Workload Balancing Using the Concept of Operation Types

András Farkas
Graduate School of Business Administration
International Management Center
1051 Budapest, Nádor utca 11, Hungary
E-mail: farkas@imc.hu

Tamás Koltai
Technical University of Budapest
Department of Industrial Management
1111 Budapest, Műegyetem rkp. 9. T. Bld., Hungary
E-mail: koltai@vvg.bme.hu

Kathryn E. Stecke
University of Michigan Business School
Department of Operations Management
Ann Arbor, Michigan 48109-1234, USA
E-mail: kstecke@umich.edu

Abstract

Operations managers need to evaluate numerous alternative uses of production capacities in modern manufacturing systems. To perform such a task, an aggregate capacity analysis approach is introduced based on the concept of 'operation type'. In order to avoid excessive capacity over- and/or under-utilization, a heuristic workload balancing procedure is developed, which can be applied to mixed-type manufacturing systems that encompass both single- and multi-purpose CNC machines. An iterative algorithm is presented that reschedules the set of orders which have been assigned to successive periods by a rough-cut production schedule. Using this procedure, both workload balancing and capacity utilization of the system can be improved considerably. These results are demonstrated with an example of a sample manufacturing system.

1. Introduction

Shop-floor scheduling can be difficult for modern manufacturing systems containing expensive, technically advanced, CNC machines and/or flexible manufacturing systems (FMSs). The system set-up of an FMS in its general form is intractable because of the complicated parts being manufactured, as well as the complexity of machines, tools, and fixture assignments, and the number of different operations. Also, there may be a large number of orders to be processed per shift. Therefore, operations managers can find good use for a capacity analysis tool that is capable of evaluating the numerous alternative uses of production capacities. Such a tool should assure that the capacity over- and under-utilization of the machines are under control.

This paper addresses such a problem, of aggregate capacity analysis (ACA) and develops a heuristic workload balancing procedure for use over a short-term planning horizon. *Aggregation* here means that similar manufacturing operations, which require the same type of machines and machine tool types, are aggregated into *operation types*. The variety of alternative uses of production capacity can be reduced by applying this concept of operation types. A systematic evaluation of capacity utilization and workload balance can be achieved. Workload balancing implies that operation types are assigned to machines such that

(i)   there is sufficient capacity to perform all of the required operation types without overloading the machines, and

(ii) enough workload is allocated to the machines to avoid idle time. Hence there is no underloading of the machines.

The workload is considered *completely balanced*, *if* both capacity over- and under-utilizations of the machines can be eliminated.

Aggregation is a widely used tool in production planning. When some elements of a production system are treated in an aggregate manner, simpler models can be applied for capacity, inventory, and production planning. Obviously, some information is lost by aggregation, but at a certain level of decision making, an aggregate treatment can be sufficient. When more detailed information is required, then the aggregated information is disaggregated and a more detailed model can be applied. In traditional production planning models, products and/or facilities can be aggregated (Thomas and McClain, 1993). Products using the same setup of the production process are aggregated into product families. Or products with similar resource consumption are aggregated into product types. When an appropriate production plan is done at an aggregate level,

a detailed production program can be prepared, in which product types and/or families are disaggregated into products (see Johnson and Montgomery, 1974; Hax and Candea, 1984). Facility-level aggregation means that several different resources of the production, such as machines, workforce, and materials are considered as a single resource or facility (see Holt et al., 1960).

The ACA approach can be utilized as a new tool for a production planning and control system. In a conventional production planning process, the master production schedule (MPS) sets the quantity of each end item to be completed in each period of a short-term planning horizon. In a subsequent planning phase, the material requirements planning explodes the end items into assemblies, components, or parts, and develops a schedule of orders for the required materials over the planning horizon. Then, the capacity requirements planning tests the MPS for capacity feasibility. Order processing is done through rough-cut routing plans and weekly load schedules. Shop-floor scheduling prepares a more detailed production plan by determining timings, sequences, and workload assignments to machines in which orders should be processed, with respect to limitations of available machine capacities.

Using ACA and the proposed heuristic workload balancing procedure in an MRP environment, shop-floor scheduling here is divided into three planning phases. First a rough-cut production schedule should be prepared. In this phase, the allocation of orders is done for each time period (usually a day) over the scheduling horizon. Several heuristics are proposed by the literature to perform this task. For example, the earliest due date rule can help to achieve an acceptable level of due date performance, or the shortest processing time rule can drive the system toward an acceptably low level of WIP inventory. Next, the workload balancing procedure takes this rough-cut schedule of the orders as given and refines it over the balancing interval by interchanging the orders that have been assigned to the periods during the previous phase. Then, this new schedule of the orders is put into operation after the machine (and tooling) assignments and the final routing plan are made.

The theoretical loading problem is to allocate the total amount of work among the machines either to maximize expected production or to maximize machine utilization. This problem can be approached in two ways: by allocating part types so as to balance workloads, or by allocating part types subject to machine capacities, both in terms of time and tool magazine. A usual balancing procedure for both conventional and flexible manufacturing systems attempts to equalize the

workload of the operations assigned to each machine. This approach has appeared in the literature by many authors, e.g., a nonlinear 0-1 mixed integer program was introduced by Stecke (1983), the optimality of the balanced workloads in FMSs under certain conditions were illustrated by Stecke and Morin (1985) and Stecke and Solberg (1985).To make the loading problem of FMSs tractable, hierarchical models were developed by Stecke (1986), Van Looveren (1986), and Jaikumar and Van Wassenhove (1989). An FMS planning model in MRP environment was presented by Mazzola et al. (1989). A capacity and lead time integrated approach was presented by Lambrecht et al., (1998). Niess (1980) introduced the concept of "functional capacity" to use as a common denominator for comparing the capacity of machines performing technologically different operations. The elements of the approach of Niess was used by Koltai et al. (1998) to reduce the complexity of the aggregate production planning problem of FMSs.

The main objective of this paper is to provide a tool for short-term production scheduling, which can cope with the complexity of an FMSs environment. The complexity of the FMS scheduling problem is a consequence of the great number of technologically different operations performed. This complexity can be reduced by aggregating operations into operation types. The management objective applied here is to achieve an acceptable level of workload balance. Workload balancing is frequently used to improve both system throughput and system utilization. This objective is fulfilled with the help of a heuristic procedure which balances workloads by reducing machine over- and under-utilization.

The paper is structured as follows. Section 2 introduces the concept of an 'operation type', and develops the aggregate capacity analysis approach. Section 3 discusses the principles of a workload balancing procedure that is based on the concept of an 'operation type'. Section 4 describes an algorithm for the workload balancing procedure and illustrates its operation through a numerical example. Finally, Section 5 presents a summary and contains conclusions. The appendix describes the necessary notation and computational formulae used in the algorithm.

2. Aggregate Capacity Analysis Based on the Concept of Operation Types

Concept and definitions of the terms required for the development of an aggregate capacity analysis (ACA) are described in this section. Notation and mathematical formulae are presented in the appendix. The implementation of ACA is explained through an example of a sample manufacturing system.

An <u>order</u> (for a single part type) consists of r_i parts of type i. Each part type has a finite

number of operations. An operation, o_p, is a machine visit, and is defined by its processing time on a given machine and by the set of cutting tools required. The number of cutting tools required for each operation j, for a particular order i, are determined by the production requirements demanded by the orders, by the tool lives of the required tools of the different types, and by the processing times of each operation. A machine visit requires a pallet, a fixture, and whatever is on the fixture. This is usually a single part, but may also be multiple parts of the same type or multiple parts of several types, in one or more mounts. Each part type may have a (partial) precedence.

An operation type, ot_h, is a fundamental type of manufacturing process. For example, possible operation types can be milling, tapping, boring, drilling, or turning. The operation type(s) which can be assigned to a machine is (are) a function of the machine's processing capabilities. For example, a horizontal milling machine cannot perform vertical operations without refixturing the part. An operation type set, S_k, consists of a single or multiple operation type(s). For our purposes here, an operation (requiring a set of cutting tools) belonging to one operation type is indivisible. Consecutive operations may (or not) be assigned to the same machine. For each operation type (e.g., horizontal milling), there is a maximum number of tools (e.g., 200 tool types) that cover all operations that are members of that particular operation type. The operations of the orders that require a particular operation type use a number of tools of different types (e.g., 84 tools). This is the tool set of the tooling requirement of a given period for those orders that require a particular operation type. This tool set is called an *operation type tool set*. The tool set of an operation type set is determined by the sum of the total number of tools (including copies) required by the operation type(s) belonging to that operation type set. This tool set is called *operation type set tool set*.

