum_{j \in S^o} MC_{i,j}^n \, z_{i,j}^n + \sum_{i \in S^c} SC_i^n \left(\delta_i^{n-1} - \delta_i^n \right) + \xi^n$$

$$(14) \qquad = B^n + \alpha^{n-1} \, \xi^{n-1}$$

$$(15) \qquad b_{\ell,p}^t \geq 0, \quad y_{\ell,p}^t \geq 0, \quad x_{\ell,\ell',p}^t \geq 0, \quad \ell, \ell' \in L, \, p \in P, \, t \in T$$

$$(16) \qquad z_{i,j}^t \geq 0, \quad i \in S^c, \, j \in S^o, \, t \in T$$

$$(17) \qquad \xi^t \geq 0, \quad t \in T$$

$$(18) \qquad \delta_\ell^t \in \{0, 1\}, \quad \ell \in S, \, t \in T$$

References

[1] T. Bender, H. Hennes, J. Kalcsics, M. T. Melo, and S. Nickel. Location Software and Interface with GIS and Supply Chain Management. In Z. Drezner and H. W. Hamacher, editors, *Facility Location: Applications and Theory*, chapter 8, pages 233–274. Springer-Verlag, Berlin, Heidelberg, 2002.

[2] Y. Hinojosa, J. Kalcsics, S. Nickel, J. Puerto, and S. Velten. Dynamic Supply Chain Design with Inventory. Technical report, Universität des Saarlands, Saarbrücken, Germany, 2004.

[3] J. Kalcsics, T. Melo, S. Nickel, and V. Schmid-Lutz. Facility Location Decisions in Supply Chain Management. In K. Inderfurth et al., editors, *Operations Research Proceedings 1999*, pages 467–472. Springer Verlag, Berlin, 2000.

[4] M. T. Melo, S. Nickel, and F. Saldanha da Gama. Large-Scale Models for Dynamic Multi-Commodity Capacitated Facility Location. Technical report 58, Fraunhofer Institute for Industrial Mathematics (ITWM), Kaiserslautern, Germany, 2003. Available at www.itwm.fhg.de

Integrated Models for Service and Inventory Management

Hans Daduna

(joint work with Maike Schwarz, Cornelia Sauer, Ryszard Szekli, Rafal Kulik)

1. The aim of our project is to construct integrated models for to assess the different policies of inventory management with respect to the quality of service (QoS) of general service systems. The central problem is: How are the classical performance measures (e.g. queue length, waiting time, etc.) influenced by the management of an attached inventory? – and vice versa: How inventory management has to react on queueing of demands, resp. customers which present that demand, which is due to incorporated service facilities?

Research concerning integrated models and their explicit analytical solution started only recently. The common approach: Define a Markovian system process and then use standard optimization methods to find the optimal control strategy of the inventory. Some contributions in this direction:

Sigman and Simchi-Levi [SSL92] with approximation procedures to find performance descriptors for models, in which the interaction of queueing for service and inventory control is incorporated,

and in a sequence of papers Berman and his coauthors ([BK99], [BK01], [BS00], [BS02]. defined for various systems a Markov process description and then used classical optimization methods to find optimal control strategies for the inventory. The integrated models found in the literature combine single server queueing systems with an attached inventory. All these models assume that the demand, which arrives during the time the inventory is depleted, is back-ordered. The models vary with respect to the lead time distribution, the service time distribution, waiting room size, order size and reorder policy.

2. Our first models are in the same range: Single server systems with an inventory under continuous-review. Both regimes are considered in case of empty inventory: Back-ordering and lost sales.

We construct Markovian models which incorporate the aspects of queueing behaviour (performance analysis) and inventory behaviour (reorder policies) in a unified framework. The first problems are to find conditions for the system to stabilize and to compute explicitly the steady state behaviour of the system process. The service systems under consideration are of standard structure such that in isolation the stationary behaviour is known. It turns out that from the modelling aspect there are fundamental differences with respect to back-ordering versus lost sales.

(i) Backorder: We find partial solutions of the steady state equations. This suffices to derive several invariance properties of the performance measures of the systems. But for the complete steady state solution of the system we have to recur to numerical procedures or approximations. It turns out that these procedures can be strongly enhanced by presolving analytically parts of the system. For more details see [SD04].

(ii) Lost sales: We find explicit solutions of the steady state equations and derive the usual performance measures of the systems and the service levels, mean inventories, service grades, etc. For more details see [SSD$^+$03b].

(iii) In both cases we are then in a position to compute explicitly standard cost functions of the systems which now incorporate costs for inventory, for waiting (both due to slow service and to depleted inventory), lost sales, rejected customers, etc.

Remarkable is that we find an asymptotic decoupling of the queueing behaviour and the storage behaviour in the lost sales case. In terms of the classical performance analysis notions this is a product form steady state distribution [Kel79]. The equilibrium state of the joint (queue length/inventory position)-vector factorizes into a product of two factors, one concerning only the steady state queue length and the other concerning the steady state inventory position.

3. The single server systems with attached inventory which have an explicitly accessible steady state (lost sales case) are then integrated as nodes into networks. We consider standard networks of queues with steady state distribution of so-called product form and integrate a server with inventory into such networks.

We assume lost sales regime when inventory is depleted, but do not assume that the lost sales at the node with inventory are lost to the complete network. This requires to redefine the routing behaviour of the customers which are rejected at the special node due to lost sales and so enabling rejected customers to return later on to that node again. We pursue three different approaches to handle routing in such cases. We derive stationary distributions of joint queue lengths and inventory processes in explicit product form. The stationary distributions are then used to calculate performance measures of the respective systems. For more details see [SSD03a].

References

[BK99]   O. Berman and E. Kim. Stochastic models for inventory management at service facilities. *Comm. Statist.– Stochastic Models*, 15(4):695 – 718, 1999.

[BK01]   O. Berman and E. Kim. Dynamic order replenishment policy in internet-based supply chains. *Mathematical Methods of Operations Research*, 53:371 – 390, 2001.

[BS00]   O. Berman and K.P. Sapna. Inventory management at service facilities for systems with arbitrarily distributed service times. *Comm. Statist.– Stochastic Models*, 16(3 & 4):343 – 360, 2000.

[BS02]   O. Berman and K.P. Sapna. Optimal service rates of a service facility with perishable inventory items. *Naval Research Logistics*, 49:464 – 482, 2002.

[Kel79]   F. P. Kelly. *Reversibility and Stochastic Networks*. John Wiley and Sons, Chichester – New York – Brisbane – Toronto, 1979.

[SD04]   M. Schwarz and H. Daduna. Queueing systems with inventory management under random lead times and with backordering. Preprint, Schwerpunkt Mathematische Statistik und Stochastische Prozesse, Fachbereich Mathematik der Universität Hamburg, 2004. forthcoming.

[SSD03a]   M. Schwarz, C. Sauer, and H. Daduna. Exponential queueing networks with an attached inventory under (r,Q)- or (r,S)-policy. Preprint 3003-09, Schwerpunkt Mathematische Statistik und Stochastische Prozesse, Fachbereich Mathematik der Universität Hamburg, 2003. submitted.

[SSD$^+$03b]   M. Schwarz, C. Sauer, H. Daduna, R. Kulik, and R. Szekli. M/M/1-systems with different inventory management policies and lost sales. Preprint 2003-07, Schwerpunkt Mathematische Statistik und Stochastische Prozesse, Fachbereich Mathematik der Universität Hamburg, 2003. submitted.

[SSL92]   K. Sigman and D. Simchi-Levi. Light traffic heuristic for an M/G/1 queue with limited inventory. *Annals of Operations Research*, 40:371 – 380, 1992.

Advanced Planning Systems

Bernhard Fleischmann

An Advanced Planning System (APS) is a software system that covers all planning tasks along the supply chain (SC) of a manufacturing enterprise. It has an hierarchical architecture composed of modules for the partial planning tasks on
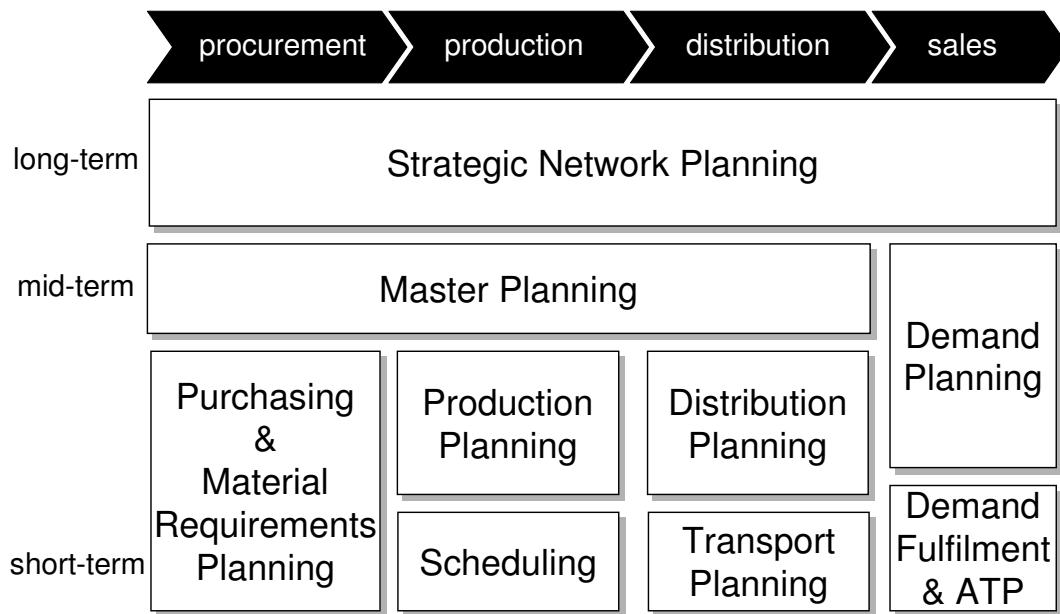
FIGURE 1. APS: General Architecture

the long-term, mid-term and short-term levels in each section of the SC (see Figure 1, taken from Meyr, Wagner and Rohde, 2002). The main "advance" of the recent APS concept consists in the fact, that it adds true planning functions to the traditional ERP systems and that it makes mathematical planning methods available in standard enterprise software.

A review of the tasks of the single modules and the models and methods used in five major APS leads to the following summary. A recent survey is given in the book by Stadtler and Kilger (2002).

The modules with the highest acceptance in practice are: The "Demand Planning" with statistical and consensus methods for sales forecasting; the "Master Planning" that coordinates the aggregate quantities and times of all flows along the SC using LP models; the "Strategic Network Planning" that optimizes the decisions on locations of factories, suppliers and warehouses on the basis of MIP models.

On the short term level, there are still considerable deficiencies of the models and methods applied in the APS: The determination of procurement orders and production lot-sizes is mostly left to the ERP system where it is based on poor single-item rules. Simultaneous lot-sizing and scheduling is missing or done with inadequate methods. In the transport planning, important mid-term decisions on the frequency of the shipments and the transport paths are not supported. The vehicle routing function, which is part of most APS, is of little use, if the transports are outsourced to a logistics service provider. The "Demand Fulfilment", i.e. the fulfilment of the known customer orders, has not found sufficient attention in the planning theory so far. It occurs in every APS, but uses only simple rules for single orders and neglects the interdependence of all previously promised orders in the shortage case.

In addition, there are still structural weaknesses of the APS. The coordination of the single planning modules can be implemented in an APS in many different ways with a strong impact on the total inventory level in the SC. The right way depends on the type of industry and is difficult to design, and there is little support by the APS providers. Stochastic factors are not considered in an adequate way, safety stock calculations, if any, are only provided for single items at single stock points.

References

[1] Fleischmann B. and H. Meyr (2003), Planning Hierarchy, Modeling and Advanced Planning Systems. In: A.G. de Kok and S.C. Graves (Eds.), Supply Chain Management. Handbooks in OR & MS 11, Elsevier, 457 – 523.

[2] Fleischmann B. and H. Meyr (2004), Customer Orientation in Advanced Planning Systems. In: H.Dyckhoff, R. Lackes and J. Reese (Eds.), Supply Chain Management and Reverse Logistics, Springer, 297 – 321.

[3] Meyr H., M. Wagner and J. Rohde (2002), Structure of Advanced Planning Systems. In: H. Stadtler and C. Kilger (Eds.), Supply Chain Management and Advanced Planning. Springer, 2nd ed.

[4] Stadtler H. and C. Kilger (Eds.) (2002), Supply Chain Management and Advanced Planning. Springer, 2nd ed.

Repeated Procurement Auctions: Theory and Computing

S. David Wu

(joint work with S. Özkan, M. Jin, and M. Erkoc)

We study repeated reverse auctions in the context of industrial procurement. A risk-neutral buyer procures materials from competing suppliers through auctions repeated over time. Auctions in this context poses two main challenges as follows: (1) the suppliers' competitive bidding strategies are significantly more complex to analyze than traditional single-round auctions; nonetheless, these strategies affect the equilibrium outcome of the auction, thus the buyer's expected procurement costs, and (2) the suppliers' costs may demonstrate both economies-of-scale and dis-economies-of-scale due to their short-term capacity limitations; this results in a U-shaped cost function for the suppliers, and a winner determination problem that is NP-hard. Motivated by the challenges, we discuss a two-part research agenda that first analyzes the bidders' competitive bidding behaviours and the expected equilibrium outcomes [1], followed by an algorithmic and computational study on the solution of the winner determination problem [2].

Competitive Bidding Analysis

We consider the optimal design of a reverse auction under symmetric incomplete information. There is one buyer and multiple suppliers (bidders) participating in the auction; each supplier is interested in maximizing his own surplus. The supplier knows his own cost structure while assuming an aggregated cost function for the

other bidders. The buyer is interested in minimizing his expected procurement costs. The buyer knows the shape of the suppliers' cost functions, but not the parameters. Defined by the information that is available to them, the suppliers' and the buyer's decision problems dictate their competitive strategies. Extending Myerson's [3] mechanism design framework for single-round auctions, we set out to examine how the bidders' competitive strategies are formed under conditions generalized in the following three dimensions:

1. *Single-Round* (SR) or *Repeated* (RP) auctions. Depending on the auction mechanism and the information released by the buyer, it might be advantages for the suppliers (bidders) to consider their bidding strategies in either a single-round or a repeated auction setting. This may significant complicate the equilibrium analysis of the auction. If not properly designed, the auction mechanisms are subject to undue instability, and the buyer's procurement cost may suffer.

2. *Long-Term Contract* (LT) or *Short-Term Contract* (ST). These are two prevailing forms of procurement auctions. In a *long-term contract auction*, the suppliers must commit to supply for an extended period of time, thus they might act strategically to optimize long-term profit; linear cost is generally considered under this setting. In a *short-term contract auction*, each supplier is only awarded a single-period contract at a time. Since capacity considerations are prevalent in the short-term setting, it is appropriate to consider a U-shaped cost structure for the suppliers.

3. *Deterministic* (DD) or *Stochastic*(SD) demand. The buyer's demand may be available to the bidders in a form that is deterministic or stochastic. This, again, has a direct impact to the bidder's competitive strategy, thus the expected outcome of the auction.

We analyze different settings of procurement auction as defined by the simple taxonomy defined by $(SR$ or $RP) \mid (LT$ or $ST) \mid (DD$ or $SD)$; but not all combinations are meaningful in the procurement context, we thus focus on the following six combinations:

- $SR \mid LT \mid (DD$ or $SD)$
- $SR \mid ST \mid (DD$ or $SD)$
- $RP \mid ST \mid (DD$ or $SD)$

We first analyze the single-round auction settings (i.e., $(SR \mid (LT$ or $ST) \mid (DD$ or $SD)))$ then extend the analysis to the repeated auction settings (i.e., $(RP \mid ST \mid (DD$ or $SD)))$ using the recurrence property. We are primarily interested in comparing the equilibrium results from the settings $(SR \mid LT \mid SD)$ and $(RP \mid ST \mid SD)$, as they represent two distinct procurement philosophies that have significant implications on the stability of supply and the procurement costs. To this end, we analyze competitive equilibrium under the construct of an *incentive compatible* and *individually rational* mechanism; we examine auction design using criteria such as *budget balancedness* and *ex-post efficiency*.

Our analysis [1] shows that the key dimensions identified above have significant impact to the bidder's competitive bidding strategies, thus the stability of

the auction mechanism and the buyer's expected procurement costs. However, the derivation of the closed-form equilibrium bidding strategies in the case of $(RP \mid ST \mid SD)$ involves the solution of differential equations that may be intractable. Using asymptotic analysis, we show that useful insights can be derived concerning the bidders' behaviours. We are in the process of extending this study to consider multi-dimensional auctions [4] where non-price attributes such as quality, delivery performance, or service level might be considered in supplier selection.

The Winner Determination Problem

In this part of the research we focus on the winner determination problem in a multi-unit reverse auctions when the suppliers' costs demonstrate a U-Shaped cost characteristic. As discussed earlier, in a short-term contract auction, the suppliers participating in the auctions are often subject to stringent capacity constraints; thus, the supplier's cost may demonstrate economies-of-scale when bidding on a quantity below his capacity, while demonstrating dis-economies-of-scale when bidding above his capacity. This motivate the study of winner determination problems with U-Shaped cost functions [5]. We show that this particular winner determination problem is NP-Hard even in the case of single-round auctions. We derive two optimality properties from the U-Shaped cost structure that leads to the design of an efficient binary tree algorithm with bounds (BTB). When tested under a variety of demand and supplier combinations, BTB significantly outperforms the general mix-integer programming solver, finding optimal solutions with, on average, 2% of the computer time. The complexity of the algorithm is linear in the number of units to be auctioned, which allows for the efficient handling of high-volume auctions.

We extend our analysis to iterative multi-unit auctions that consist of multiple rounds in which suppliers progressively update their bids [6]. Under the assumption of myopic best response strategies, we show that a slightly modified version of the winner determination model for single-round auctions can be used to address the the iterative auction problems. We discuss auction mechanisms that would improve the buyer's procurement costs in iterative settings. These mechanisms stipulate tie breaking policies that are employed when there are multiple optimal solutions for the winner determination problem. We are in the process of extending the iterative auction results to repeated auction settings.

References

[1] Özkan, S. and S. D. Wu (2004), Competitive Equilibrium Analysis of Repeated Procurement Auctions, Working Paper, Department of Industrial and Systems Engineering, Lehigh University, Bethlehem, Pennsylvania.

[2] Jin, M., S. D. Wu and M. Erkoc (2004), Multiple Unit Auctions with Economy/Diseconomy of Scale, Working Paper, Department of Industrial and Systems Engineering, Lehigh University, Bethlehem, Pennsylvania.

[3] Myerson, R. (1981), Optimal Auction Design, *Mathematics of Operations Research*, 6(1), February 58-73.

[4] Che, Y.-K., Design Competition Through Multidimensional Auctions, *Rand Journal of Economics*, 24(4),Winter, 668-80.
[5] Rothkopf, M.H. (1997), Bidding in Simultaneous Auctions with a Constraint on Exposure, *Operations Research*, 25(4), pp. 620-629.
[6] Gallien, J., L. Wein. (2000), Design and Analysis of a Smart Market for Industrial Procurement, Technical Paper No. 122, MIT Centre of eBusiness, Cambridge,MA.

