

A SYSTEM FOR MANAGING JOB-SHOP PRODUCTION

STANLEY REITER*

I. INTRODUCTION

THIS paper describes a system for planning, scheduling, and controlling production and related activities in a job-shop manufacturing facility producing to customers' orders. This system has been in use for some time in a large job shop producing gears. The basic objective and achievement of the system is to obtain in a complex job shop a degree of efficiency, coherence, and control of operations and business decisions comparable to what is possible in a well-managed line-production facility.¹

The management system presented here differs in several fundamental ways from the current practice of job-shop management. The most important differences derive from the fact that this system computes operational shop sched-

* Professor of economics and mathematics, Graduate School of Industrial Administration, Purdue University.

¹ I am pleased to be able to acknowledge the contributions of two people who were associated with me in installing this system. William K. Holstein of the Harvard Business School, then at the Krauert Graduate School of Industrial Administration, participated in the development of the system that eventually was put into operation from an early time in the undertaking, and he developed the original FORTRAN programs that were used to test the logic of the various parts of the system and that were a basis for the programs ultimately used in actual operations. In that connection he made numerous valuable contributions and in many ways helped the work go forward. Peter E. Dress, a graduate student at Purdue University, wrote most of the machine language programs used by the system now operating and also wrote some of the test programs. His knowledge of computing and programming also made a substantial contribution to the success of our work.

ules which prescribe the time (to the minute if desired) of starting and finishing every operation to be performed in the shop, taking full account of the availability of equipment and of all the work to be done. I know of no operating job-shop scheduling system that attempts this task; almost all systems currently in operation are based on machine- or shop-loading methods. Such methods do not produce feasible operational shop schedules. Therefore, the shop schedule is in practice determined by the decisions of various production supervisors or foremen, or in some cases by the workers themselves. These decisions are made more or less independently of one another on the basis of the information available to each foreman and in view of his own skills and objectives. The resulting shop schedules are not necessarily those that the management would prefer.

A second difference derives from the fact that this system calculates dates of delivery in such a way as to reflect the effect in time units of the amount of congestion or interference in the shop due to the presence of other orders and to the work-force allocation decisions affecting the capacities of work centers. These things are discussed in more detail below.

II. THE JOB SHOP'S PROBLEMS AND THE CURRENT APPLICATION

The main problems of managing a job shop derive naturally enough from the nature of job-shop production. A job shop can be thought of as a collection of

specific skills and equipment which stands ready to sell its services to customers on order. These skills and equipment are applied to a variety of different products. If the volume of production of a particular item should become large enough and persist for a long enough time to warrant creating a production facility to produce only that item, it would cease to be produced in a job shop. Variety both of items and orders is therefore an essential element of job-shop production. However, the extent to which a given business must deal with variety can be influenced by the strategic posture taken by the business. Thus, the firm might seek to reduce dependence on one customer or on one industry, with the result that the number and variety of its orders is larger; on the other hand, it might seek to reduce variety by concentrating on fewer items or customers. These strategic decisions are related to the configuration of the shop itself and apply to a relatively long time span, one measured in years rather than in weeks or months. Here the problem is, given the market for products and inputs and the present shop facilities, to find (approximately) the best mix of products; further, given the possibilities for changing the shop facilities, for example by purchase of machinery, to adapt the shop to the market so that the resulting changes in the best mix constitute good or perhaps "best" long-term improvement in the position of the business.

Given these strategic decisions, the actual flow of orders in time at best can only approximate the desired "ideal" mix. While the management can exercise some control over the mix being produced at each time, there is a fundamental variation which is uncontrollable and to which the management must adapt. The shop facilities therefore must be of such

a nature that it is possible to cope with this variation in product mix. These adaptations may involve the use of multipurpose machines; they usually also involve the ability to shift capacity from one part of the shop to another, mainly by changing the disposition of the work force. This may be done by adding to or decreasing the number of shifts worked in a certain part of the shop or by varying overtime work. But these changes generally cannot be made with high frequency. They usually are made for several weeks or even months at a time rather than for a day or a week. Thus, a second problem of managing a job shop is, given the basic strategic posture of the shop, to adapt to medium-term variations in the variety and timing of orders by changing the allocation of the work force in the shop.

A job shop sells to its customers not only an item made to desired specifications and standards but also an item delivered at a specified time. Two aspects of this delivery time are important: first, the length of time between the placing of an order and its delivery, often called "lead time," and second, the reliability of delivery at the requested or agreed time, often called "meeting due dates." Short lead time is valuable to customers because it postpones the day of commitment to decisions; reliability of delivery at the promised time is of value to customers because it enables them to plan and to carry out plans based on those delivery dates. Again, the strategic posture taken by the job-shop management can emphasize or diminish the importance of these considerations, but in each case they are a more or less important dimension of what the job-shop firm sells.

The time it takes to produce an order depends on the configuration of the shop

and its work force at the time the order is being produced, on the collection of other orders being processed during the time this order is in the shop, and on the myriad of detailed decisions to do the things required to process all these orders with one particular timing rather than another equally possible one. The particular timing of every operation carried out in the shop constitutes the shop schedule. The shop schedule not only determines the lead time of every order in the shop but, since it also determines the sequence of operations on each machine, it also determines setup costs. Moreover, by controlling the timing of production at each stage of processing, the shop schedule also determines in-process inventory costs. Here it is clear that the time scale appropriate to shop schedules is one of hours or minutes.

The shop schedule also influences the time at which each order is completed and ready to be delivered. Therefore, to achieve reliable delivery at the agreed date of delivery it is desirable that the shop schedule take effective account of the promises that have been made to customers. It is also desirable that those promises be made on the basis of information available about what the shop is likely to do as scheduled at the time the order in question will be produced.

Thus, a job shop stands ready to produce a fluctuating variety of items specified technically, in quantity and in time of delivery. Its fundamental problem is to plan, schedule, and control the use of its resources to perform that task efficiently. But the work to be scheduled and the success in carrying out the work as planned and scheduled are strongly related to basic strategic decisions affecting the number, variety, and type of customers and suppliers, and the mix of products desired; to medium-term deci-

sions adapting the shop to the flow of orders anticipated in the next few weeks or months; and to the detailed decisions governing the hour-by-hour activities in the shop.

These problems may be stated more compactly in outline form as follows:

- i) To determine the long-term strategic posture of the shop in relation to the market; in particular, to determine the target product mix and the corresponding configuration of shop capacities;
- ii) To plan and control the timing of production in detail in the shop so as to achieve efficient production, in particular so as to result in short lead times and low setup and in-process inventory costs;
- iii) To negotiate the timing of deliveries with customers on a realistic basis reflecting the presence of other orders, the capabilities of the shop, and the "costs" of achieving a certain timing, as well as the value of the timing to the customer;
- iv) To plan the configuration of the shop, in particular the allocation of man power within the shop, so as to perform the required work efficiently;
- v) To negotiate deliveries from suppliers on the basis of a consistent production plan taking account of all the orders to be produced and of shop capabilities;
- vi) To schedule and control other pre-production activities, such as engineering work; to co-ordinate these activities so as to carry out production as planned;
- vii) To perform these scheduling and planning operations on the basis of information coming to the firm irregularly through time, allowing for the uncertainty about the future and for the occurrence of uncontrolled and unpredictable disturbances in the shop, on the part of suppliers and customers, and indeed in the scheduling system itself.

At this point, it may be helpful to give a brief description of the main features of the shop in which a version of this system is currently being used.

The firm using this system manufactures high-quality gears in a shop comprising about one thousand machines.

