INT. J. PROD. RES., 2000, VOL. 38, NO. 18, 4823–4841

Cycle time reduction by improved MRP-based production planning

ASHUTOSH AGRAWAL†, IOANNIS MINIS‡ and RAKESH NAGI§*

The important practical problem of planning the production of large assemblies employing an MRP-based system is considered. The objective is to produce products on-time, with minimal cycle time and low work-in-process costs. The approach is based on the determination of accurate lead-time estimates and on the introduction and use of lead-time offsets in the solution methodology. An effective Lead-time Evaluation and Scheduling Algorithm (LETSA) is employed that can perform detailed backward scheduling of operations belonging to a large assembly on a given facility with an objective of minimizing the cycle time. A scaling procedure is used to account for capacity sharing effects by multiple products in a common facility. These scaled lead-time estimates are then employed by an MRP-based system to release work-orders on the shop-floor. The effectiveness of these lead-times and lead-time offsets are evaluated by simulating production using the MRP generated order release times and verifying on-time completion of the multiple assemblies in the common facility. Numerical experiments are presented to validate the performance of the approach. Optimized batch sizes for minimal work-in-process (WIP) costs can also be obtained using LETSA. Thus, the important objectives of minimizing cycle time for on-time delivery and minimizing schedule costs can be accomplished simultaneously.

1. Introduction

The shift from conventional mass production to batch production has accelerated in recent years. In response to continuously varying customer requirements, products are being manufactured in small batches, each with custom features. This trend is pervasive in both commercial and defence markets and has severe implications to the operations of a manufacturing enterprise. The diverse product mix being manufactured in a common facility greatly complicates both production planning and scheduling. Additional pressures on these functions are imposed by severe on-time delivery and minimal cycle time requirements placed upon manufacturers by the fiercely competitive market.

In this study, production planning is addressed in a manufacturing facility that typically produces large and complex assemblies for which cycle times range between two months to two years. Although contemporary planning techniques, such as Just-In-Time (JIT) and optimized production technology (OPT) have gained some

Revision received May 2000

† Department of Decision Sciences and Engineering Systems, Rensselaer Polytechnic Institute, Troy, NY 12180, USA.

‡ Department of Mechanical Engineering & Institute for Systems Research, University of Maryland, College Park, MD 20742, USA.

§ Department of Industrial Engineering, 342 Bell Hall, SUNY at Buffalo, Buffalo, NY 14260, USA.

* To whom correspondence should be addressed. e-mail: nagi@acsu.buffalo.edu

International Journal of Production Research ISSN 0020–7543 print/ISSN 1366–588X online © 2000 Taylor & Francis Ltd
http://www.tandf.co.uk/journals

popularity in such environments, the Materials Requirements Planning (MRP) (and Manufacturing Resources Planning, MRP II) (Orlicky 1975) philosophy is still employed by the majority of manufacturing enterprises for production planning. MRP has been found to be an effective way explicitly to consider relationships between the end items and the various components and sub-assemblies. Furthermore, it has advantages over JIT in the case of unplanned production changes, erratic demand and breakdown of key equipment. Many practitioners and researchers consider the JIT and MRP philosophies not necessarily contradictory, and seek to develop hybrid systems that take advantage of the positive aspects of the two (Vollman *et al.*, 1992). Perhaps the major drawback of MRP is the use of lead-times that are determined *a priori* from approximate considerations. These lead-times are either underestimated, causing excessive resource workloads (and production delays), or overestimated, causing excessively long cycle-times and associated high work-in-process (WIP) costs.

In this paper, we present an approach to improve MRP-based production planning targeting minimal product cycle times. The approach is based on the estimation of accurate part lead-times, and on the introduction of lead-time offsets. The Lead-time Evaluation and Scheduling Algorithm (LETSA) developed by Agrawal *et al.* (1996) is used to obtain the values of these parameters. LETSA performs detailed backward scheduling of operations that are necessary for the production of one or more large assemblies in a common facility, with an objective of minimizing the cycle time. It is anticipated that the lead-times and lead-time offsets obtained from the resulting schedule will yield minimal deviation between product due dates and will limit work-centre overloading, such that the cumulative production cycle-time is minimized.

The approach has been tested by performing production planning for random demand based on the estimated lead-times and lead-time offsets, and subsequently by using the simulation package SIMFACTORY to schedule the work orders released according to the production plan. Extensive numerical experiments were performed to assess the performance of the approach with respect to the on-time production of multiple products in a shared facility. The impact of employing lead-time offsets was studied with respect to overall cycle time, workload smoothness, and WIP costs.

The paper is structured as follows. In section 2, the relevant literature is reviewed and the problem is formulated. The proposed approach and the scheduling heuristic are presented in section 3. Software implementation issues are presented in section 4. The performance of the approach is evaluated in section 5. Finally, the conclusions of the study are presented in section 6.

2. Background

2.1. *Related work*

The following discussion of literature focuses on lead-times of multi-level products within a job-shop. Planned lead-times play a fundamental role in the operation of MRP systems. They are used both to time-phase purchase and work orders, and to maintain valid and credible priorities (Peterson 1975), so much so that the structure and level of lead-times can have a fundamental impact on virtually every aspect of the logistics system (Kanet 1982). However, as indicated in an earlier survey by Wemmerlov (1979) scant attention has been paid to setting and revising lead-time values. In addition, as acknowledged by Melnyk and Piper (1985), the task of determining even 'reasonable' planned lead-times, is not easy.

A significant body of literature suggests that the lead-times used in practice are highly inflated; 90–95% is the waiting time for shop orders and only 5–10% of the time is actual 'touch-time' (Huge 1979, Plossl and Welch 1979, Waliszewski 1979, Wight 1970). John (1985) has studied the relationship between production costs and lead-times ranging from no slack time to 95% slack (queue) time. This simulation study revealed an exponential cost increase beyond the 80–90% queue time level, and a 41% increase of cost from no slack to a 95% slack. The study concludes the need for lead-time reduction, either through downward adjustment of MRP planned lead-times or by introducing new manufacturing concepts. Waliszewski (1979) presents Hewlett-Packard's experience in reducing lead-times by 70%, which improved customer service levels by 80%, and amounted to $1.7 million savings in work-in-process inventory over three months. The above literature clearly indicates the advantage of reducing lead-times or determining accurate estimates for them, and recognizes the need for a formal method to determining them.