Designing Practical Market Mechanisms
Zuo-Jun Max Shen
(joint work with Leon Chu)

Many auctions involve the sale/purchase of a variety of distinct assets. For example, in industrial procurement auctions, a buyer may want to purchase different components needed to produce the final product; similarly, sets of furniture or airport time slots are usually offered in a *bundle* that will be sold together. de Vries and Vohra [3] provide an excellent survey on one-sided combinatorial auctions. Recently, more attention has shifted to two-sided, or *double* auctions, which allow both bids and asks. Chu and Shen [1] provide a review of double auctions and propose an Agent Competition Double Auction (AC-DA) mechanism, which is strategy-proof, weakly budget-balanced, and individual-rational for the complementarity-substitutability environment.

Usually the procurement auctions take place in a supply chain setting, since each supplier may be contacted by several buyers and each buyer may also contact several suppliers in order to get the best deal. The problem of determining the production and exchange relationships across a supply chain in response to varying needs, costs, and resource availability is called the *Supply chain formation* problem [4, 5]. This problem is different from the supply chain management problem, where the focus is on optimizing activities such as production, inventory management, and delivery in a fixed supply chain structure.

Extensive negotiations are typically involved to establish an exchange relationship in the supply chain, and the process can be very time consuming. It is difficult to simultaneously negotiate with many supply chain players to find the best business deal. To facilitate this process, online exchange marketplaces have been established recently in more than a dozen major industries. For example, in the automobile industry, a company called Covisint was formed by General Motors, Ford, and Daimler Chrysler in order to reduce the complexity and cost of communicating with customers and suppliers.

We propose two double auction mechanisms that are strategy-proof, (*ex post*) individual rational, and (*ex post*) weakly budget-balanced for a large class of supply chain formation problems that do not satisfy the complementarity-substitutability condition. These mechanisms can also be applied to more general settings with pair-related costs, which no other existing mechanism is capable of doing. Both mechanisms produce higher efficiency comparing with the existing incentive compatible and weakly budget-balanced mechanism. We also want to emphasize the

important contribution of one of the mechanisms from the implementation point of view. This mechanism is based on the solutions to a linear programming problem that possesses the following nice property: the linear relaxation has integer extreme points, thus, we only need to solve a linear programming version of a NP-hard problem.

References

[1] Chu, L. Y. and Z. -J. M. Shen (2003). "Agent Competition Double Auction Mechanism," Working paper, University of Florida.

[2] Chu, L. Y. and Z. -J. M. Shen (2004). "Dominant Strategy Double Auction Mechanisms with Application in Supply Chain Formation Problems," Working paper, University of Florida.

[3] de Vries, S. and R. Vohra (2003), "Design of Combinatorial Auctions", Chapter in *Supply Chain Analysis in the eBusiness Era*, Eds by D. Simchi-Levi, D. Wu, and Z. -J. Shen.

[4] Sadeh, N. M., D. W. Hildum, D. Kjenstad, A. Tseng (1999), "MASCOT: An agent-based architecture for dynamic supply chain creation and coordination in the internet economy", *Production Planning and Control*, Vol. 12 (3).

[5] Walsh, W. E. and M. P. Wellman (2003), "Decentralized Supply Chain Formation: A Market Protocol and Competitive Equilibrium Analysis", to appear in *Journal of Artificial Intelligence Research*.

Collaborative Planning - New Challenges for Mathematicians

Hartmut Stadtler

(joint work with G. Dudek)

SCM is concerned with the coordination of material, information and financial flows within and across often legally separated organizational units. Software vendors have developed so called Advanced Planning Systems (APS) to overcome deficiencies of traditional Enterprise Resource Planning systems and to better support the planning functions needed in SCM. However, APS are based on the principles of hierarchical planning which are well-suited for an intra-organizational Supply Chain (SC) but fall short when non-hierarchical collaboration between partners (companies) is needed. This is particularly true when a buyer and a supplier have to align their medium term order and supply plans - which is the topic of this talk.

Today inter-organizational collaboration between two SC partners - each utilizing their own APS - is enabled by an additional module, called "Collaborative Planning". After introducing its basic reasoning we will propose an alternative, namely "model supported negotiations" at the Master Planning level (Fig. 2).

Negotiations always start with pure upstream planning, i.e. the buyer generates his cost minimal master plan and from that derives the corresponding purchase plan. Then the purchase plan is evaluated by the supplier. However, one can expect that a better master plan exists for the supplier due to the fact that the purchase plan usually prevents the suppler to follow her cost minimal master plan (assuming no restrictions regarding the timing of the purchase orders).
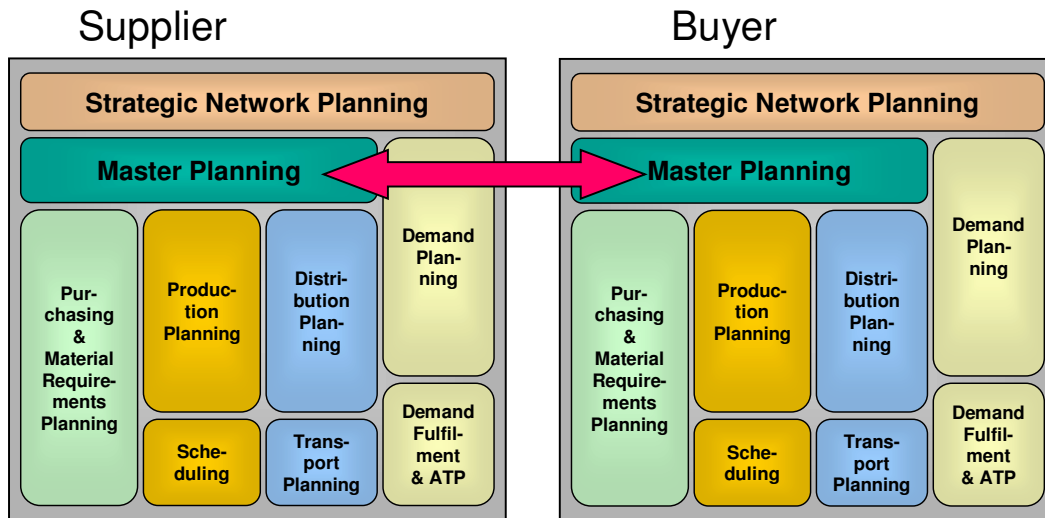
FIGURE 2. Collaborative master planning where both the supplier and buyer make use of their own APS

The main idea is to modify the master planning model of the supplier such that a counterproposal to the purchase plan of the buyer is generated which reduces the cost for the supplier drastically while modifying the purchase plan of the buyer only slightly. This is achieved by utilizing a goal programming approach (see [1] for further details).

Next, the buyer will evaluate the counterproposal of the supplier and ask for a compensation which at least covers the cost increase over the buyer's minimal cost.

It is up to the supplier to decide when to stop negotiations and which plan to accept taking into account the cost of the supplier's master plan plus the compensation to be paid to the buyer. This bilateral negotiation scheme results in near optimal solutions for the SC as a whole and a win-win situation for both partners compared with pure upstream planning.

Our solution approach is a heuristic one without any guarantee for its solution quality. In an abstract form our approach is a horizontal decomposition of a monolithic MIP model for multilevel, multi-item lot-sizing. This view might give rise to alternative (mathematical) approaches to solve the issue (see [2]).

REFERENCES

[1] Dudek, G., Stadtler, H., Negotiation-Based Collaborative Planning between Supply Chain Partners, accepted for publication in EJOR, 2004
[2] Schneeweiss, C., Distributed Decision Making, Berlin et al. (Springer), 2nd ed., 2003

Competition in the Supply Option Market

Victor Martinez de Albeniz
(joint work with David Simchi-Levi)

In Martínez-de-Albéniz and Simchi-Levi [3], we present a framework where an industrial buyer can analyze and optimize portfolios of supply contracts. In this model, the buyer can optimally structure its sourcing channels so that it takes advantage of the flexibility of different fixed-commitment or option contracts. As a result, the manufacturer purchases from each competing supplier a share of capacity that reflects the trade-off between price and flexibility offered by that supplier. Thus, flexibility and price are the two attributes that manufacturers care about. Based on this framework, we can analyze the changes in the way suppliers compete in the marketplace.

In the present research, we assume that suppliers have a differentiated cost structure. They incur an initial cost for reserving capacity and an additional cost for using the capacity to satisfy the buyer's orders. There are a number of fields where this cost specification could be used. For instance, in the electricity industry, different types of power plants exist, from nuclear to coal or gas power plants. Nuclear power plants have a relatively small degree of flexibility in adjusting production level to meet demand and hence all the incurred cost is associated with reserving capacity. On the other hand, gas power plants can adjust the production level rapidly and hence most of the cost is associated with delivering electricity. In manufacturing, and especially in the plastics, chemicals or semi-conductor industries, buyers reserve capacity with suppliers in advance of production time. Of course, different suppliers may have different costs for reserving capacity and delivering supply, depending on the type of technology (machinery) and their geographical location (labor, transportation). Finally, our model may also be relevant to the travel and tourism industry, where the service providers may have different cost structures in terms of capacity reservation cost (cost of leasing airplanes or hotels) and variable cost (operating cost).

These cost characteristics obviously impact the negotiation process. In our model, each supplier offers an option contract to the buyer characterized by two pricing parameters, a capacity reservation fee and an execution fee. Consequently, suppliers can become more competitive by pushing in two directions: either lowering the reservation price or the execution price. The trade-off is clear. A supplier that charges mainly a reservation fee (and a small execution fee) competes on price but not flexibility. On the other hand, a supplier that charges mainly an execution fee (and a small reservation fee) typically emphasizes flexibility and not price.

We describe the market equilibrium outcomes of such system, and in particular the behaviour of market prices for existing supply options. Interestingly, this model is an extension of the Bertrand price competition model to two dimensions. An important result in one dimension is that, in equilibrium, there is a unique supplier, the least costly supplier, that captures all the orders at a market price that is between its cost and the cost of the second most competitive supplier. We show

that this is not the case when two attributes are important to the buyer. Indeed, we demonstrate that in equilibrium, a variety of suppliers coexists, and these suppliers offer different prices. We call this *cluster competition*, since suppliers tend to cluster in small groups of two or three suppliers each, such that within the same group all suppliers use similar technologies and offer the same type of contract.

Most relevant to our model are papers that analyze the behaviour of suppliers in offering options to a buyer, the prelude to introducing competition between suppliers. The existing literature usually models a Stackelberg game where a single buyer is the follower and a single supplier is the leader. Typically, competition in such models is introduced by a spot market. This spot market is the buyer's sourcing alternative and a potential client for the supplier. The focus is on finding conditions for which both players are willing to sign a contract and determining option prices as the outcome of the negotiation process. The first publication in this stream of literature is by Wu et al. [6]. Motivated by electricity markets, they derive option prices as a function of the cost of the system and the elasticity of demand. Later, Spinler et al. [5] and Golovachkina and Bradley [2] analyze models similar to that of Wu et al. Interestingly, there are no papers that directly analyze competition among suppliers since this implies utilizing the notion of portfolio contracts, developed in Martínez-de-Albéniz and Simchi-Levi [3]. Here, we move from the traditional models of competition through dual sourcing, i.e., single supplier offering an option contract versus spot market, to a model of pure competition between suppliers offering different types of options.

We consider a single-period situation, where a single manufacturer looking for supply of a component that is used in the manufacturing of the final product. This component may be obtained from a pool of n suppliers, each of which offers an option contract for the component. Such a contract is defined by two parameters, $v \geq 0$, the reservation price, and $w \geq 0$, the execution price. These values are determined by the supplier based on its cost structure as well as on whether the supplier emphasizes price or flexibility. Specifically, supplier i, $i = 1, \ldots, n$, takes position in the market by offering options at a reservation price v^i and an execution price w^i.

The suppliers' cost structure is assumed to consist of two parts. Each supplier incurs a fixed unit cost for reserving capacity, f^i, $i = 1, \ldots, n$, that can be seen as the unit cost of building a factory of the appropriate size, developing the technology required to produce the component, hiring manpower, or signing its own supply contracts with its suppliers, e.g., the energy provider. In addition, the supplier pays a unit cost, c^i, $i = 1, \ldots, n$, for each unit executed by the buyer. This cost is typically the cost of raw materials and operational costs. These costs differ from supplier to supplier and may be explained by the use of different technologies or management practices.

Therefore, the profit of supplier i, $i = 1, \ldots, n$, is $(v^i - f^i)x^i + (w^i - c^i)q^i$ when a buyer reserves x^i units of capacity and executes q^i units, $q^i \leq x^i$. The objective of the suppliers is to maximize their expected profit by selecting (w^i, v^i) optimally.

On the demand side, we denote by p the selling price to the customer and assume that it is an input, not a decision variable. The total customer demand D follows a distribution with log-concave p.d.f. f, i.e., for all $x, y \in \mathbb{R}_+$, for all $\lambda \in [0, 1]$,

$$f\Big(\lambda x + (1 - \lambda)y\Big) \geq f(x)^\lambda f(y)^{1-\lambda}.$$

As mentioned in Caplin and Nalebuff [1], log-concave distributions include beta, exponential, gamma, Laplace, normal, uniform and Weibull distributions.

We analyze a two-stage model. In the first stage all the suppliers submit bids that are defined by (w^i, v^i), $i = 1, \ldots, n$. At the same time, and based on these bids, the manufacturer decides on the amount of capacity to reserve with each supplier. In the second period, demand is realized and the manufacturer decides the amount to execute from each contract. If total capacity is not enough, unsatisfied demand is lost.

This is a game a la Stackelberg in which the suppliers are leaders and the manufacturer is the follower. Thus, there are multiple leaders that compete knowing the reaction of the follower. Of course, the manufacturer's objective is to maximize expected profit based on the suppliers' bids. Suppliers have complete visibility to the manufacturer decision making process. The costs (c^i, f^i), $i = 1, \ldots, n$, are private information, i.e., each supplier knows only its own cost. In addition, we assume that the suppliers submit sealed bids simultaneously. Thus, this is a one-shot game. Every supplier submits a bid that maximizes its expected profit. We are interested in determining the Nash equilibria of this game in pure strategies, i.e., the n-uples $\Big(w^i, v^i\Big)_{i=1,\ldots,n}$ where no supplier has an incentive to unilaterally change its bid.

Define $f^{n+1} = v^{n+1} = 0$ and $c^{n+1} = w^{n+1} = p$, and let $y^0 = 0$ and $y^i = x^1 + \ldots + x^i$ for $i = 1, \ldots, n$. We denote by $V(y)$ the expected profit of the buyer as a function of y. We have that

$$\frac{dV}{dy^i}(y) = (v^{i+1} - v^i) + (w^{i+1} - w^i)Pr[D \geq y^i].$$

This provides the structure of the manufacturer's optimal portfolio which is determined by the c.d.f. of customer demand. The profit is a strictly concave separable function of $y^1, \ldots, y^n$.

Assuming that the suppliers's costs are such that $c^1 \leq \ldots \leq c^n \leq c^{n+1}$, and

$$1 > \frac{f^1 - f^2}{c^2 - c^1} > \ldots > \frac{f^n - f^{n+1}}{c^{n+1} - c^n} > 0,$$

we characterize the bidding equilibria. First, there exists a Nash equilibrium in pure strategies of the game. In addition, we show that in every equilibrium, we must have clustering of the bids, i.e., supplier i, $i = 1, \ldots, n$, places its bid (w^i, v^i):

- in the segment $[(c^{i-1}, f^{i-1}); (c^i, f^i)]$, and then $(w^i, v^i) = (w^{i-1}, v^{i-1})$;
- or in the segment $[(c^i, f^i); (c^{i+1}, f^{i+1})]$, and then $(w^i, v^i) = (w^{i+1}, v^{i+1})$.

Finally, we provide a bound on the inefficiencies created by suppliers' competition. Define the total welfare as the sum of the profits of suppliers and buyer. The

social welfare is maximized for $U = U^*$, when we consider a centralized supply chain. Of course, when the suppliers compete, the allocation of capacities is not necessarily efficient. We denote by ΔU the loss in welfare due to the suppliers' competition. We show that the distortion on the optimal decisions created by competition among suppliers is bounded: in every Nash equilibrium, we have that

$$\frac{\Delta U}{U^*} \leq \frac{1}{4}.$$

This bound is tight.

REFERENCES

[1] Caplin A. and B. Nalebuff 1991. "Aggregation and Imperfect Competition: On the Existence of Equilibrium." *Econometrica*, **59**(1), pp. 25-59.
[2] Golovachkina N. and J. R. Bradley 2002. "Supplier-Manufacturer Relationships Under Forced Compliance Contracts." Working Paper, Cornell University.
[3] Martínez-de-Albéniz V. and D. Simchi-Levi 2002. "A Portfolio Approach to Procurement Contracts." Submitted.
[4] Martínez-de-Albéniz V. and D. Simchi-Levi 2003. "Competition in the Supply Option Market." Submitted.
[5] Spinler S., A. Huchzermeier and P. R. Kleindorfer 2002. "The Valuation of Options on Capacity." Working paper, the Wharton School.
[6] Wu D. J., P. R. Kleindorfer and J. E. Zhang 2002. "Optimal Bidding and Contracting Strategies for Capital-Intensive Goods." *European Journal of Operations Research*, **137**(3), pp. 657-676.

Impact of Partial Manufacturing Flexibility on Production Variability
ANA MURIEL

As manufacturers in various industries evolve toward predominantly make-to-order production to better serve their customers' needs, increasing product mix flexibility emerges as a necessary strategy to provide adequate market responsiveness. Flexible capacity has been shown to be very effective to hedge against forecast errors at the investment stage. In a make-to-order environment, this flexibility can also be used to hedge against variability in customer orders in the short term. For that purpose, the production levels must be adjusted each period to match current demands, to give priority to the higher margin product and/or to satisfy the closest customer. However, this will result in swings in production, inducing larger order variability at upstream suppliers and significantly higher component inventory levels at the manufacturer. Through a stylized two-plant two-product capacitated manufacturing setting, in Bish, Muriel and Biller (2001) we show that the performance of the system depends heavily on the allocation mechanism used to assign products to the available capacity. While managers would be inclined to give priority to higher margin products or to satisfy customers from their closest production site, these practises lead to greater swings in production, result in higher operational costs, and may reduce profits. In Muriel, Somasundaram and

Zhang (2003) we extend the analysis to general multi-plant multi-product make-to-order manufacturing systems. We develop analytical models and an optimization-based simulation tool to study the impact of flexibility on shortages, production variability, component inventories and order variability induced at upstream suppliers. Our results show that partial flexibility leads to a considerable increase in production variability. As more flexibility is added to the system, however, the production plan will become more stable resulting in a decrease in variability and inventory in the system.

In particular, we show the following results for general multi-plant multi-product production systems. Consider a production system with n plants, each with capacity of C units, and n products with demands denoted by d_i, $i = 1, 2, n$. Consider chain flexibility configurations (see Jordan and Graves (1995)) in which each plant can build h products. We add flexibility by increasing h.