A manufacturing system is a given machine configuration constituted by a group of metal-cutting machines. The machines can be single-purpose machines (a single operation type only can be performed) or multi-purpose machines (two or more operation types can be performed) and they ordinarily are CNC machines. The manufacturing system is designed to be capable of performing certain operation types. The operation type sets are assigned to the machines. A particular value (0 or 1) of the binary variable, the assignment parameter, x_{km}, indicates whether an operation type set k is assigned to a machine m ($x_{km}=1$), or not ($x_{km}=0$). These assignments are done with respect to the technical capabilities of the machines.

Capacity of a machine, c_m, here is expressed in capacity units (CUs), over a period of, for example, a shift or two, or a day. Processing time of operation j of order i, p_{ij}, is also expressed in terms of CUs, rather than in hours or minutes. It is assumed in the ACA approach that those metal-cutting CNC machines which have the same processing capabilities are technologically interchangeable. Hence, operations of the same type on these machines have identical processing times. The capacity requirement for an operation type set is the number of CUs per period demanded by the production requirements. The production requirements for the manufacturing system are determined by the shop-floor schedule. An upper capacity bound (UB) of a particular operation type set k, u_k, is the maximum amount of available capacity for an operation type set. It is calculated as the sum of the CUs of those machines which are capable of performing any and all operations belonging to that operation type set. A lower capacity bound (LB) of an operation type set k, l_k, is the minimum amount of planned free capacity for an operation type set that is available only to the operations that belong to that operation type set. It is calculated as the sum of the CUs of those machines which are capable of performing only those operations belonging to that particular operation type set. This is free capacity because if it isn't used, there is idle time on the machines. The ideal available capacity per period for an operation type set is a range defined by the difference between the upper and the lower capacity bounds of available capacity. The capacity is sufficient if the production requirements from all operation type sets are less than their corresponding upper bounds. When all operations have been assigned to the machines and the workload is less than the lower capacity bound of any operation type set, then there is machine idle time.

The use of the concept and its constituted terms of ACA is illustrated through the following sample case. Consider a mixed type of manufacturing system consisting of five metal-cutting machines. This set of machines includes three multi-purpose machines (M1,M2,M3) and two single-purpose machines (M4,M5) as is shown in Figure 1. Single-purpose machines can either be conventional or CNC machines. The system is designed to perform three operation types (ot_1, ot_2, and ot_3), which are drilling, vertical milling, and horizontal milling, respectively. These manufacturing capabilities, which are attainable through tooling up the machines, are indicated in Figure 1. For example, ot_1 and ot_2 at machine #3 indicates that M3 is capable of performing operation types #1 and #2 (i.e., M3 does drilling and vertical milling), whereas M5 is a single-purpose machine for ot_1 (drilling). Tooling is determined by all tools that are technically specified

for the part types planned for production in a particular period. Hence, M5 requires the tool set of a single operation type set (S_1), while M3 requires all the tools of ot_1 and ot_2 which compose the tool set of a multiple operation type set (S_4). (Operation type sets, S_k, $k=1,...,7$ are defined in Figure 2).

Production requirements are given in Table 1 for one period (say for P1). Let one period correspond to one shift. The different part types manufactured in the system are termed as orders. Identification of these orders is done by the code numbers #0101 through #0110. The required operation types are indicated in the heading of Table 1, while their corresponding operations are given by the calculated total processing times expressed in CUs (i.e., order #0103 requires 0.12 CUs for drilling, 0.06 CUs for vertical milling, and 0.23 CUs for horizontal milling). 0.12 CUs means that in a single eight hour shift, 57.6 minutes of drilling operation is required. In practice, for order #0103, the 0.12 CUs of drilling may represent the drilling operations of 5 borings with different diameters and bore lengths, which can be processed by using 5 different tools during a total of 57.6 minutes. This can be done, for example, by one machine visit either on M5, or alternatively, on M3, or on M1.

The ideal available capacity range for each operation type set is displayed in Figure 2. The possible operation type sets are placed on the horizontal axis. Since the system configuration includes three technologically distinguishable operation types, there are three operation type sets (S_1, S_2, S_3) with single operation types, three operation type sets (S_4, S_5, S_6) with two operation types, and one operation type set (S_7) with three operation types. In general, the total number of the operation type sets, K, can be calculated as

$$K=\left(\sum_{h=1}^{H-1} \frac{H!}{h!(H-h)!}\right)+1,$$

where H is the total number of single (distinguishable) operation types.

On the vertical axis the lower capacity bounds (LBs) and the upper capacity bounds (UBs) of each operation type set are drawn by a dashed line and a dotted line, respectively. For example, for the operation type set S_1, LB=1, because M5 is the only machine which is dedicated for ot_1; UB=3, since M1, M3, and M5 are all capable of performing ot_1. For S_4, which is an operation type set containing operation types ot_1 and ot_2, LB=3, because M3 can perform ot_1 and ot_2, M4 can perform ot_2, and M5 can perform ot_1. For S_4, UB=5, as all of the machines can do either ot_1 or ot_2.

If the capacity requirement is smaller than the lower bound of any operation type set, then the system is underloaded, and there is free capacity on one or more machines. For example, if less than 1 CU from ot_1 is required, then the single-purpose drilling machine, M5, will have idle capacity. If the capacity requirement is greater than the upper bound of any operation type set, then the system is overloaded, i.e., there is not enough available capacity for the required operations. For instance, if more than 3 CUs are required from ot_1, then M1, M3, and M5 will not have enough capacity to perform the drilling operations for the ten orders assigned to this period. When the production requirements of each operation type set are within the lower and the upper capacity bounds (the shaded area in Figure 2), then there are neither unutilized nor excess capacity on any of the machines. In the most favorable case, the capacity requirements of a set of orders of a given period fall into the ranges defined by the lower and the upper bounds for each operation type set.

The capacity requirement for each operation type set is shown in Figure 3. It is calculated as the total processing time of all operations of the corresponding operation type set.

Figure 4 plots both the ideal available capacity range and the capacity requirement for each operation type set in one chart. The resulting graphical display is used to determine whether or not the manufacturing system is in complete technological balance, i.e., to check whether there is any excess capacity or lack of capacity from certain operation types. For example, in our case, Figure 4 illustrates that there are underloads at S_1 and S_4. The underload at S_1 indicates that if all drilling operations are processed on M5, then this machine still has idle capacity. The underload at S_4 indicates that idle capacity may exist on M3 and on M5. The underload of S_4 is higher than the underload of S_1, since the total processing time of ot_2 does not require all of the available capacities of M3 and M4. Therefore, the idle capacity of M5 is increased by the idle capacity of M3. The idle capacity of M3 can be utilized for a portion of ot_1, but then, the idle capacity of M5 increases the under-loading, as can be seen at S_4.

Figure 4 exhibits overloads at S_3, S_6, and S_7. The overload at S_3 indicates that the total capacity requirements of ot_3 are higher than the total available capacity on M1 and on M2. The overload at S_6 indicates that the capacity requirements from ot_2 and ot_3 together, are higher than the available capacities on machines M1, M2, M3, and M4. Finally, the overload at S_7 means that the capacity requirements of all of the operation types are higher than the available capacity of the manufacturing system.

In real manufacturing systems, operations managers should generally be resigned to a certain amount of idle capacity. Idle capacity is either planned and serves as a buffer capacity to absorb the effect of unexpected events (i.e., machine breakdowns, power breaks, tool breakages, etc.), or it is due to routing constraints. On the other hand, a certain amount of overtime, or other means, can be used to augment capacity. It is the competency of top management, however, to prescribe acceptable tolerances on both sides. The parameters α and β are introduced to express these tolerances. The acceptable percentage of idle capacity, α, on any machine in any period is expressed in relation to the base capacity. The acceptable percentage of excess capacity, β, on any machine in any period is also expressed in relation to the base capacity. Note that the acceptable tolerances for the machine capacities are the equivalent of the acceptable tolerances of the operation type sets as well. In accordance with the basic unit of measurement used in the ACA framework, the values of α and β should be converted to CUs for the numerical calculations. Using given values of α and β, the augmented available capacity per period for an operation type set can easily be determined. This range is defined by the difference between the increased upper capacity bound and the decreased lower capacity bound of available capacity. Figure 4 also displays the augmented available capacity range for each operation type set with the values of α=0.10 CUs and β=0.05 CUs, where the thin dotted line represents the increased upper capacity bound and the thin dashed line represents the decreased lower capacity bound for each operation type set.

3. A Heuristic Workload Balancing Procedure

In this section, a heuristic workload balancing procedure is developed. Notation and the computational formulae of the terms introduced in this session are described in the appendix. The objective of the procedure is to achieve an acceptable level of system balance. This way, the capacity utilization of the machines can also be improved. Thus the output of this procedure produces a balanced workflow of the manufacturing system in each period over the balancing interval.