The company produces gears ranging in size from small gears with diameters of less than 1 inch to large gears about 4 feet in diameter. Some gears are on shafts, some are internal rings, some external, etc. Virtually every type of gear and tooth geometry is made, and close specifications may apply to metallurgical characteristics as well as geometrical characteristics of the gears. Most gears are made from forgings purchased from outside suppliers or provided by the customers. A relatively small fraction is made from bar stock.

Gear making may involve from seven to fifty operations, done on many different machines. Roughly, these operations clean, size and shape, and heat-treat the forging or blank from which the gear will be made; they cut and smooth the teeth of the gear; and they heat-treat the finished gear to give it the desired metallurgical characteristics.

Orders vary from one piece to five thousand, with many orders in the range of fifty to three hundred pieces. The company has a relatively large number of customers and also has a shifting mix of orders, even in periods when the total volume of work remains fairly constant.

This company produces items of relatively high value for customers to whom technical quality and reliable and early delivery are important. These facts, together with the strategic decisions determining the company's posture vis-à-vis its customers, have influenced certain features of the scheduling algorithm used. These particular features may not be useful in cases where the strategic basis of the business is quite different.

III. SUMMARY OF THE SYSTEM AND ITS PERFORMANCE

The system described in this paper is designed to perform the functions de-

scribed in the preceding section. The various parts of the system may be related to the problems outlined by listing them with a corresponding numbering. Detailed discussion is reserved for the next section of the paper; not every program or report in the system appears in the following outline.

i) The problem of long-run product mix is largely a matter of management strategy calling for decisions only at long time intervals when basic market conditions change significantly. While these choices involve important considerations not dealt with in the computer programs of this system, the system does contribute two pieces of information. First, the record of load and delay reports forms part of the basis for decisions to expand or reduce shop capacity in certain directions. Second, a linear-programming formulation of the product-mix problem can be solved to provide rough but useful information for some of these decisions. This formulation is a variant of the standard linear-programming product-mix problem and so is not given explicitly in this paper.

ii) The system produces operational shop schedules showing the start and finish time of every operation to be performed in the shop during the following two schedule periods. These schedules embody *line-mode* scheduling and *lot-streaming* (concepts discussed in detail below, pp. 389 and 383, respectively) in a way that reflects comparison of the value of time in relation to other costs. In cases where it is worthwhile to do so, operations are sequenced so as to reduce setup costs. By means of a dynamic priority scheme based on relative slack, shop operations are co-ordinated with commitments to customers in the light of the totality of work to be done and resources available. A job-progress report

alerts management to jobs that for any reason are not likely to meet their promised delivery schedules. The systematic report of operations completed and related reports and procedures form part of the production-control system.

iii) The planning and promise-date system routinely calculates promised dates of delivery for each job based on the routing and processing times for the operations constituting the job, the congestion caused by other jobs in relation to resources in the work centers through which this job will pass at the times it is expected to pass there, and the delivery schedules requested by customers and promised by suppliers of materials and tools. It is also possible in the case of a prospective order to give a tentative delivery date within minutes, even if the order is for a part on which the detailed engineering work has not yet been done. The sales department at any time can tell an inquiring customer about the status of his order by consulting the order-status file for an order not yet in the shop or the summary file in the case of an order being worked on in the shop.

iv) Management decisions allocating the work force in the shop are based on reports of delays expected in each work center in each week for the following year. These delays are calculated from regularly revised forecasts of the work to be done in each work center in each week. The forecasts are made for each work center on the basis of the latest available information about orders and attempts to rely on the stability of technology and the control system rather than on any assumed constancy of the volume or mix of orders. Estimated delays are periodically revised to show the effect of management decisions as to the number of shifts to be worked in each work center, the amount of overtime

work planned for each work center, and the amount and kinds of labor assigned to each work center.

v) Pre-shop operations, such as engineering work and purchasing of materials, are not formally scheduled. However, the system signals these operations to begin and accepts information about their completion. It also produces two reports relevant to pre-shop operations. The first report is issued only to the purchasing department. It shows a list, for each supplier, of all materials that have been ordered from that supplier and identifies the jobs for which the material will be used. It also shows the adjusted delivery date of those materials by jobs. This report enables the purchasing department to communicate with each supplier with information in hand about all orders placed with that supplier and about the desired timing of delivery for each of those orders.

vi) The second is a report of adjusted completion dates for pre-shop operations for each job. This report shows each department the desired completion date for its work on every job on order but not yet issued to the shop. The adjusted completion date is based on the delivery schedule requested by the customer, the likely dates of delivery of materials from suppliers, and the estimated completion dates of all pre-shop operations carried out by other departments.

vii) From the moment a shop schedule is calculated, unforeseen incidents, such as machine breakdowns, set in motion a drift which carries the actual practice in the shop off schedule. In this system the reporting of shop operations allows a new schedule to be calculated on the basis of the actual state of the shop at the time of scheduling. This procedure makes it very easy for management to intervene in the shop at any time when, for exam-

ple, a sudden change is made by a customer on work already in the shop. In the system now in operation, a shop schedule is calculated every three days for the following six days. If the computer should fail to work during the time a schedule should be computed, the shop can proceed for another three days on the second half of the existing schedule.

The system routinely produces a number of reports not mentioned in this paper. These are mainly aggregated summaries of the condition of various files. Such reports are familiar and useful to management, but I do not feel that their description is essential to an understanding of the system. They would, in any case, vary from one company to another.

Management scientists have abstracted a "job-shop-scheduling problem" from the larger and more complex problem of managing a job shop. Anyone comparing my approach with that of most of the current scientific literature in job-shop-scheduling will see that I have chosen to enlarge the set of available actions in the model and to look for feasible schedules of the enriched kind. This may be contrasted with the effort to find a better approximation to a more narrowly conceived optimal schedule. I believe that the systematic use of line-mode scheduling and lot-streaming results in a greater contribution to payoff than could be expected from improved approximation to an "optimal" schedule that is restricted to batch processing. Experience so far indicates that the system can perform very well indeed.

The several parts of the system have been operating in the company for about a year now, the longest experience being with the shop-scheduler. A careful evaluation of the performance of the system would require either controlled experi-

mental conditions or a relatively long experience in which the uncontrolled parameters affecting the performance of the system could be observed to take a sufficient variety of different values to permit identification of effects. Was the observed experience due to the schedule or to some change in an uncontrolled factor? While we cannot yet have firm scientific evidence, experience with the system so far shows such a sharp break with the past that the effects of the system seem quite clear. What we have observed so far is (a) a reduction in the number of orders not meeting due dates to a value regarded by the management as very near the irreducible minimum; (b) a remarkable reduction in lead time per order, which is equivalent to a very large increase in shop throughput, with no substantial change in the number of machines or of men; (c) greatly increased control of shop operations and significantly smoother day-to-day operations; (d) a considerable reduction in overtime work; and (e) the emergence of routinely available information as a basis for planning such things as outside subcontracting of certain operations, something not ordinarily done by this company.

In general, the problem areas have shifted from the shop to procurement and other pre-shop paper work. The problem used to be to get the product out on time; it is now to get the material and the blue prints and tooling in on time. It is difficult to be more precise without revealing proprietary information.

IV. MAJOR SUBSYSTEMS AND THEIR INTERRELATIONSHIPS

In this section I shall first describe the system in outline by dividing it into major subsystems and describing the relationships of these subsystems to one

another. The whole system is also shown in flow-chart form in Figure 1. The reader may find it helpful to have that chart before him as he reads on. The four major subsystems are: (1) the order-status system; (2) the planning and promise-date system; (3) the shop-scheduling system; including (a) the priority system and (b) the shop-scheduler; and (4) the production- and schedule-control system. Each of these will be described in detail in subsequent sections. Following this I shall describe how the entire system works together in time. This is done in the section called "Dynamics." There is then a brief description of the computer requirements of the system and the experience so far accumulated in computing. Finally, there is a concluding remark on the impact made by the system on the size of the enterprise.

