Controlling different flow rates in job-shop like production departments

International Journal of Production Economics, 23 (1991) 239–249
Elsevier

Controlling different flow rates in job-shop like production departments

H.P.G. van Ooijen

Eindhoven University of Technology, Faculty of Industrial Engineering and Management Science, Eindhoven, The Netherlands

Abstract

We consider production departments where it is desirable that the production orders of different product types have different flow rates, independent of their production characteristics like processing times. Since flow rates are mainly determined by the work centre waiting times the product types should have different production order waiting times at the work centres. In this paper we derive a method which makes this possible in a controlled way: given certain values for some parameters the different waiting times (and thus the flow rates) are predictable. By means of simulation this method has been tested for two production situations where, per work centre, only two different flow rates are required. In the first situation both (categories of) products have the same average routing length, whereas in the second situation the average routing length for both (categories of) products differ. This simulation study shows that, using a simple balance equation and operation due date sequencing, it is possible to create different predictable flow rates. It turns out that up to a required waiting time reduction for one of the two (categories of) products of about 60% a one-to-one relation approximately exists between the scheduled (required) waiting time reduction and the, what will be called, normalized waiting time reduction.

1. Introduction

In many production situations the production departments have to live with varying routings, different batch sizes, different processing times and varying capacity utilizations at the various work centres in the department. The capacity of the production department should be used to a reasonable degree, say up to 90% of the bottleneck equipments. This leads to relatively high throughput times, often 5 up to 10 times the total processing time of a work order. Planning and scheduling techniques and priority rules can be used to reduce the average waiting times, but such reductions are limited by the requirement that the deliveries of work orders should also be reliable, that is according to the work order due dates. As has often been demonstrated, scheduling rules which achieve relatively short throughput times often perform poorly with respect to delivery reliability (see Baker [1], Conway et al. [2]).

The consequence is that in job-shop like production systems, a high capacity utilization and a short production throughput time cannot be achieved simultaneously. However, if we have a closer look at production situations in practice, we see that production departments where short throughput times are required for all products are rare. In most job-shop like production situations we can see that the work order throughput time requirements are different for different (categories of) products. For example:

- For products with a relatively low average demand, production may be to customer order and not to stock. To keep the customer order lead time within the market requirements it may be necessary to produce these products with a much shorter work order throughput time than required for the other products, which are made to stock.
- For products which use expensive materials or components it may be advantageous to keep stocks and work-in-process relatively low as compared to products which use relatively cheap materials. The work orders for the expensive

products therefore should have short work order throughput times.

- For components which are required in multi-component products and which are manufactured on a make-to-order basis the *allowed* work order throughput times are the same for all components needed in the same assembly product. However these components may strongly differ in terms of the required number of manufacturing operations. Thus the items with many operations should be produced at a higher production speed or flow rate (number of operations performed per period) than the items with only a few operations.

From these examples we conclude that a production department generally wants to realize at the same time a high capacity utilization (economic considerations) and short work order throughput times for *selected* products. The policy to achieve this is to use the items for which long work order throughput times are allowed to avoid capacity idle time, whereas the items with short work order throughput times are used to realize a high customer performance.

This policy is the subject of this paper. A limitation thereby is that we require that the throughput times are reliable (small variance in lateness). We will restrict this investigation to the following two production situations:

- The different (categories of) products have the same production characteristics; they only may have different arrival rates.
- The different (categories of) products have a different average routing length and many have different arrival rates; the other production characteristics, such as processing time distribution and transition probabilities, are considered to be the same for all products.

Since it will be difficult, it not impossible, to find answers to these questions with mathematical methods (as will be seen we deal with a kind of closed queuing network in which a dynamic priority rule is used), we have used computer simulation to study this problem.

In Section 2 the production situation is described and some potential answers are derived. These answers are checked via computer simulation. The simulation study and the experimental results are presented in Section 3. In Section 4 the results are discussed and finally, in Section 5, the findings are summarized.

2. A method for generating different flow rates

In this paper we consider discrete component manufacturing departments with a functional lay-out and a job-shop routing structure. In a functional lay-out similar machines are grouped into work centres: the job-shop routing structure implies that from each work centre work orders can flow to a number of other work centres.