The workload balancing procedure based on the concept of ACA is shown in Figure 5. The rough-cut production schedule of the orders on the shop-floor level is given. Figure 5a plots both the capacity supply and the capacity demand for each operation type set in period P1. Consider those operation type sets for which the production requirements are beyond the ideal available capacity ranges. It is apparent that this diagram reflects a quite severe workload imbalance of the

system. Therefore, the preliminary schedule of the orders is being rescheduled by an iterative algorithm in order to improve workload balance. Figure 5b displays the rescheduled set of orders for the same period. The arrows indicate that the orders are rescheduled throughout this process. If necessary, orders from period P1 are removed to relieve the workload of the operation type sets in this period and/or orders from the successive periods are inserted back to period P1 to augment the workload and to improve capacity utilization. This step-wise iteration process is repeated until the desired level of workload balance is achieved. An acceptable output of this procedure is plotted in Figure 5b for period P1. By contrasting the loading situations of the manufacturing system before (Figure 5a) and after balancing (Figure 5b), the impact of the procedure on making a drastic reduction of the overloads and the underloads is obvious. This period balancing procedure is extended to a given number of successive periods comprised by the balancing interval. The length of such a balancing interval is determined by the shop-floor scheduler with respect to the local conditions. In the balancing interval, the final schedule of the orders is frozen before the balancing procedure starts its operation for the next period. After two or three weeks, the whole procedure should be repeated for the periods of the next balancing interval. This rolling nature of the balancing procedure is similar to the rolling nature of aggregate production planning over a mid-term time horizon based on an MPS.

The following definitions describe those parameters that are used for the time scaling of the procedure (see in Figure 6). The length of a balancing period, L_p, is the principal unit of time measurement and refers to the length of time that is covered by one run of the balancing algorithm. The length of a period that can be applied in this procedure is usually a shift or two, or a day, in duration. In practice, L_t is constant for each period. A specified number of periods, T, has to be balanced. The length of the balancing interval, LB, refers to the T number of periods, which can be, ten days, two weeks, etc., over which the balancing algorithm performs T runs. Furthermore, a given number of additional periods, τ, has to be available, which contain scheduled orders that are also accessible for order interchanging purposes. The length of the scheduling horizon, LS, refers to the time horizon over which the preliminary rough-cut production schedule is prepared. An order scheduled initially into period t can be shifted to the period $t+1$. Conversely, however, any order from the τ successive periods can be moved into period t. Therefore, to balance T number of periods, the rough-cut production schedule should be available for $T+\tau$ number of periods.

As an example, the length of a period can be one day ($L_t=1$). When a particular period is balanced, the orders scheduled for the next four days ($\tau=4$) are used for interchanging purposes. If the balancing interval covers ten days ($T=10$), then the length of the scheduling horizon should cover 14 days ($T+\tau=14$). Ten days later the entire procedure is repeated again.

A scheduled order, q_{it}, consisting of r_i parts of type i, is an order that is scheduled into period t. The assignment of the orders to a given period over the scheduling horizon is done initially by the rough-cut production scheduling. Then, the assignment of an order may vary through the iteration steps of the balancing procedure. A set of orders, Q_t, is the collection of all orders that are scheduled into period t. Initially, a set of orders usually contains 8-15 orders per/shift (period). Then each of these sets is increased or decreased by one order in every iteration step of the balancing algorithm.

Five different workload categories are used. In general, the workload is defined as the total processing time that is required to perform the specified technological operations on the machines. The workload of an order, wo_{ih}, is the sum of the processing time of all operations of each operation type h required to process order i. The workload of an operation type, wot_{ht}, is the sum of the processing time of all operations of operation type h required to process all scheduled orders in period t. The workload of an operation type set, ws_{kt}, is the sum of the required machine hours of each single operation of all orders that are associated with operation type set k in period t. The relieved workload of an operation type set, ws_{ikt}^-, is the sum of the required machine hours of each single operation of all orders that are associated with operation type k in period t after excluding order i from the set of orders of period t. The augmented workload of an operation type set, ws_{ikt}^+, is the sum of the required machine hours of each single operation of all orders that are associated with operation type k in period t after adding order i to the set of orders of period t.

The overloading categories are defined as the amount of surplus capacity requirements over the available capacity determined by the upper capacity bounds. Four types of overloads are distinguished. The overload of an operation type set, ols_{kt}, is calculated for an operation type k by considering all orders in period t. The decreased overload of operation type set, old_{ikt}, is calculated for an operation type set k by considering all orders after excluding order i in period t. The decreased overload of the manufacturing system, $olmd_{it}$, is defined as the maximum of the decreased overloads of the operation type sets in period t. The overload of the manufacturing system, olm_t, is defined as the maximum of the overloads of the operation type sets in period t.

The underloading categories are defined as the amount of idle capacity (slack) under the available capacity determined by the lower capacity bounds. Four types of underloads are distinguished. The underload of an operation type set, uls_{kt}, is calculated for an operation type set k by considering all orders in period t. The decreased underload of operation type set, uld_{ikt}, is calculated for an operation type set k by considering all orders after adding order i in period t. The decreased underload of the manufacturing system, $ulmd_{it}$, is defined as the maximum of the decreased underloads of the operation type sets in period t. The underload of the manufacturing system, ulm_{t}, is defined as the maximum of the underloads of the operation type sets in period t.

The balancing algorithm attempts to implement the following managerial objective. Let the capacity requirement for each operation type set be within the range of augmented available capacity in each period of the balancing interval. This objective ensures a *satisfactory level* of workload balance on the set of machines as well as for the machine operators.

The *state of loading* of the manufacturing system can be interpreted as the difference between the capacity supply and the capacity demand of the system. In any period, five possible states of loading of the system can be distinguished:

(i) the system is *overloaded* in period t, if there is a surplus capacity requirement for at least one operation type set k. In this case a properly chosen order should be removed from period t, in order to reduce or to eliminate the overutilization of that operation type set. The removed order is shifted to period $t+1$.

(ii) the system is *underloaded* in period t, if there is idle capacity (slack) for at least one operation type set k. In this case, a properly chosen order should be inserted into period t, in order to reduce or to eliminate the under-utilization of that operation type set. The inserted order is being selected from any of the successive τ periods.

(iii) the system is in a *virtual technological balance* in period t, if the difference between the sum of the overloads and the sum of the underloads of each operation type set is equal to zero. In this case, the workload of the manufacturing system equalizes the available capacity, but certain operations cannot be performed due to shortages of the demanded type(s) of capacity, while there exists excess capacity for other operations that are not required. In such a situation either a removal or an insertion of an order may improve the systems' balance.

(iv) the system is in a *required technological balance* in period t, if the workload of each

operation type set is within the range of augmented available capacity determined by the increased upper and decreased lower capacity bounds.

(v) the system is in a *complete technological balance* in period t, if the workload of each operation type set is within the range of ideal available capacity determined by the upper and lower capacity bounds.

The *actual state of loading* of the manufacturing system is determined by the workload of that operation type set which contains *all* possible operation types (e.g., S_7 in Figure 4).

The following parameters control the process of the balancing procedure. If the system is overloaded in any period, then an order is removed from that period. The <u>removed workload</u>, $\varepsilon_{i,t}^-$ is that portion of the workload of all single operations of all operation type sets that is composed of the workload of those operation type sets which represent overloads in period t. This parameter is calculated for the removed order i, by considering all orders scheduled for period t. Conversely, if the system is underloaded in any period, then an order should be inserted into that period. The <u>inserted workload</u>, $\varepsilon_{i,t}^+$ is that portion of the workload from all single operations of all operation type sets that is composed of the workload of those operation type sets which represent underloads in period t. This parameter is calculated for the inserted order i by considering all orders that are comprised by the successive τ number of periods. The <u>remaining workload for a removal</u>, $\rho_{i,t}^-$ of order i is that portion of the workload of all single operations of all operation type sets which neither represent an overloading nor an underloading in period t. The <u>remaining workload for an insertion</u>, $\rho_{i,t}^+$ of order i is that portion of the workload of all single operations of all operation type sets which neither represent an overloading nor an underloading in period t. The value of <u>workload transfer for a removal</u>, $v_{i,t}^-$ is the change in the difference between the capacity supply and the capacity demand of the manufacturing system that is resulted in by the removal of order i, in period t. The value of <u>workload transfer for an insertion</u>, $v_{i,t}^+$ is the change in the difference between the capacity supply and the capacity demand of the manufacturing system that is the result of the insertion of order i, in period t. The latter values are calculated as the algebraic difference between the removed and the remaining workload or the difference between the inserted and the remaining workload in period t, if order i is removed or inserted into period t in a particular iteration step of the balancing algorithm. Through every iteration step of the balancing procedure, this difference in the workload of the system diminishes until the targeted tolerances determined by the values of α and β are reached in a particular period t. The algorithm

ensures that in each iteration step, the most efficient workload transfer be performed by choosing the maximum of the values of the workload transfer, called the <u>efficient workload transfer</u>, v_r In other words, that order is selected for a removal or for an insertion which produces the greatest change (decrease) in the workload difference of the system. Due to the discrete nature of the problem it may happen that no orders can be found in the sets of the orders such that a run of the balancing algorithm would result in a less workload difference of the system. In this case, the workload balancing problem has no feasible solution with respect to the current values of α and β. This situation, however, very rarely occurs in practice. Nevertheless, the probability of occurrence of such a case can be reduced by increasing the number of scheduled orders into the periods, which is attainable if the processing times of the operations belonging to a given operation type are relatively short. Then, a greater number of rescheduling alternatives can be used to improve the efficiency of the balancing procedure. The larger the set of orders in each period as well as the flexibility of the manufacturing system, the greater the chances are to produce a balanced workflow. An ultimate act of top management could be to prescribe larger values for the excess capacity β and for the idle capacity α which, however, would result in higher levels of capacity over- and under-utilization of the system.

