DEPARTEMENT BEDRIJFSECONOMIE

ACLIPS : A CAPACITY AND LEAD TIME INTEGRATED PROCEDURE FOR SCHEDULING

by

Marc R. LAMBRECHT
Philip L. IVENS
Nico J. VANDAELE

WORKING PAPER
97-256
November 1997

D/1997/2263/13

ACLIPS: A CAPACITY AND LEAD TIME INTEGRATED PROCEDURE FOR SCHEDULING [1]

MARC R. LAMBRECHT, PHILIP L. IVENS, NICO J. VANDAELE

Department of Applied Economics, Katholieke Universiteit Leuven, Belgium 3000

Department of Applied Economics, Katholieke Universiteit Leuven, Belgium 3000

Faculty of Applied Economics, UFSIA, Antwerp, Belgium 2000

Abstract

We propose a general hierarchical procedure to address real-life job shop scheduling problems. The shop typically produces a variety of products, each with its own arrival stream, its own route through the shop and a given customer due date. The procedure first determines the manufacturing lot sizes for each product. The objective is to minimize the expected lead time, and therefore we model the production environment as a queueing network. Given these lead times, release dates are set dynamically. This in turn creates a time window for every manufacturing order in which the various operations have to be sequenced. The sequencing logic is based on an Extended Shifting Bottleneck Procedure. These three major decisions are next incorporated into a four-phase, hierarchical, operational implementation scheme. A small numerical example is used to illustrate the methodology. The final objective however is to develop a procedure that is useful for large, real-life shops. We therefore report on a real-life application.

[1]The authors like to thank the Departemental Editor and three anonymous referees for their valuable comments. This research was supported by the Science Foundation of Flanders (FWO-project G.0063.98). We also thank the management team of Clark-Hurth Components for offering us the interesting case study and the support of the Tractebel Chair.

1 Introduction

The production environment is a multi-operation job shop under an assemble(make)-to-order policy. Customers arrive dynamically and each customer order is characterized by a certain volume, mix and an agreed due date. Each order requires several operations on different machines; the routes, which are characterized by a bill of processes, are not necessarily the same for each order. We moreover explicitly include the stochastic nature of the production system, on the level of the customer orders itself (the shop typically produces a large variety of products each with its own stochastic arrival stream) and on the level of the shop floor, where processing and setup times are not deterministic due to all sorts of variability and disruptions. In this paper we propose a general procedure to address this problem. The situation is similar to the ones described by Wein and Chevalier [25], Shantikumar and Sumita [21] and Zijm and Buitenhek [27] who consider due date setting whereas we assume customer confirmed due dates and determine the release date accordingly.

The methodology used is a hierarchical approach in which we link separate applications into an integrated planning and scheduling system. The hierarchical approach we propose consists of three important decisions. The first decision is a lot sizing decision. Individual customer orders for the same product are grouped into manufacturing orders, which minimize the expected total product lead time. We developed our own queueing network approach where all parameters are a function of the lot size (see section 2). This has many advantages: it explicitly includes the convex relationship between lot sizes and lead times (see e.g. Karmarkar [14] and Lambrecht and Vandaele [17]); it takes care of congestion phenomena (the impact of the utilization of the most heavily loaded machines); it quantifies the queueing delays and it takes into account the stochastic nature of the problem. At this stage we need estimates of e.g. the customer order arrival rate, the squared coefficient of the customer order interarrival times, etc. The result of the queueing network is target lot sizes which give an indication of how customer orders have to be grouped into manufacturing orders. We group booked customer orders in such a way that we approach the target lot sizes as close as possible. Given the time varying nature of the booked customer demands, the manufacturing orders may actually differ from manufacturing order to manufacturing order, but on the average we aim for lot sizes minimizing the expected lead time (and work-in-process). Most job shops have the problem of fitting in incoming customer orders quickly. The detailed real-time scheduling required to manage this, is in our approach integrated with planning through the target lot sizes and the lead time off-setting (cfr. infra).

The second major decision is the determination of the release date of the manufacturing orders. The release date is set equal to the due date minus the lead time estimate of the manufacturing order (a grouping of booked customer orders). The estimate of the lead time is equal to the expected lead time plus a safety lead time. The safety lead time depends on the customer service. The lead time estimate is such that we expect to satisfy customer orders on time, P% of the time. This of course requires knowledge of the variance and the probability distribution of the lead time.

The third major decision concerns the sequencing policy. In the previous step a time window (expected lead time plus safety time) is created for every manufacturing order. Within these time windows (one for every manufacturing order) we now have to sequence all operations in detail. We opted for the shifting bottleneck procedure (Adams, Balas and Zawack [1]) for various reasons, one being its excellent performance as described by Ivens and Lambrecht [13]. The shifting bottleneck procedure has to be adapted so that it can be used to sequence the operations for our general job shop environment including assembly operations, release dates, due dates, overlapping operations, multiple resources (machines and labour force), setup times, calendars and many other real-life features. The ESBP (Extended Shifting Bottleneck procedure) is described in section 3. This sequencing application can clearly be interpreted as a deterministic real-time scheduler. There is no conflict with the two previously described stochastic applications. The stochastic applications result in realistic estimates of time windows. Real-time scheduling is a very dynamic process which will need (due to many changes) frequent rescheduling. It is hoped for that the estimates of the time windows are robust so that most of the due dates are finally met.

This methodology, based on three major decisions (the lot size decision, the release decision and the sequencing decision), is next transformed into a hierarchical, four phase, operational implementation scheme as summarized in figure 1. Phase one is the lead time estimation and lot sizing step. In this phase the manufacturing system is transformed into a queueing network. The outcomes are lot sizes and lead time estimations. The second step is a tuning phase where management intervention is required. Management may consider the lead times as unacceptable and may decide to adjust the capacity structure (e.g. overtime, capacity expansion), to off-load heavily loaded resources, to consider alternative routings, etc. The adjustments may result in a new run of the queueing model. The actions to be taken here depend upon the practical situation on hand. The next phase is the scheduling phase, including (a) the grouping of customer orders into manufacturing orders; (b) determining the release date for each manufacturing order; and (c) the detailed sequencing of all operations. In the final phase, the detailed plans are transferred to the shop floor on a real-time basis.