Order-status system.—When an order is received, it is identified with a number and entered into the order-status file. Before the item can be manufactured, the company must decide how to make it. That is, a routing must be determined for it and blueprints made for it; tools must be ordered, made, or verified to be on hand; and materials must be obtained. Each department or group involved in these pre-shop operations receives notification of the order and is required to make two responses. First, possibly after communicating with outside suppliers, it responds with an estimate of the date by which it expects its work to be completed or delivered. These estimated dates are recorded in the order-status file. Second, it responds when the work is actually completed, for example, when the material on order has been received. This information is also recorded in the order-status file. When all necessary pre-shop operations have been completed, the order is ready to be issued to the

shop. An order remains on the order-status file until it has been issued to the shop, at which time it enters the summary file, which contains all orders issued to the shop.

Planning and promise-date system (PPDS).—An order on the order-status file enters the PPDS as soon as it has a routing (and alternative routings, if any), estimates of the completion times of all pre-shop operations, and a schedule of release dates requested by the customer. The PPDS does the following:

1. It decides or provides information and presents alternatives for a management decision on how many sublots the order shall be divided into.

2. It chooses the routing to be used

3. It decides whether a production lot shall be "streamed" in the shop, which includes deciding at which operations streaming is to occur and into how many sublots.²

4. (a) It calculates a promised delivery date called "promise date" for each production lot. (b) It calculates by how many days the estimated completion time of each pre-shop operation is early or late in relation to the release dates (dates at which the customer desires or permits delivery) requested by the customer. Two reports are prepared. One report presents the information by order number, the second by the vendor, that is, for each outside source of supply for materials, tools, etc., the report shows the number of days each order from that supplier can be delayed without endangering meeting the promise date or should be advanced in order to meet that date. This report is made the basis of communication with suppliers, thus avoiding piecemeal alterations in suppliers' delivery schedules which may advance de-

² See p. 383 for a detailed discussion of "lot streaming."

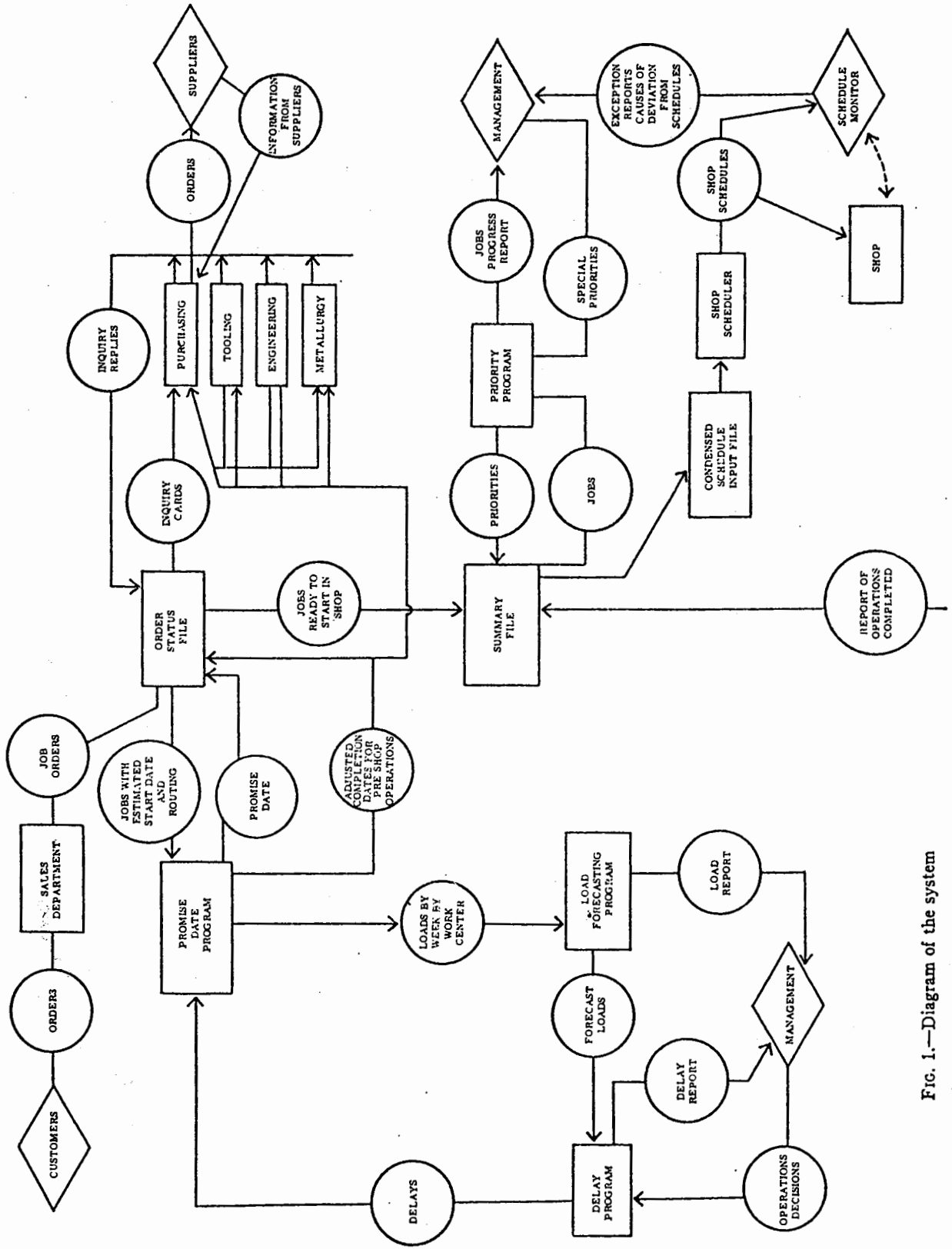


FIG. 1.—Diagram of the system

livery of one order at the expense of inadvertently delaying delivery of another and perhaps more urgent one. This report also can be combined with inventory information in the case of materials that can be used for several different items or orders.

5. It calculates each production lot's contribution to the load in each work center in the weeks during which the job will appear in the work center. This contribution is measured both in units of number of jobs and in hours of processing required.

6. It uses the information available about future work loads to bring up to date a forecast of loads in each work center in each week for 52 weeks in the future. Based on these forecasts and on certain decisions about the allocation of

enter the shop for processing. It then leaves the order-status file and enters the shop-scheduling system for the first time via the summary file and remains there until completed and shipped.

The first step in this system consists of making a revised routing decision based on the current estimate of the current state of the shop. This uses the vector of delays currently in force. Second, if the look-ahead feature (see p. 387) is used, the relevant quantities are calculated and attached to the operations. Those lot-streaming decisions not already made are made now. Finally, the priority indicators are calculated. There is a choice of priority functions to be made here among several possibilities. The system now in operation uses relative slack S as a priority indicator.

$$S = \frac{\text{Promise Date} - \text{Adjusted Processing Time Remaining}}{\text{Promise Date} - \text{Current Date}}$$

labor among work centers, it calculates the expected waiting time or "delay" of orders in each work center in each future week. These delays are a representation of the state of the entire shop. The state of the shop is thus given by a matrix of delays, the dimension of which is the number of weeks by the number of work centers—in our case, 52×230 .

7. It produces periodic reports showing the anticipated state of the shop for several periods in the future. These reports are a basis for: (a) planning the allocation of labor among work centers and (b) evaluating prospective orders, given the anticipated state of the shop. This evaluation can be used to direct sales effort and in connection with pricing in some circumstances.