4. Algorithm of the Workload Balancing Heuristic and a Numerical Example

The simplified flowchart of the workload balancing algorithm is presented in Figure 7. Assumptions underlying the use of this heuristic are the following. The machine configuration of the manufacturing system and the available operation types that the machines are capable of performing are known. The acceptable percentages of excess and idle capacities of the machines are specified. Furthermore, the rough-cut production schedule of the orders over the scheduling horizon is available. Given these inputs, the major steps of the algorithm are the following:

Step 0. Generate the initial set of orders for all the periods and the corresponding capacity requirement for each operation type set. Compute the range of ideal available capacity and the range of augmented available capacity per period for each operation type set.

Step 1. Begin a run of the balancing algorithm for period t ($t=1,...,T$).

Step 2. Begin an iteration step of the balancing algorithm for period t. Compute the overload and the underload of each operation type set in period t. Evaluate the state of loading. If the system is in the state of the required technological balance, or in a complete

technological balance, then go to Step 1 and begin the balancing of the next period ($t=t+1$). If the system is overloaded, or is in a virtual technological balance, then go to Step 3. If the system is underloaded, then go to Step 4.

Step 3. Determine the workload release of each operation type set in period t. Compute the removed workload and the remaining workload of the operation type sets for each order in period t. Then determine the value of the workload transfer of the manufacturing system. In the set of orders of period t, find that order i for which the value of the workload transfer is maximal. Choose order i. Remove order i from the set of orders of period t. Go to Step 5.

Step 4. Determine the augmented workload of each operation type set in period t. Compute the inserted workload and the remaining workload of the operation type sets for each order in period t. Then determine the value of the workload transfer of the manufacturing system. In the sets of orders of τ successive periods, find that order i for which the value of the workload transfer is maximal. Choose order i. Insert order i into the set of orders of period t.

Step 5. Fix the new set of orders for period t at the current iteration step n. Compare the current overload or underload of the system with the corresponding threshold of α or β, representing the acceptable levels of excess and idle capacity, respectively. If the actual capacity requirement is within the range of augmented available capacity, i.e., the system is in the state of the required technological balance, then go to Step 1 and begin the balancing routine for the next period ($t=t+1$). Otherwise go to Step 2 and begin the next iteration step n for the same period t.

The balancing algorithm stops when all periods of the balancing interval are in the required technological balance ($t=T$). This procedure is repeated again when the scheduling of the orders of the next balancing interval is requested by the management of the production system.

The output of the algorithm produces the final schedule of the orders in each period t ($t=1,..,T$) of the balancing interval. The actual values of the overload and underload of the manufacturing system resulting from the last iteration step of the consecutive runs of the algorithm are also provided.

The five-machine sample manufacturing system displayed in Figure 1 is used to illustrate the

performance of the procedure and to provide numerical results. The rough-cut production schedule of the orders over the scheduling horizon is given in Table 2 for $T+\tau$ (10+4=14) consecutive periods, i.e., for P1, ..., P14. In this table, the four digit numbers identify the orders. The first two digits of these numbers refer to the period to which these orders have been assigned by the rough-cut production schedule, and the last two digits of these numbers are the code numbers of the orders. There are a total of 140 orders which are distributed uniformly among the periods. It can also be seen from this table that overloads are dominant in the system. There are some very large overloads in several periods and one large underload in period P2.

Let the acceptable levels of excess capacity (threshold for overload that can be covered by overtime) and of idle capacity (threshold for machine idle time) be equal to 5 percent (β=0.05 CU) and 10 percent (α=0.10 CU), respectively. This conforms to a management policy of when production managers are more sensitive to the more expensive overtime than to the less costly capacity under-utilization of the machines. Table 3 exhibits the rescheduled sets of orders over the ten-period balancing interval. The computed new overloads and underloads of the manufacturing system show that a feasible solution is obtained for the workload balancing problem after performing 10 runs of the algorithm. Interchanges of the orders triggered by the algorithm can also be detected from this table. Orders with boldfaced code numbers have been rescheduled by the algorithm. For example, order #0505 was initially scheduled into period P5 (see the number 0505 in Table 2), but the balancing algorithm assigned it to period P2 as is seen from the second column of Table 3. Now, the system is in a required technological balance, since the overloads and the underloads are within the specified tolerance limits in each period. Examining the new values of overloads and underloads it is apparent that the capacity utilization of the manufacturing system has also improved considerably. However, the total number of orders scheduled into P1, ..., P10 has been reduced to 96 from 100. As a consequence, presumably, schedulers should reckon with an overloading of the system to a lesser extent over the next balancing interval unless the next scheduling horizon would be less overloaded.

A more detailed illustration of the results is shown in Table 4, where the new capacity requirement of each operation type set is given for each period t, t=1,...,10. In the heading of this table, the upper (UB) and the lower (LB) capacity bounds of each operation type set are indicated. By comparing the new capacity requirement and the upper and the lower capacity bounds for each operation type set, the balance between the capacity demand and the capacity

- 15 -

supply of the system can easily be studied. The overloads of the operation type sets are indicated in the bracketed superscripts with a positive sign, whereas the underloads of the operation type sets are indicated in the bracketed subscripts with a negative sign. Each of them correspond to the appropriate operation type set and a particular period. Obviously, all of these overloads and underloads are within the range of augmented available capacity for each operation type set.

5. Summary and Conclusions

This paper presents an aggregate production planning concept and a computational tool for the short-term workload allocation problem of manufacturing systems. The aggregation of operations into operation types provides a main idea to help address this problem. This way, the complexity of the balancing problem can significantly reduced and the capacity utilization of the machines can be improved. Using the aggregate capacity analysis concept, two type of operations management questions can be formulated:

(i) How to allocate operation types to machines in order to achieve an acceptable level of system balance when a set of orders is given.

(ii) What is the subset of a set of orders which provides an acceptable level of system balance when the operation type allocation to machines is given.

Problem (i) can be solved by a 0-1 mathematical programming model (see Koltai et al., 1998), while problem (ii) is discussed in this paper.

The presented heuristic workload balancing procedure, based on the aggregate capacity analysis, develops a smoothed schedule of the orders over the short-term time horizon such that both the overloading and the underloading of the manufacturing system remain within acceptable tolerance limits in each period.

The balancing algorithm introduced in this paper utilizes the machine capabilities and system flexibility. Therefore, it can be considered as an alternative approach beyond the traditionally-used loading strategies. Nevertheless, some other issues, for example, finding a feasible machine tooling or an optimal solution to the part type routing problem are beyond the scope of this paper. Also, it should be noted that if the number of operation types is large, then the aggregate model is difficult to handle. This is, however, a general problem that may arise in the case of any known aggregate production planning model. Hence, the development of a adequate method for disaggregating the operation types into operations will be the subject of future research.