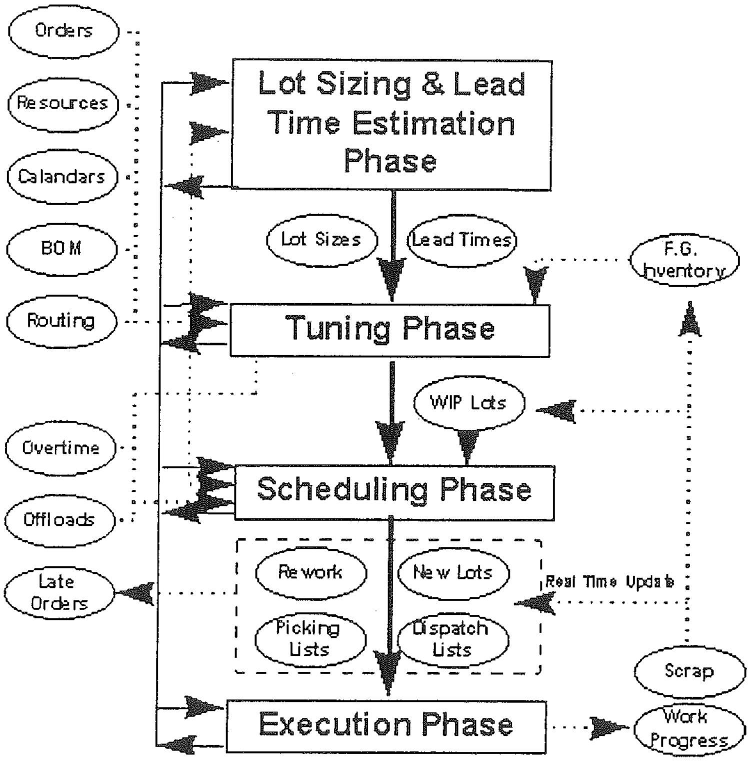

Figure 1: The four phase hierarchical approach of ACLIPS

Through electronic data captation, information concerning the execution of the detailed plan is fed back so that rescheduling can be done. The nature and frequency of rescheduling heavily depends on the dynamics of the situation and the level of responsiveness required. The system described above was named ACLIPS (A Capacity and Lead Time Integrated Procedure for Scheduling) and has been fully implemented in a metal working company. The application will be described in section 4.

The objective of our hierarchical approach is to obtain an integrated planning and scheduling system. Lead times are estimated through a queueing model taking into account congestion phenomena and the queueing impact of lot sizing (which results in a simultaneous treatment of both capacity and material flow). Standard lead time off-setting (as is e.g. done in MRP) is replaced by realistic estimates of order release dates. The lead time estimates include a safety margin so that customer service targets can be specified. The detailed real-time scheduler operates within time windows allowing to deal explicitly with the dynamics of the floor. The tuning phase allows a management intervention to cope with the capacity/inventory(lead time) trade-off.

The remainder of the paper is organized as follows. The lead time estimation and lot sizing phase is discussed in section 2. The scheduling phase is explained in section 3. We discuss a real-life application in section 4. We draw conclusions in section 5.

2 Lead time estimation and lot sizing phase

2.1 Introduction

Our model clearly builds on well known queueing approximations found in the literature, but our approach adds some additional features making it more suitable for practical applications. The most significant differences with existing approximations are the following. First, our expressions for the expected lead time and the variance of the lead time are written as a function of the lot size. This allows us to use an optimization routine to find the optimal lot sizes for all products simultaneously. Second, our model explicitly includes setup times. Exact analytic solutions have been proven to be mathematically untractable except for some very specific problem settings (for more details, see Lambrecht, Chen and Vandaele [16]). Third, we further extended and adapted existing approximations by including the following refinements: the average aggregate batch processing time, the scv of the aggregate batch processing time, the selection of more appropriate weights in the objective function; the determination of the squared coefficient of variation of the (aggregated) external arrival stream; a modified approximation for the variance of the lead time; the use of the lognormal distribution to approximate the lead time distribution function and the introduction of the concept of customer service. All these refinements enable us to better describe the stochastic nature of the production environment. Similar approaches are described by Shantikumar and Buzacott [20] Buzacott and Shantikumar [9], Shantikumar and Sumita [21], Bitran and Tirupati [6], [7]. Software pacckages include, among others, MPX (Suri and DeTreville [23]) and QNA (Whitt [26]).

In our approach, equations (for the expected lead time and the variance of the lead time) are derived that capture the dynamics of the system in an aggregate way. The arrival process for each product is characterized by the expected customer demand and the average and variance of the order interarrival times. The exogenous arrival rate can be estimated from historical data or from demand forecasts or even confirmed orders depending on the availability of data. The other parameters are: the service times (average and variance of both setup and unit processing times) and shop parameters such as routings and calendars. The outcome of the model are expressions for the expected lead time and variance of the lead time as a function of the lot size. Although we rely on approximations, simulation studies (for small examples) turned out that the approximations behave satisfactory (see Vandaele [24]). The deviations between the approximations and the simulation results are in line with other results available in the literature. We assume a constant lot size per product over the entire routing. Next an optimization routine is used to find the lot sizes that minimize the expected lead time. We call these lot sizes 'target lot sizes'. A lognormal distribution is postulated to characterize the lead time distribution. This in turn allows the user to specify a lead time, satisfying a predetermined customer service (lead time percentile).

Throughout the paper a small example will be used to numerically illustrate the various steps of our procedure. The shop, a small metal shop, fabricates two products, P and S, and has three machine (centers) types: a cutter (C), a grinder (G) and a lathe (L). Product P has three stages on its route (on machine C, G and L) and product S has two stages (on machine L and G). The shop runs three shifts per day, seven days a week. There is one machine available of each type. The customer demands for both products are summarized in table 1.

	Product P	Product S
Average interarrival time (hours)	144	48
Variance interarrival time	3744	494
Average order quantity (units)	3	2

Table 1: Demand characteristics for products P and S

Table 1 is interpreted as follows: for product P we expect a customer order every 144 hours, while the average order size equals 3 units. The processing and setup times are summarized in table 2 (all times are expressed in hours). In table 2 it can be seen that both the cutter and the lathe have deterministic setup and processing times. The grinder faces exponentially distributed setup and processing times.

product	operations	machine	setup average	setup variance	processing average	processing variance
Product P	3	cutter	20	0	30	0
		grinder	20	400	10	100
		lathe	24	0	12	0
Product S	2	lathe	16	0	8	0
		grinder	20	400	10	100

Table 2: Production characteristics of the metal shop (expressed in hours)

2.2 Model derivation

We will now discuss the formal treatment of the lead time estimation and lot sizing phase. Assume k to be the product index ($k = 1 \ldots K$), m the machine index ($m = 1 \ldots M$) and o the operation index for product k ($o = 1 \ldots O_k$), where O_k is the number of operations for product k. Each product k is characterized by an average order quantity $\overline{OQ}_k$, an average order interarrival time $\overline{Y}_k$, the variance of the order interarrival time $s^2_{Y_k}$, the squared coefficient of the order interarrival time $c^2_{Y_k}$ and the arrival rate $\lambda_k = 1/\overline{Y}_k$. For the small metal shop we assume the following characteristics: $\overline{Y}_P = 144$ and $\overline{Y}_S = 48$, $s^2_{Y_P} = 3744$ and $s^2_{Y_S} = 494$, $c^2_{Y_P} = 13/72$ and $c^2_{Y_S} = 3/14$, $\lambda_P = 1/144$ and $\lambda_S = 1/48$, $\overline{OQ}_P = 3$ and $\overline{OQ}_S = 2$.