Priority subsystem or slack system.—When all pre-shop operations on an order are reported to have been completed, the order is actually ready to

All operations to be scheduled are then listed by jobs in order of decreasing priority in the condensed schedule input file. The condensed schedule input file is the basic input to the shop-scheduler.

Shop-scheduler.—The shop-scheduler produces a schedule for each work center showing the start and finish times, to the nearest hour, of each operation to be processed in that work center. It also shows the scheduled finish time of the preceding operation and the work center in which that operation is carried out, and the scheduled start time of the next operation and the work center in which that operation is to be carried out.

A typical page from a shop schedule used in operation is shown in Figure 2. The first column identifies the work center for which this is the schedule. The second column lists the part number, identifying the item.³ The third column

³ The part numbers shown in Fig. 2 have been changed.

WORK PART CNTR NUMBER	SHOP ORDER	SEQ DPER LOT NO. CODE SIZE P U	C S	START TIME	FINISH TIME	PREV FINISH + WORK CNTR	NEXT START + WORK CNTR	ORDERED QUANTITY	CUSTOMER CODE
2235 1079	273202	12 0078 49		1TH 08	1TH 12		2483 1TH 12	107	042
2235 1838	528901	15 0078 117		1TH 11	1F 04		2645 1TH 20	117	118
2235 1164	304001	10 0024 111		1TH 14	2M 00	2333 1TH 14	2235 2M 02	111	060
2235 5055	556301	06 0050 28		1F 06	1F 08	2654 1TH 14	2235 1F 10	28	043
2235 5055	556301	07 0052 28		1F 10	1F 12		2645 1F 15	28	043
2235 9220	519801	10 0028 47		1F 14	2M 03	2351 1F 08	2341 2M 06	47	085
2235 1164	304001	11 0027 111		2M 02	2M 10		2342 2M 07	111	060
2235 3608 FORGING- 36081	513901	05 0034 25 ANALYSIS-		2M 05	2M 06	2633 1F 11	2510 2M 10	25	028
2235 325719 FORGING- 325719	693101	10 0042 105 ANALYSIS- 1045		2M 08	2M 14	2522 1F 22	2645 2M 15	105	183
2235 9220	519801	12 0050 47		2M 12	2M 15	2341 2M 09	2359 2TU 05	47	085
2235 2224	570401	10 0024 100		2TU 00	2TU 11	2333 2M 03	2342 2TU 05	100	074
2235 3623	609501	10 0050 26		2TU 00	2TU 00	2236 2M 12	1611 2TU 04	26	043
2235 2224	570402	08 0050 97		2TU 02	2TU 13	2234 2TU 02	2359 2TU 10	101	074
2235 2710 FORGING- 2774C	611701	09 0073 20 ANALYSIS- 8620		2TU 13	2TU 15	2522 1F 03	2641 NT SCD	21	069
2235 1164	304001	14 0032 111		2TU 15	2M 12	2342 2M 01	2342 NT SCD	111	060
2235 5012	568901	32 0131 25		2W 00	2W 05	3053 1TH 14	1614 NT SCD	26	087

Fig. 2.—Sample page from shop schedule

gives the shop order number identifying the production lot. The fourth column gives the operation sequence number, for example, the thirteenth operation in the routing for this job. The next column gives an operation code that identifies for the foreman or operator the operation to be performed, for example, drill and tap holes. The next information given is the number of pieces in this production lot or subplot. The next two items identify special information connected with pricing. Then follows the start time of this operation, its finish time, the number of the work center from which this lot is coming, and the scheduled finish time of the last piece in that work center; next comes the number of the work center to which this lot goes next and its scheduled start time there. Comparison of the various start and finish times tells the foreman of this work center whether he is part of a line with either the preceding or the following work center or both. Then follows the quantity ordered, which may differ from the lot size, and a code identifying the customer. Some part numbers have been changed to preserve proprietary information.

The shop-scheduler uses line-mode processing whenever that is advantageous; it schedules the sublots to be streamed as called for by earlier decision, or by a built-in decision criterion, and schedules desirable sequences of setups in those work centers in which such scheduling is desired. In work centers where significant setup advantages are relatively rare it provides for foremen to take advantage of such opportunities for significant setup savings when they arise without incurring the costs of scheduling them.

Along with the shop-scheduling system goes the production- and schedule-control system. There is first of all a report-

ing system by means of which all operations completed or partially completed are reported at certain specified times. Second, although the shop-scheduler employs a dynamic priority system, that is, one in which the priority attached to a production lot may change from time to time, some jobs may fall too far behind to allow even the highest priority to get them through in time to meet the promise date. To meet this possibility the system produces a job-progress report that identifies those jobs in danger of not meeting their promised dates of completion, even allowing for dynamic adjustment of priorities. Finally, the schedule-monitor subsystem samples work-center performance to compare actual operations with scheduled operations and reports large departures from the schedule.⁴ These are then investigated in detail and, if necessary, reported to management.

We turn now to a detailed description of each subsystem.

Planning and promise-date system.—The planning and promise-date system is based on a statistical view of the shop as scheduled by the shop-scheduler. Such a view rather than one that leads to detailed scheduling is appropriate here, because this system deals with relatively far future times, about which the best available information is inherently uncertain.

In order to know when a job will be finished it is desirable to know how long it will take to perform all the work to be done on it. This in turn involves knowing how long it will spend waiting to be worked on. We use a "delay," the expected waiting time generated by a statistical model of the work center as an estimate of waiting time.

⁴ See p. 391 for a detailed description of the schedule-monitor system.

In order to plan the disposition of the work force, it is desirable to know the amount of work there will be at each work center. However, the number of setups and the amount of processing is the same no matter how many men are used to perform that work. Therefore, the effect of varying the amount of labor devoted to a work center is to change the time required to perform the work, thereby affecting waiting time of jobs passing through that work center during the period of time in question. Thus, decisions to allocate more or less labor to a work center during a certain period of time, or to operate fewer or more shifts, turn on the effect of such decisions on lead time and in-process inventory cost. This effect can be measured by the "delay." It is perhaps worth pointing out that work load measured in processing hours to be performed or in jobs is the usual basis for these decisions. Neither processing hours nor jobs, however, is a very good measure of work load because the relationship between labor input and waiting time is highly non-linear. The burden of estimating the effect of varying the amount of labor is left to the intuition of the manager when his decision is based on load measured in processing hours or similar units. This may or may not lead to wise decisions, but it is in any case difficult and may require several iterations of trial-and-error decisions to arrive at a suitable result.

Certain other decisions, for example, the choice of a routing among alternatives, likewise depend on comparison of time costs and values with other costs, and so information about waiting times can be quite useful in making them.

At this point the system has on file for each work center: (a) a table of expected delays for each of the next 52

weeks; (b) an estimate of the fraction of operations in that work center which are scheduled in line with the next operation scheduled; alternatively, this estimate may be expressed as a fraction of processing time in the work center in question; (c) a control value of the total processing time; this value triggers a lot-streaming decision starting at the next work center. For most work centers this control value is infinite, that is, only at a few work centers is this decision ever made. These items of information are used to make several types of decisions.