Appendix

Subscripts:

balancing period	$t = 1,...,T$
subset of periods	$t' = t+1,...,t+\tau$
order (part type)	$i = 1,...,I$
operation	$j = 1,...,J$
operation type	$h = 1,...,H$
set of operation types	$k = 1,...,K$
subset of a set of operation types	$k' = 1,...,K'$
machines	$m = 1,...,M$
iteration step of a run of the algorithm	$n = 1,...,N$

Parameters:

r_i = production requirements of part type i, $i=1,...,I$

o_j = operation j, $j=1,...,J$

ot_h = operation type h, $h=1,...,H$

S_k = operation type set k, $S_k = \{ot_h, h=1,...,H_{Sk}\}$, $h=1,...,H$; $k=1,...,K$,

where H_{sk} is the number of operation types in operation type set S_k, $k=1,...,K$

S_k' = set of all possible subsets of S_k, $k=1,...,K'$

x_{km} = operation type set k such that (i) if it can be performed on machine m, then $x_{km}=1$, or (ii) if it cannot be performed on machine m, then $x_{km}=0$, $k=1,...,K$; $m=1,...,M$

c_m = production capacity of machine m (expressed in CUs), $m=1,...,M$

p_{ij} = processing time of operation j of order i (expressed in CUs), $i=1,...,I$; $j=1,...,J$

u_k = upper capacity bound of operation type set k (expressed in CUs),

$$u_k = \sum_{\forall S_k | ot_h \in S_k} \sum_{m=1}^{M} c_m x_{km}, \quad k=1,...,K; \quad h=1,...,H$$

l_k = lower capacity bound of operation type set k (expressed in CUs),

$$l_k = \sum_{\forall S_k' \in S_k} \sum_{m=1}^{M} c_m x_{km}, \quad k=1,...,K \quad k'=1,...,K'$$

α = acceptable percentage of idle capacity on any machine in any period (its value is converted to CUs)

β = acceptable percentage of excess capacity on any machine in any period (its value is converted to CUs)

T = number of periods to be balanced

τ = number of additional periods which also contain scheduled orders

L_t = length of balancing period t (the principal unit is one period)

$$L_t = \frac{L}{T}, \qquad t=1,..,T$$

LB = length of the balancing interval,

$$LB = \sum_{t=1}^{T} L_t$$

LS = length of the scheduling horizon,

$$LS = LB + \sum_{t=T+1}^{T+\tau} L_t$$

q_{it} = order i scheduled into period t, $i=1,...,I$; $t=1,..,T+\tau$

Q_t = set of orders scheduled into period t, $t=1,...,T+\tau$

wo_{ih} = workload of order i of all operations of each operation type h (expressed in CUs),

$$wo_{ih} = \sum_{\forall o_j | o_j \subset ot_h} p_{ij}, \qquad i=1,...,I$$

wot_{ht} = workload of an operation type h for all orders in period t (expressed in CUs),

$$wot_{ht} = \sum_{\forall q_{it} \in Q_t} p_{ij}, \qquad o_j \in ot_h, \quad h=1,...,H; \quad t=1,...,T+\tau; \quad i=1,...,I$$

ws_{kt} = workload of operation type set k, for all orders in period t (expressed in CUs),

$$ws_{kt} = \sum_{ot_h \in S_k} wot_{ht}, \qquad k=1,...,K; \quad t=1,...,T+\tau$$

ws^-_{ikt} = relieved workload of operation type set k for all orders after excluding order i in period t (expressed in CUs),

$$ws^-_{ikt} = ws_{kt} - wo_{ih}, \qquad q_{it} \in Q_t, \quad t=1,...,T; \quad k=1,...,K$$

ws^+_{ikt} = augmented workload of operation type set k for all orders after adding order i in period t (expressed in CUs),

$$ws^+_{ikt}=ws_{kt}+wo_{ih} \, , \quad q_{it}\in Q_{t'}, \quad t'=t+1,...,t+\tau; \quad k=1,...,K$$

ols_{kt} = overload of operation type set k for all orders in period t (expressed in CUs),

$$ols_{kt}=\max\{ws_{kt}-u_k;0\}, \quad k=1,...,K; \quad t=1,...,T$$

uls_{kt} = underload of operation type set k for all orders in period t (expressed in CUs),

$$uls_{kt}=\max\{l_k-ws_{kt};0\}, \quad k=1,...,K; \quad t=1,...,T$$

old_{ikt} = decreased overload of operation type set k for all orders after excluding order i in period t (expressed in CUs),

$$old_{ikt}=\max\{ws^-_{ikt}-u_k;0\}, \quad q_{it}\in Q_t; \quad t=1,...,T \quad k=1,...,K$$

uld_{ikt} = decreased underload of operation type set k for all orders after adding order i in period t (expressed in CUs),

$$uld_{ikt}=\max\{l_k-ws^+_{ikt};0\}, \quad q_{it}\in Q_{t'}, \quad t'=t+1,...,t+\tau \quad k=1,...,K$$

$olmd_{it}$ = decreased overload of the manufacturing system for all orders after excluding order i in period t (expressed in CUs),

$$olmd_{it}=\max_i\{old_{ikt}\}, \quad q_{it}\in Q_t, \quad t=1,...,T \quad k=1,..,K$$

$ulmd_{it}$ = decreased underload of the manufacturing system for all orders after adding order i in period t (expressed in CUs),

$$ulmd_{it}=\max_i\{uld_{ikt}\}, \quad q_{it}\in Q_{t'}, \quad t'=t+1,...,t+\tau \quad k=1,..,K$$

olm_t = overload of the manufacturing system for all orders in period t (expressed in CUs),

$$olm_t=\max_k\{ols_{kt}\}, \quad t=1,...,T \quad k=1,..,K$$

ulm_t = underload of the manufacturing system for all orders in period t (expressed in CUs),

$$ulm_t=\max_k\{uls_{kt}\}, \quad t=1,...,T; \quad k=1,...,K$$

ε_{it}^{-} = removed workload of the operations of all operation type sets of order i in period t (expressed in CUs),

$$\varepsilon_{it}^{-}=olmd_{it}-olm_{p} \quad q_{it}\in Q_{p} \quad t=1,...,T$$

ε_{it}^{+} = inserted workload of all operations of all operation type sets of order i in period t (expressed in CUs),

$$\varepsilon_{it}^{+}=olmd_{it}-olm_{p} \quad q_{it}\in Q_{t'}, \quad t'=t+1,...,t+\tau$$

ρ_{it}^{-} = remaining workload for the removal of order i of all operations of all operation type sets in period t (expressed in CUs),

$$\rho_{it}^{-}=ulmd_{it}-ulm_{p} \quad q_{it}\in Q_{p} \quad t=1,...,T$$

ρ_{it}^{+} = remaining workload for the insertion of order i of all operations of all operation type sets in period t (expressed in CUs),

$$\rho_{it}^{+}=ulmd_{it}-ulm_{p} \quad q_{it}\in Q_{t'}, \quad t'=t+1,...,t+\tau$$

v_{it}^{-} = value of workload transfer for the removal of order i of all operations of all operation type sets in period t (expressed in CUs),

$$v_{it}^{-}=\varepsilon_{it}^{-}-\rho_{it}^{-}, \quad q_{it}\in Q_{p} \quad t=1,...,T$$

v_{it}^{+} = value of workload transfer for the insertion of order i of all operations of all operation type sets in period t at (expressed in CUs),

$$v_{it}^{+}=\varepsilon_{it}^{+}-\rho_{it}^{+}, \quad q_{it}\in Q_{t'}, \quad t'=t+1,...,t+\tau$$

v_{t} = efficient workload transfer of all operations of all operation type sets in period t (expressed in CUs),

$$v_{t}=\max\{v_{it}^{-};v_{it}^{+}\}, \quad q_{it}\in Q_{p} \quad t=1,..,T, \quad \text{and} \quad q_{it}\in Q_{t'}, \quad t'=t+1,...,t+\tau$$

References

Hax, A.C. and Candea, C.(1984), *Production and Inventory Management*, Prentice-Hall, Englewood Cliffs, NJ.

Holt, C.C., Modigliani, F., Muth, J.F., and Simon, H.A. (1960), *Planning Production, Inventories and Work Force*, Prentice-Hall, Englewood Cliffs, NJ.

Jaikumar, R. and van Wassenhove, L.N. (1989), "A Production Planning Framework for Flexible Manufacturing Systems", *Journal of Manufacturing and Operations Management*, Vol. 2, pp. 52-79.

Johnson, L.A. and Montgomery, D.C. (1974), *Operations Research in Production Planning, Scheduling and Inventory Control*, John Wiley and Sons, NY.

Koltai, T., Farkas, A., and Stecke, K. E. (1998), *Aggregate Production Planning of Flexible Manufacturing Systems using the Concept of Operation Types*, Working Paper #98007, University of Michigan, Ann Arbor.

Lambrecht, M.R., Ivens, P.L., and Vandaele, N.J. (1998), "ACLIPS: A Capacity and Lead Time Integrated Procedure for Scheduling", *Management Science*, Vol.44, No.11.

Mazzola, J.B., Neebe, A.W., and Dunn, C.V.R. (1989), "Production Planning of a Flexible Manufacturing System in a Material Requirements Planning Environment", *International Journal of Flexible Manufacturing Systems*, Vol.1, pp. 115-142.