As far as the production characteristics are concerned, the following is defined for product k and operation o, expressed in hours: T_{ko}, setup time random variable; X_{ko}, unit processing time random variable; $\overline{T}_{ko}$, expected setup time; $\overline{X}_{ko}$. expected unit processing time; μ_{ko}, unit processing rate ($= 1/\overline{X}_{ko}$); $s^2_{T_{ko}}$, variance of the setup time; $s^2_{X_{ko}}$, variance of the unit processing time; $c^2_{T_{ko}}$, scv of the setup time; $c^2_{X_{ko}}$, scv of the unit processing time. In addition, define $\delta_{kom} = 1$ if operation o for product k is on machine m and 0 otherwise. The routing of the metal shop consequently results in: $\delta_{P1C} = 1$, $\delta_{P2G} = 1$, $\delta_{P3L} = 1$, $\delta_{S1L} = 1$, $\delta_{S2G} = 1$ and all other δ_{kom}'s equal 0.

At this point all the input parameters are given. We use a queueing network approach to model the job shop. Each machine is modelled as a multi product lot sizing model with queueing delays. The multiple arrival processes of the k products are superposed into one aggregate arrival process. All characteristics of the aggregate arrival process and the aggregate production process are functions of the lot sizes Q_k. Note that we express Q_k as a multiplier of the average order quantity $\overline{OQ}_k$. For each machine m we have to obtain: l_m, the aggregate batch arrival rate; ca^2_m, the scv of the aggregate batch interarrival time; ca'^2_m, the scv of the external aggregate batch interarrival time; μ_m, the aggregate batch processing rate; cs^2_m, the scv of the aggregate batch processing time; ρ'_m, the adapted traffic intensity.

The aggregate arrival process at machine m is characterized by the average and the scv of the aggregate batch interarrival times. Note that the batch arrival rate of product k at the first machine of its routing equals $\lambda_{b_k} = \lambda_k/Q_k$ which is a result of grouping the order quantities into a manufacturing batch of size $Q_k\overline{OQ}_k$ (expressed in units). The aggregate batch arrival rate of product k at machine m equals $l_{mk} = \sum_{o=1}^{O_k} \lambda_{b_k}\delta_{kom}$. Then the aggregate batch arrival rate at machine m equals $l_m = \sum_{k=1}^{K}\sum_{o=1}^{O_k} \lambda_{b_k}\delta_{kom}$ which includes both the internal and the external batch arrivals at machine m. The external aggregate batch arrival rate at machine m equals $l'_m = \sum_{k=1}^{K} \lambda_{b_k}\delta_{k1m}$. For our numerical example we obtain $\lambda_{b_P} = 1/144Q_P$, $\lambda_{b_S} = 1/48Q_S$ and $l_C = 1/144Q_P$, $l_G = 1/144Q_P + 1/48Q_S$, $l_L = 1/144Q_P + 1/48Q_S$ and

$l'_C = 1/144Q_P$, $l'_G = 0$, $l'_L = 1/48Q_S$.

We now turn to the production process at machine m. The aggregate batch processing time on machine m equals $1/\mu_m = \sum_{k=1}^{K} \frac{l_{mk}}{l_m} \sum_{o=1}^{O_k} \frac{\lambda_{b_k}\delta_{kom}}{l_{mk}} (\overline{T}_{ko} + Q_k\overline{OQ}_k\overline{X}_{ko})$ where l_{mk}/l_m is the probability that a randomly picked product in front of machine m is of product type k. The expression for $1/\mu_m$ is a weighted average over product batch processing times, which are in turn weighted averages of the operations on machine m for the same product. For the numerical example we obtain $\frac{1}{\mu_C} = 20+90Q_P$, $\frac{1}{\mu_G} = \frac{48Q_S}{48Q_S+144Q_P}(20+30Q_P)+\frac{144Q_P}{48Q_S+144Q_P}(20+20Q_S)$ and $\frac{1}{\mu_L} = \frac{48Q_S}{48Q_S+144Q_P}(24+36Q_P) + \frac{144Q_P}{48Q_S+144Q_P}(16+16Q_S)$.

Along the same lines, we obtain the scv of the aggregate batch processing time

$$\begin{aligned} cs^2_m &= \left[\sum_{k=1}^{K} \frac{l_{mk}}{l_m} \sum_{o=1}^{O_k} \frac{\lambda_{b_k}\delta_{kom}}{l_{mk}} [\overline{T}_{ko} + Q_k\overline{OQ}_k\overline{X}_{ko}]^2\right] \mu^2_m - 1 \\ &\quad + \sum_{k=1}^{K} \frac{l_{mk}}{l_m} \sum_{o=1}^{O_k} \frac{\lambda_{b_k}\delta_{kom}}{l_{mk}} \frac{[s^2_{T_{ko}} + Q_k\overline{OQ}_k s^2_{X_{ko}}]}{[\overline{T}_{ko} + Q_k\overline{OQ}_k\overline{X}_{ko}]^2} \end{aligned} \tag{1}$$

Applied for the small metal shop

$$\begin{aligned} cs^2_C &= \frac{(20+90Q_P)^2}{(20+90Q_P)^2} - 1 + 0 = 0 \\ cs^2_G &= \frac{\frac{48Q_S}{48Q_S+144Q_P}(20+30Q_P)^2 + \frac{144Q_P}{48Q_S+144Q_P}(20+20Q_S)^2}{\left[\frac{48Q_S}{48Q_S+144Q_P}(20+30Q_P) + \frac{144Q_P}{48Q_S+144Q_P}(20+20Q_S)\right]^2} - 1 \\ &\quad + \frac{48Q_S}{48Q_S+144Q_P}\left[\frac{400+300Q_P}{(20+30Q_P)^2}\right] + \frac{144Q_P}{48Q_S+144Q_P}\left[\frac{400+200Q_S}{(20+20Q_S)_2}\right] \\ cs^2_L &= \frac{\frac{48Q_S}{48Q_S+144Q_P}(24+36Q_P)^2 + \frac{144Q_P}{48Q_S+144Q_P}(16+16Q_S)^2}{\left[\frac{48Q_S}{48Q_S+144Q_P}(24+36Q_P) + \frac{144Q_P}{48Q_S+144Q_P}(16+16Q_S)\right]^2} - 1 + 0 \end{aligned}$$

When setup times are included in the machine utilization, we define the adapted traffic intensity ρ', which includes both the utilization due to setups and the utilization due to processing. The utilization without setup is the traditional traffic intensity ρ. Now we can determine the adapted traffic intensity for machine m

$$\rho'_m = \frac{l_m}{\mu_m} = \sum_{k=1}^{K}\sum_{o=1}^{O_k} \lambda_{b_k}\delta_{kom}[\overline{T}_{ko} + Q_k\overline{OQ}_k\overline{X}_{ko}] = \sum_{k=1}^{K}\sum_{o=1}^{O_k} \lambda_{b_k}\delta_{kom}\overline{T}_{ko} + \rho \tag{2}$$

and applied for the metal shop: $\rho'_C = \frac{5}{36Q_P} + \frac{5}{24}$, $\rho'_G = \frac{5}{36Q_P} + \frac{5}{12Q_S} + \frac{5}{8}$ and $\rho'_L = \frac{1}{6Q_P} + \frac{1}{3Q_S} + \frac{7}{12}$.