The difference between the planned and actual lead-times has also been referred to as lead-time error. Several methods to minimize the error have been proposed in the literature (Melnyk and Piper 1985): (i) adjusting the master production plan until the materials plan is consistent with capacity utilization, (ii) lead-time management, or manual intervention with pre-emption, lot-splitting and a short-term capacity addition, (iii) tracking lead-time error and updating planned lead-times (Schuchts 1979, Steele 1975), and (iv) adding to the average observed lead-time a multiple of the standard deviation of the lead-time error. The first two approaches are rather arbitrary and are not based on a formal methodology. The third method is practised often in industry. However, extensive data collection and the need for an established history are essential, while, often, it may be too late to update planned lead-times as the mistakes have already been made. The last method is formal and is based on simulation and statistical analysis. Simulation of a multi-product, multi-stage production environment is used to determine the lot-sizing schemes that minimize the lead-time error. Although the major objectives of the above mentioned methods is to minimize the deviations between planned and observed lead-times, they do not, in general, attempt to determine the lead-times that would result in minimal product cycle-times. Besides, simulation can be computationally expensive.

2.2. *Problem description*

A classic job-shop environment, consisting of a large number of work-centres, each of which may comprise functionally identical machines, is considered. It is assumed that the facility manufactures complex assemblies in small batches. Each assembly has numerous levels in its Bill-Of-Materials (BOM), and each manufactured (or make) part is produced following a unique sequence of operations (called routing). Production planning is performed by an MRP-based system, for the given demand and order quantity of each product. It is noted that alternative routings for make items, although sometimes generated by MRPII, are rarely implemented, and this methodology does not consider them. The problem is to determine: (i) a fixed lot-size for each end-item that minimizes the Work-In-Process (WIP) costs using a lot-for-lot strategy, (ii) the lead-times and lead-time offsets for all make parts of these end-items, for the above determined lot-sizes and for an average product mix, such that cycle-times of end-items are minimized.

Note that since optimality, or even feasibility, of a production plan can only be assessed once it is finally realized through the shop-floor schedule, we evaluate the production plans that employ the resulting lead-times and lead-time offsets by the shop performance at the operations scheduling stage. In addition to cycle time, performance metrics include WIP costs and work-centre smoothness (see Appendix B for the definitions of these metrics).

For the operations scheduling stage, used to evaluate the suggested lead-times and lead-time offsets, the following assumptions are adopted. A batch production strategy is employed, i.e. (i) the entire batch of a part is processed by a certain machine before the subsequent operation is commenced; (ii) batches are not distributed amongst functionally identical machines. This strategy is believed to be most common in industrial practice. Furthermore, the following assumptions are made:

- a lot-for-lot strategy is employed for make items;
- extended WIP costs are important compared with production costs (set-up and processing) (because the assemblies may have significantly long cycle times);
- machines are reliable;
- a machine can perform only one operation at a time;
- an operation can be performed on at most one machine at a given time;
- pre-emption of operations is not permitted; and
- processing times of all operations and due-dates of all final products are deterministic and known *a priori*.

3. Approach

As acknowledged earlier, lead-times are dynamic and they consist of the following components: (i) set-up time (per batch), (ii) run time (or processing time for a batch), (iii) move times between operations, and (iv) queue times. While the first two components are known and the third is somewhat easy to determine (especially when material handling systems are not the bottlenecks), the queue time is dependent on the availability of resources and is stochastic in general. The resource availability is, in-turn, a function of the demand and the scheduling strategy. This study attempts to assess the queue time for average resource availability such that the cycle time is minimized while respecting scheduling constraints. Our approach consists of the following steps.

Step 1. For each end item, determine the optimal batch size, assuming that it is the only end item in the system.

Step 2. For the batch size computed in step 1, perform detailed scheduling of operations using LETSA. From this schedule, calculate the lead-times and lead-time offsets of all the make items of the final assembly. This process can be repeated for each end item.

Step 3. Perform capacity scaling for end items based on average product mix and their utilization of critical work-centres. This involves determining the proportion of capacity of critical work-centres that is required by each end-item, based on the product mix. Average product mix can be obtained from historical data and forecasts.

Step 4. Inflate all part lead-times of an end item employing the capacity scaling factors computed in step 3.

Steps 1 and 2 are accomplished by using the LETSA algorithm (Agrawal *et al.* 1996). It uses a precedence network of operations, which incorporates lead-time offsets, and employs an effective critical path heuristic to schedule these operations. Note that backward scheduling is used in an attempt to minimize the WIP costs.

This section is organized as follows. In section 3.1, we present the concept of lead-time offsets and highlight the benefits expected from using them. We briefly describe the Lead-time Evaluation and Scheduling Algorithm in section 3.2. Step 1 of our approach, optimal batch-sizing, using LETSA is described in section 3.3; this concurrently accomplishes step 2, lead-time and lead-time offset estimation for single assemblies. The capacity scaling and inflation of lead-times (steps 3 and 4) are described in section 3.4. A flow chart of our overall approach is presented in figure 1. A detailed account of the inputs and outputs of the software system is presented later in section 4.

This approach has been verified using the predicated lead-times to develop a production plan given the demand and the due-dates. Subsequently, the shop-floor simulation package SIMFACTORY is employed to schedule the work-orders that have been released according to the production plan. This process emulates an MRP-based planning and the shop-floor schedule resulting from it. Numerical studies verify the validity of the approach as well as the benefit of using lead-time

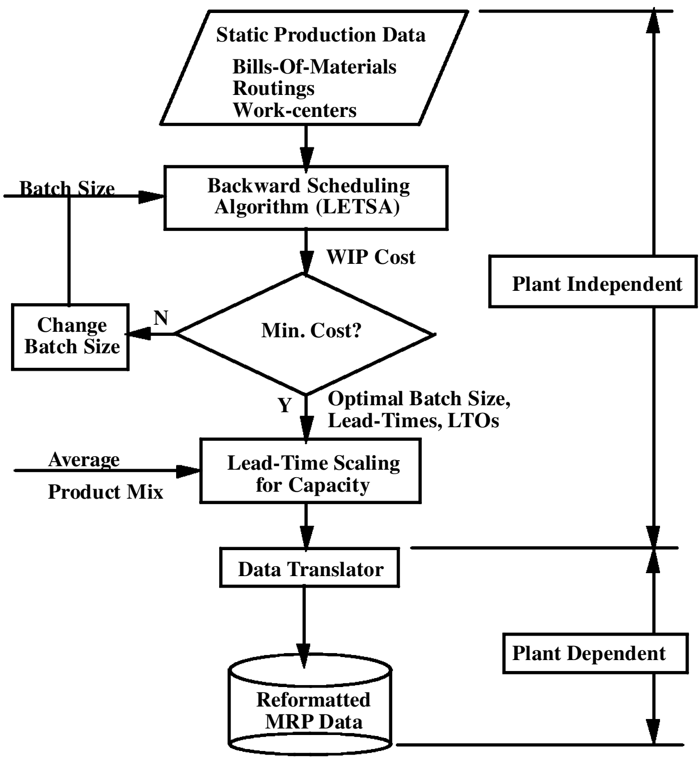


Figure 1. Overall system approach.

offsets, their impact on reducing overall cycle time, improving workload smoothness, and reducing WIP costs. These results are presented in section 5.