Theorem 1. *If management commits to having sufficient component inventory never to cause shortfall for any potential demand realization, then:*

(1) *System inventory is increasing in the level of flexibility.*
(2) *The system inventory for a chain flexibility configuration in which each plant can build h products equals nhC.*

Theorem 2. *If management commits to having sufficient component inventory never to cause shortfall as long as demand is in a bounded set B characterized by:*

$$\sum_{k=0}^{l} d_{(i+k) \bmod n} \le C \min\{n, l+2\} \text{for } l = 1, 2, \ldots, n-1 \text{ and } i = 1, 2, \ldots, n,$$

$$d_n \le C,$$

then:

(1) *For the dedicated system, system inventory is nC but shortfall will be positive for some $\mathbf{d} \in B$.*
(2) *For the $h = 2$ chain, system inventory is $2nC$ and capacity does not cause shortfall for all $\mathbf{d} \in B$.*
(3) *For the $h = 3$ chain, inventory is $2nC - C$ and capacity does not cause shortfall for all $\mathbf{d} \in B$.*
(4) *The minimum system inventory needed is $2nC - C$ for any flexibility configuration that results in no shortfall for all $\mathbf{d} \in B$.*

Another interesting research question is how the installed flexibility should be managed, that is, how to allocate the demand to the available capacity and flexibility to maximize sales and minimize variability. We show that allocation policies that evenly distribute plant capacity to product demands lead to consistently better performance, since they avoid the misplacement of inventories by replicating the performance of a single-plant system.

Theorem 3. *Consider a fully flexible production system with n plants, each with capacity of C units, and m products with current demands of d_i, $i = 1, 2, m$. Let*

I_i be the on-hand system inventory of component i, for $i = 1, 2, m$, and I_{ij} be the portion of that inventory located in plant j, for $j = 1, 2, n$.

(1) *The distributed policy will have I_i/n components of product i in each plant and will be able to produce a total of*

$$S_1 \equiv \min\{nC, \sum_{i=1}^{m} \min\{d_i, I_i\}\}.$$

(2) *Any other policy will have $I_{ij} \in [0, I_i]$ components of product i at plant j and thus the amount of demand satisfied is at most*

$$S_2 \equiv \min\{nC, \sum_{i=1}^{m} \min\{d_i, \sum_{j=1}^{n} \min\{I_{ij}, C\}\}\} \leq S_1.$$

References

[1] BISH, E., A. MURIEL, and S. BILLER. 2001. "Managing Flexibility in a Make-to-order Environment." Technical report, University of Massachusetts Amherst.

[2] JORDAN, W. C. and S. C. GRAVES. 1995. "Principles and Benefits of Manufacturing Process Flexibility." Management Science 41:4, 577-594.

[3] MURIEL, A., A. SOMASUNDARAM and Y. ZHANG. 2003. "Impact of Partial Flexibility on Production Variability." Technical report, University of Massachusetts.

Polynomial Time Algorithms for Multi-Level Lot-Sizing Problems with Production Capacities

H. Edwin Romeijn

Co-authors: Stan van Hoesel, Dolores Romero Morales, Albert P.M. Wagelmans.

We consider a problem in which production, inventory, and transportation decisions in a basic supply chain are integrated. Traditional models usually consider only one or two of these aspects in isolation from the other(s). Substantial evidence exists (see, for instance, Arntzen et al. [1], Chandra and Fisher [2], Geoffrion and Powers [5], and Thomas and Griffin [8], as well as the references therein) that shows that integrating these decisions can lead to substantial increases in efficiency and effectiveness. Integrating different decisions in the supply chain are particularly important when resources are limited, and when costs are nonlinear, e.g., exhibit economies of scale.

We will consider a serial supply chain for the production and distribution of a product. Such a supply chain will for instance occur when value is added to a product in a sequence of production facilities, and intermediate goods need to be transported between these facilities. Kaminsky and Simchi-Levi [6] describe an example of such a chain as it arises in the pharmaceutical industry. Another example is the third-party logistics industry. In this case, a downstream distribution centre that satisfies demands in a certain geographical area may employ the

services of a third-party warehouse before products are transported to the actual distribution centre for distribution to its retailers. A serial supply chain model can then be used to represent part of a supply chain that is relevant to the distribution centre (see Lee et al. [7]). A final example is a situation in which production takes place at a manufacturer. The items that are produced are then stored at the manufacturer level or transported to the first warehouse level. At each of the warehouse levels, again products are either stored or transported to the warehouse at the next level. From the final warehouse level products are then, after possibly having been stored for some periods, transported to a retailer (possibly allowing for early deliveries, i.e., inventories at the retailer level). Such a structure may arise if a retailer actually represents an entire market, and the supply chain from manufacturer to this market is very long. This could make it advantageous to, in several stages, employ economies of scale by transporting larger quantities over long distances to intermediate storage facilities before being distributed in the actual market.

All situations described above can be represented by a generic model consisting of a manufacturer, several intermediate production or distribution levels, and a level where demand for the end product takes place which we will often refer to as the retailer level (although this does not necessarily represent the level at which actual demand consumption takes place). In fact, in such a model the intermediate production and transportation stages are indistinguishable from one another, so that we will usually simply refer to all intermediate stages as transportation stages between warehouses.

The serial supply chain model sketched above can be viewed as a generalization of a fundamental problem, which in fact is one of the most widely studied problems in production and inventory planning, the *economic lot-sizing problem (ELSP)*. The basic variant of this problem considers a production facility that produces and stores a single product to satisfy known demands over a finite planning horizon. The problem is then to determine production quantities for each period such that all demands are satisfied on time at minimal total production and inventory holding costs. The cost functions are non-decreasing in the amount produced or stored, and are usually assumed to be linear, fixed-charge, or general concave functions. The production facility may or may not face a capacity constraint on the amount produced in each period.

To model the serial supply chain, the classical ELSP can be extended to include transportation decisions, as well as the possibility of holding inventory at different levels in the chain. In addition to production and inventory holding costs, we then clearly also need to incorporate transportation costs, which adds the problem of the timing of transportation to the problem of timing of production. The objective will be to minimize the system-wide cost while satisfying all demand. Even if the manufacturer and retailer are in fact distinct participants in the supply chain, each of which faces a part of the supply chain costs, this problem will be relevant. In this case, the participants clearly still need to decide how to distribute the minimal total costs, which is a coordination problem that is outside the scope of

this talk. But alternatively, we may interpret the holding costs at the retailer level as a penalty or a discount on the purchasing price of an item, which is given by the manufacturer to the retailer if items are delivered early. In this case the costs minimized by our optimization model are all incurred by the manufacturer. As in standard lot-sizing problems, all cost functions are assumed to be non-decreasing in the amount produced, stored, or shipped. In addition, we will assume that all cost functions are concave.

In general, all levels in a serial supply chain, regardless of whether they correspond to production or transportation decisions, may face capacities. We will concentrate on serial supply chains with capacities at the first (production) level only, as a first step towards the study of more general capacitated supply chains. Adding capacities at additional levels appears to significantly change the structure of the problem and thereby the problem analysis. Therefore, such problems remain a topic of ongoing research. Note that, under certain cost structures, it may be possible to eliminate capacitated levels from the supply chain. One such example is provided by Kaminsky and Simchi-Levi [6], who transform a 3-level serial supply chain model with, respectively, a capacitated production level, an uncapacitated transportation level, and a capacitated production level to a 2-level serial supply chain model with capacities at the first level only.

We will call the problem of determining optimal production, transportation, and inventory lot sizes in a serial supply chain under production capacities at the first level the *multi-level capacitated lot-sizing problem (MCLSP)*. In the presence of non stationary production capacities this problem is NP-hard, as it is a direct generalization of the ELSP with general production capacities which itself is NP-hard (see Florian et al. [4]). The ELSP with stationary production capacities, however, is solvable in polynomial time (see Florian and Klein [3]). Since our goal is to identify polynomially solvable cases of the multi-level lot-sizing problem, we will mainly focus on cases where the production capacities are stationary.

We study problems with general concave production, inventory holding, and transportation costs, as well as problems with linear inventory holding costs and two different transportation cost structures. In particular, we will consider special cases with (i) linear transportation costs; and (ii) fixed-charge transportation costs without speculative motives, which means that with respect to variable costs holding inventory is less costly at higher levels than at lower levels in the supply chain. Our solution methods are based on a dynamic programming framework that uses a decomposition principle that generalizes the classical zero-inventory ordering (ZIO) property of solutions to uncapacitated lot-sizing problems as described in Zangwill [10] for the multi-level case, and, for instance, in Wagner and Whitin [9] for the single-level case. Our algorithms all run in polynomial time in the planning horizon of the problem. Moreover, while our algorithm for the case of general concave cost functions is exponential in the number of levels in the supply chain ($O(T^{2L+3})$), it is remarkably insensitive to the number of levels for the two specific cost structures mentioned above ($O(T^5 + LT^2)$ and $O(T^7 + LT^4)$, respectively, where the last running time can be reduced to $O(T^6)$ when $L = 2$).

Open issues for future research in this area can be divided into three general directions. Firstly, the complexities, although polynomial in the planning horizon, are of relatively high order: $O(T^5)$ to $O(T^7)$ for the two-level cases. It would be interesting if the order of the running time could be reduced. In addition, although the number of levels will generally be relatively small, it would nevertheless be interesting to determine if the multi-level case with general concave cost functions can be solved in polynomial time in both the time horizon and the number of levels. A second direction is the study of serial supply chains in the presence of capacities at other or additional levels in the chain. Finally, it would be interesting to consider more complex supply chain structures, including, for example, product assembly structures at the producer level, or multiple retailers.

<div align="center">References</div>

[1] B.C. Arntzen, G.G. Brown, T.P. Harrison, and L.L. Trafton. Global supply chain management at Digital Equipment Corporation. *Interfaces*, 25(1):69–93, 1995.

[2] P. Chandra and M.L. Fisher. Coordination of production and distribution planning. *European Journal of Operational Research*, 72:503–517, 1994.

[3] M. Florian and M. Klein. Deterministic production planning with concave costs and capacity constraints. *Management Science*, 18:12–20, 1971.

[4] M. Florian, J.K. Lenstra, and A.H.G. Rinnooy Kan. Deterministic production planning: algorithms and complexity. *Management Science*, 26(7):669–679, 1980.

[5] A.M. Geoffrion and R.F. Powers. Twenty years of strategic distribution system design: an evolutionary perspective. *Interfaces*, 25(5):105–127, 1995.

[6] P. Kaminsky and D. Simchi-Levi. Production and distribution lot sizing in a two stage supply chain. 35(11):1065–1075, 2003.

[7] C.Y. Lee, S. Çetinkaya, and Jaruphongsa. A dynamic model for inventory lot sizing and outbound shipment scheduling at a third-party warehouse. *Operations Research*, 51(5):735–747, 2003.

[8] D.J. Thomas and P.M. Griffin. Coordinated supply chain management. *European Journal of Operational Research*, 94(1):1–15, 1996.

[9] H.M. Wagner and T.M. Whitin. Dynamic version of the economic lot size model. *Management Science*, 5:89–96, 1958.

[10] W.I. Zangwill. A backlogging model and a multi-echelon model of a dynamic economic lot size production system – a network approach. *Management Science*, 15(9):506–527, 1969.

<div align="center">**Production Planning Models and MIP**

Laurence A. Wolsey</div>

This survey is based on work carried out by many people, but Yves Pochet and Mathieu van Vyve in particular have made major contributions.

To introduce the decomposition approach used, we first present a formulation of the most basic "hard" problem, the multi-item, multi-period lot-sizing problem with a joint capacity constraint in each period. This can be viewed as the intersection of single item lot-sizing sets, and single period single node flow sets. The approach is then to characterize exactly or approximately the convex hull of these "simpler" sets. As we wish to solve problems directly by mixed integer programming (MIP) using a default branch-and-cut approach, we either need to represent

the convex hulls by valid inequalities and a separation algorithm, or through an extended formulation involving additional variables. The latter, when it can be found, typically has the advantage of being "polynomial" in size, but it can still be very large, i.e. $O(n^2)$ or $O(n^3)$ where n is the number of periods. The uncapacitated lot-sizing problem is used to provide examples of a convex hull reformulation both with valid inequalities and with an extended formulation. Details of many of the results can be found in [3, 4].

To effectively make use of the results on formulations available in the literature, we propose a classification of lot-sizing problems [5].

Single item problems are classified according to three parameters: $PROB-CAP-VAR$ where
$PROB$ denotes the problem type: standard, with Wagner-Whitin (non-speculative) costs, or full capacity (discrete) production
CAP indicates the item capacity constraints: uncapacitated, constant capacity, or arbitrary, and
VAR indicates variants: backlogging, start-up cost or times, safety stocks, sales, etc.
Here many tight formulations both with inequalities and extended formulations are known in the case of constant or unlimited capacity. Tables are presented based on the single-item classification showing precisely what results are available.

Single period joint resource constraints are classified in a similar fashion, so as to allow access to improved formulations. A distinction is made between the mode constraints (MC) which determine the constraints on the discrete set-up, start-up and changeover variables, such as a restriction to one start-up per period, and the joint production capacity constraints (RC) in which start-up or cleaning times, changeover times, etc., may reduce the available joint production capacity in the period.

In the past, we have proposed several prototype systems, general purpose BC-OPT [2], and specific to production planning BC-PROD [1]and BC-LS, providing certain automatic reformulation possibilities. What we now propose for production planning MIPs is a library of reformulations written in the modelling language used to represent the underlying production planning problem as a mixed integer program. Thus using the XPRESS modelling language MOSEL, the user is just required to model his problem instance, to classify it, and then we provide a library of black-box subroutines allowing him to automatically improve his formulation either with cutting planes or with an extended formulation. An effective relax-and-fix (or time decomposition) primal heuristic can also be called in a similar fashion.

Finally computational results using this black-box approach are presented for

three applied problems demonstrating both the reformulation and heuristic routines. Two of the problems are multi-level, a typical two-level mixing and packing model, and a six level assembly model, whereas the third problem concerns the packing line for a well-known supermarket product.

Several important open questions are concerned with **Modelling:**
Finding tighter formulations for Resource Capacity RC constraints?
Using approximation parameters based on product characteristics?
Extensions to include transportation to client demand areas - path inequalities, mixing inequalities - how to make good choices?
Using Discrete Lot-sizing or Big Bucket RC Models?
and others with solving **MIPs:**
How to use good heuristic solutions to improve branch-and-cut lower bounds?
How to combine relax-and-fix heuristics with RINS?

REFERENCES

[1] G. Belvaux and L.A. Wolsey, BC-PROD: A specialized branch-and-cut system for lot-sizing problems, Management Science 46, 724-738 (2000).
[2] C. Cordier, R. Laundy, H. Marchand and L.A. Wolsey, bc-opt: a Branch-and-Cut Code for Mixed Integer Programs, Mathematical Programming 86, 335-354 (1999).
[3] Y. Pochet and L.A. Wolsey, Algorithms and Reformulations for Lot-Sizing Problems, in Combinatorial Optimization, eds. W. Cook, L. Lovasz and P. Seymour, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, Vol 20, American Mathematical Society 1995.
[4] M. Van Vyve, A solution approach to production planning problems based on compact formulations for single-item lot-sizing models, Ph.D. thesis, Faculté des Sciences Appliquées, Université catholique de Louvain, 2003.
[5] L.A. Wolsey, Solving Multi-Item lot-sizing problems with an MIP solver using classification and reformulation, Management Science 48, 1587-1602 (2002).

New Modelling Approaches for Process Industries

CHRISTOPHER SUERIE

Process industries differ from discrete manufacturing industries in many aspects. These aspects, e.g. campaign production, production in batches or long setup times have to be considered when planning for production. This is especially true if time-indexed modelling approaches are used which are commonplace, because time-indexing often introduces a representation defect into the planning system.

Representation defect of a time-indexed model means that the optimal solution with respect to a continuous time scale is not feasible in the time-indexed setting. Representation defects occur within standard lot-sizing models (e.g. CLSP; [1]) with respect to (a) setup states at period boundaries, (b) lot sizes which span over two (or more) periods and (c) setup times which do not lie completely in one period.

We consider the representation defects due to (a) first. In a standard big-bucket lot-sizing model (e. g. CLSP) an implicit assumption is made, that whenever there is production in a certain period a corresponding setup operation has to be performed. In reality, this is not true, because in many cases production of a certain product can continue at the beginning of period $t+1$ if it was going on at the end of period t without the necessity of a new setup. This is obvious in a 24/7 environment where the start of the week is a rather arbitrary point in time.

Consequently, models have been tackled in literature which explicitly consider the setup state at each period boundary (e. g. CLSPL; [5]). On the other hand, small-bucket models (e. g. DLSP, CSLP, PLSP; [2]) do not require any new modelling feature here, because there the preservation of setups from one period to the next is a fundamental modelling paradigm.

With respect to lot sizes both small-bucket and big-bucket models require alterations. With the data given in Table 1, Figure 3 shows optimal solutions, if different restrictions on lot sizes are imposed. In this example, the PLSP (e. g. [2]) is chosen as a basic model. In a time-indexed setting, the difficulty arises that, for example, it is not known that the first lot size of product $j = 2$ is 45, but rather, that production of $j = 2$ is 20 in $t = 1$ and 25 in $t = 2$ (Figure 1, PLSP).

TABLE 1. Data for example ([3])

Prod.	Demand d_{jt}				Prod. coeff.	Setup time	Avail. cap.	Hold. cost	Setup cost
j	$t = 1$	$t = 2$	$t = 3$	$t = 4$	a_j	st_j	c_t	h_j	sc_j
$j = 1$	0	25	30	35	1	10	80	1	10
$j = 2$	20	25	30	35	1	10		1	10

Figure 1 shows how the optimal solution changes, if a minimal lot size requirement of 50 units is introduced (MIN), a maximal lot size requirement of 60 units is introduced (MAX), production is only allowed in multiples (batches) of 20 units (BATCH) or production is only allowed in multiples of 20 units and demand can only be fulfilled from completed batches (BATCH flow). Note, that in this illustrative example the proposed restrictions are not applied to the last lot within the planning horizon, as it is assumed that this lot fulfils the requirement later.

In [3] a MIP model formulation based on time discretization for these kind of model is proposed, which is shown to clearly outperform similar approaches from literature.

The last representation defect concerning setup times which do not lie completely in one period is tackled in [4]. There, a MIP model formulation is presented which allows for setup times which span over (multiple) period boundaries.