Further define f_{0n}, the proportion of batches from outside and going to machine n, f_{mn}, the proportion of batches leaving machine m and going to machine n, f_{m0}, the proportion of batches leaving machine m and going outside, and $\mathbf{F}$, the transition matrix of f_{mn} $(m, n = 0 \ldots M)$. Solving the following set of linear equations yields the M unknowns ca_m^2, $m = 1, \ldots, M$:

$$-\sum_{n=1}^{M} l_n f_{nm}^2 (1 - \rho'_n) ca_n^2 + l_m ca_m^2 = \sum_{n=1}^{M} l_n f_{nm} (f_{nm} {\rho'_n}^2 cs_n^2 + 1 - f_{nm}) + l'_m {ca'}_m^2 \qquad (3)$$

Equations (3) are a slightly adapted version (in terms of general exogenous arrivals instead of Poisson arrivals) of the results obtained by Shantikumar and Buzacott [20]. The entrances of the transition matrix $\mathbf{F}$ are obtained as follows: $f_{0n} = l'_n / \sum_{m=1}^{M} l'_m$, $f_{mn} = (1/l_m) \sum_{k=1}^{K} \sum_{l=1}^{O_k - 1} \lambda_{b_k} \delta_{kom} \delta_{ko+1n}$, $f_{m0} = (1/l_m) \sum_{k=1}^{K} \lambda_{b_k} \delta_{k O_k m}$ for $n = 1 \ldots M$ and $m = 1 \ldots M$. Note that in our model, due to the fact that the routings are given, we face deterministic routing. Therefore, the transition matrix $\mathbf{F}$ can be derived in the way described above.

Returning to the small metal shop we have the following transition matrix $\mathbf{F}$:

	0	C	G	L
0	0	$\frac{48Q_S}{48Q_S+144Q_P}$	0	$\frac{144Q_P}{48Q_S+144Q_P}$
C	0	0	1	0
G	$\frac{144Q_P}{48Q_S+144Q_P}$	0	0	$\frac{48Q_S}{48Q_S+144Q_P}$
L	$\frac{48Q_S}{48Q_S+144Q_P}$	0	$\frac{144Q_P}{48Q_S+144Q_P}$	0

To obtain ca'^2_m we use the following approximation: if $\sum_{k=1}^{K} \delta_{k1m} \geq 2$, then

$$ca'^2_m \approx \frac{1}{3} + \frac{2}{3} \sum_{k=1}^{K} \frac{\lambda_{b_k} \delta_{k1m}}{\sum_{k=1}^{K} \lambda_{b_k} \delta_{k1m}} \frac{c^2_{Y_k}}{Q_k} = \frac{1}{3} + \frac{2}{3} \sum_{k=1}^{K} \frac{\lambda_{b_k} \delta_{k1m}}{l'_m} \frac{c^2_{Y_k}}{Q_k}$$

If $\sum_{k=1}^{K} \delta_{k1m} = 1$ then $ca'^2_m = \frac{c^2_{Y_k}}{Q_k}$.

The approximation for ca'^2_m is the sum of a constant and a weighted average of the scv's of all the external batch arrivals at machine m. It is an interpolation between complete deterministic arrivals (where the aggregate scv is approximated by the scv of a uniform distribution $U[0, 2/l'_m]$ and the scv of Poisson arrivals (where all scv's equal one). The latter is the only known exact result in the literature for the superposition of arrival processes. The weights 1/3 and 2/3 in the expression for ca'^2_m are a particular instance of a general approximation described by Albin [2].

For our illustrative example we obtain $ca'^2_C = \frac{13}{72} \frac{1}{Q_P}$, $ca'^2_G = 0$, $ca'^2_L = \frac{3}{14} \frac{1}{Q_S}$. Then finally the lead time for product k on machine m for operation o is

$$E(W_{ko}) = \sum_{m=1}^{M} E(Wq_m) \delta_{kom} + \overline{T}_{ko} + Q_k \overline{OQ}_k \overline{X}_{ko}$$

with

$$\begin{aligned} E(Wq_m) &= \frac{{\rho'_m}^2 (ca^2_m + cs^2_m)}{2 l_m (1-\rho'_m)} \exp\left\{ \frac{-2(1-\rho'_m)(1-ca^2_m)^2}{3\rho'_m (ca^2_m + cs^2_m)} \right\} & \text{if } ca^2_m \leq 1 \\ E(Wq_m) &= \frac{{\rho'_m}^2 (ca^2_m + cs^2_m)}{2 l_m (1-\rho'_m)} & \text{if } ca^2_m > 1 \end{aligned}$$

This approximation is based on the well-known Kraemer-Lagenbach-Belz [15] approximation, which has been tested widely in the literature (see e.g. Shantikumar and Buzacott [19]). Although some authors suggest modifications to this basic approximation (see e.g. Bitran and Tirupati [6]), we found that the baisc Kraemer-Lagenbach-Belz approximation performs well for our purposes.

The aggregated objective function for machine m can be stated as follows

$$E(W_{M_m}) = E(Wq_m) + \sum_{k=1}^{K} \frac{\sum_{o=1}^{O_k} \lambda_k \overline{OQ}_k \delta_{kom}}{\sum_{k=1}^{K}\sum_{o=1}^{O_k} \lambda_k \overline{OQ}_k \delta_{kom}} \left(\sum_{o=1}^{O_k} \frac{\lambda_k \overline{OQ}_k \delta_{kom}}{\sum_{o=1}^{O_k} \lambda_k \overline{OQ}_k \delta_{kom}} [\overline{T}_{ko} + Q_k \overline{OQ}_k \overline{X}_{ko}] \right)$$

This objective function for machine m is the weighted average over the products visiting machine m, which on their turn are weighted averages over the operations on machine m for product k. Note that weight $\sum_{o=1}^{O_k} \lambda_k \overline{OQ}_k \delta_{kom} / \sum_{k=1}^{K}\sum_{o=1}^{O_k} \lambda_k \overline{OQ}_k \delta_{kom}$ is independent from the manufacturing lot size multiplier. It measures the relative importance of product k for machine m.