3.1. *Lead-time offsets*

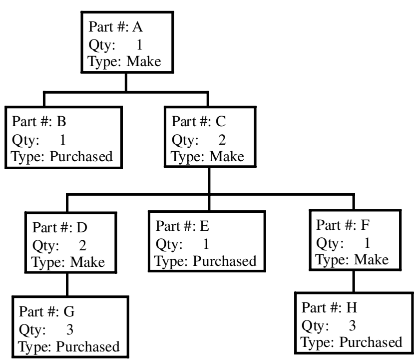
Conventionally, the components of an assembly or sub-assembly are available in inventory before an assembly work-order is released. In contrast, using lead-time offsets permits the manufacture of such an assembly to proceed concurrently with the manufacture of some of its component parts.

Consider the Bill of Materials (BOM) structure and the routing information for the final product A shown in figure 2. Lead-time offsets take advantage of the fact that production of a component p_k of part p_i may be delayed until p_k is actually required in a certain operation of the routing of p_i. Thus, the lead-time offset of p_k is computed as follows: if p_k is required for the first time for operation o_{ij} of its parent part p_i, its lead-time offset is the difference between the start times of operations o_{ij} and o_{i1}.

In figure 2, it is obvious from the routing of part A that availability of C can be delayed until it is actually required for operation A.20. Thus, operation C.10 may be performed concurrently with operation A.10 (assuming another machine in work-centre 1 is available), thereby improving resource utilization as well as reducing the cycle-time. In the manufacture of large and complex assemblies, a large number of component lead-time offsets may be realized. The introduction of these lead-time offsets into the planning and scheduling functions yields a number of benefits, as will be shown in section 5.2. It is interesting to note that most current MRP systems do provide the capability of using lead-time offsets.

3.2. *The lead-time evaluation and scheduling algorithm*

The determination of the minimum cycle-time production schedule is a complex problem; optimal methods can, at best, solve small dimensional cases. To address practical problems, an effective heuristic has been developed by Agrawal *et al.* (1996). It exploits the critical path of the network of required manufacturing operations to generate a near optimal schedule. This network represents the precedence relationships among operations and may be obtained from the BOM structure and the routings of all make items of an assembly, as shown below. This network can also incorporate the component lead-time offsets. Figure 3 shows the



Part	Operation	Components Required	Processing Time*	Workcenter
A	A.10	B	5	WC #1
	A.20	B, C	2	WC #2
C	C.10	D, E, F	1	WC #1
D	D.10	G	3	WC #1
F	F.10	H	2	WC #2
	F.20	H	3	WC #1

*Processing time includes the setup and run times for the entire batch

Figure 2. Bill-of-Materials (BOM) of the final product A and routings of its make parts.

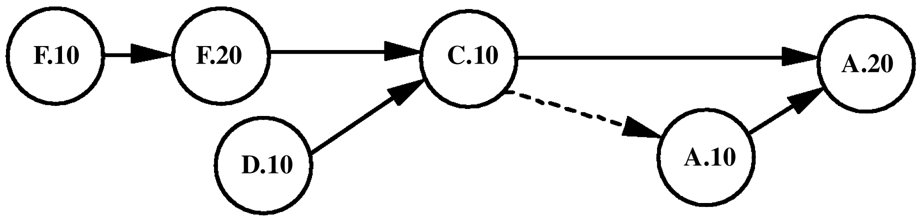


Figure 3. Operation network corresponding to the example of figure 2.

precedence network of operations obtained by the BOM and the routing information for the product structure shown in figure 2.

The network of figure 3 shows that operation A.20 can be initiated only after operations C.10 and A.10 are completed. Similarly, operation C.10 can be initiated only after operations D.10 and F.20 are completed. The above network explicitly represents all the precedence relationships required to manufacture final product A. The operation network of figure 3 also incorporates lead-time offsets. If such offsets were not used, the network would need to be modified to include the dotted arrow to indicate that operation C.10 must be performed before A.10, which represents an additional precedence constraint since operation A.10 is required to be performed before operation A.20.

Given such a network of a final assembly, a continuous sequence of operations that starts from the final operation of the finished item and terminates at a purchased item is defined as a BOM path. The critical path is defined as the BOM path along which the sum of all operation times is maximal. The length of the critical path would be the time needed to produce the final product if resources were available when required. However, due to limited resource availability, the actual product cycle time is, in general, larger than the length of the critical path. This 'infinite capacity' critical path is a fixed property of a given BOM structure and determines the lower bound of the product's cycle time.

The LETSA algorithm generates a feasible schedule with a near-optimal cycle time. It proceeds in a backward scheduling manner similar to MRP II, in which the last operation is scheduled first and the remaining operations are scheduled subsequently while respecting all precedence constraints.

The inputs to the algorithm: the delivery schedule, product structures and routing data, are used to construct an integrated operation network. Given this network, a set F is defined to include those operations that do not have a succeeding operation. Generating the schedule comprises four phases: (i) select an operation from the set of feasible operations F, (ii) select a machine from the required work-centre, (iii) schedule the selected operation, and (iv) update the operation network and the set of feasible operations.

In step (ii), the operation to be scheduled is selected from F as follows: the processing times (set-up plus run) that correspond to each path of the existing network are computed by summing the processing times of all operations along this path. The critical path is determined and its first operation, which also belongs to F, is selected for scheduling. The starting and completion times of the selected operation are determined from: (i) the starting time of its successor operation, or the due-date; (ii) the first available time of the machine of the corresponding work-centre. The operation is then scheduled on that machine. Subsequently, the operation network is modified by deleting the operation just scheduled. Its predecessors are included in the feasible set of operations F, and the process is repeated until F is empty, i.e. the schedule is complete. The flowchart of the algorithm is presented in Appendix A (see also Agrawal *et al.* 1996). Later, Anwar and Nagi (1997) showed that an equivalent, yet more efficient way to perform this backward scheduling is first to compute early finish times of the operations network (similar to the critical path method in project networks). Then, in step (i), select from F, the operation with the highest early finish time.

3.3. *Optimal batch sizing using LETSA*

LETSA is used to perform the trade-off analysis between processing costs and WIP costs in batch sizing decisions implemented within an overall lot-for-lot production strategy. The function of LETSA in this mode is to evaluate the WIP cost, and once the batch size corresponding to the minimal WIP costs has been determined, the scheduling results of LETSA also provide lead-times and lead-time offsets of each make item of the assembly. To this end, the algorithm is run iteratively, varying only the batch size and the resulting WIP cost is evaluated from the schedule of operations as described in Appendix B. The optimum batch size is selected to be the one that corresponds to the lowest cost per unit. A typical plot of the relationship between batch size and cost per unit is shown in figure 4. The assembly tested in this case is a large industrial product. It includes 200 make items and the lead-times ranged from 57 days (batch size of one) to 522 days (batch size of 15). Additional details of the run and set-up costs terms are similar to those in section 5.1. The two curves in the figure correspond to the conditions during scheduling; the top curve was obtained when no lead-time offsets were used, while the lower curve was obtained using lead-time offsets. Figure 4 clearly illustrates the benefits of introducing lead-time offsets towards lowering unit cost.