*Lot-size decision.*⁵—Certain orders may require or permit partial delivery at several dates. If an order for three thousand pieces is to be delivered in sublots of one thousand pieces each at March 1, June 1, and September 1, there is an opportunity for splitting the total order into sublots and therefore of shifting production in time to some extent. To make this decision, a program first generates certain possible combinations of lots that permit meeting required release dates; for example, we can produce: Pattern I, one lot of three thousand pieces by March 1; Pattern II, one lot of one thousand pieces by March 1 and one lot of two thousand pieces by June 1; Pattern III, one lot of one thousand by March 1, one of one thousand by June 1, and one of one thousand by September 1. (Still further subdivision is possible, but a limit, depending on proximity in time and set-up costs, is certainly reached. We consider only these three possibilities here.) Calculate:

1. Incremental interest costs on work in process; for example, Pattern I requires holding one thousand units for 3 months and one

⁵ When production for inventory is involved, this decision is more complicated, and the approach taken here must be modified.

thousand units for 6 months as compared to Pattern II, which requires holding one thousand units for an extra 3 months, and Pattern III, which incurs no extra holding costs.

2. Additional setup costs; for example, Pattern I requires no additional setup; Pattern II, one additional setup per operation; Pattern III, two additional setups per operation, with appropriate adjustments for streamed sublots.
3. Finally, we need a measure of the contribution to the delay of other orders made by each pattern of production. This can be given by: (a) the sum of delays in each work center in the weeks in question for each pattern and (b) the incremental change in delays due to that pattern. Both (a) and (b) are needed because the effect of load on delays is strongly non-linear.

There are two types of decision procedures for lot size:

- a) Delays may be evaluated and converted into time costs by means of evaluation coefficients applying those delays to all contemporary orders.
- b) Control limits can be set on delays for certain work centers and certain preferred patterns used unless some control limit is exceeded. Management attention is then called to situations in which a non-routine decision is called for.

Choice of routing.—The routings for each order are arranged in decreasing order of preference. Using the estimated start time, the promise-date program yields: (1) a total elapsed time for each routing, (2) delays in work centers on the routing. One of the following criteria may be used: (a) minimum elapsed time; (b) control limits on delays in work centers; that is, the next preferred alternate routing is chosen if it avoids work centers where the delay is larger than its control limit; (c) some combination of (a) and (b). Choice of routing may be revised at any shop-schedule cutoff time, if desired, on the basis of current delays. Control limits on delays may be used to indicate

cases in which subcontracting of operations can be advantageous.

Lot-streaming.—Certain operations may have extremely long total processing times and be followed by operations with short processing times. In such cases line scheduling will not occur; yet it may be advantageous to shorten the time-to-completion by processing the job in sublots accumulated as the long-running operation completes pieces. This we call “streaming the lot ahead” or “lot-streaming.” The time advantage in time-to-completion is offset, of course, by the additional setup costs incurred. Alternatively, if tooling is available, two or more machines may be used simultaneously in the long-running operation instead of lot-streaming.

Lot-streaming is built into the planning and promise-date system as follows. Each work center is given a critical total processing time. If the total processing time (TPT) on an operation exceeds its critical value k , the lot is split into $TPT/k + 1$ sublots,⁶ each of which is scheduled separately in succeeding operations. The start time in the next operation of a subplot is the finish time of the last piece in that subplot in the current operation. Sublots may be further streamed at subsequent operations.

Promise dates.—Having decided the number and size of each production lot and its routing, and having built in the decision criteria for streaming at each work center, a promise date for each lot is calculated as follows.

The estimated start time of the order is given. This becomes the finish time of the zeroth operation. An adjusted lot size is computed by means of a function giving the fraction of operations that is likely to be scheduled in line mode.

⁶ The symbol $[A]$ denotes the largest integer less than A .

Where an assembly is involved, estimated start time for the assembly is the latest completion time of the items to be assembled.

The completion time in any work center is the finish time in the preceding work center plus transport time, plus a delay, plus setup time, plus total processing time calculated using the adjusted lot size.

Where the customer's release date is later than the promise date calculated, notice is given that the start date could be advanced by the amount of the difference, and the effect on the timing of delivery of the necessary materials, tooling, and blueprints is calculated and reported.

As each job is scheduled by the promise-date program, the calculated arrival of that job at each work center is recorded in the appropriate week, and the total processing hours are also recorded in the appropriate weeks. Thus, scheduled loads, measured both in number of jobs and in hours, accumulate by weeks by work center. Sometimes estimated start dates or desired delivery dates of jobs change, leading to the necessity to calculate a new promise date. If available computer storage permits these components to be identified with a job number, rescheduling of jobs can be done "with unloading"; that is, the contribution of that job, as originally scheduled to load in each work center, is canceled when the job is rescheduled. If not, then rescheduling is done "without unloading," relying on the statistical stability of the system to avoid gross errors. The promise date is firm in any case once a commitment has been made to the customer.

Delays.—Delays are calculated for each work center for each week for some number of weeks, say 52, into the future. These delays are calculated using a sta-

tistical queuing model of the work center. In the system in operation we use a single-phase multichannel queuing model with Poisson arrival distribution, exponential holding time, and first-come, first-served queue discipline. Work centers are treated as statistically independent. In certain cases this model is capable of introducing systematic distortions in the distribution of waiting times.⁷ The main error is that the variance may be understated. But this is not excessively troublesome to the system and is compensated for by the simplicity of the model. Compensating adjustments in parameter values can easily be made to reduce the error to a level acceptable for the purpose at hand.

The queuing model requires three inputs for each week and work center, an arrival rate, a clearing rate, and the number of servers.

The arrival rate is based on the number of jobs scheduled to arrive in that work center in that week. The clearing rate is based on the number of hours of processing together with the number of processing hours produced per clock hour by the work center. (This is called "work center efficiency.") The clearing rate also depends on the number of shifts operated. The number of servers is the number of machines.

These parameters are estimated from forecast loads as scheduled by the promise-date program. This is described on p. 385 under "Load-Forecasting."

Delay and load reports.—The delay for a given work center is a function of two time variables. There is a delay corresponding to the time point at which the

⁷ See Roy D. Harris, "An Empirical Investigation and Model Proposal of a Jobshop-like Queuing System" (Working Paper No. 84 [Los Angeles: Western Management Science Institute, University of California]). Also of interest is the work of J. R. Jackson referred to there.

delay is calculated and the week to which the delay refers. Thus,

$$\delta_{\omega}(\tau, t)$$

is the delay for work center ω in week t , estimated at time τ .

For each ω , the program can at time τ report $\delta_{\omega}(\tau, t)$ for several values of t . It also can be made to report at τ for a given t , $\delta_{\omega}(\tau', t)$, for several values of $\tau' \leq t$.

The program also can be made to call attention only to those work centers where (a) $\delta_{\omega}(\tau, t)$ falls outside specified limits; (b) $\delta_{\omega}(\tau, t)$ exhibits certain trends either in t or in τ for fixed t .

When decisions are made to allocate labor to a group of work centers or a department rather than to a single work center, the report can be made to show the maximum and minimum delay for work centers in that group or department. It is also possible for management to receive reports of the effect on delays of any of several decisions. These include (a) decisions to change the labor force in a given work center, expressed as a change in "efficiency," and (b) decisions to change the number of shifts operated. These reports involve evaluating one formula, namely, the algebraic expression for expected delay. This may be done conveniently by calculating these delays on the computer. Alternatively, a table or perhaps even a crude graphical display of these effects might be more useful in some circumstances. The manager may also receive load reports directly, that is, hours and number of jobs by work center, by weeks. These are more familiar but may be expected to be less directly useful because of the strongly non-linear effect of load on delay.

Delay reports also can be used in evaluating the effects of prospective orders on the plant, particularly whether

the prospective order brings the mix closer to the mix best suited to the shop's projected configuration. These reports provide a sound basis from which to negotiate delivery dates of prospective orders in cases where time does not permit calculating them routinely.

Delay reports give the anticipated state of the shop in future weeks and thus are useful when management is considering whether to make price concessions based on quantity or timing or other considerations.