The objective function for the total job shop becomes

$$\begin{aligned} E(W) = {} & \sum_{m=1}^{M} E(Wq_m) + \sum_{k=1}^{K} \frac{\lambda_k \overline{OQ}_k}{\sum_{k=1}^{K} \lambda_k \overline{OQ}_k} \frac{[Q_k \overline{OQ}_k - 1]\overline{Y}_k}{2\overline{OQ}_k} + \\ & \sum_{m=1}^{M}\sum_{k=1}^{K} \frac{\sum_{o=1}^{O_k} \lambda_k \overline{OQ}_k \delta_{kom}}{\sum_{k=1}^{K}\sum_{o=1}^{O_k} \lambda_k \overline{OQ}_k \delta_{kom}} \left(\sum_{o=1}^{O_k} \frac{\lambda_k \overline{OQ}_k \delta_{kom}}{\sum_{o=1}^{O_k} \lambda_k \overline{OQ}_k \delta_{kom}} [\overline{T}_{ko} + Q_k \overline{OQ}_k \overline{X}_{ko}] \right) \end{aligned} \quad (4)$$

The weight $\lambda_k \overline{OQ}_k / \sum_{k=1}^{K} \lambda_k \overline{OQ}_k$ takes care of the relative importance of product k for the total job shop. The second sum of equation (4) measures the average waiting time of finished batches until their due date. For the metal shop the objective function (as a function of the lot size multiplier Q_k) for the entire job shop equals

$$\begin{aligned} E(W) = {} & E(W_{q_C}) + E(W_{q_G}) + E(W_{q_L}) + 20 + 90Q_P + \frac{1}{3}(20 + 30Q_P) + \frac{2}{3}(20 + 20Q_S) \\ & + \frac{1}{3}(24 + 36Q_P) + \frac{2}{3}(16 + 16Q_S) + 8(3Q_P - 1) + 8(2Q_S - 1) \end{aligned}$$

At this point the formulation of the job shop is complete.

2.3 Optimization and decomposition

The minimization problem involves a non-linear objective function and a set of simultaneous, non-linear constraints. A dedicated optimization routine has been developed to solve the problem and is described in Vandaele [24]. The optimal lot sizes for the small metal shop are $Q^*_P\overline{OQ}_P = 4$ and $Q^*_S\overline{OQ}_S = 6$. Note that when we use the term optimal, we refer to the values minimizing the objective function (the objective function of course being an approximation). The decomposition, after the optimization, can be summarized as follows. The optimal multipliers Q^*_k (or the vector $\mathbf{Q}^*$) for each product are used to calculate the expected lead time of operation o of product k on machine m, $E(W_{ko}) = \sum_{m=1}^{M} E(Wq_m(\mathbf{Q}^*))\delta_{kom} + \overline{T}_{ko} + Q^*_k\overline{OQ}_k\overline{X}_{ko}$. The first term is clearly common for all products using machine m. The total lead time of product k (for the whole routing) is given by

$$E(W_k) = \frac{(Q^*_k\overline{OQ}_k - 1)\overline{Y}_k}{2\overline{OQ}_k} + \sum_{o=1}^{O_k}\left(\sum_{m=1}^{M} E(Wq_m(\mathbf{Q}^*))\delta_{kom} + \overline{T}_{ko} + Q^*_k\overline{OQ}_k\overline{X}_{ko}\right) \tag{5}$$

The numerical outcomes are summarized in table 3. From this table it can be seen that there is a small queue in front of the cutter. On the other hand, both the grinder and the lathe face long waiting times compared to their processing times. This is mainly due to the high adapted traffic intensities. The waiting time for the grinder is even larger. This is due to the stochastic nature of that machine. The operation 'stock' is the average time that a particular customer order (as part of a manufacturing batch) has to wait until its due date.

The variance of the total lead time of product k is approximated by

$$\begin{aligned} V(W_k) = {} & \frac{Q^*_k\overline{OQ}_k - 1}{2\overline{OQ}_k^2}s^2_{Y_k} + \frac{(Q^*_k\overline{OQ}_k - 1)(Q^*_k\overline{OQ}_k + 1)}{12\overline{OQ}_k^2}\overline{Y}_k^2 + \\ & \sum_{o=1}^{O_k} V(Wq_m(\mathbf{Q}^*))\delta_{kom} + \sum_{o=1}^{O_k} s^2_{T_{ko}} + \sum_{o=1}^{O_k} Q^*_k\overline{OQ}_k s^2_{X_{ko}} \end{aligned} \tag{6}$$

product	optimal lot size	operation	adapted traffic intensity (%)	waiting time	setup time	processing time	lead time
P	4	cutter	73	7	20	120	147
		grinder	87	109	20	40	169
		lathe	82	42	24	48	114
		stock					72
		total					502
S	6	lathe	82	42	16	48	106
		grinder	87	109	20	60	189
		stock					60
		total					355

Table 3: Optimal lot size and lead time for the metal shop

The term $V(Wq_m)$ is given in Vandaele [24]. For our example, the standard deviation of the total lead time is 158 hours for product P and 154 hours for product S which suggests that the lead times are highly variable. If the lognormal distribution is assumed, then the parameters are $\beta_k = \ln\left(\frac{E(W_k}{\sqrt{\frac{V(W_k)}{E(W_k)^2}+1}}\right)$ and $\gamma_k^2 = \ln\left(\frac{V(W_k)}{E(W_k)^2}+1\right)$. The lead times, including safety time, are obtained in the following way. W_{P_k} is the total lead time guaranteeing a service of P_k%. This means that the manufacturer will satisfy this lead time P_k% of the time for product k. Then

$$W_{P_k} = \exp\{\beta_k + z_{P_k}\gamma_k\} \tag{7}$$

where z_{P_k} can be obtained from the standard normal table (P_k is the required percentile for product k). Subsequentially we will call W_{P_k} the planned lead time, because it will be used to fix the planned release date. For our example, we show the planned lead time for some values of P_k in table 4.

P_k	80%	90%	95%	99%
Product P	621	710	794	980
Product S	463	554	644	855

Table 4: Some lead time percentiles, including stock time

3 Scheduling Phase

3.1 Grouping of Customer Orders into Manufacturing Orders

The problem addressed here is the grouping of C_k customer orders of product k, characterized by an order quantity OQ_{kc} $(1 \leq c \leq C_k)$ and a due date DD_{kc} $(1 \leq c \leq C_k)$, into a number of manufacturing orders $L_{kl}(l = 1, \dots, S_k)$ of which the number of units ideally approach the previously fixed target lotsize Q_k^*. In table 5 we give the various booked customer orders, covering roughly a time period of one month. As can be seen, we have 5 customer orders for product P and 15 orders for product S. Each order is characterized by an order quantity and a due date. Table 5 has to be interpreted as follows: one unit of product P has to be delivered at day 22, 5 units at day 28, 3 units at day 37, etc.