3.4. *Capacity scaling approach*

Based on the schedule generated, the lead-time of each make part can be determined. Obviously, if production planning is performed using these lead-times, the schedule provided by LETSA can be replicated in push production. However,

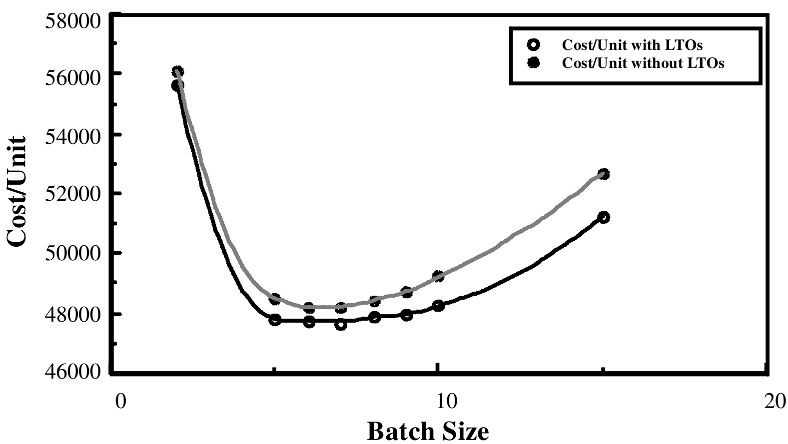


Figure 4. Implementation of LETSA to optimize batch sizing.

these lead-time estimates would be valid if the assembly under consideration was the only product in the system. The presence of other products is considered by inflating the lead-times of parts to account for the resource capacity consumed by these products.

In order to determine a robust scaling method that can provide realizable lead-times and, at the same time, keep cycle-times minimal, we attempted two capacity scaling procedures to account for multiple products within a shared facility: (i) Exact Capacity Scaling (ECS), and (ii) Bottleneck Capacity Scaling (BCS). To define these we will use the following indexes, i for product, j for a make part of product i, and k for work-centre. For the ECS method, the workload requirements of each product on each work-centre were computed by summing, for all operations (requiring that work-centre) belonging to the routings of all make parts of the product, the processing times (derived from set-up and run times) for the given long-term product demand. The total workload requirement for each work-centre was determined by summing the workload requirements over all the products. The ratio of the total workload requirement of a work-centre to a product's workload on that work-centre was defined as the scaling factor S_{ik}. The routing length of each make part was computed by adding, over its operations (on different work-centres), the processing times for the given long term demand, and the ratio of this part's processing on work-centre k to the part's routing length was determined as L_{jk}. Then the lead-time of the make part j was inflated by a factor $\sum_k (S_{ik} \times L_{ik})$.

For the BCS method, the work-centre with the highest workload requirement for each product for the given long-term demands was determined individually as the bottleneck for that product. Then the lead-time of each part of a particular product was inflated by the ratio of the cumulative workload (over all products) of the bottleneck to the product's workload requirement from the bottleneck.

As indicated in the results in section 5.3, we recommend the BCS method to be superior to the ECS. BCS suggested shorter lead-times that resulted in corresponding shorter cycle-times than ECS in all cases. However, in 3% of the cases it underestimated the lead-time. In these cases the actual lead-times were 1.5% longer, and in no case were they more than 2.5% longer than the suggested lead-time. ECS-suggested lead-times were 21% higher than that of BCS, but they were realizable in all cases (with an acceptable minor elongation of less than 0.03% in a few cases).

4. Software implementation

The overall approach for the determination of accurate lead-times and lead-time offsets, and optimal batch sizes, has been implemented in a software system. The inputs and outputs of this system are detailed in this section. These data were obtained from an MRP II system and translated automatically to conform to the system input requirements. Some degree of automation has been incorporated into the user interface to facilitate data input by providing a menu to select the necessary components for each operation.

4.1. *System inputs*

Two types of inputs are provided to the system: (i) MRP data, and (ii) user data.

MRP II data

MRP data include complete production information for manufactured products as well as detailed information about the manufacturing environment. Specifically, the inputs are:

- item master records;
- bills-of-materials (BOM) for all final products;
- routing information for all make items;
- manufacturing facility information, including a list of work-centres, the number of identical machines per work-centre, work-centre efficiency and capacity, set-up and run rates for each work-centre.

User inputs

In addition to the MRP II data, the user is required to provide:

- the batch sizes of all end-items;
- the percentage workload contributed by each product to the manufacturing resources. This is provided as the long-term production requirements of each product, and accounts for sharing of the resources amongst multiple products. Note that it can be computed from historical data and projected forecasts;
- the interest rate at which the work-in-process (WIP) costs are to be compounded;
- component requirements for each operation in a routing. This information is necessary primarily for the accurate computation of lead-time offsets and WIP costs.

4.2. *System outputs*

- Operations dispatch list: generated for each work-centre, it provides the start and finish times of each operation performed at the work-centre within the scheduling horizon.
- Schedule of operations by make item: generated for each make item, it provides the start and finish times for each operation in the corresponding routing.
- Constrained capacity lead-times and lead-time offsets for all make items: these are calculated from the schedule of operations.
- Cost of each make item: it combines raw material, labour and WIP costs.
- Order times and quantities for raw materials: since LETSA operates in a backward scheduling mode, it provides the appropriate release dates of purchase orders for raw materials in order to meet the due date of the end items.
- Daily workload for each machine: it is computed in a straightforward manner from the generated schedule of operations. The work-centre smoothness profiles (see Appendix B) can then be calculated from this output.
- Scaled lead-times and lead-time offsets for all manufactured parts to be employed in MRP.

5. Performance

This section evaluates the effectiveness of the proposed method. The first section examines the benefits of using lead-time offsets. In the second section, the production plans that are generated based on the system's lead-times are examined under realistic production conditions, assuming more than one final assembly is being manufactured concurrently in the facility. For the performance of LETSA and its effectiveness in minimizing cycle time, the reader is referred to Agrawal *et al.* (1996).

5.1. *Benefits of using lead-time offsets*

A numerical study was conducted to assess the benefits of using lead-time offsets using an example set of 200 randomly generated BOMs. The part, routing, work-centre, and BOM structure data for all examples were generated randomly. For each BOM structure, there was an of average of four work-centres in the facility, each with one to two identical machines. This particular set of examples constituted large and wide BOMs, characteristic of large assemblies. Average parameters for the BOMs for these examples were 5 levels, 130 make parts, 300 operations, and about 200 branches. Routings for the make parts were made up of one to four operations, chosen uniformly. Each operation had a set-up time per batch between 10 and 30 minutes, and a processing or run time per part between 5 and 10 minutes, both chosen uniformly. Since the labour rates for set-up and processing were comparable, the ratio of set-up cost to run cost was between 1 and 6. For lead-time offsets, if a make part was assembled from two or more components, its last operation in the routing required all component parts, but the preceding operations required one or two less components. The offset for a component would then be generated from the time difference between the first operation of the parent make part and the first operation where this specific component was required.