Load-forecasting.—The model used to estimate work-center delays requires forecasts of two quantities: (1) the number of jobs that will arrive in the work center during each week for the next 52 weeks; (2) the number of hours of processing required in order to clear those jobs. These quantities are forecast from data accumulated in the load files of the promise-date program. Recall that these files contain for each work center and each future week the number of jobs scheduled to arrive in that work center during that week. This file also contains the number of hours of processing contributed by those jobs to the work-center load during the weeks they are scheduled to be processed. I shall describe in detail the method of forecasting the load in jobs; a strictly analogous procedure is used to forecast the load in hours.

Let $N_j^{(k)}$ denote the number of jobs actually arriving in work center k during week j , and let $N_{ij}^{(k)}$, $i \leq j$ denote the number of jobs accumulated prior to week i and scheduled to arrive in work center k during week j .

Now, $N_{jj}^{(k)}$ is an estimate of $N_j^{(k)}$, based on the planning and promise-date schedule, and it is in a certain sense the best estimate the system produces on the basis of information routinely generated. However, we need estimates of $N_j^{(k)}$ at

earlier times than j . Hence, we consider the problem of estimating from $N_{ij}^{(k)}$. That is, we use the load scheduled prior to week i for work center k in week j as a basis for our estimate of the load $N_{jj}^{(k)}$ which will be scheduled at week j for week j and indirectly as an estimate of the actual load $N_j^{(k)}$ in week j .

Define

$$\pi_{ij}^{(k)} = \frac{N_{ij}^{(k)}}{N_{jj}^{(k)}}, \quad N_{ij}^{(k)} \neq 0.$$

I assume that

$$\pi_{ij}^{(k)} = \pi_{j-i}^{(k)}; \quad i \leq j,$$

which is to say that the fraction of all jobs ultimately scheduled to arrive in work center k in week j which have already been scheduled by week i depends only on the lag, $j - i$.

These functions $\pi_{ij}^{(k)}$, considered as functions of i for fixed j or under the assumption made as functions of the lag $\alpha = j - i$, have the following properties:

1. $\alpha < \beta$ implies $\pi_{\alpha}^{(k)} \geq \pi_{\beta}^{(k)}$.
2. $\lim_{\alpha \rightarrow \infty} \pi_{\alpha}^{(k)} = 0$.
3. There is a number $\alpha \geq 0$ such that $\pi_{\alpha}^{(k)} = 1$; in particular, $\pi_0^{(k)} = 1$.

If the function $\pi_{\alpha}^{(k)}$ is known, the obvious estimate $N_{jj}^{*(k)}$ of $N_{jj}^{(k)}$ from $N_{ij}^{(k)}$ is:

$$N_{jj}^{*(k)} = \frac{1}{\pi_{j-i}^{(k)}} \cdot N_{ij}^{(k)}$$

for i and j such that $\pi_{j-i}^{(k)} \neq 0$.

This estimate is self-correcting in the sense that, as $i \rightarrow j$, $N_{ij}^{(k)} \rightarrow N_{jj}^{(k)}$, and $\pi_{j-i}^{(k)} \rightarrow 1$, so that $N_{jj}^{*(k)} \rightarrow N_{jj}^{(k)}$. That is, as we get closer to the week j our estimate of the load scheduled for week j converges to the load scheduled at all times prior to week j for week j .

Estimates $\pi_{\alpha}^{*(k)}$ of the functions $\pi_{\alpha}^{(k)}$ are obtained by using the promise-date scheduling program with suitable delays,

together with a suitably chosen arrival process for jobs, to simulate the scheduled work-center arrivals $N_{ij}^{(k)}$.

In the system now in operation we used a Poisson arrival process for orders. We have observed the time pattern of lags $\pi_{ij}^{(k)}$ to be stable.

The sources of error in these forecasts derive from the effect of variations in the volume and mix of orders on work-center delays and on the distribution of routings of jobs arriving in a given work center. Changes in the volume and mix of orders have a direct effect on the scheduled arrivals $N_{ij}^{(k)}$ and hence on the forecast. The effect on delays is mitigated by the fact that management acts so as to keep delays close to desired values and therefore, to the extent that the system is in stable control, to keep the variation of delays within small bounds. The effect of variation in the mix of orders, apart from its action via delays, depends to some extent on the type of work center considered. Broadly speaking, we can describe the degree of specialization of a work center by the subset of routings that pass through that work center. A work center associated with a small set of parts, for example, spiral-gear cutters, will not be affected by variations in mix that do not involve spiral gears. At the other extreme, a work center associated with a large set of parts, for example, miscellaneous grinding, will be only little affected by changes in the mix, since the mix of routings involved may be expected to approximate random samples from the population of all routings. In any case, a change in the mix must appear at some point in the form of orders. When those orders are scheduled by the promise-date program and so appear among the $N_{ij}^{(k)}$, the forecast automatically reflects the mix as it actually exists at that time; that is, the forecast reflects changes

in the mix at the earliest time they are actually observable in the system.

Priorities.—All jobs issued or ready to be issued to the shop, and therefore listed in the summary file, are given priorities. The operations not yet completed are listed; those operations in process at cutoff time are identified; and the number of pieces completed is given.

Priorities are assigned as follows. The relative slack S is computed for each job:

$$S = \frac{\text{Promise Date} - \text{Adjusted Processing Time Remaining}}{\text{Promise Date} - \text{Current Date}}$$

Total processing time remaining is calculated from an adjusted lot size; it takes account of lot-streaming and line-mode scheduling, so that a non-linear function of lot size may be introduced here. The system now operating uses a function depending on the square root of the lot size.

Each uncompleted job is assigned a relative-slack value. Total remaining delay (the sum of delays in all work centers through which the remaining operations of the job pass) is then attached to the first unperformed operation of each job.

The total delay applicable to any unscheduled operation can be calculated by subtracting work-center delays for scheduled operations from the job's total delay as operations are scheduled.

Scheduling priority then can be any well-defined function (including lexicographic-ordering principles) of job sequence (reflecting relative-slack priority) and total remaining delay (reflecting congestion to be found down the line in the routing). Use of total remaining delay to determine priorities involves "looking ahead" in the routing. We call this possibility the "look-ahead feature."

In the system actually in operation, scheduling priority is determined by rela-

tive slack alone. The rest of this discussion assumes that priority is determined in this particular way. A control limit on the value of relative slack is used to select those jobs whose slack is too low. These jobs are listed in the job-progress report as jobs unlikely to meet their promise dates in the ordinary course of events.

Shop-scheduler.—For each order of job whose production is to be scheduled, a

unique routing is chosen. This routing lists operations to be performed in specified *work centers* (a work center is a group of similar machines) in a prescribed order. Thus, a particular operation cannot be performed until all preceding operations have been performed. We note three types of precedence relations.

First is the "normal" case in which the first operation of a job precedes the second and so on to the end of the job.

Second, there is a "mating-part" precedence. This case arises in gear manufacturing when two parts must be made to fit one another to close tolerances, for example, a gear and pinion that must run together or "mate." Here care is taken to cut the teeth of the pinion to fit the teeth of the gear. This means that before the teeth of the pinion can be cut we must have one sample gear whose teeth have been cut and subjected to the distortions of heat-treating.

The third case is called "assembly" precedence. In an assembly operation two or more parts produced as different jobs must come together in a single operation. Hence, preceding operations from two or more jobs must be finished before the assembly operation can be carried

out. It is also possible for a single operation to immediately precede several different operations.