Combining all these features (modelling setup states at period boundaries, considering period overlapping lot sizes and period overlapping setup times), the resulting model formulation is capable of representing any plan that is possible on a
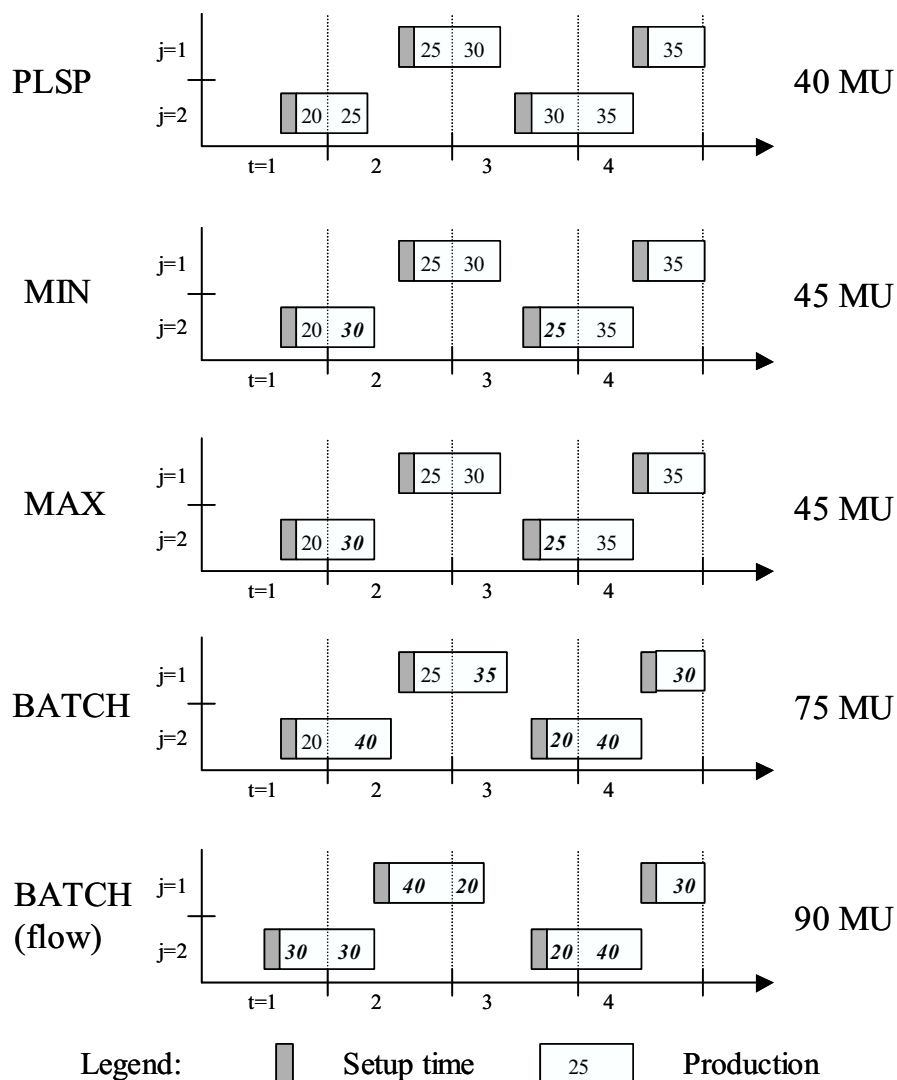
FIGURE 3. Optimal solutions for example ([3])

continuous time scale within a time-discretized setting. Thereby the representation defect of time-indexed model formulations is overcome.

REFERENCES

[1] Dixon, P.S., Silver, E.A. (1981): A heuristic solution procedure for the multi-item, single level, limited capacity lot sizing problem, Journal of Operations Management, 2, 23-39.

[2] Drexl, A., Kimms, A. (1997): Lot sizing and scheduling – survey and extensions. European Journal of Operational Research, 99, 221-235.

[3] Suerie, C. (2004): Campaign planning in time-indexed model formulations, International Journal of Production Research, accepted for publication.

[4] Suerie, C. (2004): Modelling of period overlapping setup times. Working paper, Schriften zur Quantitativen Betriebswirtschaftslehre 1/2004, Technische Universität Darmstadt, Institut für Betriebswirtschaftslehre.

[5] Suerie, C., Stadtler, H. (2003): The capacitated lot-sizing problem with linked lot-sizes, Management Science, 49, 1039-1054.

Characterizing Flows through Components in Assembly Networks

S. Thomas McCormick

(joint work with Maurice Queyranne)

Dedicated to Alan J. Hoffman's 80th birthday

Suppose that our company builds personal computers (PCs) from various components. Often manufacturers have a choice of several mostly equivalent components such as hard disk drives, motherboards, CD burners, etc. Most components are compatible with each other, but there are a few incompatible components which cannot feasibly be assembled into a final product.

Ball et al. [2] propose a graph model for this where each component is a node, and there is an edge between incompatible components. We denote the set of such *incompatibility edges* by I. The components are grouped into *layers* $L_1, L_2, \ldots L_k$ such that a final product must contain exactly one component from each layer. Denote the set of all edges from L_i to L_{i+1}, $i = 1, \ldots, k-1$, by A'. Then $A \equiv A' - I$ represents the set of components on successive layers which are compatible.

Now consider L_1 to be a *source* set S, and L_k to be a *sink* set T. If each edge of I connects components of successive layers, then the set of feasible final products corresponds exactly to the set of paths $\mathcal{P}$ from a node in S to a node in T using only edges of A. Denote the number of components by $n = \sum_j |L_j|$.

Then a vector $y \in \mathbf{R}^{\mathcal{P}}$, where for $P \in \mathcal{P}$ we have $y_P \geq 0$ giving the production for final product P, completely specifies a production plan. The usage of component i in this production plan is $x_i = \sum_{P \in \mathcal{P} : i \in P} y_P$. If we define $M \in \mathbf{R}^{n \times \mathcal{P}}$ to be the 0–1 node-path incidence matrix of the graph, then we have that $x = My$. When $|I|$ is small, $|\mathcal{P}|$ is almost $\prod_j |L_j|$, which can be exponential in n. For modelling purposes, it would be much more convenient to work with the variables x, whose size is only n.

But this raises the question: Given $x \in \mathbf{R}^n$, when is there a $y \in \mathbf{R}^{\mathcal{P}}$ such that $x = My$? For the given structure, Ball et al. [2] give a complete set of inequalities answering this question.

However, not every PC contains a CD burner. Thus Queyranne [5] proposed generalizing this model to allow edges in A that skip layers. In this case it is simpler to forget the layer structure entirely, but to now insist that the connections between component nodes are directed, and that the directed graph is *acyclic*, i.e.,

contains no directed cycles. We continue to have a set S of source nodes, and a disjoint set T of sink nodes, and we interpret the set A of arcs such that every path P from a source node to a sink node (we again denote the set of such paths by $\mathcal{P}$) represents a feasible final product.

Keeping M as the node-path incidence matrix, we can again ask the question: Given $x \in \mathbf{R}^n$, when does there exist a $y \in \mathbf{R}^{\mathcal{P}}$ such that $x = My$? This question was partially answered by [5]. A preliminary version of the current paper giving a full answer is in [4].

It is natural to want to extend this model in two directions: (1) What if we allow the graph to contain directed cycles? (2) What if we have upper bounds u_{ij} on the flow through arc $i \rightarrow j$, i.e., require that $\sum_{P \in \mathcal{P}: i \rightarrow j \in P} y_P \leq u_{ij}$, and/or require that $x_i \leq u_i$? This paper extends [4] in both of these directions.

To deal with graphs with directed cycles, for each $j \in T$ and $i \in S$ add a *return arc* $j \rightarrow i$ to A, thereby completing a cycle C for each path from i to j. Since the original graph was acyclic, these are the only cycles created by the return arcs. Note that this device allows some modelling flexibility: we need not put in return arc $j \rightarrow i$ if final products containing both j and i are not feasible. We now denote the set of directed cycles by $\mathcal{C}$ and index y by $C \in \mathcal{C}$, and consider M to be the node-cycle incidence matrix.

Define $Q^u = \{x \in \mathbf{R}^n \mid 0 \leq x \leq u, \ \exists y \in \mathbf{R}^{\mathcal{C}}$ with $\sum_{C \in \mathcal{C}: i \rightarrow j \in C} y_C \leq u_{ij}$, and $x = My\}$. We use Q^∞ when all u_{ij} and u_i are ∞. Our task is to characterize Q^u. To do this we consider and solve each of the following problems:

1. Separation Problem: Given some $\bar{x} \in \mathbf{R}^n$, either prove that $\bar{x} \in Q^u$, or find some *cut* $\alpha^T x \geq \beta$ that is satisfied by every $x \in Q^u$, but violated by $\bar{x}$, i.e., such that $\alpha^T \bar{x} < \beta$.

We solve this by splitting each node i into i and i' with a new arc $i \rightarrow i'$ with bounds $l_{ii'} = u_{ii'} = \bar{x}_i$. Then it is easy to see that $\bar{x} \in Q^u$ iff there is a feasible flow in this extended network. This can be decided using one max flow (see, e.g., [1]). When there is no feasible flow then as in Hoffman's Circulation Theorem [3] the associated min cut gives us disjoint subsets (possibly empty) of the (original) nodes J and K, and a subset of the (original) arcs U, defining the *Hoffman cut* $x(J) \leq x(K) + u(U)$. This proves a conjecture of [5] that every facet of Q^∞ has $0, \pm 1$ components.

2. Validity Problem: Given a proposed cut $\alpha^T x \geq \beta$, either prove that it is valid for all $x \in Q^u$, or find some $\bar{x} \in Q^u$ such that $\alpha^T \bar{x} < \beta$.

We solve this by using the same split-node network. This time we put cost α_i on new arc $i \rightarrow i'$ and cost 0 on all other arcs, and we do a min-cost circulation on the network that effectively minimizes $\alpha^T x$. If the optimal value $z^* \geq \beta$, then the cut is valid, else the optimal flow yields an $\bar{x}$ violating the cut. Note that we can really restrict our attention to just the $0, \pm 1$ Hoffman cuts, in which case there are min-cost flow algorithms with faster run times [1].

3. Dimension of Q^∞: Compute the dimension of Q^∞. Then it is easy to see that, when $u > 0$, $\dim(Q^u) = \dim(Q^\infty)$.

We again use the node-split graph, but this time we delete all the $i \to i'$ arcs. It can be shown that any implicit equality of Q^u induces a node subset W of this graph with no arcs crossing its boundary in either direction. Hence each such W must be a union of connected components of the graph. Conversely, each connected component yields an implicit equality satisfied by every $x \in Q^u$, and if there are q connected components, then $q - 1$ of these equalities are linearly independent. This shows that $\dim(Q^\infty) = \dim(Q^u) = n - (q - 1)$.

4. Dimension of induced face: Given some valid cut $\alpha^T x \geq \beta$, compute the dimension of the face F of Q^u it induces. Solving this allows us to characterize which valid cuts are facets of Q^u.

Here we combine the techniques of the previous two items, and we restrict w.l.o.g. to Hoffman cuts $x(J) \leq x(K) + u(U)$. We again use the node-split graph, and put cost -1 on arcs $i \to i'$ with $i \in J$ and costs $+1$ on arcs $i \to i'$ with $i \in K$, costs 0 elsewhere, and do a (cheap) min-cost flow. We can determine which arcs have their flows fixed at a bound in every optimal solution. Delete all such arcs except those needed to keep the same number of connected components. Then delete all $i \to i'$ arcs. Finally, compute the number of connected components q_F of the remaining graph. Then the dimension of F is $n - (q_F - 1)$. Thus $x(J) \leq x(K) + u(U)$ is a facet iff $q_F = q + 1$.

5. Separation to a facet: Given some $\bar{x} \notin Q^u$, find a cut $\alpha^T x \geq \beta$ separating x from Q^u that induces a facet of Q^u.

Hoffman cut $x(J) \leq x(K) + u(U)$ reveals that it is not a facet by having too many connected components in the previous algorithm. In this case we can use these connected components to decompose the cut into a sum of Hoffman cuts on smaller node subsets (at least one of which must also be violated), and recurse until we obtain a facet. We conjecture that in fact each cut from the decomposition must already be a facet.

Finally we extend our results also to cases with lower bounds l. In this case a Hoffman cut can have a subset L of original arcs and becomes $x(J) + l(L) \leq x(K) + u(U)$.

References

[1] Ahuja, R.K., T.L. Magnanti, and J.B. Orlin (1993). *Network Flows: Theory, Algorithms and Applications.* Prentice Hall, New York, NY.
[2] Michael O. Ball, Chien-Yu Chen, and Zhen-Ying Zhao (2003). Material compatibility constraints for make-to-order production planning. *OR Letters*, **31**, 420–428.
[3] Hoffman, A.J. (1960). Some Recent Applications of the Theory of Linear Inequalities to Extremal Combinatorial Analysis. In Bellman, R. and Hall, Jr., M. (eds.) *Proceedings of Symposia in Applied Mathematics*, Vol. X, *Combinatorial Analysis*. American Mathematical Society, Providence, RI, pp. 113–127.
[4] S. T. McCormick and M. N. Queyranne (2003). Le Cône des Flots aux Noeuds dans un Réseau Acyclique. Proceedings of *Journées sur les Polyèdres en Optimisation Combinatoire* at Clermont-Ferrand (in French).
[5] M. N. Queyranne (2003). On the Node Flow Cone of an Acyclic Directed Graph. Seventh Aussois Conference of Combinatorial Optimization.

Optimization in SAP Supply Chain Management

Heinrich Braun

In the first part I introduce the planning models of Supply Network Planning and Detailed Scheduling. For mastering the algorithmic complexity I present the various implemented strategies for aggregation and decomposition integrated in a modular optimization architecture like a tool box. The main focus of our development was to solve larger problems with higher quality and more functionality whereas now the focus changed in favour to master the solution complexity: Given the planners detailed explanation about reasons of delays, non-deliveries, additional shifts/overtime etc.

In the second part I discuss the various difficulties in practise in order to achieve the acceptance by the customers, including the different involved rules: OR-Specialists, IT-Department, Consulting and last but not least the planner as the end user.

Integrated Optimization of School Starting Times and Public Bus Services

Armin Fügenschuh

(joint work with Alexander Martin and Peter Stöveken)

The optimization of public bus services in rural areas is mostly an optimization of the traffic caused by pupils on their ways to school and back home, because they are the largest group of customers. Beside the morning and afternoon peaks, there is a much lower demand for public buses over the rest of the day. The counties in Germany we focus on are rural, for their population density is rather low. The biggest city has no more than, say, 30,000 inhabitants, around 150,000 people live in an area of about 1,000 square kilometres. More than half of all pupils are coming to school by public transit, that is, about 10,000 pupils take the bus to 100 different schools. The average way to school has a length of around 10 km, a few pupils travel even more than 30 km twice a day.

It was noted by the consulting company *BPI-Consult*, a subsidiary of the Finnish *Jaakko Pöyry*, that a significant lower number of buses is needed, if the bus scheduling problem is solved together with the starting time problem, i.e., the simultaneous settlement of school and trip starting times [11]. Since then, BPI successfully consulted several counties, where they were able to find solutions which reduce the number of buses by 15 – 20%. Moreover, BPI does not only present a solution, instead they accompany the whole embedding process, including negotiations with all participating groups (bus companies, pupils, parents, teachers, schools, and the county government). Within this process it is sometimes necessary to re-optimize the problem, when new, previously unknown constraints emerge. However, their solutions are currently generated manually. Generating solutions manually is a difficult task, even if we only concentrate on the morning

peak, which gives a small planning horizon from 5:30 – 9:00 a.m.. Thus, the idea of an automatic planning tool was born.

A wide range of transportation problems involving public bus transit, pupils and/or schools were already studied before, see [5, 7, 6, 8, 10], to name just a few. However, none of the presented models completely fits to our problem, mainly for some or all of the following reasons. The time windows of school starting times are fixed and cannot be changed to save buses. In all modelling approaches, pupils are always transported directly to school, and changing the bus is not allowed. Locating bus stops, designing routes (trips) and assigning pupils to routes is sometimes part of the optimization, but for us these are input figures. Finally, scheduling drivers is not an issue for us: Since our time horizon is small (from 6:00 till 8:30), no planning of breaks is needed.

Turning the laws and administrative regulations (see [1, 2, 3], for instance) into an optimization model, we identified (after several discussions with BPI-Consult) the following variables, constraints and objectives. We focus on the following degrees of freedom (variables):

- The schedules of the buses,
- the starting times of the bus trips, and
- the starting times of the schools.

No other possible variables are issued, for example, planning the routes of the bus trips, or locating the bus stops. Moreover it is required that all pupils are using the same bus trips for their ways to school as they do it today. For BPI-Consult, the decision on which variables the focus should lie, is mainly a political one. Changing the starting times of schools and bus trips causes already enough public opposition. For example, if some school doesn't start at, say, 7:40, but at 8:30, then the pupils leave home nearly one hour later, which might cause troubles for working parents. Therefore, changing school starting times in a whole county at once is a very delicate issue, and political skills are needed to implement any given solution.

The decision variables are not independent from each other, they are coupled by the following constraints:

- The (legal) bounds on the school start (7:30 – 8:30 a.m.),
- lower and upper bounds on the waiting time for pupils at the school,
- bounds on the waiting time for pupils while transferring from one bus trip to another,
- bounds on the starting time of trips.

There are several conflicting goals that have to be addressed by the optimization. In particular, we want to minimize the following:

- The total number of deployed buses,
- the time for driving deadhead-trips,
- the standing times of buses between two trips,
- the absolute change of the schools' starting times,
- the absolute change of the starting times of the bus trips,

- the waiting times for pupils at their schools, and
- the waiting times at a transfer bus stop.

Thus it turns out that the integrated optimization of bus schedules and school starting times is a multicriterial discrete linear optimization problem.

The problem can be modelled as a mixed-integer programming problem, based on well-known models for the classical vehicle routing problem with time windows (see [9], for instance). Due to the strict time limits on the minimum and maximum waiting times for pupils at their school, additional coupling constraints are introduced to this model.

Unfortunately, even state-of-the-art MIP solvers are not able to solve this model to optimality. Feasible solutions found by the MIP solver are of poor quality (i.e., a high number of deployed buses). Thus, we developed a greedy-type heuristic, that uses techniques from mixed-integer preprocessing as a main feature. The heuristic repeats the following steps to fix the binary variables of the model to their lower or upper bounds.

In each step, a local-best deadhead trip is selected. The criterion which deadhead trip is more promising than the others is based on its length (shorter deadhead trips are preferred) and also on the change of time windows, once the two corresponding trips are connected (deadhead trips that do not change time windows are preferred). After selecting a deadhead trip, it must be checked, whether the connection is feasible. In mathematical terms, this question can be decided by checking feasibility of a suitable integer program with at most two non-zeros per inequality. By a result of [4], this problem can be solved in pseudo-polynomial time $O(mU)$, where m is the number of inequalities (i.e., the number of schools plus the number of trips) and U is the largest difference between upper and lower bound on each variable. In our application, every school and every trip must start some-when between 5:30 and 9:00 a.m., therefore we can check feasibility even in polynomial time $O(m)$. When the deadhead trip turns out to be feasible, further strenghtenings on the bounds of time variables and fixings of binary variables can be deduced.

In total, our heuristic produces a feasible solution in time $O(|A| \cdot m)$, where A is the set of all deadhead trips. In practice, the presented heuristic is able to produce good feasible solutions (in terms of few buses) after a very short amount of time (typically, less than 3 seconds on a modern computer).

References

[1] Schulgesetz Mecklenburg-Vorpommern vom 15.05.1996. Stand Feb. 2000, 68. (In German).

[2] Satzung über die Anerkennung der notwendigen Kosten für die Schülerbeförderung des Landkreises Demmin. Aus: Kreisanzeiger des Landkreises Demmin, Nr. 6/97. Amtliche Mitteilungen und Informationen, 7 – 8. (In German).

[3] Hinweise zur Schulorganisation für allgemein bildende Schulen. Verwaltungsvorschrift des Ministeriums für Bildung, Wissenschaft und Kultur vom 21. Juli 2000. Aus: Mitteilungsblatt des Ministeriums für Bildung, Wissenschaft und Kultur Mecklenburg-Vorpommern, Nr. 8/2000, 362 – 365. (In German).

[4] Bar-Yehuda R., Rawitz D. (2001), Efficient algorithms for integer programs with two variables per constraint. *Algorithmica* **29** (4), 595 – 609.

[5] Bodin L., Berman L. (1979), Routing and Scheduling of School Buses by Computer. *Transportation Science* **13** (2), 113 – 129.

[6] Bowerman R.L., Hall G.B., Calamai P.H. (1995), A multi-objective optimisation approach to school bus routing problems. *Transp. Research A* **28** (5), 107 – 123.

[7] Braca J., Bramel J., Posner B., Simchi-Levi D. (1997), A Computerized Approach to the New York City School Bus Routing Problem. *IIE Transactions* **29**, 693 – 702.