LETSA was applied to determine the schedule of operations with and without LTOs and the results were compared with respect to the following production performance metrics of Appendix B. When LTOs are not used, all components of a make part are made to be required at the first operation of its routing. It is noted that the above parameters are conservative and that improved performance characteristics can be obtained by increasing the number of identical machines, operations per routing, and the number of levels of the BOM.

The following measure was used to quantify the benefits of LTOs in terms of product cycle time or makespan (m) and cost/unit (c). In $\mathbf{v}^*$ below, the asterisk is $\mathbf{m}$ and $\mathbf{c}$ for percentage improvement in cycle time and cost respectively.

$$\text{Percentage variation} = \mathbf{v}^* = \frac{(Result)_{(\text{Without LTOs})} - (Result)_{(\text{With LTOs})}}{(Result)_{(\text{Without LTOs})}} \times 100. \quad (1)$$

For workload smoothness equation (B.1) was used; see Appendix B.

$$\text{Percentage variation} = \mathbf{v}^{\mathbf{ws}} = \frac{(\mathbf{WS})_{(\text{With LTOs})} - (\mathbf{WS})_{(\text{Without LTOs})}}{(\mathbf{WS})_{(\text{With LTOs})}} \times 100. \quad (2)$$

$\mathbf{P}^* = E\{\mathbf{v}^*\}$ (mean percentage variation) was evaluated considering all examples within the example set. Note that for both cases the larger this measure, the greater the benefit of using LTOs.

In figures 5 through 7, the relationship between $\mathbf{P}^* = E\{\mathbf{v}^*\}$ versus the batch size is shown. Note that the batch size was used as the independent variable, since it is provided by the user and directly impacts upon the benefits resulting from the use of LTOs. The relationship **P*** = **P*** (batch size) is represented by the continuous line in figures 5 through 7. In addition, the bounds within which 95% of the cases were contained (i.e. 95% confidence intervals) are represented by the closed and open circles, above and below the continuous line respectively.

Figure 5 shows the effect of LTOs on product cycle time. It can be seen that for small batch sizes, the improvement is higher and stabilizes to about 7.2% for the large batch size cases. The improvement in cycle time with the use of LTOs is expected because operations with LTOs are delayed with respect to the due-dates specified by conventional planning, without affecting the product cycle-time. This increases the degree of flexibility while assigning the operations to work-centres, facilitating better resource utilization. The effect of batch size on $\mathbf{P}^{\mathbf{m}}$ is attributed mainly to the fact that smaller batch sizes require lower processing times. In this case, there exists a higher degree of flexibility while assigning operations to work-centres, since the shorter operations can better fill idle times between scheduled work-centre operations. However, as the batch size increases, processing times

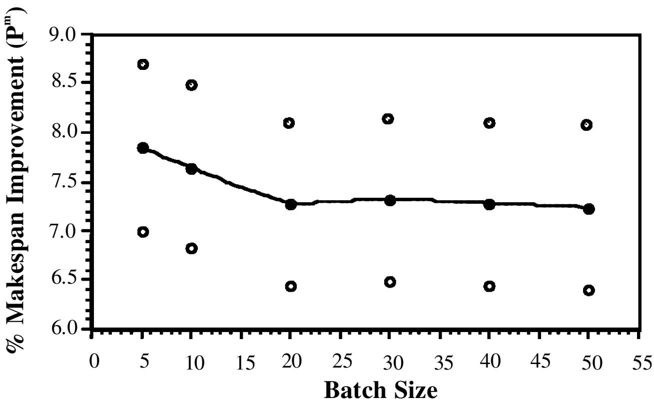


Figure 5. Effect of using lead-time offsets on product cycle time: $\mathbf{P}^{\mathbf{m}}$ versus batch size.

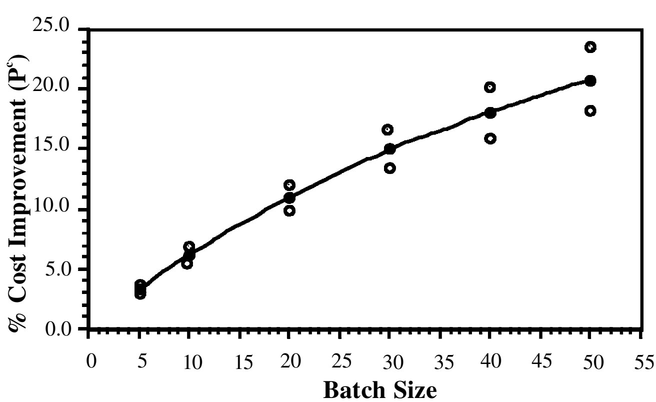


Figure 6. Effect of using lead-time offsets on product cost/unit $\mathbf{P}^{\mathbf{c}}$ versus batch size.

increase almost proportionally, making it more difficult to utilize inter-operation idle time.

Figure 6 shows that for a batch size of 5, the improvement in product cost/unit is approximately 2.5% but increases to about 21% for a batch size of 50. Component parts with LTOs can be introduced after the assembly of the parent part has started. Thus, expensive parts do not have to wait for extended periods of time in inventory, reducing WIP costs. As seen from the figure, savings in WIP costs realized in our experiments continue to increase proportionately with batch size. This is due to the compounding of interest accrued on delayed value addition. When LTOs are utilized, early savings in offsetting a component's value addition continue to be compounded over the remaining duration of the makespan. Further, the makespan increase is roughly proportional to batch size because the processing times of operations as well as the work-centre idle times increase almost proportionally (NB processing time for an operation is set-up time $+$ batch size $\times$ run time). Consequently, the savings in extended WIP costs continue to increase with batch size. It has to be noted that the slope of the savings curve would depend on the raw-material costs and rate of compounding. Our experiments use some exaggerated values for material costs and 20% annual percentage rate compounded hourly, which may be high for some industrial situations. This is simply for illustrative purposes and to amplify the benefit to WIP costs when using LTOs. Finally, an alternative view of these increasing savings in unit cost due to LTOs can be seen in the diverging right halves of the unit cost curves shown in figure 4.

Figure 7 shows that the introduction of lead-time offsets improves the smoothness of work-centre load. This is due to staggering the requirements of the components of an assembly such that they are not all due at the same time (which causes an unnecessary peak in the workload). Note that with increasing batch size the processing times increase and, thus, the LTOs increase almost proportionally. Consequently, the flexibility in filling the idle times, which results from the introduction of LTOs, is unchanged. Figure 7 confirms this hypothesis, showing that despite the variation in product batch size, the advantage of using LTOs remains almost constant.