Precedence relations are represented as follows. Operations to be scheduled are numbered in sequence, starting with the first operation of the first job to be scheduled and ending with the last operation of the last job to be scheduled. (The order in which jobs are arranged is discussed below.) We associate with each operation an integer, giving the number of operations immediately preceding. Thus, in the normal case, this number will be 1, except for the first operation of a job, in which case it will be 0. In the case of an assembly of r parts, the number will be r . Mating-part precedence is indicated by a tag. Where the number of preceding operations is greater than 1, the operation sequence number of each exceptional preceding operation also is given. Similarly, the number and identity of all succeeding operations are given.

Whenever an operation is scheduled, its precedence variable and that of every immediately succeeding operation are reduced by one. At any stage of scheduling, all operations whose precedence indicator is zero—and only such operations—are ready to be scheduled: "ready to go."⁸

The summary file, containing all operations of each job issued or ready to be issued to the shop, together with its relative slack, is first stripped of all completed operations and then sorted by relative slack, those with high priority coming first. If a lot is streamed, all sublots appear in sequence with the same relative slack as the original lot. These

⁸ This procedure is based directly on that of T. Heller and G. Logemann, "An Algorithm for Construction and Evaluation of Feasible Schedules," *Management Science*, Vol. VIII, No. 2 (1962).

remaining operations are numbered sequentially from the first uncompleted operation of the first job to the last operation of the last job and are listed, omitting information irrelevant to scheduling, on the condensed schedule input file.

The initial reference time for the schedule is the final time at which we receive reports of operations completed or running in the shop, called "cut time."

For partially completed operations the report shows the number of pieces completed at cut time. These operations are identified, and the number of pieces remaining to be processed is recorded. All partially completed operations are scheduled first with due allowance for the number of pieces already processed at cut time using the calculations explained below.

All operations whose precedence index is zero are collected in a so-called ready-to-go vector. Each such operation has an earliest possible start time ϵ calculated as follows. For simplicity of exposition I use the case of a normal operation, i , having only one immediately preceding operation, $i - 1$.

Suppose the lot to be processed in operation i has $N(i)$ pieces and that $N(i - 1) = N(i) \equiv N$.

The following dictionary of symbols is useful.

- i = Operation sequence number
- $t_1(i)$ = Finish time of first piece in operation i
- $T(i)$ = Finish time of last piece in operation (i)
- $N(i)$ = Number of pieces in lot (i)
- $\tau(\omega)$ = Earliest time at which machine ω is available
- $\tau\{W(i)\}$ = $\text{Min } \tau(\omega) =$ earliest time at which a machine $\omega \in W(i)$ is available in the work center $W(i)$ specified for operation i
- $\epsilon(i)$ = Earliest potential start time of

operation i , defined for operations that are ready to go

$p(i)$ = Processing time per piece in operation i

$\sigma(k,i)$ = Setup time required for operation i to follow operation k on the same machine

$s(i)$ = Scheduled start time of operation i

$T(i) \equiv S(i)$ = Scheduled completion time of operation i

$\pi(i)$ = Precedence index associated with operation i

$L(j)$ = Slack associated with (every operation of) job j

Assume for simplicity that all setup and transport times are zero, and let $x(i)$ be the smallest non-negative number such that

$$x(i) + t_1(i - 1) - T(i - 1) + p(i)(N - 1) \geq 0. \quad (1)$$

Note that (1) can be rewritten as

$$x(i) + p(i)(N - 1) \geq p(i - 1)(N - 1).$$

Then

$$\epsilon(i) = \max\{t_1(i - 1) + x(i), \tau[W(i)]\}. \quad (2)$$

Among all ready-to-go operations we select the *first* operation whose earliest possible start time is a minimum. (See comments on the look-ahead feature in priorities for possible variants of this rule.) Because of sequencing jobs in order of increasing relative slack, this in effect gives priority to those operations belonging to jobs with the smallest relative slack. But we do not keep a machine idle to wait for a higher priority job not yet ready to go if another job can start earlier. Then

$$s(i) = \epsilon(i) \quad (3)$$

and

$$T(i) = s(i) + p(i)N(i) \quad (4)$$

(ignoring, as we said, setup and transport times).

The smallest numbered machine in work center $W(i)$, say ω , now has $\tau(\omega) = T(i)$. And the precedence index of i and of every operation of which i is an immediate predecessor is decreased by one. Finally, the earliest possible start time of each operation now in the ready-to-go vector is revised to take account of the finish times $t_1(i)$ and $T(i)$ of the newly scheduled operation i .

In the case in which setup time is a function only of the operation i —that is, $\sigma(k,i) = \sigma(i)$ or, more specifically, a constant for the work center in which operation i is performed—the setup and transport times contribute additive terms as follows: Let the transport time be $R(i)$ and the setup time $\sigma(i)$; then expression (1) becomes

$$x(i) + t_1(i - 1) + R(i) + \sigma(i) + p(i)(N - 1) - T(i - 1) \geq 0, \quad (1a)$$

and

$$\epsilon(i) = \max\{t_1(i - 1) + R(i) + x(i) + \sigma(i), \tau[W(i)] + \sigma(i)\}. \quad (2a)$$

Here we require that both the machine and one piece from the lot must be available in order to carry out the setup. An alternative requiring only that the machine be available to be set up, possibly in advance of the receipt of the first piece for operation i , would result in

$$\epsilon(i) \max\{t_1(i - 1) + R(i) + x(i), \tau[W(i)] + \sigma(i)\}. \quad (2b)$$

Sequencing of setups.—Where it is desirable to take account of sequencing of setups, the following modified procedure is used.

Suppose the machines of work center $W(i)$ can be in one of a finite number of states $c = 1, \dots, C$. These states correspond to setup conditions. Thus, if oper-

ation k is scheduled to be carried out in machine $\omega(k)$, it determines both the earliest availability of that machine, namely, $T(k)$, and the state of that machine at that time, namely, $c(k)$. We also write c_ω to indicate the state of machine ω .

But operation i requires the machine to be in state $c(i)$ and, hence, determines a time $\sigma(k, i)$ needed to convert the machine from $c(k)$ to $c(i)$. Now

$$\tau(\omega) = \min\{\tau(\omega) + \sigma[c_\omega, c(i)]\} \omega \in W(i),$$

and we proceed as before, taking due account of $\sigma[c_\omega, c(i)]$ as our setup time.

Non-machine operations.—Certain operations occur in job-shop manufacturing that are not ordinary machine operations. An ordinary machine operation is one in which each piece or group of pieces being processed occupies the capacity of the machine from the time processing starts on them until it is over. Certain operations do not have this property. A typical example is heat treating in a continuous furnace. Suppose we have a furnace fed by a conveyor belt with ten trays or positions. Consider an item that must spend 10 hours in the furnace. Thus, the conveyor advances at the rate of one position per hour. If the first trayload enters the furnace at time 0, remaining there until time 10, the "machine" is *not* occupied until time 10 but, rather, the second trayload can enter at time 1. The time accounting needed to provide for this is an obvious modification of what is done in the case of machines.

Certain other features of scheduling both machine and non-machine operations may be illustrated by the continuous heat-treating furnace.

Suppose each furnace consists of two conveyor belts, each consisting of twelve positions and traveling through three

zones. Temperature is constant in the first two zones but can be varied in the third. Finally, there is a quenching bath in the furnace, which some, but not all, items use. Certain large items require quenching in a special press located near the furnace exit. There is just one press for each furnace, so that only one conveyor of the furnace can be occupied by items requiring press quenching. Further, an untreated (so-called green) piece is needed to set up the quenching press.