[8] Corberan A., Fernandez E., Laguna M., Marti R. (2000), Heuristic Solutions to the Problem of Routing School Buses with Multiple Objectives. Technical Report TR08-2000, Dep. of Statistics and OR, University of Valencia, Spain.

[9] Cordeau J.-F., Desaulniers G., Desrosiers J., Solomon M., Soumis F. (2002), VRP with Time Windows. In: Toth P., Vigo D. (Eds.), The Vehicle Routing Problem. SIAM Monogr. on Disc. Math. and Applications. SIAM, Philadelphia.

[10] Desaulniers G., Lavigne J., Soumis F. (1998), Multi-depot vehicle scheduling with time windows and waiting costs. *European Journal of Operational Research* **111**, 479 – 494.

[11] Stöveken P. (2000), Wirtschaftlicherer Schulverkehr: ÖPNV-Optimierung mit erfolgsabhängiger Honorierung. *Der Nahverkehr* **3**, 65 – 68. (In German).

Connecting mrp, MRP II and ERP - Supply Chain Production Planning via Optimization Models

Stefan Voss

(joint work with David Woodruff)

Supply chain management (SCM) rose to prominence as a major management issue in the last ten to fifteen years. While the focus of managing supply chains has undergone a drastic change as a result of improved information technology, production planning and logistics remain critical issues. Recently, we have investigated optimization models that correspond to mrp (materials requirements planning) and MRP II (manufacturing resources planning) and end up as a basis for useful SCM planning models. In order to develop those models, we need to consider extensions leading away from mrp to models that bear little resemblance to mrp but are able to model reality in a more appropriate way. One of the not yet fully understood extensions in this respect are load dependent lead times.

Enterprise resource planning (ERP) may be referred to as an integrated software application designed to support a wealth of business functions. After the early focus on mrp switched over to MRP II, later it was extended to cover enterprise-wide business functions from finance, human resources and the like. The basic functions incorporated into ERP, however, still heavily rely on mrp and MRP II. And even if so-called advanced planning systems (APS), incorporating powerful planning procedures and methodologies together with concepts for dealing with exceptions and variability, are being developed this has not yet changed (for popular books on ERP and APS see, e.g., [1, 2]).

The focus of managing supply chains has changed considerably due to an increased availability of information technology. However, basic issues such as production planning remain critical. Long before the words "Supply Chain" were popular, mrp was seen as the best possible way to do production planning. While mrp and MRP II may be regarded as philosophy, it is important to bridge the

gap between such philosophy and advanced mathematical programming models in production planning and SCM. In the last couple of years we have investigated related mathematical models in the context of supply chain planning [3]. In this respect, a reasonable way is to begin with optimization models that map to mrp and MRP II and end up as some basis for advanced planning models. Based on this, connections with hierarchical systems for planning and detailed scheduling are possible.

Extensions of our models incorporate, e.g., load dependent lead times. Let us define the *lead time* as an estimate of the time between the release of an order to the shop floor or to a supplier and the receipt of the items. Typically, these times are composed of processing, waiting and transportation times, and they are assumed as given data [4]. It is useful to make a distinction between lead times for due dates and lead times for planning as is done by [5]. They are concerned primarily with the former, which is the set of times published for use by customers. We are concerned primarily with the latter, which are used for planning production. These may also be referred to as *waiting times*. That is, as organizations move from creating plans for individual production lines to entire supply chains it is increasingly important to recognize that decisions concerning utilization of production resources impact the lead times that will be experienced. One may gain insights into why this is the case by looking at the queueing that results in delays. Then these insights may be used to investigate optimization models that take into account load dependent lead times and routing alternatives. Our work can be seen as making use of so-called clearing functions for the trade-off between loading and waiting time with extensions to include multiple routings or subcontractors. Alternatively, our work can also be seen as extending the category denoted as multiple-stage production planning with limited resources to include routing flexibility and load dependant lead times.

Planning systems such as mrp or MRP II are usually based on an assumption of lead times that depend only on the stock keeping unit (SKU) and have been roundly criticized in the academic literature (see, e.g., [6]). In fact, we can see that under reasonable assumptions the lead time depends less and less on the SKU as the throughput grows. The expected lead time depends more and more on the expected length of the queues.

In a complex supply chain using modern practises, which include small move batches, one can reasonably justify a Poisson model of the arrival process, at least as a rough approximation to gain insight. Work arrives from a variety of sources in a fairly random pattern. We adopt the convention that the processing time for the homogeneous family is one time unit, so that the expressions are a bit simpler. If a particular SKU, i, is added to the routing we can analyze the expected queueing time under the assumption of a Poisson arrival process. Suppose that the expected processing rate for the new SKU i (using the specified time units) is μ_i, then since it is well known that Poisson arrivals see time averages we can make use of standard
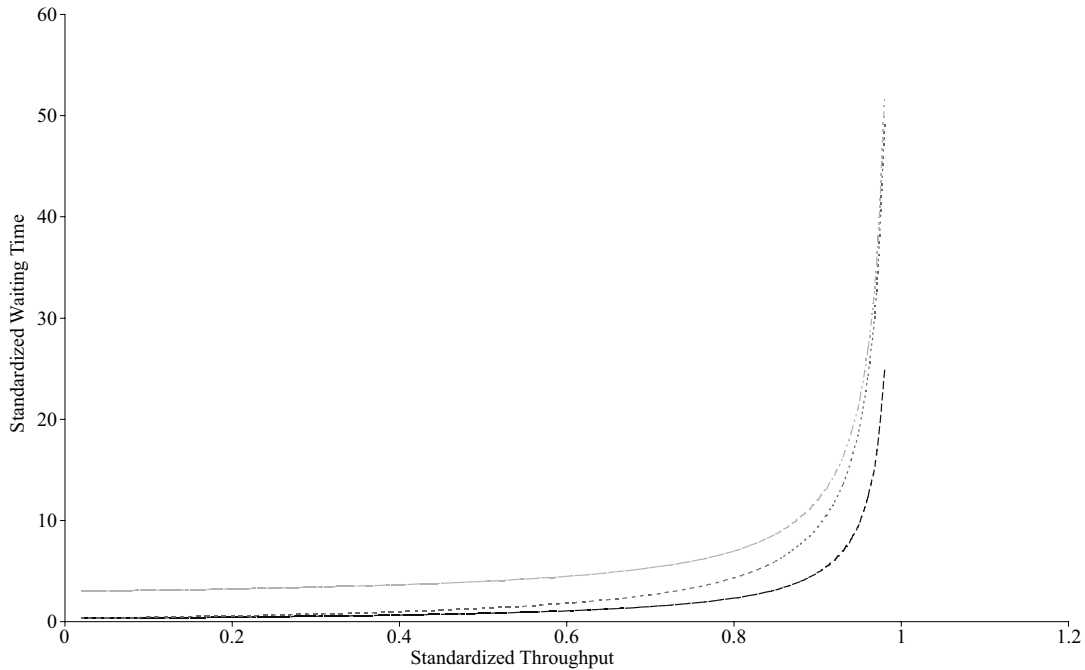
FIGURE 4. Waiting Time as a function of arrival rate.

queueing formulas to write the expected waiting time for SKU i as

$$(19) \qquad \frac{1}{\bar{\mu}_i} + \frac{\bar{\lambda}\sigma^2 + \bar{\lambda}}{2(1 - \bar{\lambda})}$$

where σ is the standard deviation of the service times.

Figure 4 shows waiting time as a function of the arrival rate for four different situations. The value of σ is set to zero or one and μ_i is set to $1/3$ and 3. The queueing depends both on the part (the value of μ_i), the nature of the server (the value of σ) and the loading (the value of λ). But for high loadings, the value of λ dominates as can be seen by inspecting Expression (19) or the figure. This is why we need to include waiting time effects in the supply chain planning model along with routing choices.

Note that generally the qualitative conclusions remain if the arrivals are not Poisson, but the mathematics does not. An additional qualitative conclusion is that if there is one resource with a much higher utilization value than all others in the supply chain, then it will be responsible for much of the queueing. We would call such a server the *bottleneck* or *critical resource*.

That is, lead times impact the performance of the supply chain significantly. Although there is a large literature concerning queueing models for the analysis of the relationship between capacity utilization and lead times, and there is a substantial literature concerning control and order release policies that take lead times into consideration, there have been only a handful of papers that describe models at the aggregate planning level that recognize the relationship between the planned utilization of capacity and lead times. In [7] we provide an in-depth

discussion of the state-of-the art in this literature, with particular attention to those models that are appropriate at the aggregate planning level. The biggest issue is that planning models typically treat lead times as static input data, but in most situations, the output of the planning model implies capacity utilizations which, in turn, imply lead times.

The approach of modelling clearing functions in order to account for load dependent lead times is considered very promising and should be implemented in a stochastic framework by using queueing models with the purpose to integrate the problem of variable demand patterns and to analyze the behaviour of load dependent lead times. This could be used as a starting point for additional modelling of the production system. It is evident that not all concepts and methods are well suited for each type of production system (job shop, flexible manufacturing system, etc.) with one or more production stages or products (one level - multi level production systems, one item - multi items or products), respectively. In case of load dependent lead times and their integration into mathematical programming models for the aggregate production planning process it is wise to begin with the most simple production system and then proceed to more complex systems and situations in order to arrive at overall supply chain networks with the aim to find general principles of modelling and accounting for load dependent lead times.

References

[1] Knolmayer, G., Mertens, P. and Zeier, A., "Supply Chain Management based on SAP Systems", Springer, Berlin, 2002.

[2] Stadtler H. and C. Kilger (Eds.), "Supply Chain Management and Advanced Planning". Springer-Verlag, 2nd ed., 2000.

[3] S. Voß and D.L. Woodruff, "Introduction to Computational Optimization Models for Production Planning in a Supply Chain", Springer-Verlag, 2003.

[4] Karmarkar, U.S., "Lot Sizes, Lead Times and In-Process Inventories", Mangement Science, 33:409–418, 1987.

[5] Matsuura, H. and Tsubone, H., "Setting Planned Leadtimes in Capacity Requirements Planning", Journal of the Operational Research Society, 44:809–816, 1993.

[6] Spearman, M.L., Woodruff, D.L. and Hopp, W.J., "CONWIP: A Pull Alternative to Kanban", International Journal of Production Research, 28:879–894, 1990.

[7] J. Pahl and S. Voß and D.L. Woodruff, "Production planning with load dependent lead times: A survey", Technical Report, University of Hamburg, 2004.

Current IBM Research in Supply Chain Modelling and Optimization

Brenda Dietrich

The Mathematical Sciences department at IBM T J Watson Research Centre has worked with IBM's various manufacturing units for over 20 years, providing models and software tools that are used to improve the efficiency of manufacturing and related planning processes. Activities have included product and process design, demand forecasting, inventory planning, procurement support, production planning, and pricing. Improvements in the underlying information technology, including availability of real-time data and rapid computation have allowed us to

deploy new applications of optimization techniques, with much of our current emphasis on operational, rather than strategic, decision making. In recent years our modelling capability and tools have also been made available to a select set of IBM customers and partners, typically through consulting and services engagements.

This talk discusses two optimization projects, implosion and fleet dispatching, which are illustrative of the range of applications of optimization by IBM.

We began studying the "implosion" or resource-allocation problem in 1989. At that time our group provided modelling and analysis support and software to IBM manufacturing sites. In the late 1980's the PS/2 card plant in Austin experienced shortages of electronic components that were required for the production of several different PS/2 cards. IBM's world-wide material planning process required that the card plant commit an availability schedule for PS/2 cards within a few days of receiving a forecast of card requirements and an availability schedule for components. The card volume planners had no tools, other than simple spreadsheets, to aid in determining how to allocate the limited availability of the scarce components to cards, and on occasion had produced infeasible committed availability schedules. This lead to both lowered revenue, and increased costs resulting from excess inventory of other computer parts (disk drives, power supplies) that could not be used.

The difference between resource-allocation-based planning and traditional material requirements planning (MRP) approaches can be understood in terms of the inputs, assumptions, and outputs of these two methods. MRP considers the top-level demand (MPS) to be fixed input data and assume infinite material and capacity availability. MRP calculates required supply quantities and generates recommendations for changes to supply orders. In contrast, resource-allocation models take the material availability to be known input data, and treat the top-level demand as a desirable but not necessarily attainable target. These models calculate a feasible production schedule in a manner that ensures feasibility and optimizes specific economic criteria. In addition, resource-allocation models can be extended to consider factors that cannot be represented with traditional MRP methods, such as allocation of production to customers or demand classes, use of substitute material, and allocation of production to alternative manufacturing sites.

In resource-allocation-based planning, the production quantities for each product in each period are decision variables. Constraints on the decision variables are determined from the bill-of-capacity structure, the material-availability limits, the bill-of-material structure, and the original demand schedule. These constraints limit the values that can be simultaneously taken by the decision variables. Profit or serviceability maximization is often used as an objective function, and the problem is solved through the use of heuristics or linear programming algorithms.

During the early years of the implosion project, the problem size (number of parts times number of time periods), the computational speed of the available hardware, coupled with limitations of the available software, restricted the practicality of using of linear programming in implosion applications. Therefore, an

alternate solution method, known as the "implosion heuristic" was developed. The implosion heuristic is intended to quickly produce feasible, near optimal solutions. It takes as input standard MRP data: bills of material, supply of parts, and demand; it and produces as output a feasible production plan and shipment schedule. It can also provide reports on backlog and inventory levels of parts.

The idea behind the implosion heuristic is quite simple, and is based on the fact that given a finite supply of resources, for a single demand element (that is, a part, time period, quantity triple p,t,N) there is a maximum quantity of p that can be completed in time period t. By considering the demands in some specified order, and for each demand determining the maximum quantity of the demand that can be met in the prescribed time period and subtracting the resources used to meet this demand from the supply, one can produce a feasible production plan and corresponding shipment plan. Computational efficiency and solution quality can be achieved through extensions of this simple approach. When all resources are readily available, this approach produces a production plan that meets all demands through just in time production. If there are shortages of some resources, the production of parts that require those resources will be limited by the availability of the resources. If a scarce resource is used in only one part, the production of that part will be reduced so that it does not exceed the availability of that resource. In the more typical case, where a scarce resource is used by multiple parts, the order in which the demands are considered is the primary factor determining the allocation of the scare resource. The demands that are considered first will be met and will consume the resources, leaving none for the demands that are considered later. By ordering the demands according to business objectives, high quality solutions can be obtained. If equitable solutions, which evenly share scarce resources, are desired, each demand can be broken into several smaller demands, and these demands interleaved in the order. This basic heuristic approach can be extended to deal with substitutions, various forms of build-ahead, and some reallocation of stock. In all of the heuristic extensions, the "no-backtracking" principle has been maintained. That is, once a production quantity has been determined to be feasible, that quantity is never later reduced. Adhering to this design principle has allowed us to customize the heuristic to address a number of complex scenarios without compromising execution speed. One particularly important form of customization addresses business rules related to backlog. If a demand cannot be completely met in the requested period, the remaining quantity can be ignored (appropriate for the case where customers will substitute a competitor's product), it can be added to demand for that part in the following period, or it can be used to create a new, high priority demand in the following period.

The Watson Implosion Tool (WIT), which includes both an LP based solver and the heuristic solver described above has been in use within IBM for well over a decade. WIT has been continually enhanced in response to requirements from internal and external customers. It has been applied to a wide range of planning problems, including corporate wide constrained material requirements planning,

reverse logistics, distribution planning, and services capacity planning. Most recently it has been used as the basis for a dynamic available to promise/available to sell application for IBM's highly configured products such as main frames and high-end servers. In the available to sell application, current inventory and component purchase contracts are evaluated against multiple demand scenarios to determine the expected usage and the resulting expected overage, of components and sub-assemblies. The expected overage is then imploded, together with information on the cost of purchasing additional parts, against numerous sales opportunities to identify those that are most profitable to pursue. As new sales are closed, the opportunity list is updated to reflect usage of parts. This process has enabled IBM to better manage its inventory and improve our demand fulfilment processes.

WIT now uses the COIN interface to access the open source COIN LP and IP solvers, as well as OSL. (See www.coin-or.org.)

In contrast to the WIT project, which has continued with essentially the same staff for over ten years, much of our current work involves rapid development and deployment of optimization models to support customer needs. We recently developed the optimization component of a decision support system for a provider of car service Fleet scheduling and dispatching involves allocating vehicles and drivers to meet customer demands within tight time windows. Cars are fitted with two-way data terminals allowing the dispatch centre to maintain knowledge of drivers' states and positions at all times. Modestly priced compute servers allow us to solve and re-solve the scheduling and routing problem nearly optimally throughout the day as demands change. Despite highly dynamic data, our optimization tool is able to solve the scheduling and dispatching problem quickly enough to provide a timely schedule, i.e., before the demands have changed so much that the schedule is largely invalid. Combined with a mechanism that locally updates a schedule within seconds in response to single a new input, this yields a system in which the car service provider operates more efficiently than it could with its former manual scheduling system. The system was installed in March, 2003 and is currently in use 24 hours a day. Productivity has increased significantly.

The basic problem is formulated using shortest path based column generation. The base problem is solved to optimality a few times each day to determine a starting plan for the next several shifts. However, data is constantly changing due to flight delays, customer requests, and traffic conditions. Columns are updated, and new columns are generated, as real-time data is received. A sub-problem, corresponding to the current and next shift, is updated and resolved several times each hour by the "continual solver". Since the dispatcher must be able to over-ride solver recommendations, must always have access to a feasible schedule, and requires near instantaneous response, a variety of heuristic methods are employed in an "instant solver" to produce and maintain good feasible solutions. The fleet scheduling and dispatching application is remarkably similar to a service technician scheduling application developed by the same team in 1997-98 for IBM internal use.

Optimization-Based Order Promising and Fulfilment
Michael O. Ball

The Available to Promise (ATP) business function is the set of capabilities that support responding to customer order requests. Traditionally ATP refers to a simple database lookup into the Master Production Schedule. With the advent of assemble-to-order (ATO) and configure-to-order (CTO) production environments, the ATP function requires advanced real-time decision support systems with underlying model support. ATP research can be classified into two categories: push-based models, which allocate resources and prepare information based on forecasted demand, and pull-based models, which generate responses to actual customer orders. ATP systems operate within a short-term operational environment where most resource availability is considered fixed because of procurement lead-time limitations. This distinguishes both push-based and pull-based ATP models and systems from traditional planning, scheduling and inventory management processes. Reference [2] provides an overview of ATP research.

This presentation covers a series of papers [1, 3, 4, 5], which describe mixed integer programming (MIP) pull-based ATP models. These models allocate available resources to a batch of order requests that arrive within a pre-determined batching interval. They have a strong temporal component in that both the component availability and production capacity vary over time and the orders have constraints on possible delivery dates. They can be viewed as both order-promising and order-fulfilment models, since they specify a schedule for the use of production capacity.

In order to model the batch ATP problem as a decision problem, we first must characterize the decision space. We classify decisions as either front-end (order related) or back-end (production related). The front-end decisions include: 1) whether to accept or reject an order, 2) determining the order quantity, 3) determining the order delivery date and 4) deciding whether to split an order satisfied by multiple separate deliveries. Depending on the application, any of these decisions might be fixed a-priori and thus not addressed within the model. For example, if a company had a policy of accepting all orders then there would be no category 1) variables. Similarly, if the order quantity or delivery date was fixed within the customer order then there would be no category 2) or 3), resp, decision variables. Finally, a company could have a policy against splitting orders eliminating category 4) variables. Back-end decisions involve the allocation of production resources, including assignment of components to orders, assignment of orders to factories and production lines and order scheduling. Generally speaking, ATP models emphasize front-end considerations and as such they do not include detailed models of production/factory resources. In a typical business implementation, detailed factory scheduling would be addressed by other models and systems.