From the above numerical results and discussion it can be clearly seen that the introduction of lead-time offsets in planning and scheduling results in significant advantages in terms of product cycle-time and cost per unit. In addition, the result-

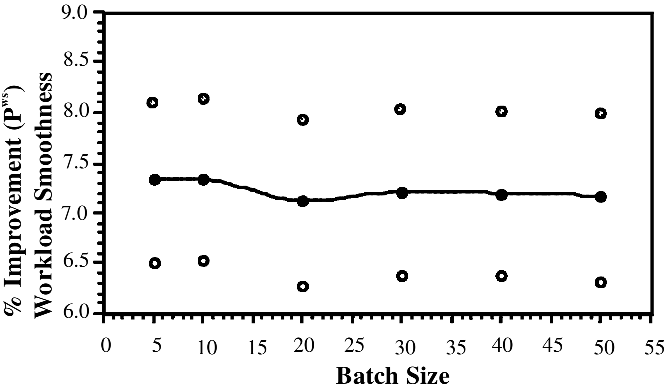


Figure 7. Effect of using lead-time offsets on work-centre load profile: $\mathbf{P}^{ws}$ versus batch size.

ing improvement of workload profiles is invaluable in the case of alleviation of peaks and valleys on bottleneck resources.

5.2. *Overall system evaluation: validation of proposed lead-times*

The proposed approach has been verified by performing production planning based on the derived lead-times, and then using a discrete-event simulation tool (SIMFACTORY) to release work-orders into the system according to the suggested lead times. This production planning and scheduling system emulates MRP II-based planning. To restate this in detail, this is what our simulation study attempts to do. We obtain our lead time and offset information from the approach developed in this work. We then generate an arbitrary shipping schedule for a mix of end products. Due dates in this shipping schedule are generated randomly, while the delivery quantities match the respective optimal batch sizes. Random due-dates create a real-life situation of arbitrary shop loading in response to customer orders and permit a study of the robustness of our derived lead times. Now, an MRP run is emulated using the BOM, routing and lead time information. Material is released into the simulation system using the backward MRP calculated release dates. Manufacturing operations on the work-centres are scheduled using the commonly used first-come first-serve (FCFS) shop-floor dispatching rule. We let the production proceed in time as the simulation clock advances, and observe the completion times of the different end items, and compare them to the desired due dates given in the original shipping schedule.

It is important to note that discrete-event simulation involves construction of an elaborate model, which is a labour intensive and error-prone process. To overcome this limitation, a translator routine was developed to automatically generate the model from the available input data. This is of significant practical importance, rendering simulation a useful tool in large job-shop applications. Further, this discrete event simulation approach can be readily used to study additional stochastic events like machine failures and rework/scrap rates. We will, however, remain focused on our validation under a deterministic shop operation.

This numerical study was performed considering 16 sets of examples, with each set consisting of 10 unique end products. The first step of our overall approach, determining optimal batch sizes, was performed by the simple enumeration procedure detailed in section 3.3. In the second step, each product was considered individually and LETSA was employed to provide independent lead-times and lead-time offsets for the make parts. The two capacity scaling procedures described in section 3.4 were employed to derive estimated lead-times for the parts in a shared facility; this concludes steps 3 and 4 of our procedure. We considered the shipping schedule to be generated arbitrarily where one batch of each product in its respective optimal batch size was assumed due at a specific due-date. An MRP II-type algorithm was employed to schedule the release of purchased orders and make orders in our simulated factory. Push simulation was performed using the FCFS rule, but respecting order release times. The completion times of the 10 end-products were observed from the resulting simulation run, and compared with the planned lead-times to assess the on-time performance.

Table 1 presents the results of example set 7, consisting of 10 end products. Column 1 contains the example product identification number. Column 2 contains the lead-times that are determined from LETSA's schedule when all products are considered in a combined manner. Column 3 contains the estimated lead-time using

1 Ex. ID	2 LETSA combined	3 LT/EC	4 SF/EC/PP	5 SF/EC/MP	6 LT/BC	7 SF/BC/PP	8 SF/BC/MP	9 PUSH SF/0
7.1	12 904	18 881	17 679	18 882	16 638.9	16 220.5	16 637.9	16 684.1
7.2	15 997	18 268	17 975	18 268	17 279.6	17 139.3	17 279.8	16 275.4
7.3	12 984	21 818	21 454	21 818.2	22 270.7	21 907	22 271.1	16 573.3
7.4	15 253	18 806	18 358	18 806.9	16 939.1	16 636.6	16 939.2	15 344
7.5	16 182	16 907	16 746	16 906.6	16 203.1	16 045.4	16 203	16 755.5
7.6	5563	20 462	18 982	20 458.5	18 391.8	17 335.1	18 391.6	16 655.2
7.7	16 132	22 159	18 092	22 158.8	16 802	15 816.7	16 802	16 582
7.8	16 058	23 987	21 655	23 987.4	16 558.1	15 840.1	16 558.2	15 832
7.9	16 113	20 503	19 970	20 503.1	17 065.7	16 534.3	17 066.6	16 400.9
7.10	16 420	19 849	19 115	19 849	15 168.2	14 876.7	15 168.3	16 619.2
Makespan	17 190.9		23 826.6	23 987.4		22 130.6	22 271.1	16 755.5

Table 1. Estimated lead-times and realized lead-times by MRP planning with FCFS dispatching.

the exact capacity scaling procedure. Column 4 contains the lead-times that are realized by SIMFACTORY with FCFS dispatching when purchased parts (raw materials) are released according to the lead-time estimates of column 3. Column 5 contains similar information to column 4, but with both purchased as well as make part orders released according to column 3 lead-times. This emulates a true MRP schedule. Columns 6 through 8 have similar information as columns 3 through 5 respectively, only that the bottleneck capacity scaling procedure is employed in lead-time estimation. Finally, column 9 contains a push schedule when all purchased parts are released at time ‘0’.

Similar results were obtained for all 16 sets of examples. Without presenting the details of these rather similar results, we summarize our findings as follows. LETSA combined and PUSH resulted in very close makespans (within 0.1% on the average), which were the smallest among the production scheduling methods. This is again consistent with the results found by Agrawal *et al.* (1996). In these examples, LETSA combined was never more than 5.6% higher than PUSH, while, in one case, PUSH was about 48% higher than LETSA. More importantly, even with comparable makespans, PUSH resulted in significantly higher WIP costs, making it an undesirable production strategy. This is further substantiated from the results that the individual lead-times in a PUSH schedule were about 20% higher than the ones derived from a LETSA schedule, on average.