Niess, P.S. (1980), "Kapazitätsabgleich bei Flexiblen Fertigungssytemen", *IPA Forschung und Praxis*, Springer Verlag, Berlin, Germany.

Stecke, K.E. (1983), "Formulation and Solution of Nonlinear Integer Production Planning Problems for Flexible Manufacturing Systems", *Management Science*, Vol. 29, No. 3, pp. 273-288.

Stecke, K.E. and Morin, T.L. (1985), "The Optimality of Balancing Workloads in Certain Types of Flexible Manufacturing Systems", *European Journal of Operational Research*, Vol. 20, pp. 68-82.

Stecke, K.E. and Solberg, J.J. (1985), "The Optimality of Unbalancing Both Workloads and Machine Group Sizes in Closed Queueing Networks of Multiserver Queues", *Operations Research*, Vol. 33, No. 4, pp. 882-910.

Stecke, K.E. (1986), "A Hierarchical Approach to Solving Machine Grouping and Loading

Problems of Flexible Manufacturing Systems", *European Journal of Operational Research*, Vol. 24, pp. 369-378.

Thomas, L.J., and McClain, J.O. (1993), "An Overview of Production Planning", In: *Logistics of Production and Inventory* (Editors, Graves, S.C., Rinnooy Kan, A.H.G., and Zipkin, P.H.), North-Holland, Amsterdam, The Netherlands.

Van Looveren, A.J., Gelders, L.F., and van Wassenhove, L.N. (1986), "A Review of FMS Planning Models", In: *Modelling and Design of Flexible Manufacturing Systems, (Editor, A. Kusiak)*, Elseviers Science Publishers B.V., Amsterdam, The Netherlands, pp. 3-31.

Captions and Table Headings

Figure 1. Sample manufacturing system with the allocation of the operation types

Figure 2. Ranges of ideal available capacity for the operation type sets

Figure 3. Capacity requirements for the operation type sets

Figure 4. Ranges of ideal available capacity and augmented available capacity with the production requirements for the operation type sets (α=0.10 CUs, β=0.05 CUs)

Figure 5. Illustration of order rescheduling of the workload balancing procedure

Figure 6. Time scaling of the balancing procedure

Figure 7. The simplified flowchart of the balancing algorithm

Table 1. Production requirements in CUs for period (P1)

Table 2. Rough-cut production schedule of the orders over the scheduling horizon

Table 3. Rescheduled sets of orders over the balancing interval (α=0.10 CUs, β=0.05 CUs)

Table 4. The capacity requirement of the operation type sets for the rescheduled sets of orders in each period over the balancing interval

Table 1. Production requirements in CUs for P1

Order No.	ot_1	ot_2	ot_3
0101	0.28	0.83	0.34
0102	0.04	0.00	0.40
0103	0.12	0.06	0.23
0104	0.00	0.00	0.35
0105	0.00	0.46	0.35
0106	0.01	0.00	0.17
0107	0.19	0.15	0.73
0108	0.00	0.00	0.10
0109	0.33	0.05	0.01
0110	0.00	0.35	0.12
Total	**0.97**	**1.90**	**2.80**

Table 2. Rough-cut production schedule of the orders over the scheduling horizon

Periods	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14
Overload [CU]	0.67	0.00	0.33	0.93	2.11	0.66	1.03	0.73	0.84	0.33	0.83	1.63	1.63	1.63
Underload [CU]	0.00	2.78	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Orders	0101	0201	0301	0401	0501	0601	0701	0801	0901	1001	1101	1201	1301	1401
	0102	0202	0302	0402	0502	0602	0702	0802	0902	1002	1102	1202	1302	1402
	0103	0203	0303	0403	0503	0603	0703	0803	0903	1003	1103	1203	1303	1403
	0104	0204	0304	0404	0504	0604	0704	0804	0904	1004	1104	1204	1304	1404
	0105	0205	0305	0405	0505	0605	0705	0805	0905	1005	1105	1205	1305	1405
	0106	0206	0306	0406	0506	0606	0706	0806	0906	1006	1106	1206	1306	1406
	0107	0207	0307	0407	0507	0607	0707	0807	0907	1007	1107	1207	1307	1407
	0108	0208	0308	0408	0508	0608	0708	0808	0908	1008	1108	1208	1308	1408
	0109	0209	0309	0409	0509	0609	0709	0809	0909	1009	1109	1209	1309	1409
	0110	0210	0310	0410	0510	0610	0710	0810	0910	1010	1110	1210	1310	1410
No. of orders	10	10	10	10	10	10	10	10	10	10	10	10	10	10

Table 3. Rescheduled sets of orders in the balancing interval ($\alpha=0.1$; $\beta=.05$)

Periods	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
Overload [CU]	0.05	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.01
Underload [CU]	0.00	0.00	0.09	0.01	0.04	0.08	0.00	0.03	0.04	0.00
Orders	0101	0106	0206	0303	0404	0602	0401	0803	0908	0702
	0102	0107	0209	0402	0409	0603	0510	0804	0909	0802
	0103	0201	0301	0403	0502	0604	0703	0806	1002	0810
	0104	0202	0302	0406	0503	0605	0704	0808	1005	0902
	0105	0203	0304	0407	0504	0606	0705	0809	1006	0907
	0108	0205	0305	0408	0507	0607	0708	1107	1202	1003
	0109	0207	0306	0410	0508	0608	0709	1110	1206	1010
	0110	0208	0307	0706	0509	0609	0710	1210	1302	1104
	0204	0210	0308	0807	0801	0610	1001		1310	1106
		0505	0309			1007	1102			1410
		0506	0310							
No. of orders	9	11	11	9	9	10	10	8	9	10

Table 4. The capacity requirement of the operation type sets for the rescheduled sets of orders in the balancing interval

Operation type sets	S_1	S_2	S_3	S_4	S_5	S_6	S_7
Capacity upper bound (u_k)	3	4	2	5	4	4	5
Capacity lower bound (l_k)	1	1	0	3	1	2	5
P1	1.28	1.79	1.98	3.07	3.26	3.77	$5.05^{(+0.05)}$
P2	2.17	1.02	1.81	3.19	3.98	2.83	5.00
P3	1.66	1.29	1.96	$2.95_{(-0.05)}$	3.62	3.25	$4.91_{(-0.09)}$
P4	1.81	1.20	1.98	3.01	3.79	3.18	$4.99_{(-0.01)}$
P5	1.65	1.40	1.91	3.05	3.56	3.31	$4.96_{(-0.04)}$
P6	1.34	1.58	2.00	$2.92_{(-0.08)}$	3.34	3.58	$4.92_{(-0.08)}$
P7	1.71	1.29	$2.03^{(+0.03)}$	3.00	3.74	3.32	$5.03^{(+0.03)}$
P8	1.64	1.54	1.79	3.18	3.43	3.33	$4.97_{(-0.03)}$
P9	1.99	$0.99_{(-0.01)}$	1.98	$2.98_{(-0.02)}$	3.97	2.97	$4.96_{(-0.04)}$
P10	1.58	1.48	1.95	3.06	3.53	3.43	$5.01^{(+0.01)}$

M1
ot_1, ot_2, ot_3

M2
ot_2, ot_3

M4
ot_2

M3
ot_1, ot_2

M5
ot_1

Figure 1. Sample manufacturing system with the allocation of the operation types

Capacity [CUs]