The case depth required, the metallurgical characteristics of the pieces being processed, and their size determine how long they must remain in the furnace and also determine the temperature setting in the third zone. Since both conveyors pass through this zone, only jobs requiring the same temperature can be in that zone of the furnace at the same time. Given these complexities, the essence of the scheduling problem is as follows.

If two items require the same time in the furnace, they can follow one another on the same conveyor without loss of time. If, however, the time in the furnace of the second job is different from that of the first, the conveyor speed must be changed. In order to do this without damaging the first job, the conveyor must be partially or completely cleared. For example, if each piece of job 1 must spend 12 hours in the furnace and job 2, 6 hours, then twelve trays of the conveyor belt must move empty through the furnace. The conveyor will move at the rate of one position per hour until the first job clears the furnace and then at the rate of two positions per hour. Thus, 6 hours of furnace time will be wasted. These 6 hours are in the nature of a setup time required to go from job 1 to job 2.

We handle this type of problem in ex-

actly the same way as the scheduling of setups described above.

Each conveyor or track of the furnace is in a certain state, namely, its speed or rate of advance expressed in positions per hour. Each job requires a track in a certain state. The transition time from the existing state to the required state can be calculated and added to the earliest possible start time of the job, as described above for setups, and scheduling proceeds in the same way. This results in schedules with low idle-furnace time. The other complexities of time accounting and the various restrictions on temperature and quenching are easily handled in the obvious way and do not warrant detailed description here. Indeed, several furnaces of different sizes can be scheduled as a single group, allowing for the possibility that some jobs must go into one specified furnace while others may go into one of two or more furnaces. For example, suppose there are three furnaces and that the dimensions of the door of one of the furnaces is too small to admit a certain job. Then it must be scheduled in either of the other two, but not the first.

Production control—schedule control.—It is desirable for management to get information about the progress of jobs through the shop. While the scheduling algorithm gives effect to changes in priority resulting from the passage of time and from the performance of work, it may happen that some jobs fall behind or, indeed, start out too far behind. It also may happen that schedules are not followed closely enough in work centers. We consider two types of control reports serving two different purposes.

1. *Schedule monitor.*—Random samples are taken from reports of work performed and are compared with scheduled operations. They may be taken either by

a production controller who examines a work center during a short period of time or, if the shop reporting system is sufficiently precise to include timing or sequencing, by using the written shop reports themselves.

2. *Job-progress report—an exception report.*—Consider the distribution of slack in a particular work center or group of work centers. By relating shipping dates to promise dates and to slack associated with a job at the time it was in this work center, it is possible to establish critical values for slack rather like statistical control limits. Then any job whose slack falls outside the critical limits has a high likelihood of failing to meet its promise date, despite the priority arrangements in scheduling. All such jobs are listed and reported as exceptions.

V. DYNAMICS

In this section I shall describe the way the whole system works in time.

At a given time the system has:

1. An order-status file containing orders received but not yet ready to be issued to the shop. Some of these orders have promise dates, some do not because they lack either a routing or an estimated date of completion of one or more pre-shop operations.

2. A summary file containing all jobs issued or ready to be issued to the shop, including completed operations, together with the current relative slack for each job.

3. A matrix of delays, one for each work center in each week for, say, the next 52 weeks.

4. A matrix whose entries are the number of jobs for which promise dates have been calculated according to the promise-date program estimated to arrive in each work center in each of the next 52 weeks.

5. A matrix whose entries are the number of processing hours by work center by week required by the jobs for which promise dates have been calculated.

6. Jobs corresponding to those listed in the summary file at various stages of processing in the shop.

7. Newly arrived orders.

Each day:

1. Newly arrived orders are entered in the order-status file, and inquiry cards are sent to the various departments engaged in pre-shop operations.

2. Inquiry cards returned from pre-shop operations are used to update the order-status file.

3. The promise-date program is used to calculate promise dates for all orders not already having promise dates and for which the requisite information is available. The delays currently in force are used. The jobs are added to the jobs and hours totals in the appropriate weeks in the respective matrices.

4. Promise dates are entered in the order-status file.

5. Reports are issued to various departments performing pre-shop operations showing allowable variation in dates of completion of those operations.

6. Reports of operations completed in the shop are entered in the summary file.

7. If the day is one on which a new shop schedule is calculated, the number of pieces completed on operations in process at cut time is reported from the shop.

Every third day:

1. Relative slack is recalculated for every job in the summary file. Management decisions to give some jobs extraordinarily high priority are given effect by entering a fictitious promise date for the job in question.

2. The summary file is sorted according to the newly calculated relative slack, and a new condensed schedule input file is prepared, which also contains the operations in process at cut time.

3. The job-progress report is issued.

4. A shop schedule is prepared for the next six days, starting with completion of the scheduled operations in process at cut time. If there should be a breakdown in the computer, the shop thus could continue to operate for one more schedule period using the old schedule.

Every 2 weeks:

1. Accumulated loads in jobs and hours are used to forecast anticipated loads in each work center in each of the next 52 weeks. These forecast loads, together with the current decisions regarding work-force allocation and the

number of shifts operated, are used to calculate revised delay for each work center in each of the next 52 weeks.

2. All information referring to past weeks is dropped from the load files but may be retained for a time in another form to provide a historical record.

3. The current estimates of future delays are entered as the delay parameters in the promise-date program. For the next two weeks all promise dates will be calculated on the basis of the revised delays.

4. Delay reports are issued to management to be used to help guide work-force allocations and other longer-term operations decisions in the future. Some changes in work-force allocation are likely to be planned more than 2 weeks in advance.

We may note that this system adjusts to the inevitable deviations from the shop schedule by picking up the actual state of the shop at cut time and creating a new schedule starting from the existing condition rather than from the scheduled terminal conditions. Thus the shop cannot get "further off schedule" than can happen in 3 days. In practice, in this shop the drift off schedule has not been much. In other shops a longer or shorter period between schedules might prove to be desirable.

VI. COMPUTING

A version of the system described here has been programmed for an IBM 1440 system with two 1311 disk drives. The various files used in the system are maintained on disk packs. At the present time the programs are maintained on punched cards. The reporting system from the shop is essentially a hand system requiring key punching of the necessary information. Various devices have been used to achieve the degree of accuracy necessary to operate the system properly. It would be possible to use a more highly automated data-collection system for reporting, but the cost of such a

system is not warranted at present in this particular application.

The shop-scheduler typically works with a condensed schedule input file containing about fifteen thousand operations to be considered for scheduling. This is very close to the storage capacity (one hundred thousand locations) of the present computer system. From 14 to 18 hours are required to generate a single shop schedule making use of eleven different program phases. The critical elements of the computer system are disk storage capacity, which determines the number of operations that can be considered, and disk access speed, which determines running time of the program. If a disk system with high speed should be available, or in the case of a small shop in which the available internal core storage would be adequate, the running time would be bound by the internal speed at which core can be searched. Tape systems are impractically slow for this program.

At present the system is being reprogrammed for an IBM 360/30 system. Other than the shop-scheduler, the pro-

grams of the system do not tax the capacities of existing computers in any unusual way.

VII. CONCLUDING REMARKS

The obvious benefits afforded by this system are to increase the efficiency of utilization of the shop, to reduce lead time, to plan well for and meet the flow of work to be done, and to make realistic delivery promises to customers and keep them. However, there is one effect of the system which may be less obvious but not less important. For each type of shop and type of business there is a size beyond which the shop becomes unmanageable using the usual methods. When job shops exceed that size, they have chronic problems symptomatic of the difficulty. With an effective computerized management system like this one, the size of shop that a given management can control is significantly larger than its former limiting size. This enables the technical skills and special knowledge available to the enterprise to be used effectively and profitably on a larger scale than hitherto has been possible in job-shop production.