For some problem classes the models must include complex material availability constraints related to consideration of a flexible bill-of-material (BoM). The materials needed in assembling end-items are grouped into different component types.

The manufacturer may have multiple suppliers that provide materials of the same type, which differ in features such as quality, price and technology. We call the combination of a component type and a supplier as a component instance, which represents the very basic material element in the models. Each customer order has an associated BoM that specifies the quantity of each component type required to build a specific product. Furthermore, for each selected component type, a customer may specify a set of preferred suppliers, implying that the manufacturer must choose a component instance from one of the preferred suppliers. Thus, component substitution at component-instance level is allowed but subject to certain restrictions. However, there can be incompatibilities between certain pairs of component instances, e.g. a hard disk assembly from supplier A is incompatible with a printed circuit board from supplier B. The derivation of constraints to represent these incompatibilities presented an interesting problem in polyhedral projection that was solved for a special case in [1]. However, more general instances of this problem remain open.

We now describe the specific ATP decision problem associated with a particular electronic product (denoted by EP) manufactured by Toshiba Corporation and the MIP ATP model we developed. This model has been implemented by Toshiba and is used to support daily ATP decision-making. The order promising process proceeds by iteratively collecting and processing batches of orders. The ATP model is used to determine delivery dates, a decision on whether to split the order and the production schedule for each order. The model must balance available resources relative to a batch of orders requesting multiple products that share certain common components. The objective function criteria include minimization of due date violation, inventory holding cost and a day-to-day production smoothness measure. The due date violation is computed as the sum, over all orders, of the amount delivered late times the number of days late. The holding cost contains both a material holding cost and a finished product inventory holding cost. Production smoothness is based on a measure of day-to-day variation in the production amount of each assembly line at each factory.

The EP supply chain consists of multiple final assembly and testing (FAT) factories all located in Japan, which provide EPs delivered directly to both domestic and international business customers. Due to high product mix, an ATO production framework is employed to increase the degree of product flexibility. The order promising and fulfilment process involves in total several thousand-product models. Order sizes range from a very small number of units to a few hundred. Orders are generated by one of several sales units and are processed by a single central order processing system in Toshiba headquarters. The ATP system collects orders over a 1/4 hour time interval and returns commitments to the sales offices at the end of each ATP run (1/4 hour interval). Order commitments are booked up to approximately ten weeks in advance of delivery.

In the order promising process for EP, Toshiba employs the business practice of never denying an order. If an order cannot be fulfilled before its requested

due date, then a promise date beyond the requested date is given, i.e. it is back-ordered, or the order is split with a portion given an early promise date, e.g. before the due date and a portion given one or more later promise dates. However, an order cannot be split among different factories, namely, one order can only be committed in one factory. In order to emphasize customer satisfaction for EP, Toshiba weights due date violation higher than any holding costs and production smoothness penalty in its order fulfilment decision models. Occasionally, the sales staff will book "pseudo orders" based on enquiry orders from customers to reserve critical resources for anticipated future high priority demands.

A particularly interesting aspect of this problem was the manner in which resource constraints varied across the order-promising time horizon. For the fixed product interval, which spans from approximately the present time to two weeks into the future, resources, in the form of manufacturing orders (MO) are fixed. An MO specifies the production quantity for each product at each assembly line in each factory. That is, a fixed production schedule is set, which takes into account both production capacity availability and critical material availability. Having a fixed schedule stabilizes production dynamics in the near term and allows for the required materials to be set up and put in place. Any order commitments made for this time interval must fit within the fixed production schedule. In the flexible product interval, two kinds of resources, capacity and material, are considered in order promising. The production capacity is given daily at the factory level in terms of machine-hour and manpower availability. The weekly availability of individual critical materials is aggregated into finished goods level availability grouped based on the bill of material (BoM), balance-on-hand inventory, pipeline inventory and scheduled receipts. It is defined as a Production Capability (PC). Any order commitments made for this time interval must satisfy the capacity and material availability constraints. The flexible product interval spans from approximately two weeks to two months into the future. For the flexible resource interval, which covers due dates three weeks into the future, the only constraint considered is production capacity, which is specified daily at factory level in terms of machine-hour and manpower availability. This interval starts beyond the material resource lead times so any resource commitments can be met. The resource allocation variables within the MIP model varied by time interval in order to accommodate the variation in resource constraints just described. In the actual implementation, in order to solve the largest real instances an approximate aggregation procedure was used. In particular, an aggregate model was first solved to assign orders to weekly resource availability. Subsequently, for each week a version of the model was solved that assigned that week's orders to daily resource availability. For a typical problem instance, there were 4355 different EP product models. In one batch run there were 1162 newly-arrived orders which were combined with 3834 previously-promised customer orders. Both classes of orders generated appropriate resource assignment variables because since the production timing and factory assignment could be changed for the previously promised customer orders (even though the

deliver date commitments were fixed). The associated MIPs were solved within 3 minutes of computing time using a commercial MIP solver.

References

[1] Ball, M.O., Chen, C-Y. and Zhao, Z., 2003, "Material Compatibility Constraints for Make-To-Order Production Planning", Operations Research Letters, 31, 420-428.
[2] Ball, M. O., C.-Y. Chen and Z.-Y. Zhao 2003, "Available to Promise", to be published in Simchi-Levi, D., D. Wu, and M. Shen, eds., Handbook of Supply Chain Analysis in the eBusiness Era
[3] Chen, C-Y. , Zhao, Z. and Ball, M.O, 2001, "Quantity and Due Date Promising Available to Promise", Information Systems Frontiers, 3, 477-488.
[4] Chen, C-Y. , Zhao, Z. and Ball, M.O., 2002, "A Model for Batch Available-to-Promise", Production and Operations Management. 11, 424 - 440
[5] Zhao, Z., M. Ball and M. Kotake, "Optimization-Based Available-to-Promise and Multi-Stage Resource Availability", Annals of Operations Research, forthcoming.

Performance Measurement for Inventory Routing

MARTIN W.P. SAVELSBERGH

(joint work with Jin-Hwa Song)

Vendor managed inventory resupply (VMI) has become a popular strategy to reduce inventory holding and/or distribution costs. In environments where VMI partnerships are in effect, the vendor is allowed to choose the timing and size of deliveries. In exchange for this freedom, the vendor agrees to ensure that its customers do not run out of product. In a more traditional relationship, where customers call in their orders, large inefficiencies can occur due to the timing of customers' orders, i.e., high inventory and high distribution costs. By employing VMI partnerships companies may be able to reduce demand variability and therefore their inventory holding and distribution costs. Realizing the cost savings opportunities of VMI partnerships, however, is not an easy task, particularly with a large number and variety of customers. The inventory routing problem (IRP) seeks to do exactly that: determining a distribution strategy that minimizes long term distribution costs. A large body of literature on the IRP exists; Campbell et al. [1] and Kleywegt et al. [2], among others, contain an overview of the major research activities in this area.

We do not focus on developing distribution strategies, but instead on measuring the effectiveness of distribution strategies. A popular performance measure used in practice to evaluate distribution strategies in an environment where VMI partnerships are in effect is the volume delivered per mile or *volume per mile* for short. As the volume that needs to be delivered by the vendor over a given period of time is determined by the total usage of its customers, and not under the control of the vendor, the vendor strives to minimize the total mileage required to deliver product. However, volume per mile by itself is not a meaningful number, because it is impacted by many factors, such as the geography of customer locations and

customer usage patterns, but it is valuable for comparing performance in consecutive periods of time. If a company has a stable customer set and customer usage patterns do not fluctuate much, then an increase (decrease) in volume per mile indicates that distribution planning is improving (worsening).

The above discussion shows that volume per mile is a useful measure for monitoring relative distribution strategy performance. However, volume per mile cannot be used to determine, in an absolute sense, the quality of a distribution strategy. We develop a linear programming based methodology that allows the computation of tight lower bounds on the total mileage required to satisfy customer demand over a period of time (and thus upper bounds on volume per mile).

The approach is based on solving the following variant of the inventory routing problem. A single product has to be distributed from a single facility to a set I of n customers over a period of time of length T. Each customer $i \in I$ has the capability to maintain a local inventory of product up to a maximum of C_i. In the period of interest customer i consumes an amount u_i of product. A fleet of homogeneous vehicles, with capacity Q, is available for the distribution of the product. We assume an unlimited supply of product and an unlimited number of vehicles in the fleet. We denote the travel distance between two locations i and j by t_{ij}. The objective is to determine the minimum total distance required to satisfy all demand.

Observe that when $C_i \geq Q \ \forall i \in I$, then the optimal distribution strategy is to always deliver a full truck load to a customer right when the customer's storage tank becomes empty. The resulting total distance is $\sum_{i \in I} \frac{u_i}{Q} 2t_{0i}$, where 0 denotes the plant. Therefore, a simple lower bound on the minimum total distance required to satisfy all demand is obtained by assuming that all customers' storage capacities are greater than the truck capacity, i.e.,

$$LB_1\text{:} \ \sum_{i \in I} \frac{u_i}{Q} 2t_{0i}$$

In practice, deliveries to customers with storage capacity less than the truck's capacity, i.e., $C_i < Q$, are usually combined with other deliveries to ensure a high utilization of the truck's capacity.

Define a feasible delivery pattern $P_j = (d_{j1}, d_{j2}, ..., d_{jn})$ to be a delivery pattern that satisfies $\sum_{i \in I} d_{ji} \leq Q$ and $0 \leq d_{ji} \leq C_i \ \forall i \in I$. Let $\delta(P_j) = \{i \in I : d_{ji} > 0\}$ denote the set of customers visited in delivery pattern P_j. The cost of delivery pattern P_j, denoted as $c(P_j)$, is the value of an optimal solution to the travelling salesman problem involving the plant and the customers in $\delta(P_j)$. Let $\mathcal{P}$ be the set of all feasible delivery patterns and let x_j be a decision variable indicating how many times delivery pattern P_j is used. Then the optimal objective function value of the following linear program, called the pattern selection LP, provides a lower

bound on the total distance required to satisfy the demand

$$D^* = \quad min \quad \sum_{P_j \in \mathcal{P}} c(P_j) x_j$$

$$s.t. \quad \sum_{P_j \in \mathcal{P}} d_{ji} x_j \geq u_i, \quad \forall i \in I$$

$$x_j \geq 0$$

There are two major obstacles to using this linear program:
- The number of feasible delivery patterns is prohibitively large.
- The calculation of the cost of each delivery pattern involves the solution of a travelling salesman problem.

We show that only a finite set of delivery patterns needs to be considered and that the size of the linear program can further be reduced by pattern dominance and approximation ideas.

Using this methodology companies will be able to gain insight into the effectiveness of their distribution strategy.

References

[1] A.M. Campbell, L. Clarke, A. Kleywegt, and M.W.P. Savelsbergh(1998). Inventory Routing. In *Fleet Management and Logistics*, T. Crainic and G. Laporte (eds.), Kluwer Academic Publishers.
[2] A. Kleywegt, V. Nori, and M.W.P. Savelsbergh (2002). The Stochastic Inventory Routing Problem with Direct Deliveries, *Transportation Science* 36, 94–118.

Primal-Dual Approximation Algorithms for Deterministic Inventory Models

Retsef Levi

(joint work with Robin Roundy and David Shmoys)

In this talk, we consider several classical models in deterministic inventory theory: the single-item lot-sizing problem, the joint replenishment problem (JRP) and the multi-stage assembly problem. These inventory models have been studied extensively over the years, in a number of different settings, and play a fundamental role in broader planning issues, such as the management of supply chains. We shall consider the variants in which there is a discrete notion of time with a finite planning horizon, and the demand is deterministic (known in advance) but dynamic, i.e., it varies over the planning horizon.

Each of the inventory models that we consider has the following characteristics. There are N commodities (or equivalently, items) that are needed over a planning horizon consisting of T time periods; for each time period and each commodity, there is a demand for a specified number of units of that commodity. To satisfy these demands, an order may be placed in each time period. For each commodity i

ordered, a *fixed ordering cost* K_i is incurred, which is independent of the number of units ordered from that commodity. The order placed in time period t may be used to satisfy demand in time period t or any subsequent point in time. In addition, the demand in time period t must be satisfied completely by orders that have been placed no later than time period t. (In the inventory literature, these assumptions are usually referred to as "neither back orders nor lost sales are allowed".) Since the cost of ordering a commodity is independent of the number of units ordered, there is an incentive to place large orders, to meet the demand not just for the current time period, but for subsequent time periods as well. This is balanced by a cost incurred for holding inventory over time periods. We will let h^i_{st} denote this *holding cost*, that is, the cost incurred by ordering one unit of inventory in period s, and using it to meet the demand for item i in period t. We will assume that h^i_{st} is non-negative and, for each (i,t), is a non-increasing function of s. (Note that in particular, we do not require sub-additivity; we could have that $h^i_{rt} > h^i_{rs} + h^i_{st}$ for some $r < s < t$.) The goal is to find a policy of orders that satisfies all demands on time and minimizes the overall holding and ordering cost.

The details of the three inventory models are as follows. In the *single-item lot-sizing problem*, we have a single item ($N = 1$) with specified demands over T time periods $(d_1, .., d_T)$.

In the *joint replenishment problem* we have N commodities, where for each commodity $i = 1, \ldots, N$, and for each time period $t = 1, \ldots, T$, there is a specified non-negative demand d_{it}. In addition to the item ordering costs, K_i, $i = 1, \ldots, N$, *any* order incurs what we call *a joint ordering cost* K_0, independent of the (nonempty) subset of commodities that are included in the order (and again, independent of the (positive) number of units for each commodity included). The joint ordering cost creates a dependency between the different commodities and complicates the structure of the optimal policy. The holding cost follows the same structure described above.

In the *assembly problem*, we have a somewhat more involved structure. As part of the input, we also specify a rooted directed in-tree, where each node in the tree corresponds to an item, and we assume that the items are indexed so that $i > j$ for each edge (i, j) in the tree. Node (or item) 1, the root of the tree, is facing external demands over T time periods $(d_1, .., d_T)$. A unit of item i is assembled from one unit of each of its predecessor items in the tree. Thus, any unit of item 1 consists of one unit of each of the other items. We again have an ordering cost and holding cost for each item.

We note that the way we model the holding cost is much more general than the most common setting, in which each item i has a linear holding cost, so that the cost of holding one unit from time period s to time period t is equal to $(t - s)h_i$, for some choice of $h_i > 0$ (or to $\sum_{l=s}^{t} h^i_l$ in the more general case). By allowing the more general structure described above, we can capture other important phenomena, such as perishable goods, where the cost of holding an item longer than a specified interval is essentially infinite. The strength of the general holding cost structure is also demonstrated, where we show how to apply the algorithm to the

more general JRP model with back-orders. As for the ordering cost, we note that our algorithms are applicable also in the presence of time dependent cost parameters. Furthermore, in addition to the (fixed) ordering cost that is independent of the order size, one can incorporate a per-unit ordering cost into the holding cost term (as long as we preserve the monotonicity).

In this talk, we describe a unified novel primal-dual algorithmic framework that provides optimal and near-optimal solutions to the three inventory models described above. The algorithms improve known results in several ways: the performance guarantees for the quality of the solutions improve on or match previously known results; the performance guarantees hold under much more general assumptions about the structure of the costs, and the algorithms and their analysis are significantly simpler than previous known results. Finally, our primal-dual framework departs from the structure of previously studied primal-dual approximation algorithms in significant ways, and we believe that our approach may find application in other settings.

Our main result is a 2-approximation algorithm for the joint replenishment problem. By this we mean that for any instance of the problem, our algorithm computes a feasible solution in polynomial-time, with cost that is guaranteed to be no more than twice the optimal cost. The joint replenishment problem is NP-hard, but it can be solved in polynomial-time by dynamic programming for a fixed number of commodities, or for a fixed number of time periods, (by fixing the times at which joint orders are placed the problem decomposes by item). LP-based techniques have not previously played a significant role in the design of approximation algorithms for NP-hard deterministic inventory problems with constant performance guarantee. LP-rounding was applied to a more general problem by Shen, Simchi-Levi, and Teo, but this yielded a guarantee of only $O(\log N + \log T)$. This absence of results is particularly surprising in light of the fact that it has long been understood that these problems admit integer programming formulations with strong linear programming relaxations, i.e., that provide tight lower bounds. These formulations are closely related to formulations that have been studied for the facility location problem, which has also been a source of intense study for approximation algorithms.

The single-item lot-sizing problem was shown to be solvable in polynomial time by dynamic programming in the landmark paper of Wagner & Within. Furthermore, Krarup & Bilde showed, in this case, that the facility location-inspired LP has integer optima by means of a primal-dual algorithm, and Bárány. Van Roy, and Wolsey gave yet another proof of this by means of an explicitly generated pair of primal and dual optima (that are computed, ironically, via a dynamic programming computation). Finally, Bertsimas, Teo and Vohra gave a proof, which is based on LP rounding. If we consider our joint replenishment algorithm as applied to the special case of the single-item lot-sizing problem (where, since there is only one item, one can merge the joint ordering cost and the individual item

ordering cost into one new ordering cost), then we obtain a new, extremely simple, primal-dual optimization algorithm that also proves the integrality of this LP formulation.

Finally, with some modifications, our primal-dual algorithm can also be applied to the assembly problem to yield a 2-approximation algorithm. Here, we achieve the same approximation ratio as Roundy, who gave a 2-approximation algorithm (again for the case where all cost parameters are fixed over time) using a non-linear relaxation and ideas borrowed from continuous-time lot-sizing problems. Although we only match the previous performance guarantee, our approach is much simpler, and it yields the performance guarantee under a much more general cost structure. In particular, under our assumptions on the cost structure, it is easy to show that the assembly problem is NP-hard by a reduction from the joint replenishment problem. However, for the variant of the problem considered by Roundy, it is still not known whether it is NP-hard or not.

As a byproduct of our work, we prove upper bounds on the integrality gap of the corresponding LP relaxations, the worst-case ratio between the optimal integer and fractional values; for both the JRP and the assembly problem, we prove an upper bound of 2.

We note that our algorithms have their intellectual roots in the seminal paper of Jain & Vazirani, which gives a primal-dual approximation algorithm for the uncapacitated facility location problem. Nonetheless our algorithms depart from their approach in rather significant ways, as we shall describe in detail in the next section. We believe that this new approach may find applications in other settings.

An extended abstract of this paper will appear in STOC 2004. The paper with all relevant references can be found in *http://www.orie.cornell.edu/~levi/Index.html.*

Toward Robust Revenue Management

MAURICE QUEYRANNE
(joint work with Michael O. Ball)

We develop robust revenue management policies by viewing the customer acceptance problem from the perspective of competitive analysis on on-line algorithms. Specifically, the on-line revenue management algorithm seeks to maximize the revenue associated with the accepted customers, where a decision has to be made to accept or reject each customer at the time of the customer arrival. The quality of such algorithms is measured against an optimal offline solution that would result from perfect knowledge of the whole input sequence and determine which customers to accept after all customers have arrived. The online approach allows for the definition and analysis of algorithms without the need for demand forecasts or a risk-neutrality assumption. The policies that emerge from this analysis have worst-case guarantees on their performance, and appear to be appealing from a practical standpoint. In particular we derive optimal booking policies for

multiple-fare booking problems, defined by nested protection levels, which fit current industry practice, but with protection levels derived from very passimonions input. We also derive optimal algorithms of new types for bid control problems.