$S_1 = \{ot_1\}$

$S_2 = \{ot_2\}$

$S_3 = \{ot_3\}$

$S_4 = \{ot_1, ot_2\}$

$S_5 = \{ot_1, ot_3\}$

$S_6 = \{ot_2, ot_3\}$

$S_7 = \{ot_1, ot_2, ot_3\}$

Operation type set

S_1 S_2 S_3 S_4 S_5 S_6 S_7

•••••• Upper capacity bounds
━ ━ Lower capacity bounds
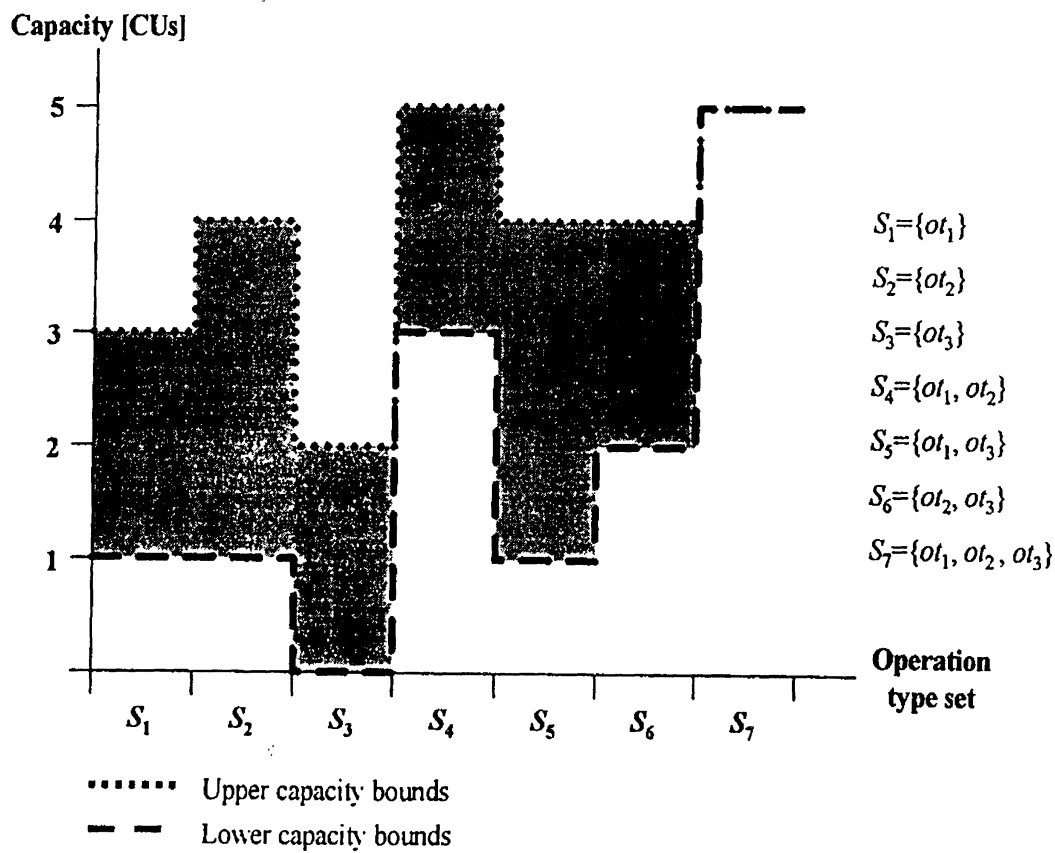
Figure 2. Ranges of ideal capacity for the operation type sets
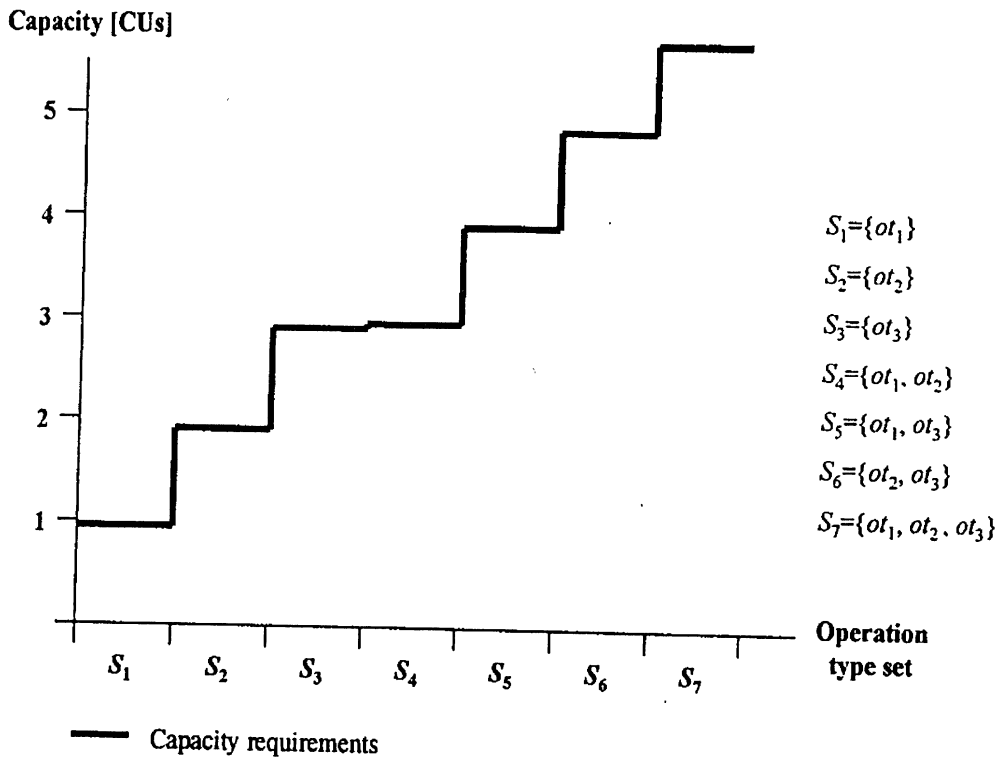
Capacity [CUs]

5 —
4 —
3 —
2 —
1 —

S_1 S_2 S_3 S_4 S_5 S_6 S_7

Operation type set

$S_1=\{ot_1\}$

$S_2=\{ot_2\}$

$S_3=\{ot_3\}$

$S_4=\{ot_1, ot_2\}$

$S_5=\{ot_1, ot_3\}$

$S_6=\{ot_2, ot_3\}$

$S_7=\{ot_1, ot_2, ot_3\}$

—— Capacity requirements

Figure 3. Capacity requirements for the operation type sets

Capacity [CUs]

6 —
5 —
4 —
3 —
2 —
1 —

S_1 S_2 S_3 S_4 S_5 S_6 S_7

Operation type set

$S_1=\{ot_1\}$

$S_2=\{ot_2\}$

$S_3=\{ot_3\}$

$S_4=\{ot_1, ot_2\}$

$S_5=\{ot_1, ot_3\}$

$S_6=\{ot_2, ot_3\}$

$S_7=\{ot_1, ot_2, ot_3\}$

▪▪▪▪▪▪▪▪ Upper capacity bounds   ‑‑‑‑‑‑ Increased upper capacity bounds

▬ ▬ ▪ Lower capacity bounds   ‑‑‑‑‑‑‑ Decreased lower capacity bounds
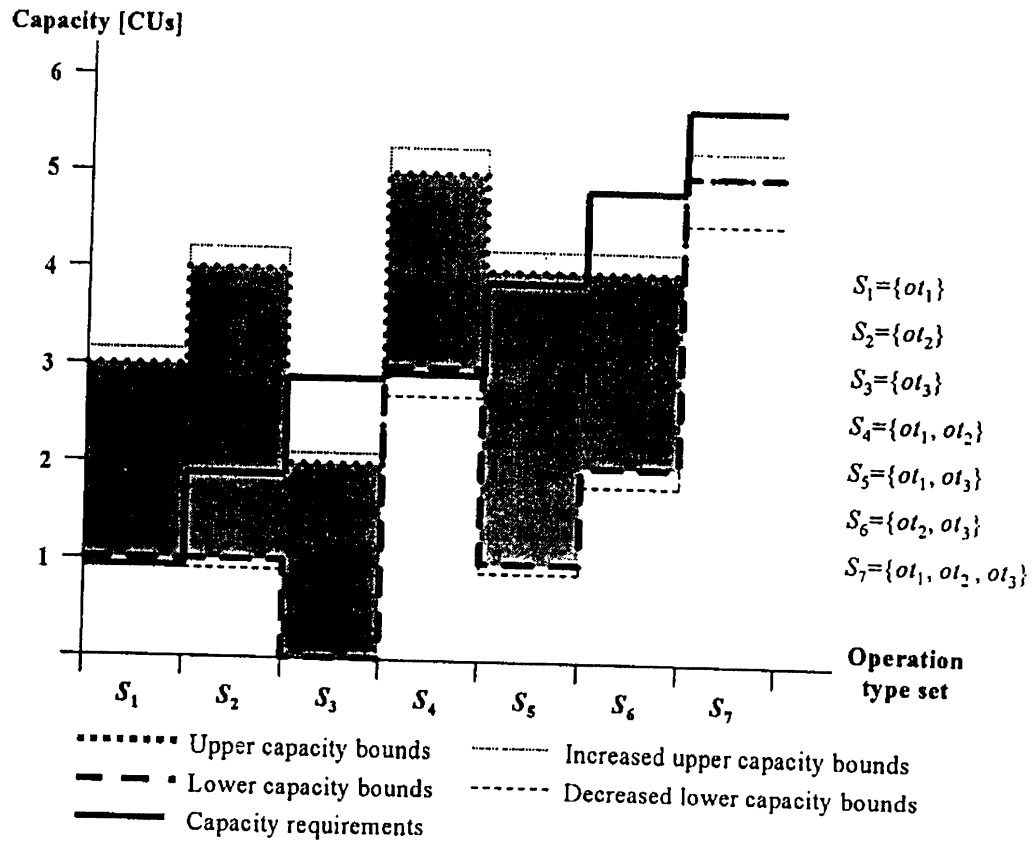
—— Capacity requirements