The estimated lead-times due to the exact capacity scaling (ECS) method were always realized very accurately, when purchased as well as make part orders were released on time (column 5) and were never underestimated by more than 0.02%. When these estimated lead-times were used to release only purchased parts while make orders were released as early as possible due to FCFS and PUSH production (column 4), the realized lead-times were always smaller, and smaller by 6% on average. This indicates that this capacity scaling method is too conservative.

Along similar lines, the estimated lead-times due to the bottleneck capacity scaling (BCS) method were always realized very accurately when purchased as well as make part orders were released on time (column 8). In addition, the estimated lead-times were only underestimated by 0.07% on average and never more than 2.4%. When these estimated lead-times were used to release only purchased parts (column 7), the realized lead-times were generally smaller, and smaller by 3.5% on average. The realized lead-time was never more than 2.6% higher than the estimated one. We believe that the BCS capacity scaling method is not overly conservative because any further stringency will only cause realized lead-times to exceed estimated ones. Further, given that the lead-times estimated by the BCS method were 21% lower than the ECS method, BCS is clearly superior to the ECS method.

Even while we have presented an accurate estimator for lead-times that are closely realizable and not overly conservative, it has to be recognized that MRP planned production in this manner results in overall makespans that are about 28% higher than that of LETSA combined (or PUSH). This is not surprising in that static lead-time-based planning and FCFS dispatching is not expected to outperform a specialized scheduling method such as LETSA, yet the difference may be significant. In summary, since MRP-based planning and scheduling remains a popular tool in industry, we have proposed a systematic method to determine accurate lead-time data that will result in low product cycle times.

6. Conclusions

Facing global competitiveness, manufacturers are hard pressed to minimize product cycle times and costs, and meet deliveries on time. In the batch production environment for assembled products, MRP/MRP II still remains a popular tool in industry to accomplish production planning and scheduling. In addition, most MRP systems perform planning and scheduling using a priori lead-times (although some recent systems include limited finite scheduling capabilities). Underestimating and overestimating these lead-times can lead to a variety of undesirable circumstances.

In this paper we have revisited this classical and industrially relevant problem of determining accurate estimates for item lead-times because it is at the heart of a successful MRP schedule. At the backbone of our proposed approach is a recent JIT production methodology for assembled products, LETSA (Agrawal *et al.* 1996) that performs detailed backward scheduling of one or more large assemblies in a common facility, with an objective of minimizing the cycle time. It is obvious that, now, if lead-times for items are determined from this detailed schedule and employed in an MRP system to plan the release of raw-materials and manufacturing orders, push production can accomplish on-time delivery at minimal cycle time. That is, a backward generated schedule can be replicated in a forward production run. Despite this, there is an important impediment: dealing with multiple end-items in a dynamic shipping schedule while the lead-times in MRP are static. It would obviously not be practical to repeat this process for every shipping schedule. To overcome this remaining problem, we proposed a capacity scaling approach that inflates, according to a long-term capacity utilization, the LETSA lead-times determined for each end-item separately. Many numerical studies demonstrate the efficacy of this approach.

Apart from a methodology to determine realistic lead-times that minimize product cycle times, this paper makes the following contributions. It demonstrates, in quantitative terms, the benefits of introducing lead-time offsets in their ability to minimize cycle time and cost, and to result in more balanced workloads on resources. It also presents an enumeration approach to determine the optimal batch sizes of end-items for minimal WIP costs using LETSA. These lot sizes strike a balance between the set-up cost and WIP inventory costs. Thus, important objectives of minimizing cycle time for on-time delivery and minimizing schedule costs can be accomplished simultaneously.

Appendix A: Lead-time Evaluation and Scheduling Algorithm (LETSA)

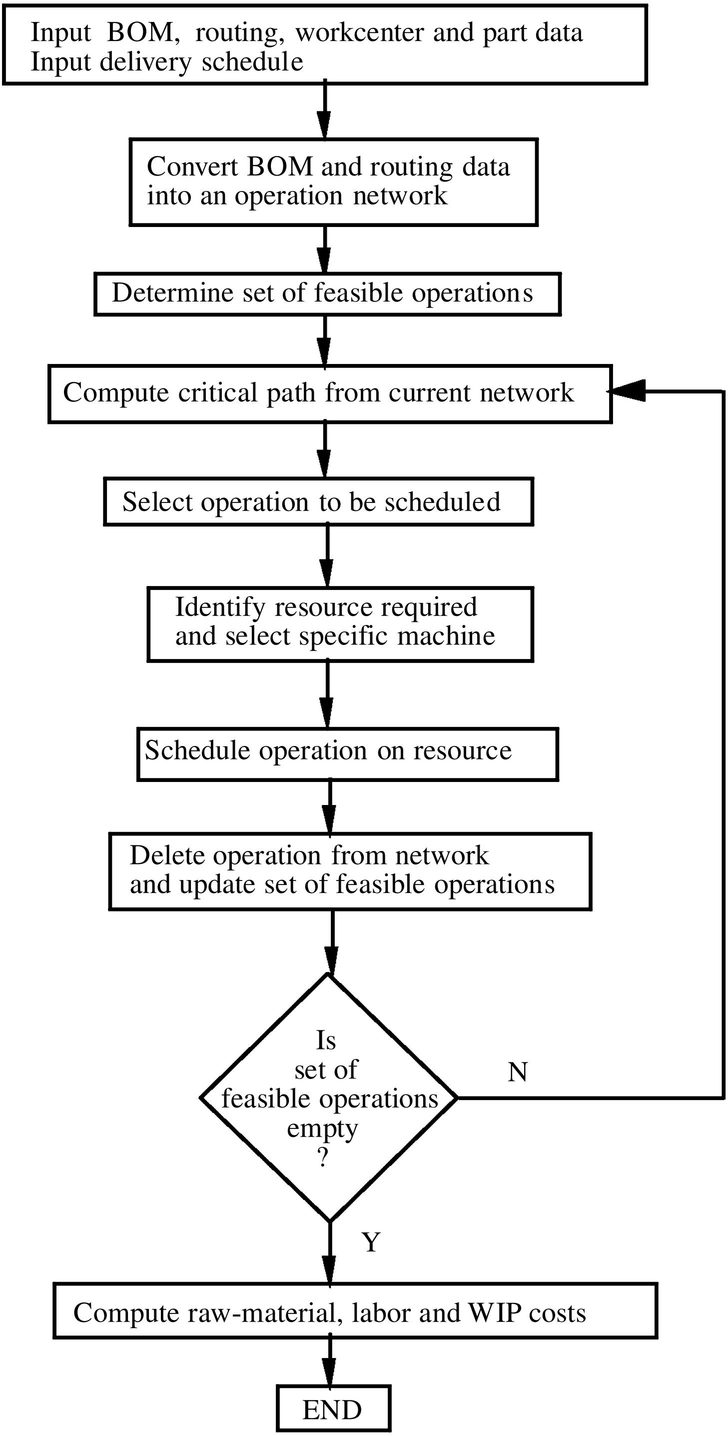


Figure A1. LETSA flow chart.