Pricing and Manufacturing Decisions when Demand is a Function of Prices in Multiple Periods

Philip Kaminsky

(joint work with Hyunsoo Ahn and Mehmet Gumus)

In recent years, as manufacturing and supply chains have become more and more efficient, the conflict between production planning and marketing has become more apparent. There is a growing research literature focusing on joint marketing decision-making and production planning. The overall objective of this literature is to develop approaches that avoid the possible conflicting consequences of marketing and operations planning decisions, by integrating marketing/pricing decisions and manufacturing decisions to jointly achieve a common objective. Virtually all of this research considers demand at each period to be a function of price in that period. However, in reality, in many cases customers may consider making a purchase for several periods, and actually make the purchase in either the first period that the price is below the reservation price, or in the period that they believe will have the lowest price. Hence, demand realized in each period is not only a function of current price but also of past and future prices. In this work, we focus on models that combine elements of both marketing and production planning, designing a profit-maximizing production schedule and product pricing schedule in the face of inter-temporal demand-price interactions.

A variety of aspects of joint pricing and manufacturing models have been analyzed in the operations management and marketing management literature. The models vary from constant demand, EOQ-like frameworks, to non-stationary discrete and continuous-time frameworks. To the best of our knowledge, the earliest paper where price and production quantity are both decision variables is by [1], who extends the basic EOQ model to include a revenue term, and finds the optimal price and lot size using a calculus-based approach. This work was extended by many authors, including [2], [3], [4], and [5]. A variety of authors have also considered the deterministic discrete-time production framework, including [6], [7], [8], [9], and [10]. For models of dynamic pricing in the continuous-time deterministic setting, see the comprehensive survey of joint pricing and production literature by [11]. A variety of authors have also considered joint pricing and production models with stochastic demand (see [12] and the references therein), as well as *revenue management* models, as surveyed in [13].

In this work, we consider two models that differ from much of the previous work by allowing demand to remain in the system for more than one period. In the first model, a potential customer has the patience to wait until the price drops to a level that she can afford. This assumption is somewhat analogous

to the assumptions used in inter-temporal price discrimination models in micro-economics (c.f., [14]). However, while most related literature, with the exception of [15], focuses on the analysis of models with a fixed number of customers of unknown valuation, we allow new customers to be introduced into the system in each period. Furthermore, our model differs from a few models that allow the entry of new customers by assuming a finite lifetime of potential demand (i.e., an unfulfilled customer will remain only for a finite number of periods) as well as non-stationarity of demand and parameters. In this first model, customers make a purchase the first time that the asking price drops below their reservation price. In our second model, customers are assumed to be aware of the optimization problem being solved, and so they make a purchase when their discounted utility is maximized.

In most prior work, consumers in a system focus on the price in the period in which they enter the system. If the price is below their reservation price, they make a purchase. If not, they exit the system. However, in many cases consumers consider the price in more than the current period. In our own experience, we have decided how much we were going to spend on a car or computer, and then waited for the price to fall within our budget. Therefore, it is not always realistic to assume that customers whose reservation prices do not exceed the current price of the product set by manufacturer just leave the market. In many cases, some proportion, α ($0 \leq \alpha \leq 1$) of the customers in this system will wait for the next period, and check if the next period's price is less than or equal to their reservation price. Hence, one way to extend the traditional model of demand realization is to to allow some portion of unsatisfied customers to stay in the system for at most some number of periods K, during which they continue to examine price levels to purchase the product. This extension to the traditional model is our first model. It is reasonable to assume that, for any period, only a portion of remaining customer will wait for one or more periods and the number of customers waiting decreases in time. To describe such consumer behaviour, we use α_i, $i = 0, 1, \ldots K, K + 1$ to represent the proportion of customers that will wait for at least i periods such that $1 = \alpha_0 \geq \alpha_1 \geq \alpha_2 \ldots \geq \alpha_K \geq \alpha_{K+1} = 0$.

By implementing this extension, we model two different streams of demand at each period t. The first one consists of the customers who acquire their reservation price using the current demand function and whose reservation price is higher than the current price, and the second one consists of the customers who have entered the system at previous time periods, and whose reservation prices have been less than prices in previous periods, but exceed the product price at period t. We call the former source of demand *current demand*, and the latter source *residual demand*.

For general demand functions, we formulate the two-step demand realization process described above as follows. First, we define some notation. Suppose that $P = (p_1, p_2, \ldots, p_T)$ represents prices for a product over the planning interval T. Let $d_t = D_t(p_t)$ be the demand function at period t and $D = (d_1, d_2, \ldots, d_T)$ be demand induced by pricing sequence, P. Let r_t^k represent the portion of demand

in period t originating from period $t - k$. In this case, r_t^k can be formulated as follows.

$$(20) \quad r_t^k = \begin{cases} D_t(p_t), & \text{if } k = 0; \\ \alpha_k \left[D_{t-k}(\wedge_{i=1}^k p_{t-i}) - D_{t-k}(p_t) \right]^+), & \text{if } 1 \le k \le \min(K, t-1). \end{cases}$$

We introduce a minimization operator to select the minimum price between periods $t - k$ and $t - 1$, since this gives the leftover residual demand after observing the actual prices from periods $t - k$ to $t - 1$. Hence, α_k of this leftover residual demand consists of the customers whose reservation prices have not exceeded the product price from period $t - k$ to $t - 1$.

In this work, we consider a linear demand curve at each period, and consider this demand model within a standard discrete-time multi-period production system where at each period we decide both price of the product, p_t and the production quantity of product, x_t. Our objective is to maximize the net profit subject to inventory balance, production capacity, and demand realization constraints.

One reasonable critique of this model is that it assumes that consumers aren't aware of impending price decreases. Indeed, this model is intended to model situations in which customers place a high value on a good's availability, and tend to buy it as soon as their budget constraint (i.e., reservation price) is met. Alternatively, we introduce our second model, in which customers are aware of the pricing pattern. In this case, they enter the system and stay for at most K periods, and actually make a purchase when their discounted utility is maximized. We model utility as the difference between modified price and reservation price, where $\beta_j \ge 1$ is a factor that we multiply by price in the future to represent the disutility of waiting j periods. We consider this alternate demand model within the same capacitated production framework.

We consider a variety of special cases of these models. Although both models are in general neither concave nor convex, we characterize their structure, and use this structure to design optimal algorithms for both models. For special cases including stationary parameters, we design more efficient algorithms, and for the special one-period interaction uncapacitated stationary case, we provide a closed form expression of the optimal solution in both cases. Finally, we complete an extensive computational investigation to determine the value of considering demand interaction in a variety of situations.

References

[1] Whitin, T.M. Inventory Control and Price Theory. *Management Science*, Vol. 2, No. 2, 61-68, 1955.
[2] Kunreuther H. and J. F. Richard. Optimal Pricing and Inventory Decisions for Non-seasonal Items. *Econometrica* , Vol. 2, 193-205, 1971.
[3] Abad, P.L. Determining Optimal Selling Price and Lot Size when the Supplier offers All-Unit Quantity Discounts. *Decision Sciences* , Vol. 19, No. 3, 622-634, 1988.
[4] Lee, W. J. Determining Order Quantity and Selling Price by Geometric Programming: Optimal Solution, Bounds, and Sensitivity. *Decision Sciences* , Vol. 24, No. 1, 76-87, 1993.

[5] Kim, D., and W. J. Lee. Optimal Joint Pricing and Lot Sizing with Fixed and Variable Capacity. *European Journal of Operational Research* , Vol. 109, 212-227, 1998.

[6] Thomas, L.J. Price-Production Decisions with Deterministic Demand. *Management Science*, Vol. 16, No. 11, 747-750, 1970.

[7] _____, and L. Schrage. Joint Pricing and Inventory Decisions for Constant Priced Item. *Management Science* , Vol. 19, No. 7, 732-738, 1973.

[8] Gilbert, S. M. Coordination of Pricing and Multi-period Production for Constant Priced Goods. *European Journal of Operational Research*, Vol. 114, No. 2, 330-337, 1999.

[9] L. M. A. Chan, D. Simchi-Levi, J. Swann. Flexible Pricing Strategies to improve supply chain performance. *working paper, Northwestern University*, 2000.

[10] Deng, S.,and Yano C. Joint Production and Pricing Decisions with Setup Costs and Capacity Constraints. *working paper*, 2000.

[11] Eliashberg, J.,and R. Steinberg. Marketing-Production Joint Decision-Making. Chapter 18 in *Handbooks in Operations Research and Management Science:Marketing*, Vol.5, North Holland, 827-880, 1993.

[12] _____, and A. Heching. Combined Pricing and Inventory Control Under Uncertainty. *Operations Research*, Vol. 47, No. 3, 454-475, 1999.

[13] J. I. McGill, G. J. Van Ryzin. Revenue management: Research overview and prospects. *Transportation Science*, Vol. 33, No. 2, 233-256, 1999.

[14] Conlisk, J., E. Gerstner, and J. Sobel. Cyclic Pricing by a Durable Goods Monopolist. *The Quarterly Journal of Economics*, Vol. 99, Issue 3, 489-505, 1984.

[15] Sobel, J. Durable Goods Monopoly with Entry of New Customers. *Econometrica*, Vol. 59, Issue 5, 1455-1485, 1991.

Experiments with Cooperative Optimization Algorithms for Production Scheduling

Claude Le Pape

(joint work with Emilie Danna)

Industrial optimization applications must be both "efficient", *i.e.*, provide "good" solutions within reasonable time, and "robust" with respect to variations in the problem instances: variations in problem size, variations in numerical characteristics, and addition of side constraints. It should in most cases be easy to include additional constraints without re-designing the overall problem-solving strategy. Cooperative optimization algorithms, which combine different techniques, can often be used to improve either efficiency or robustness (or both) without sacrificing the other ([3]).

For example, mixed integer programming, constraint programming, and local search techniques, provide good results under very different assumptions. Mixed integer programming is especially efficient whenever the continuous relaxation of the problem model is a good approximation of the convex envelope of the solutions (at least around the optimal solution) or when the relaxation can be iteratively tightened (by adding cuts) to improve this approximation. Constraint programming provides good results when critical constraints propagate well and, in an optimization context, if a tight bound on the optimization criterion results (by propagation) in constraints that effectively guide the search toward a good solution. Local search is efficient when good solutions share characteristics (*e.g.*, the

ordering of two tasks in a scheduling problem) that can be compactly represented and that are likely to be kept when local search operators proceed from a solution to the next. If each of these techniques encounters difficulties on one aspect or on some instances of a problem, a hybrid algorithm may give on average better results than pure mixed integer programming, constraint programming, or local search algorithms.

Production scheduling with earliness and tardiness costs is interesting in this respect. In its purest version, an $n \times m$ job-shop problem consists of a set J of n jobs and a set R of m machines of capacity one. Each job $j \in J$ consists of a set of m non-preemptive operations ordered according to a given permutation $\sigma_j = (\sigma_{j1}, \sigma_{j2}, \ldots, \sigma_{jm})$ of the machines: Job j must be executed first on machine σ_{j1}, then on σ_{j2}, and so on. Let p_{ji} be the processing time of job j on machine i. Job j cannot start before its release date r_j and is additionally characterized by its due date d_j and two nonnegative cost factors, earliness α_j and tardiness β_j. Let C_j be the completion date of the last operation of job j. The cost incurred by job j is $\alpha_j(d_j - C_j)$ if $C_j \le d_j$ and $\beta_j(C_j - d_j)$ if $C_j > d_j$. The objective is to minimize the sum of the costs incurred for each job. Real-life problems tend to be more complex with the addition of other constraints and cost factors (*e.g.*, setups, secondary resources, calendars). In the following, we consider the case in which it is not certain that operation σ_{ji}^{-1} of job j will use machine i, either because there is a choice between different machines, or because the operation can be left *unperformed*. An *unperformed* operation must be assigned a start time and will require its usual processing time to be completed, but it will not use capacity on any machine. This corresponds for example to sub-contracting that operation, with an additional cost v_j^i.

This problem is difficult for all of the optimization techniques mentioned above. Mixed integer programming is a good candidate for representing the cost function, but no good model is known to state that a machine can only perform one operation at a time. Constraint programming usually deals well with precedence and resource constraints, but adding an upper bound on the weighted-sum optimization criterion does not result in effective constraint propagation. Local search operators based on permuting operations are easy to design, but the impact of a permutation on the total cost is hard to estimate. In the following, cooperative optimization algorithms centred on a mixed integer programming model are proposed and compared with a pre-existing combination of constraint programming and local search.

The mixed integer programming model relies on the disjunctive model of [1]. Let continuous variable x_{ji} represent the start time of job j on machine i. Minimizing the sum of earliness and tardiness costs is modelled by $\min \sum_{j \in J} z_j$ where variable z_j represents the cost incurred by job j:

$$\begin{cases} z_j \ge \alpha_j(d_j - x_{j\sigma_{jm}} - p_{j\sigma_{jm}}) \\ z_j \ge \beta_j(x_{j\sigma_{jm}} + p_{j\sigma_{jm}} - d_j) \end{cases} \quad \forall j \in J$$

Release dates constraints are simply:

$$(21) \qquad\qquad x_{j\sigma_{j1}} \geq r_j, \ \forall j \in J$$

Precedence constraints between operations in each job are modelled by:

$$(22) \qquad x_{j\sigma_{j,t+1}} \geq x_{j\sigma_{jt}} + p_{j\sigma_{jt}}, \ \forall j \in J, \ \forall t = 1 \ldots m-1$$

Resource constraints are modelled by the constraints: $\forall \ p < q \in J, \forall \ i = 1 \ldots m$,

$$(23) \qquad\qquad x_{pi} \geq x_{qi} + p_{qi} - My_{pq}^i$$

$$(24) \qquad\qquad x_{qi} \geq x_{pi} + p_{pi} - My_{qp}^i$$

$$(25) \qquad\qquad y_{pq}^i + y_{qp}^i = 1$$

$$y_{pq}^i \in \{0,1\} \text{ and } y_{qp}^i \in \{0,1\}$$

where M is some large constant. The interpretation is that $y_{pq}^i = 1$ if job p is scheduled before job q on machine i, and $y_{pq}^i = 0$ otherwise. This type of model is known as a *big-M formulation*. Its advantage is to be simple but it only weakly links the decision variables y that appear in the resource constraints and the secondary variables x that appear in the precedence constraints. It is hence expected to behave poorly, *i.e.*, usually to have a loose continuous relaxation, or as shown in [4] to make it difficult for a mixed integer programming solver to find good integer solutions.

Unperformed operations are modelled by introducing binary variable $u_j^i = 1$ if operation σ_{ji}^{-1} of job j is unperformed, and $u_j^i = 0$ otherwise. The objective function becomes $\min \sum_{j \in J} z_j + \sum_{j \in J} \sum_{i=1}^m v_j^i u_j^i$. To link resource constraints and unperformed variables, Equation 25 is changed to $y_{pq}^i + y_{qp}^i \leq 1 + u_p^i + u_q^i$.

The following table provides the results obtained with five algorithms based on this model, all developed by Emilie Danna and further detailed in [2].

- The MIP algorithm is the default search strategy of CPLEX 9.0.
- The IS+MIP algorithm consists in using constraint programming to construct an initial solution to the problem. This solution is then used both to reduce the value of the M constant in each resource constraint and as a starting point for CPLEX.
- The IS+MIP+RINS algorithm is similar to IS+MIP but activates the relaxation induced neighbourhood search option of CPLEX ([5]). Relaxation induced neighbourhood search is a form of local search which relies on the continuous relaxation to define a neighbourhood of the incumbent solution: the integer variables that have the same values in the solution of the continuous relaxation and in the incumbent are fixed to these values and a sub-MIP on the remaining variables is solved (with a limit on the number of nodes explored).
- The IS+MIP+RINS+GD algorithm adds the "guided dives" option of CPLEX ([5]) to the IS+MIP+RINS algorithm. When a binary variable is selected for branching, the "guided dives" strategy will explore first

the node in which this variable is fixed to the value that it takes in the incumbent.

- The IS+MIP+RINS+GD+MCORE algorithm adds to the IS+MIP+RINS +GD algorithm another form of local search which defines a neighbourhood by heuristically reducing the values of the M coefficients. See [2] for details.

These algorithms are tested on 22 job-shop instances from the Manufacturing Scheduling Library (MaScLib) ([6]), with up to 260 operations. For each instance, 4 problem variants are tested: optimizing the sum of weighted earliness and weighted tardiness with no unperformed operation (ET); optimizing weighted tardiness with no unperformed operation (T); optimizing the sum of weighted earliness, weighted tardiness, and non-performance costs (ET_UNP); optimizing the sum of weighted tardiness and non-performance costs (T_UNP). Each run is limited in CPU time, with a limit dependent on the number of operations. The measure of performance of each run is defined as the ratio of the cost of the solution found by the algorithm divided by the cost of the best-known solution. For each algorithm and each optimization criterion, the table below provides the geometric mean of the ratios obtained for the 22 instances under consideration.

The algorithms are also compared with an unpublished constraint programming and local search algorithm CP+LS developed at ILOG. The results show the interest of all the components that have been added to the initial MIP algorithm. They also show that on pure problems, hybrid algorithms based on mixed integer programming can compete with state-of-the-art techniques. Open issues include the representation of additional constraints such as setup time and costs, calendars, etc., as proposed in ([6]).

Algorithm	Optimization criterion				Overall
	ET	T	ET_UNP	T_UNP	mean
MIP	1.626	1.708	1.370	1.774	1.612
IS+MIP	1.494	1.415	1.208	1.313	1.353
IS+MIP+RINS	1.378	1.339	1.144	1.194	1.260
IS+MIP+RINS+GD	1.192	1.188	1.088	1.094	1.140
IS+MIP+RINS+GD+MCORE	1.130	1.155	**1.060**	**1.056**	1.099
CP+LS	**1.107**	**1.015**	1.148	1.082	**1.087**

REFERENCES

[1] D. Applegate and W. Cook (1991), A Computational Study of the Job-Shop Scheduling Problem. ORSA Journal on Computing, 3(2), 149-156.
[2] E. Danna (2004), Intégration de techniques de recherche locale à la programmation linéaire en nombres entiers. PhD Thesis, University of Avignon.
[3] E. Danna and C. Le Pape (2004): Two Generic Schemes for Efficient and Robust Cooperative Algorithms. In: M. Milano (editor), Constraint and Integer Programming, Kluwer Academic Publishers.

[4] E. Danna, E. Rothberg, and C. Le Pape (2003): Integrating Mixed Integer Programming and Local Search: A Case Study on Job-Shop Scheduling Problems. *CP-AI-OR'03 Proceedings*, 65-79.

[5] E. Danna, E. Rothberg, and C. Le Pape (2004): Exploring Relaxation Induced Neighbourhoods to Improve MIP Solutions. *Mathematical Programming*, to appear.

[6] W. Nuijten, T. Bousonville, F. Focacci, D. Godard, and C. Le Pape (2004): Towards an Industrial Manufacturing Scheduling Problem and Test Bed. Proceedings of the Ninth International Conference on Project Management and Scheduling, 162-165.