Figure 4. Ranges of ideal available capacity and augmented available capacity with the production requirements for the operation type sets (α=0.1; β=0.05)
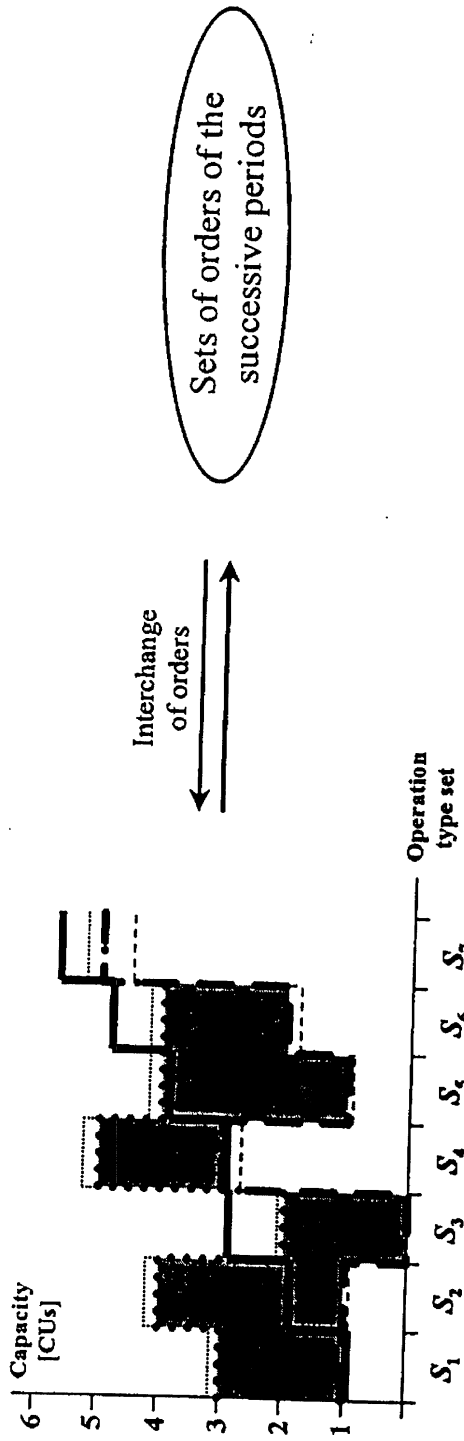
Sets of orders of the successive periods

Interchange of orders

Capacity [CUs]

Figure 5/a. Production requirements in P1 before balancing

S_1 S_2 S_3 S_4 S_5 S_6 S_7

Operation type set

Capacity [CUs]
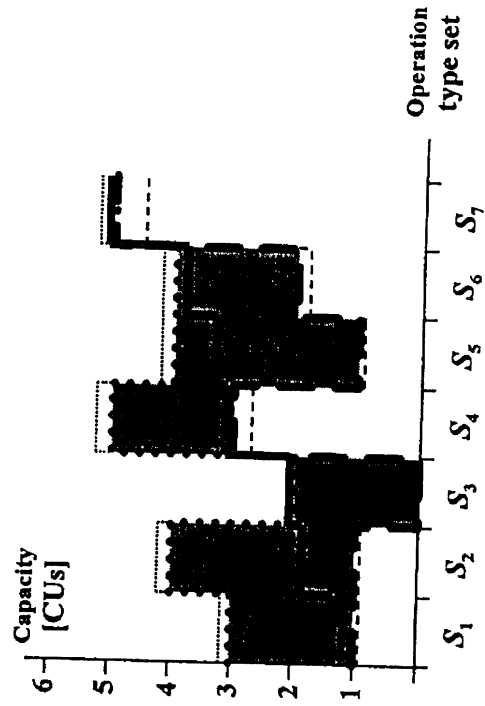
Figure 5/b. Production requirements in P1 after balancing

S_1 S_2 S_3 S_4 S_5 S_6 S_7

Operation type set

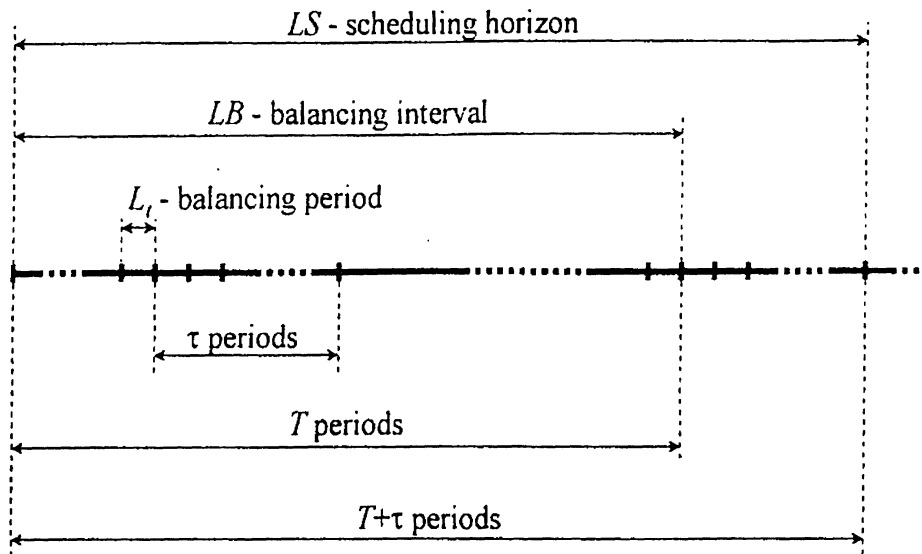Figure 5. Illustration of order rescheduling of the workload balancing procedure
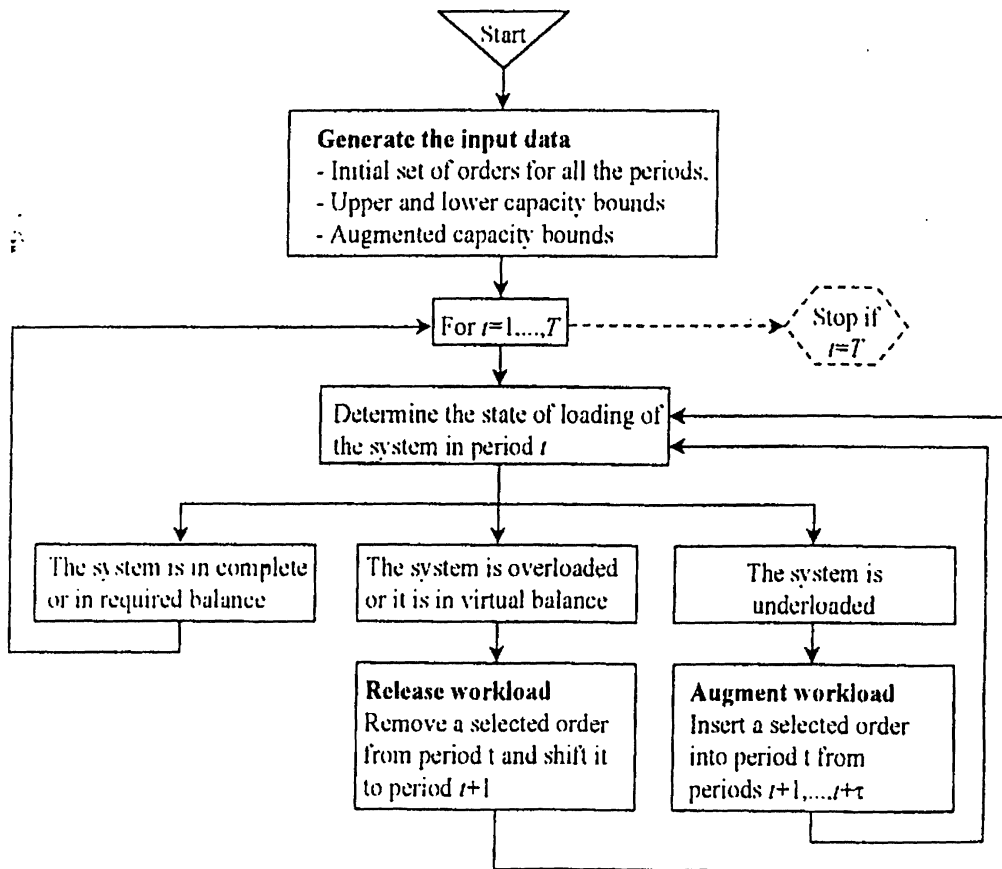
Figure 6. Time scaling of the balancing procedure

Figure 7. The simplified flowchart of the balancing algorithm