In such production systems throughput time control is a very important issue because of the interaction between throughput time, capacity utilization and batch sizes. A generally accepted way to *control* the *average* work order throughput time is to control the workload in the shop, which is known as Input/Output planning (see Bechte [3], Bertrand and Wortmann [4] and Kingsman and Tatsiopoulos [5]). By controlling the workload in the shop all products will have the same flow rate, which depends on the average work centre waiting times.

To create different flow rates for different categories of products we need to create different waiting times for the different categories of products. How can we realize this? An obvious way to do this is via operation due date assignment: for work orders which belong to the fast category we assign operation due dates which imply that there is only a little allowance for waiting, whereas for work orders which belong to the slow category we assign operation due dates which imply a large allowance for waiting. Using the operation due date sequencing rule than "forces" the individual work orders to flow at the rates implied by the due date with a small variance in work order lateness (see Kanet and Hayya [6]).

Of course this procedure for setting flow rate norms only makes sense if the actual flow rates are in accordance with the flow rate norms used for setting the due dates. This implies that there is a restriction on the combination of flow rate norms which we will allow. In general terms this restriction can be stated as follows:

Consider a production department with M work centres. Suppose we have k different product categories, that the average routing length of the orders in category i equals L_i, $i=1,...,k$ and that the relative shop arrival rates of the order streams are given by

$$\Lambda_i/\Lambda,\ i=1,...,k$$

where Λ_i is the average number of arrivals of orders for product category i per unit of time and Λ is the total average number of arrivals of orders per unit of time.

Furthermore suppose that the number of orders in the shop is controlled via Input/Output planning, such that the total average waiting time per work centre is equal to a value S_m.

If we denote the work centre arrival rates by λ_i, $i=1,...,k$, then realistic values for the average waiting times per category (which imply the flow rates) obey the following equation:

$$\lambda_1\times S_{m,1}+\lambda_2\times S_{m,2}+\cdots+\lambda_k\times S_{m,k}$$
$$=(\lambda_1+\cdots+\lambda_k)\times S_m \qquad (1)$$

where $S_{m,k}$ is the average waiting time for product category k and work centre m, $m=1,...,M$.

If all transition probabilities are equal then

$$\lambda_1=L_i\times\Lambda_i/(\text{number of work centres})$$

In that case (1) becomes:

$$L_1\times\Lambda_1\times S_{m,1}+L_2\times\Lambda_2\times S_{m,2}+\cdots+L_k\times\Lambda_k$$
$$\times S_{m,k}=(L_1\times\Lambda_1+\cdots+L_k\times\Lambda_k)\times S_m \qquad (2)$$

If we want the actual waiting times to be in accordance with the required flow rates it will be clear that the waiting time norms used in the process of setting the due dates must obey eqn. (1).

Now the remaining question is: If we control the workload in the production department such that the average waiting time norm for work centre m equals the actual total average waiting time S_m and if we use operation due date sequencing with due dates based on waiting time norms which obey eqn. (1).
is then the actual average waiting time per category equal to the waiting time norm, or will there be a difference? If there is a difference, how large is it, and how can it be explained and controlled?

The next section deals with finding an answer to these questions, using the results of a simulation study.

3. Experimental settings and results

As in many other simulation studies on production control, the production model used is a pure job shop (Conway et al. [2]). The shop consists of five unique machines. In this study we considered a production situation with two product categories. For products in category 1 another throughput time is required than for products in category 2. For both product categories processing times are generated from a negative exponential probability density function with a mean value of 1 unit. Processing times larger than 5 are set equal to 5 time units and hence the expected processing time will be 0.993 time units.

In our first set of experiments we used the *same* routing characteristics for both order categories. Job routings are determined upon arrival. They are generated so that each machine has equal probability of being selected as the first machine and that after the first operation the probabilities for the job of either going to each other machine or leaving the shop are 0.2. We will not allow more than 10 operations, so the generating process will be aborted after the tenth operation. Hence the expected number of operations per job equals 4.46. In this case, with the same average routing length for both product categories, relation (2) becomes:

$$\Lambda_1\times S_1+\Lambda_2\times S_2=(\Lambda_1+\Lambda_2)\times S$$

Because all work centres are identical this relation is valid for all work centres.