For each product k, we first fix the number of setups $S_k = \left\lfloor \frac{1}{Q_k^*} \sum_{c=1}^{C_k} OQ_{kc} \right\rfloor$ where $\lfloor x \rfloor$ is the largest integer smaller than or equal to x. This is a conservative rounding precluding infeasibility. In our case, S_P equals 3 ($\lfloor 15/4 \rfloor$) and S_s equals 5 ($\lfloor 30/6 \rfloor$). The grouping into manufacturing lots can be done in several ways. It is clear that this problem can

Product P	quantities	1	5	3	2	4										
	due dates (days)	22	28	37	41	44										
Product S	quantities	1	3	2	3	1	1	3	2	3	1	2	3	3	1	1
	due dates (days)	17	18	19	22	24	26	27	30	33	34	35	36	39	42	44

Table 5: Booked orders for products P and S

be formulated as an integer programming model or, more elegantly, transformed into a dynamic program. Given the standard nature of this problem we omit the formulation. It is however important to mention that the objective function we used minimizes the number of inventory-days. In table 6 we summarize the results for our illustrative case where QL_{kl} stands for the lot size of the new manufacturing orders. It should be clear that as long as the manufacturing lots are not physically released, the optimal grouping can be recalculated when new information becomes available. In this way we are able to react fast to changing circumstances, enabling the planner to fit in quickly new incoming orders. However, if the changes are drastic, such as a significant increase or decrease in the size and the number of customer orders, we opt for re-optimizing the target lot sizes, so that these changes will be reflected in the lead time estimates.

product	L_{kl}	Grouped Customer Orders (QL_{kl})
P	L_{Pl}	1-2 (6), 3-4 (5), 5 (4)
S	L_{Sl}	1-2-3 (6), 4-5-6 (5), 7-8 (5), 9-10-11-12 (9), 13-14-15 (5)

Table 6: The manufacturing lot sizes

3.2 Release of New Manufacturing Orders

For the newly determined manufacturing order quantities, QL_{kl}, we have to compute the corresponding expected lead time and the planned lead time (expected lead time plus safety time). Because each manufacturing order is due at the due date of the first customer order from this manufacturing order, we have to remove the terms accounting for the stock time from the equations (5) and (6) as $E(W_k) = \sum_{o=1}^{O_k} \left(\sum_{m=1}^{M} E(Wq_m(\mathbf{Q}^*))\delta_{kom} + \overline{T}_{ko} + Q_{kl}\overline{X}_{ko} \right)$ and $V(W_k) = \sum_{o=1}^{O_k} V(Wq_m(\mathbf{Q}^*))\delta_{kom} + \sum_{o=1}^{O_k} s^2_{T_{ko}} + \sum_{o=1}^{O_k} Q_{kl}s^2_{X_{ko}}$ respectively.

Next, the planned lead time (70% service) is obtained using expression (7) from section 2.3. Subsequently, the planned lead times are deducted from the due dates to obtain the release dates for each manufacturing order. These results are summarized in table 7. The due date for L_{P1} is day 22 (it includes customer orders 1 and 2) so the due date equals 528 hours from now. The expected lead time is calculated for each manufacturing batch. Across manufacturing batches for the same product, the waiting time and setup time are equal but the batch production times differ due to the different manufacturing quantities. The same is true for the lead time variance so that each manufacturing batch of a given lot size ends up with its own planned lead time (these planned lead times do not coincide with the lead time percentiles from table 4 because at this point now we do not include the stock time). The planned lead time is substracted from the due date and the release date is obtained. Negative release dates mean that the guaranteed service level will not be reached because the batch can only be released at the current moment. Due to the fact that the planned lead time incorporates both waiting time and safety time it is still possible, but less likely, that the manufacturing order is finished before the due date.

To conclude, this phase has set the time windows (planned lead time between release date and due date) for each manufacturing order. The various operations of each manufacturing order will be sequenced within these time windows. This will be covered in the next phase, the detailed scheduling phase.

Manufacturing Lot	Due Date	Expected Lead Time	Planned Lead Time	Release Date
$L_{P1}(6)$	528	534	593	-65
$L_{P2}(5)$	888	482	539	349
$L_{P3}(4)$	1056	430	484	572
$L_{S1}(6)$	408	295	338	70
$L_{S2}(5)$	528	277	318	210
$L_{S3}(5)$	648	277	318	330
$L_{S4}(9)$	792	349	398	394
$L_{S5}(5)$	936	277	318	618

Table 7: Release Dates of the Manufacturing Orders, expressed in hours

3.3 Detailed Scheduling of the Operations

At this stage of our procedure, all non-completed operations of manufacturing orders are scheduled between the release date (or the current moment if the order is overdue) and the due date of the order. Detailed scheduling requires to specify for each operation of each manufacturing order when it has to be performed and by what resource, explicitly taking into account the limited availability of the various resources and many other constraints such as precedence among operations, release dates and due dates. A schedule needs to optimize a predetermined objective. Many production managers strive for due date performance, short lead times and low in-process inventory levels.

The well known job shop scheduling problem is the theoretical abstraction of this problem and has been subject of numerous research efforts. Both optimal and heuristic solution procedures are proposed in the literature. Recent integer programming based models can be found in Applegate and Cook [3]. Among others, Carlier and Pinson [11] and Brucker, Jurisch and Sievers [8] propose implicit enumeration methods for solving the job shop problem. Unfortunately, the job shop scheduling problem is NP-hard in the strong sense. This implies that there is little hope to find optimal solutions to large real-life scheduling problems within reasonable computer time. For practical applications heuristic schedule generation procedures with priority dispatching rules are often used (First Come First Served, Shortest Processing Time, Earliest Due Date, Most Work Remaining, Critical Ratio,...). Other approaches exist such as tabu search, genetic algorithms and simulated annealing.

Adams, Balas and Zawack [1] introduced the Shifting Bottleneck Procedure (SBP), a new, powerful heuristic for the job shop scheduling problem. Dauzère-Péres and Lasserre [12] and Balas, Lenstra and Vazacopoulos [4] increased its performance and their experiments indicated that the SBP offers exceptionally good results compared to other heuristics such as priority dispatching rules. Other work on the shifting bottleneck procedure can be found in the recent book of Ovacik and Uzoy [18], which covers the research progress on the shifting bottleneck method. Because of the SBP's good balance between computational complexity and the quality of the generated schedules, we have chosen this procedure as the engine of our detailed scheduling phase.

3.3.1 The Extended Shifting Bottleneck Procedure (ESBP)

However, the scope of the theoretical job shop scheduling problem is far too limited to be applicable in practical environments. We therefore extended the SBP so that non-standard features such as release dates, due dates, assembly structures, split structures, overlapping operations, setup times, transportation times, parallel machines and in-process inventory can be modelled. For an in-depth treatment of these extensions we refer to Ivens and Lambrecht [13]. Other work in this area is offered by Schutten, Van de Velde and Zijm [22]. The extensions are modelled by an Extended Disjunctive Graph (EDG). This representation is similar to the disjunctive graph (Adams, Balas and Zawack [1]), but arcs can have labels to represent general precedence relationships, e.g. to allow overlapping, forced delays and waiting times. Product assemblies and splits can easily be modelled by allowing multiple predecessors and successors. Customer or manufacturing lots have a release date and a due date. We also allow restrictions on starting times or finishing times of individual operations (e.g. due to temporary unavailability of raw materials, labor, tools, etc...). Some resources are available in multiple units (i.e. parallel machines). Thus, in addition to sequencing, an assignment of operations to resources has to be done. Recent extensions include the use of resource calendars and the possibility that operations require more than one resource at a time. In addition, other performance criteria could be considered. In this paper we use the minimal maximum lateness criterion. For many practical applications, we found out that computer time needed for the machine problems constitutes no problem.