Stochastic Lot-Sizing Polyhedra

George L. Nemhauser

(joint work with Yongpei Guan, Shabbir Ahmed)

We study a multi-stage stochastic integer programming model of the uncapacitated lot-sizing problem (ULSP) under uncertainty. The deterministic ULSP is to determine a minimum cost production and inventory holding schedule for a single product so as to satisfy its known demand over a finite discrete-time planning horizon. Integer variables are needed to model the fixed cost of production. The deterministic problem is well studied. It is solvable in polynomial time by dynamic programming and it is also possible to give a linear programming formulation whose solution is integral, by adding a simple, but exponential, family of so-called (l,S) inequalities to the standard formulation. This LP formulation is very useful as a relaxation in solving more complicated lot-sizing problem with, for example, capacity constraints. In the stochastic version of ULSP, demand in each period is given by a discrete probability distribution function. This information structure can be interpreted as a scenario tree where a node j at level (period) t represents the state of the system that can be distinguished by information available up to period t. In other words, each path in the tree from its root to a leaf corresponds to a deterministic realization of the problem and has a known probability. The objective now is to minimize the expected costs over all of the paths. We first extend the deterministic integer programming model to the stochastic case. Then we show that the (l,S) inequalities, which are derived from sub-paths of the tree, are valid for the stochastic problem as well. We then give our main contribution, which is an extension of the (l,S) inequalities to sub-trees. We establish necessary and sufficient conditions for the sub-tree inequalities to be facet defining. A computational study verifies the usefulness of these inequalities as cuts in a branch-and-cut algorithm. Finally, we show how these inequalities can be applied to other stochastic integer programming problems by giving a general way of extending inequality descriptions of polyhedra for multi-stage deterministic problems to stochastic versions of these problems.

References

[1] I. Barany, T. Van Roy and L.A. Wolsey, "Uncapacitated Lot Sizing: The Convex Hull of Solutions," Mathematical Programming Study 22, 32 -43, 1984.

[2] Y. Guan, S Ahmed and G.L. Nemhauser, "A Branch-and-Cut Algorithm for the Stochastic Uncapacitated Lot-Sizing Problem," TLI 04-01, Industrial and Systems Engineering, Georgia Institute of Technology, 2004.

Solving Multi-Objective Models

ED ROTHBERG

(joint work with B. Bixby, M. Fenelon, Z. Gu, and R. Wunderling)

Models involving a large set of linear constraints and a set of conflicting objectives arise in a number of application areas. Such models are particularly common in supply chain planning. A number of practical issues arise when solving such models. Through extensions to CPLEX presolve and careful orchestration of the various steps of the solution process, we are able to obtain substantial runtime reductions.

The first question that must be addressed when solving a multi-objective model is how to trade off the various objectives. The obvious approach of forming a single blended objective from a linear combination of the original objectives can produce undesirable results. Modest multipliers on objectives involving different units will typically produce non-intuitive trade-offs. Very large multipliers, typically chosen to achieve a strict prioritization among the objectives, quickly runs into numerical difficulties due to the limitations of finite-precision arithmetic. A common approach to combining objectives is therefore to choose a lexicographic ordering, where an objective is treated as being infinitely more important than objectives of lower priority.

A model involving a set of lexicographic objectives can be solved by optimizing the highest priority objective, then fixing all variables with non-zero reduced costs, and then continuing in the same fashion with the next highest priority objective. At each stage in this multi-stage solution process, the LP solver must solve a model with a static set of linear constraints, a strictly growing set of fixed variables, and a new objective vector. Note that the solution vector for the previous objective is also primal feasible for the next objective.

Given the limited set of changes to the model from one stage to the next (new variable fixings and a new objective vector), one may wonder whether it would be possible to reuse information computed during the solution of the previous model to solve the next. In particular, CPLEX presolve performs a set of logical reductions that decrease the size and computational difficulty of the model without removing the optimal solution. Given the set of accumulated variables fixings in this context, the scope for presolve reductions is substantial.

While these presolve reductions would ideally carry over from one stage to the next, an important detail prevents this from happening. Specifically, presolve performs reductions that depend on the objective vector, so the presolved model may no longer contain the optimal solution when given a new objective vector. CPLEX presolve therefore has been extended to classify individual reductions based on the model properties on which they depend. Presolve now allows

the user to shut off those reductions that depend on model properties that will change. Presolve groups reductions into two categories — primal reductions and dual reductions. Primal reductions are those that depend only on the primal constraints and are independent of the objective function. They are compatible with adding additional constraints or changing the objective function. Dual reductions depend only on dual constraints, and are compatible with adding variables and changing the right-hand-side vector. In our context, setting CPLEX parameter CPX_PARAM_REDUCE to value $CPX\_PREREDUCE\_PRIMALONLY$ allows us to make changes to the objective vector while retaining the presolved model.

Another important presolve extension whose usefulness in our context will hopefully be clear shortly is the ability to transform a basic solution on the original, unpresolved model into a basic solution on the presolved model. This is done through a process of converting the basis on the original model to a primal and dual solution vector, then applying presolve reductions on these vectors to obtain solution vectors on the presolved model, and finally using these vectors to guess a basis for the presolved model. The seemingly more straightforward path of performing presolve reductions on the basis itself turns out not to be possible. This solution transformation process is encapsulated in a single parameter change in CPLEX ($CPX\_PARAM\_ADVIND = 2$), whereby the next presolve invocation will transform a model and basis into a presolved model and an associated basis.

Given these presolve capabilities, the solution of the lexicographic multi-objective model can be performed using a straightforward sequence of steps. As each objective is considered, we maintain a current presolved model, a set of fixings on the original and presolved models, and current basic solutions for the original and presolved models. The following steps are then performed before moving on to the next objective:

- If the number of fixed variables in the presolved model is large (thus suggesting that further presolve reductions would be beneficial), perform a new primal-only presolve on the original model, and transform the current solution basis on the original model into a basis on the new presolved model.
- Transform the current objective vector on the original model into an objective vector on the presolved model.
- Solve the LP using primal simplex, starting from the available start basis.
- Use the computed reduced costs to update the set of fixings on the original and presolved models. Also update the basic solution on the original model.

Computational experiments with this approach show runtime reductions whose magnitude depends on the size of the model and the number of objectives. A model with roughly one million constraints, 1.5 million variables, and 20 objectives achieved a 2X performance gain. A model with roughly 2.5 million constraints, 3.5 million variables, and 20 objectives achieved a nearly 5X performance gain.

The extensions to CPLEX presolve described here are discussed in more detail in the CPLEX reference manual [1].

References

[1] ILOG (2004), CPLEX 9.0 Reference Manual, ILOG, Inc., Mountain View, CA.

Time-Line Network Based Optimization Models for Vehicle and Crew Scheduling
Leena Suhl

Within planning processes in public transport, especially vehicle and crew scheduling in airline, bus, and railway traffic, many tasks can be expressed as optimization problems and modelled with techniques of mixed-integer programming. Although the simple vehicle scheduling problem with one depot is polynomially solvable, there are numerous aspects implying that the corresponding optimization models become np-hard, so that problems of practical size are hardly solvable with current state-of-the-art technologies.

We have investigated time-space-network based modelling techniques which may drastically reduce the number of arcs within a network model. The modelling approach has been applied to rotation building problems of airlines and railways as well as to vehicle scheduling problems of bus transit. We take into account aspects such as dead-heading and periodicity. Through an aggregation technique we may implicitly include all possible deadhead trips into the model. Further aspects such as maintenance routing of vehicles or break routing of crews, multiple depots, and multiple vehicle types may be modelled as well.

Especially, the multiple-depot, multiple-vehicle-type problem of bus transit can be modelled with a multi-layer multi-commodity aggregated flow network. Aggregation of potential deadhead trips enables large-scale practical problems (with thousands of trips) to be solved directly using standard optimization software.

The integrated rotation building and maintenance routing problem for both airlines and railways can be modelled with an aggregated time-space flow network as well, if we introduce maintenance states for a given maintenance rule. Furthermore, we may build rotations for groups of locomotives and carriages using special techniques to deal with shared vehicle capacities and to aggregate "equivalent" vehicle groups.

Analogously to maintenance routing, airline crew scheduling problems with special rules guaranteeing night and weekend breaks ("maintenance" of human beings) can be modelled with maintenance states as well.

All model types mentioned above have been implemented together with industrial partners using a standard MIP solver such as CPLEX or MOPS. Because very large instances still cannot be solved to optimality, the approach has been enhanced with heuristics in order to guarantee a feasible solution in any case.

Supply Chain Optimization and Approximate Dynamic Programming
Daniel Bienstock

This work concerns ongoing joint work between Daniel Bienstock and Guillermo Gallego and two Ph.D. students (N. Ozbay and O. Sahin), all from Columbia University), and an industrial partner. This company has clients throughout the world, and has subcontracted a large amount of its manufacturing to other companies, which in turn subcontract key work to component suppliers. Our partner company assembles the components into finished products and ships them to customers through distribution centres (DCs) that are co-located with the customers. The suppliers are located throughout the world.

The company faces a demand profile that is seasonal, with sharp end-of-quarter effects. It produces forecasts that are propagated to its manufacturing subcontractors (and their subcontractors) in order to mitigate shortages. A key issue is that the manufacturing leadtime for some of the more important components can be extremely long, when the order quantity is in substantial excess of forecast amounts.

Our task is to produce a good buffering policy, taking into account all relevant data (manufacturing leadtimes, shipping delays, information flow delays, batching effects, demand profiles, and so on). Our initial work has been in the direction of formulating the problem as a Markov Decision Process, with the aim of understanding key structural properties of optimal policies and how these can be approximated with "reasonable" (implementable) policies.

In order to find the optimal solution to a Markov Decision Process one needs to solve a Linear Program. This linear program will generally be quite large – the number of variables will equal the number of states, and the number of constrains will equal the number of (state, decision) pairs. For this reason it has generally been believed that realistic instances will be unsolvable.

The table below presents computational experience solving LPs arising from Markov Decision Process models. The models we consider are a generalization of the model in [3]. In [3] the authors characterize optimal policies for the following inventory problem: at each time period we face a constant deterministic demand λ, which must be satisfied, and a stochastic demand that does not need to be satisfied (when it is not satisfied we pay a penalty). In addition, at each time period we can produce, and the production cost will include a fixed cost. The model we consider is just like the model in [3] except that there is a positive production leadtime – specifically, if we start production at time t the work will not be done until time $t + 2$.

This change greatly increases the complexity of the model (and of the optimal policies). The running times in the table below were obtained on a 3 GHz Xeon machine, with 6 GB of physical RAM. The LPs were solved using the interior point solver in CPLEX 8.1, which greatly outperformed all simplex variants.

As we can see from the table, the running times are substantial but, perhaps, not as discouraging as one might first expect, given the size of the problems.

n	m	nonz	sec
6336	458304	5498372	434
10901	1023264	12277232	1112
13596	1728958	20745428	2401
16379	1609014	19305490	2467

Nevertheless, it is clear that a model with (say) ten times the number of variables will prove beyond the reach of even the best LP solver.

To this effect, we are launching an effort to develop a good implementation of the ideas in [2] for approximately solving dynamic programs, together with classical ideas from the dynamic programming literature (such as Policy Iteration and Successive Approximation) and modern ideas from computer science (see [1]) related to the approximate solution of Linear Programs.

Based on prior work, we envision a hybrid algorithm that alternates between approximation techniques (such as those described above) and calls to an exact LP solver.

References

[1] D. Bienstock (2002), Potential Function Methods for Approximately Solving Linear Programming Problems, Theory and Practice. Kluwer (Boston).
[2] D. P. de Farias and B. Van Roy (2003), The Linear Programming Approach to Approximate Dynamic Programming. Operations Research **51**, 850 – 865.
[3] K.C. Frank, R.Q. Zhang, and I. Duenyas(2003), Optimal Policies for Inventory Systems with Priority Demand Classes. Operations Research **51**, 993 – 1002.

Reporter: Armin Fügenschuh and Jan van der Veen

Participants

Prof. Dr. Michael O. Ball

mball@rhsmith.umd.edu

Department of Decision and
Information Technologies
RHSmith School of Business
YMH Building, University of MD
College Park MD 20742-1815 – USA

Dr. Daniel Bienstock

dan@ieor.columbia.edu

dano@columbia.edu

Dept. of Industrial Engineering and
Operations Research, Columbia Univ.
Seeley W., Mudd Building
500 West 120th Street
New York, NY 10027 – USA

Prof. Dr. Robert E. Bixby

bixby@caam.rice.edu

bixby@ilog.com

6020 Annapolis Street
Houston TX 77005 – USA

PD Dr. Heinrich Braun

Heinrich.Braun@sap.com

SAP AG
BSG M 3
Neurottstr. 16
D–69190 Walldorf

Prof. Dr. Hans Daduna

daduna@math.uni-hamburg.de

Fachbereich Mathematik - ST
Universität Hamburg
Bundesstr. 55
D–20146 Hamburg

Prof. Dr. Brenda L. Dietrich

dietric@watson.ibm.com

IBM Corporation
Thomas J. Watson Research Center
P. O. Box 218
Yorktown Heights, NY 10598 – USA

Prof. Dr. Awi Federgruen

af7@columbia.edu

Graduate School of Business
Columbia University
Mail code 9142
101 Uris Hall
New York NY 10027 – USA

Prof. Dr. Bernhard Fleischmann

Bernhard.Fleischmann@wiwi.uni-augsburg.de

Betriebswirtschaftslehre mit
Schwerpunkt Produktion und Logistik
Universität Augsburg
D–86135 Augsburg

Armin Fügenschuh

fuegenschuh@mathematik.tu-darmstadt.de

Diskrete Optimierung
FB Mathematik
Technische Universität Darmstadt
Schlossgartenstr. 7
D–64289 Darmstadt

Prof. Dr. Vincent Gosselin

gosselin@ilog.com.sg

ILOG (S) Pte Ltd
#03-24/27 Cintech III
77 Science Park Drive
Singapore 118256 – Singapore

Prof. Dr. Martin Grötschel

groetschel@zib.de

Konrad-Zuse-Zentrum für
Informationstechnik Berlin (ZIB)
Takustr. 7
D–14195 Berlin

Prof. Dr. Karla Hoffman

khoffman@gmu.edu

Dept. of Systems Engineering and
Operation Research
George Mason University
4400 University Drive
Fairfax, VA 22030 – USA

Prof. Dr. Phil Kaminsky

Kaminsky@ieor.berkeley.edu

Department of Industrial
Engineering and Operations Research
University of California
Etcheverry Hall
Berkeley, CA 94720-1777 – USA

Prof. Dr. Claude Le Pape

clepape@ilog.fr

ILOG SA
BP 85
9, rue de Verdun
F-94253 Gentilly Cedex

Retsef Levi

levi@orie.cornell.edu

rl227@cornell.edu

OR&IE Department
College of Engineering
Cornell University
206 Rhodes Hall
Ithaca NY 14853 – USA

Prof. Dr. Alexander Martin

martin@mathematik.tu-darmstadt.de

FB Mathematik, AG 7
TU Darmstadt
Schlossgartenstr. 7
D–64289 Darmstadt

Victor Martinez de Albeniz

albeniz@MIT.EDU

MIT
Room E40-130
77, Massachussets Avenue
Cambridge MA 02139 – USA

Prof. Dr. S. Thomas McCormick

stmv@adk.commerce.ubc.ca

Sauder School of Business
University of British Columbia
463 Henry Angus Building
2053 Main Mall
Vancouver, B.C. V6T 1Z2 – Canada

Prof. Dr. Ana Muriel

muriel@ecs.umass.edu

MIE Department
University of Massachusetts
220 ELAB
Amherst MA 01003-2210 – USA

Prof. Dr. George L. Nemhauser

george.nemhauser@isye.gatech.edu

Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30332-0205 – USA

Prof. Dr. Stefan Nickel

s.nickel@wiwi.uni-sb.de

s.nickel@orl.uni-saarland.de

Lehrstuhl für Operations Research
und Logistik
Universität des Saarlandes
Postfach 151150
D–66041 Saarbrücken

Prof. Dr. Maurice Queyranne

Maurice.Queyranne@sauder.ubc.ca

Maurice.Queyranne@imag.fr

Sauder School of Business
University of British Columbia
463 Henry Angus Building
2053 Main Mall
Vancouver, B.C. V6T 1Z2 – Canada

Prof. Dr. H. Edwin Romeijn

romeijn@ise.ufl.edu

Dept. of Industrial & Systems Eng.
University of Florida
303 Weil Hall
P.O.Box 116595
Gainesville FL 32611-6595 – USA

Dr. Ed Rothberg

rothberg@ilog.com

ILOG, Inc.
1080 Linda Vista Ave.
Mountain View CA 94043 – USA

Prof. Dr. Martin W.P. Savelsbergh

mwps@isye.gatech.edu

martin.savelsbergh@isye.gatech.edu

Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30332-0205 – USA

Prof. Dr. Jeremy F. Shapiro

jshapiro@mit.edu

jshapiro@slimcorp.com

SLIM Technologies
LLC
226 Commonwealth Avenue
Boston MA 02116 – USA

Prof. Dr. Zuo-Jun Max Shen

shen@ise.ufl.edu

Dept. of Industrial & Systems Eng.
University of Florida
303 Weil Hall
P.O.Box 116595
Gainesville FL 32611-6595 – USA

Prof. Dr. David Simchi-Levi

dslevi@mit.edu

levi@iems.nwu.edu

Dept. of Civil and Environmental
Engineering
Massachusetts Inst. of Technology
77 Massachusetts Ave.
Cambridge, MA 02139 – USA

Prof. Dr. Hartmut Stadtler

stadtler@bwl.tu-darmstadt.de

FB 1, Institut für Betriebs-
wirtschaftslehre, TU Darmstadt
FG Produktion & Supply Chain Manag.
Hochschulstr. 1
D–64289 Darmstadt

Dipl.-Wir.-Ing. Christopher Sürie

suerie@bwl.tu-darmstadt.de

FB 1, Institut für Betriebs-
wirtschaftslehre, TU Darmstadt
FG Produktion & Supply Chain Manag.
Hochschulstr. 1
D–64289 Darmstadt

Prof. Dr. Leena Suhl

suhl@uni-paderborn.de

FB 5

Universität Paderborn

Warburger Str. 100

D–33098 Paderborn

Dipl. Math. Jan van der Veen

J.van-der-Veen@tu-bs.de

Abteilung für Mathematische
Optimierung

TU Braunschweig

Pockelsstr. 14

D–38106 Braunschweig

Prof. Dr. Stefan Voß

stefan.voss@tu-bs.de

stefan.voss@uni-hamburg.de

Institut für Wirtschaftsinformatik

Universität Hamburg

Von-Melle-Park 5

D–20146 Hamburg

Prof. Dr. Laurence A. Wolsey

wolsey@core.ucl.ac.be

Center for Operations Research and
Econometrics

Universite de Louvain

34 Voie du Roman Pays

B-1348 Louvain-la-Neuve

Prof. Dr. S. David Wu

david.wu@lehigh.edu

Department of Industrial and
Systems Engineering

Lehigh University

200 West Packer Avenue

Bethlehem PA 18015-1582 – USA

Prof. Dr. Uwe Zimmermann

u.zimmermann@tu-bs.de

Abteilung für Mathematische
Optimierung

TU Braunschweig

Pockelsstr. 14

D–38106 Braunschweig