Appendix B: Performance metrics

Product cost per unit

In general, a product's cost is composed of material cost, direct labour, work-in-process costs, and overhead costs. In this metric, we consider a subset of production costs that are a function of the operations schedule, i.e. work-in-process costs—we disregard costs that cannot be controlled by the production planning and scheduling function. The material and direct labour costs can also be computed, but they are constants in our case because both the material and work (labour) content of parts remains unchanged with the actual production schedule.

The cost of production as well as work-in-process (WIP) costs are also computed from the schedule of operations. For calculating WIP costs, a rate of interest per annum and a certain compounding period is assumed. Figure B1 illustrates the method used to compute the cost of a part processed through two operations with idle time between them. Note that the accumulation of labour cost during an operation is assumed to be linear, and interest during the operation is compounded assuming that the principal consists of the mean value of the part(s) and labour costs. Thus, for operation 10 shown in figure B1 the total cost comprises three components, the initial raw material cost, the labour cost (linear component), and the interest (non-linear component). During the idle time between operations 10 and 20, the value of the part is incremented by the amount of interest accrued on the value of the part at the end of operation 10, for the duration of the idle time. This procedure is repeated for all parts to obtain the total WIP cost.

Work-centre utilization—smoothness of work-centre load profile

The measure of workload smoothness can be evaluated from the load profile of each work-centre. Figure B2 shows such a profile over the scheduling horizon. The daily load requirement is plotted against the number of days in the scheduling horizon and the average workload requirement is also indicated.

From this profile, the workload smoothness is calculated as

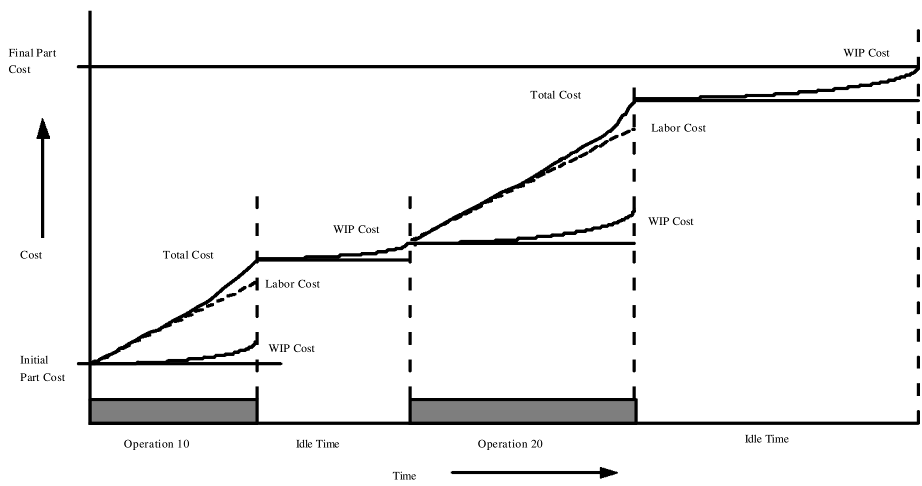


Figure B1. Method used for computing part cost (material, labour and WIP costs).

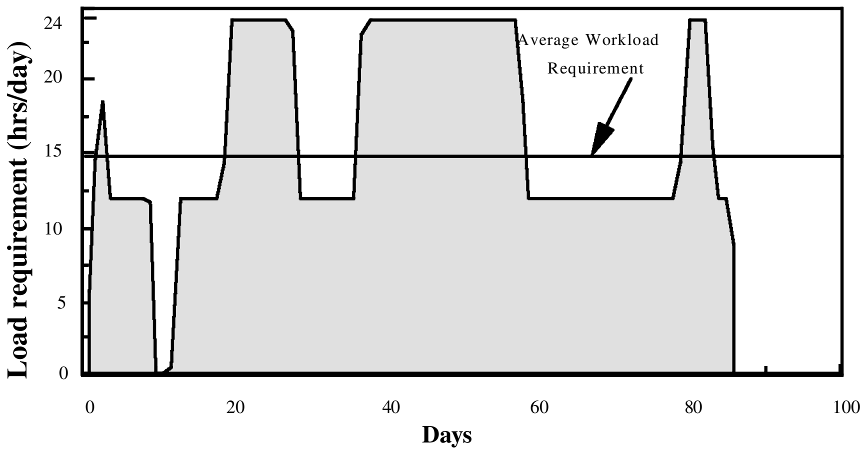


Figure B2. Work-centre load profile.

$$\mathbf{WS} = 1 - \frac{\text{Area above average workload}}{(\text{Average workload})*(\text{Length of scheduling horizon})}. \tag{B1}$$

The closer this value is to 1, the smoother the work-centre loading.

References

Agrawal, A., Harhalakis, G., Minis, I. and Nagi, R., 1996, Just-In-Time production of large assemblies. *IIE Transactions on Scheduling and Logistics*, **28**, 653–667.

Anwar, M. F. and Nagi, R., 1997, Integrated lot-sizing and scheduling for just-in-time production of complex assemblies with finite set-ups. *International Journal of Production Research*, **35**, 1447–1470.

Huge, E. C., 1979, Managing manufacturing lead-times. *Harvard Business Review*, **57**, 116–123.

John, R. S., 1985, The cost of inflated planning lead-times in MRP systems. *Journal of Operations Management*, **5**, 119–128.

Kanet, J. J., 1982, Towards understanding lead-times in MRP systems. *Production and Inventory Management*, **23**, 1–14.

Melnyk, S. A. and Piper, C. J., 1985, Leadtime errors in MRP: the lot-sizing effect. *International Journal of Production Research*, **23**, 253–264.

Orlicky, J., 1975, *Material Requirements Planning* (New York: McGraw Hill).

Peterson, L. D., 1975, Design considerations for improving the effectiveness of MRP. *Production and Inventory Management*, **16**, 48–68.

Plossl, G. W. and Welch, W. E., 1979, *The Role of Top Management in the Control of Inventory* (VA: Reston Publishing).

Schuchts, D. R., 1979, Give MRP a kiss: keep it smooth and simple. *Production and Inventory Management*, **20**, 68–80.

Steele, D. C., 1975, The nervous MRP system: how to do battle. *Production and Inventory Management*, **16**, 83–88.

Vollman, T. E., Berry, W. L. and Whybark, D. C., 1992, *Manufacturing Planning and Control Systems*, Third edition (Richard D. Irwin).

Waliszewski, D. A., 1979, Lead-time reduction in multi-flow shops. *APICS 22nd Annual Conference Proceedings*, 209.

Wemmerlov, U., 1979, Design factors in MRP systems, a limited survey. *Production and Inventory Management*, **20**, 15–35.

Wight, O., 1970, Input/output control: a real handle on lead-time. *Production and Inventory Management*, **11**, 9–31.