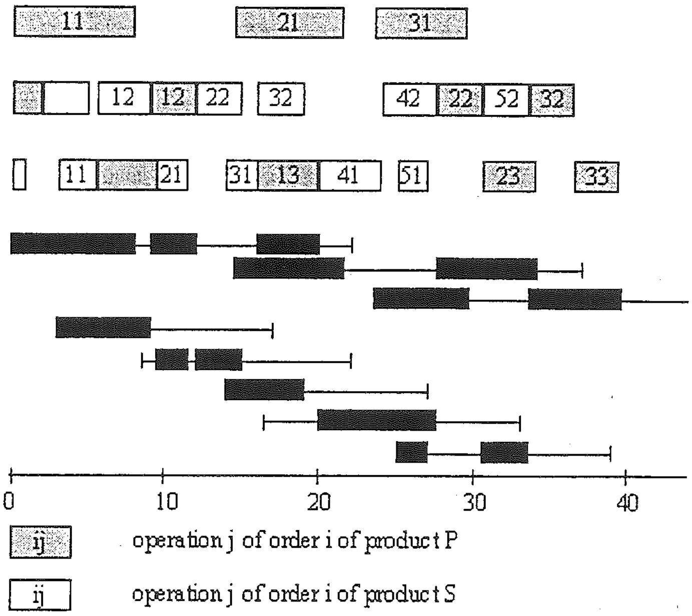

Figure 2: The Gantt chart of the detailed schedule

3.3.2 The Metal Shop Example

We will now illustrate the ESBP for the small metal shop example. The eight manufacturing orders from table 6 have to be scheduled. Currently, two previously released manufacturing orders reside on the shop floor. In addition to the two in-process orders, there are three manufacturing orders P with three operations and five manufacturing orders S with two operations. The output of the ESBP is visualized in figure 2 as a Gantt-chart. The Gantt-chart shows the two manufacturing orders already in process (the unlabeled blocks) and the eight newly released manufacturing orders. The shaded blocks stand for product P while the blank blocks visualize product S. Next, for each manufacturing order the time window (planned lead time) is shown. The release and due date for each order can be seen. Although order P1 could not start on its release date (as it is already late), due to the safety time, the order can still be finished on time. From the Gantt chart it can also be seen that the orders are scheduled as early as possible in their respective time windows (cf. minimal lateness criterion). Of course the sequencing of the various operations causes some operations to start later than the release date (S2). These time windows have been designed is such a way that the deterministic schedule will not be useless as soon as disruptions occur; the in-build slack (safety) time will absorb the impact of variability. The in-build slack is clearly visualized in figure 2. A re-optimization may be required in case there are too many changes.

3.3.3 The Execution Phase

In this phase, the detailed schedule will be executed and dispatching and picking lists can be obtained. A data captation system can transmit information concerning work progression back to ACLIPS. From time to time, a recalculation of the detailed schedule will be necessary because of the numerous changes on the shop floor. The frequency of recalculation is of course a function of the dynamics of the shop. A re-optimization of the lot sizing and lead time estimation phase will also be required now and then but, of course, less frequently. The lot sizing and lead time estimation phase is on an aggregate level and focuses more on a long term planning horizon. The information needed for this phase does not change very quickly. It also allows the user to evaluate what-if questions. Of course we have to avoid rescheduling each time that a single disruption occurs. That's where we benefit from the safety time included in the time windows.

4 An Application

The methodology outlined in this paper is well suited for real-life applications. In this section we report on an implementation in a medium-sized metal working company. We will stress the applicability of our approach and focus on the resulting benefits and savings. The quality of our approach can be demonstrated by comparing current lot sizing practice and the current scheduling practice with the lot sizes and schedules proposed by ACLIPS.

The metal working company we consider produces transmissions for off-road vehicles. Our application focuses on the shop where raw steel components are transformed into shafts and gears. The shop orders stem directly from the MRP requirements demanded by the next department, a furnace. In the future ACLIPS will be expanded to all shops, making the current MRP system of the company superfluous. The subsystem we consider consists of 70 machines and 556 different components which results in 3,484 different operations. On a yearly basis, this metal shop handles about 10,000 customer orders.

We realize that it is difficult to verify the performance of the ACLIPS approach. We therefore conducted the following experiments. In the first experiment we used ACLIPS to simulate the impact of the current planning practice. Current practice involves the use of heuristically determined company lot sizes (fixed at 1, 2, 4, 6, 8, 12 or 16 weeks of supply) and second it involves the use of a total slack based priority rule for scheduling. Feeding these parameters in ACLIPS resulted in the performance indicators given in column 1 of table 8.

In the second experiment, we again used our ACLIPS model but now fed with parameter values obtained through the methodology explained in this paper. That means, we searched for the target lot sizes (minimizing the expected lead time) and we applied the ESBP (section (3.3.1)). Consequently we optimized the lot sizes of the 556 products on a Pentium 60 Mhz PC (the optimization was stopped whenever the objective function did not change by more then 10^{-10}, clearly an accuracy level not required for a practical application). We used the minimal maximum lateness criterion for the ESBP routine. The results of this experiment are displayed in column 2 of table 8. So, for both experiments we predicted the performance based on the ACLIPS routine but each time with another parameter setting. We did not use a discrete event type of simulation tool given the extremely high dimension of the problem at hand.

The summary of the performance is given in table 8. The upper part of the table refers to the expected lead time performance, and the lower part refers to the projected scheduling performance. We refer to current practice when we discuss experiment 1, and to the optimized (improved) environment when we discuss experiment 2. The average lead time per operation decreased from 68 hours to 22 hours. In our our case study there are on average 3.4 operations per product (order), so we project an average lead time per order of 227 hours in experiment 1 and 78 hours in experiment 2. This 66% lead time reduction clearly illustrates the significance of the lot sizing and the lead time estimation phase. The average planned lead time (including safety time to guarantee a 95% service level) per order equals 486 hours in experiment 1 and 163 hours in experiment 2. This measure gives us an idea of the width of the time windows.

In the scheduling phase, we group the various customer orders and forecasts (for a one-year time horizon) into manufacturing orders for both experiments. After determining the release dates for each manufacturing order we ran a total slack rule (experiment 1) and an ESBP rule (minimize maximum lateness, experiment 2). This resulted in a huge detailed scheduling problem of approximately 30,000 operations. The results are shown in the lower part of table 8. Based on a deterministic schedule we expect an average shop lead time of 139 hours in experiment 1 (note an improvement compared to the 227 hours, due to the slack based priority rule which outperforms the First Come, First Serve rule implied in the queueing model). The average shop lead time in experiment 2 equals 55 hours. This is again a significant improvement compared to 78 hours (also the Shifting Bottleneck Procedure is performing much better the FCFS rule embedded in the queueing model). Three numbers are of importance. The current practice lot sizes (experiment 1) result in an expected lead time of 227 hours, this can be reduced to 78 hours by using better target lot sizes (experiment 2). The 78 hours can be further improved to 55 hours by using the ESBP. In this experiment the improvement from 227 hours to 55 hours is for 87% due to a better lot sizing policy and for 13% due to improved scheduling. In table 8 we further give details on the maximum lateness, the average lateness and the average tardiness.