In the second set of experiments jobs of category 1 were given a maximum value of 10 operations and jobs of category 2 were given a maximum value of 22 operations. Hence L_1, the expected number of operations per job for category 1, equals 4.46 and L_2, the expected number of operations per job for category 2, equals 9.02. For this set of experiments the following relation must hold for all work centres (see (2)):

$$\Lambda_1\times S_1+2\times\Lambda_2\times S_2=(\Lambda_1+2\times\Lambda_2)\times S$$

Although the pure job shop is not the kind of production situation often encountered in practice it is very useful to give insight into the usefulness of our method. If our method turns out to be usable in this kind of situation (which is very complex because of the immense number of routings and variety of processing times) then it must also be possible to use the method in less complex (more realistic) production situations.

The shop is loaded via Input/Output planning using a job release rule based on the total number of orders in the shop: as soon as a job leaves the shop another job is released. The category this new job belongs to depends on the relative arrival rates. We assume that for both categories always a number of jobs (at least one) is present. For most production systems this will not be an unreasonable assumption. The relative arrival rates can then be interpreted as relative entry rates.

For each set of experiments simulations were performed for three average shop utilization rates: ~85%, ~90%, and ~95%. For each utilization rate we used three ratios of the relative arrival rates (category 1 is the *fast* category):

$\Lambda_1 = \Lambda_2$; $\Lambda_1 = \frac{1}{2}\Lambda_2$; $\Lambda_1 = \frac{1}{4}\Lambda_2$

For each combination of utilization rate and ratio of relative arrival rates we gradually reduced the norm waiting time for the fast category from W (the overall norm waiting time) to zero in steps of $0.2\times W$. Using equation (2) we determined realistic norm waiting times for category 2 (the *slow* category).

The results obtained from the simulations are listed in Tables 1, 2 and 3 for $L_1 = L_2$ and in Tables 4, 5 and 6 for $L_1 = 2\times L_2$. The first column in each table gives the percentage of norm waiting time reduction that has been used for determining the norm waiting time for the fast category. For example 40% reduction means that the norm waiting time for the fast category equals $0.60\times W$, here W is the overall FCFS waiting time. The total norm waiting times based on the work centre norm waiting times which were used to calculate the operation due dates for the category 1 production orders are given in the second row.

4. Discussion of results

Because there is a lower bound >0 for the minimum waiting time that can be achieved for category 1 we compared the actual reduction in waiting time to the maximum reduction that is possible: the FCFS waiting time minus the minimum waiting time that can be achieved. The minimum waiting time that can be achieved has been determined by simulation by giving absolute priority to category 1: products in category 1 always have priority over products in category 2.

We adapted the results of the simulations to account for this effect comparing the *actual reduction* in waiting time to the *maximum reduction* that can be obtained. From this the "normalized" waiting time reduction was obtained:

$$100\times\frac{(W_{\text{fcfs}} - W_1)}{(W_{\text{fcfs}} - W_{\text{min}})}$$

TABLE 1

Average shop waiting times for different norm waiting times and different ratios Λ_1/Λ_2; $L_1 = L_2$, utilization rate 84%, prio = absolute priority waiting time for category 1

W_1	Norm waiting time, category 1	$\Lambda_1 = \Lambda_2$		$\Lambda_1 = \frac{1}{2}\Lambda_2$		$\Lambda_1 = \frac{1}{4}\Lambda_2$	
		Category 1	Category 2	Category 1	Category 2	Category 1	Category 2
1.0 W	17.63	17.75	17.89	18.16	18.09	17.65	18.12
0.8 W	14.10	14.82	20.78	15.33	19.44	15.01	18.71
0.6 W	10.58	12.25	23.47	12.51	20.82	12.31	19.27
0.4 W	7.05	9.83	25.88	10.10	22.02	9.82	19.88
0.2 W	3.53	8.04	27.93	8.03	23.12	7.72	20.33
0.0 W	0.00	6.99	29.02	6.57	23.82	6.28	20.68
Prio		5.48		4.56		4.13	