On all performance measures, the ACLIPS routine (with the optimized parameter settings) outperformed current practice. Consequently the company decided to implement the ACLIPS model and we observed the behavior of the system during a 3 month period. During this 3 month period several other tests were performed, one involving the installation of a second shift for one of the bottleneck machines. This can be interpreted as an application of the tuning phase discussed earlier. This additional shift had the potential to reduce the

		Experiment 1	Experiment 2
Lead Time	Average lead time per operation	68 hours	22 hours
Performance	Average lead time per order	227 hours	78 hours
(queueing)	Average planned lead time per order (95%)	486 hours	163 hours
	Average (deterministic) shop lead time	139 hours	55 hours
Schedule	Maximum lateness	470 hours	89 hours
Performance	Average lateness	- 477 hours	- 120 hours
	Average tardiness	5 hours	3 hours

Table 8: Summary of the ACLIPS Performance

average lead time per order to 60 hours (as compared to the original 78 hours).

In order to fully obtain the benefits of the proposed hierarchical approach suggested in this paper one has to focus on data accuracy and on some behavioral aspects of scheduling. Current practice is based on priority rules. This myopic approach has to be replaced by a computerized scheduler looking at all machines simultaneously. This results in a dispatch sequence which is not always preferred by the operators who are used to set priorities autonomously. In order to overcome this, one has to spend a lot of time on the floor to introduce this overall approach and to make sure everybody is confident with the proposed priorities.

5 Conclusion

In this paper we proposed a general methodology to analyze and to schedule a job shop. A four phase methodology is proposed including a lot sizing and lead time estimation phase, a tuning phase, a scheduling phase and an execution phase. In each phase we use analytic approaches which are suitable for practical applications. The methodology is illustrated with an example and an application is given. The ACLIPS methodology is embedded in a software package. Our experience indicates that our approach has great potential both in terms of computational effort required and in terms of the quality of the generated schedules.

References

[1] Adams J., E. Balas, and D. Zawack. The shifting bottleneck procedure for job-shop scheduling. *Management Science*, 34(3):391–401, 1988.

[2] Albin S.L. *Approximating queues with superposition arrival processes.* PhD thesis, Department for Industrial Engineering and Operations Research, Columbia University, 1981.

[3] Applegate D. and W. Cook. A computational study of the job-shop scheduling problem. *ORSA Journal of Computing*, 3(2):149–156, 1991.

[4] Balas E., J.K. Lenstra, and A. Vazacopoulos. The one-machine problem with delayed precedence constraints and its use in job shop scheduling. *Management Science*, 41(1):94–109, 1995.

[5] Barnes W. and J. Chambers. Solving the job shop scheduling problem with tabu search. *IIE Tranactions*, 27:257–263, 1995.

[6] Bitran G.R. and D. Tirupati. Multiproduct queueing networks with deterministic routing: decomposition approach and the notion of interference. *Management Science*, 34(1):75–100, 1988.

[7] Bitran G.R. and D. Tirupati. Approximations for product departures from a single-server station with batch processing in multi-product queues. *Management Science*, 35(7):851–878, 1989.

[8] Brucker P., B. Jurisch, and B. Sievers. A branch-and-bound algorithm for the job-shop scheduling problem. *Discrete Applied Mathematics*, 49:107–127,1994.

[9] Buzacott J.A. and J.G. Shantikumar. On approximate queueing models of dynamic job shops. *Management Science*, 31(7):870–887, 1985.

[10] Carlier J. The one-machine scheduling problem. *European Journal of Operational Research*, 11:42–47, 1982.

[11] Carlier J. and E. Pinson. An algorithm for solving the job-shop problem. *Management Science*, 35(2):164–176, 1989.

[12] Dauzère-Péres S. and J.B. Lasserre. A modified shifting bottleneck procedure for job-shop scheduling. *International Journal of Production Research*, 31(4):923–932, 1993.

[13] Ivens P.L. and M.R. Lambrecht. Extending the shifting bottleneck procedure to real-life applications. *European Journal of Operational Research*, 90:252–268, 1996.

[14] Karmarkar U.S. Lot sizes, lead times and in-process inventories. *Management Science*, 33(3):409–423, 1987.

[15] Kraemer W. and M. Lagenbach-Belz. Approximate formulae for the delay in the queueing system $GI/GI/1$. In *Congressbook*, pages 235–1/8, Melbourne, 1976. Eighth International Teletraffic Congress.

[16] Lambrecht M.R., Chen Shaoxiang, and N.J. Vandaele. A lot sizing model with queueing delays: the issue of safety time. *European Journal of Operational Research*, 89:269–276, 1996.

[17] Lambrecht M.R. and N.J. Vandaele. A general approximation for the single product lot sizing model with queueing delays. *European Journal of Operational Research*, 95:73–88, 1996.

[18] Ovacik I.M. and R. Uzoy *Decomposition methods for complex factory scheduling problems.* Kluwer academic publishers, 1996.

[19] Shantikumar J.G. and J.A. Buzacott. On the approximations to the single server queue. *International Journal of Production Research*, 18(6):761–773, 1980.

[20] Shantikumar J.G. and J.A. Buzacott. Open queueing network models of dynamic job shops. *International Journal of Production Research*, 19(3):255–266, 1981.

[21] Shantikumar J.G. and U. Sumita. Approximations for the time spent in a dynamic job shop with applications to due-date assingment. *International Journal of Production Research*, 26:1329–1352, 1988.

[22] Schutten J.M.J., S.L. Van de Velde and W.H. Zijm. Single machine scheduling with release dates, due dates and family setup times *Management Science*, 42(8):1165–1174, 1988.

[23] Suri R. and S. DeTreville. Rapid modelling: The use of queueing models to support time based competitive manufacturing. In *Proceedings of the German/US Conference on Recent Developments in Operations Research*, pages 34–42, Berlin, 1992. Springer.

[24] Vandaele N.J. *The impact of lot-sizing on queueing delays: multi product, multi machine models.* PhD thesis, Department of Applied Economics, Katholieke Universiteit Leuven, 1996.

[25] Wein L.M. and P.B. Chevalier. A broader view of the job shop scheduling problem. *Management Science*, 38:1018–1033, 1992.

[26] Whitt W. The queueing network analyzer. *The Bell System Technical Journal*, 62(9):2779–2815, 1983.

[27] Zijm W.H.M. and R. Buitenhek. Capacity planning and lead time management. *International Journal of Production Economics*, 46-47:165–179, 1996.