TABLE 2

Average shop waiting times for different norm waiting times and different ratios A_1/A_2; $L_1=L_2$; utilization rate 89%; prio=absolute priority waiting time for category 1

W_1	Norm waiting time, category 1	$A_1=A_2$		$A_1=\frac{1}{2}A_2$		$A_1=\frac{1}{4}A_2$	
		Category 1	Category 2	Category 1	Category 2	Category 1	Category 2
1.0 W	29.52	28.90	29.11	29.37	29.77	29.10	30.13
0.8 W	23.62	24.07	34.52	24.59	32.17	24.65	31.16
0.6 W	17.71	19.24	39.36	19.83	34.50	19.74	32.10
0.4 W	11.81	15.03	43.82	15.37	36.73	15.42	33.21
0.2 W	5.90	11.39	47.56	11.44	38.72	11.48	34.10
0.0 W	0.00	9.02	50.45	8.68	40.25	8.62	34.85
Prio		6.46		5.26		4.77	

TABLE 3

Average shop waiting times for different norm waiting times and different ratio A_1/A_2, $L_1=L_2$, utilization rate 95%; prio=absolute priority waiting time for category 1

W_1	Norm waiting time, category 1	$A_1=A_2$		$A_1=\frac{1}{2}A_2$		$A_1=\frac{1}{4}A_2$	
		Category 1	Category 2	Category 1	Category 2	Category 1	Category 2
1.0 W	54.63	54.31	55.11	54.01	55.19	53.05	55.15
0.8 W	43.70	44.69	54.81	44.49	59.96	43.74	57.57
0.6 W	32.78	35.35	74.13	35.19	64.69	34.84	59.77
0.4 W	21.85	26.39	82.68	26.28	69.27	26.52	61.93
0.2 W	10.93	18.57	90.49	18.15	73.21	18.62	63.84
0.0 W	0.00	13.14	96.43	12.43	76.29	12.41	65.30
Prio		7.48		6.15		5.48	

TABLE 4

Average shop waiting times for different norm waiting times and different ratios A_1/A_2; $L_1=2\times L_2$; utilization rate 85%; prio=absolute priority waiting time for category 1

W_1	Norm waiting time, category 1	$A_1=A_2$		$A_1=\frac{1}{2}A_2$		$A_1=\frac{1}{4}A_2$	
		Category 1	Category 2	Category 1	Category 2	Category 1	Category 2
1.0 W	35.70	35.64	17.26	35.24	17.20	36.20	17.40
0.8 W	28.56	30.01	22.77	29.39	20.01	30.20	18.83
0.6 W	21.42	24.80	27.87	23.77	22.71	24.26	20.19
0.4 W	14.28	20.31	32.64	18.84	25.17	19.01	21.51
0.2 W	7.14	17.03	35.74	14.93	27.14	14.63	22.57
0.0 W	0.00	15.52	37.20	13.16	27.99	11.99	23.25
Prio		14.21		11.48		9.61	

TABLE 5

Average shop waiting times for different norm waiting times and different ratios A_1/A_2, $L_1=2\times L_2$, utilization rate 90%; prio=absolute priority waiting time for category 1

W_1	Norm waiting time, category 1	$A_1=A_2$		$A_1=\frac{1}{2}A_2$		$A_1=\frac{1}{4}A_2$	
		Category 1	Category 2	Category 1	Category 2	Category 1	Category 2
1 0 W	55 21	55 50	27 02	55 12	26 90	56 59	27 09
0 8 W	44 17	46 14	36 26	45 63	31 58	46 76	29 45
0 6 W	33 13	37 35	45 03	36 50	36 21	37 12	31 78
0 4 W	22.08	29 04	52 86	27 72	40 42	27.76	33.90
0.2 W	11.04	22.47	59.46	20 23	44 16	19 66	35.83
0.0 W	0.00	19.00	62.93	16 10	46 10	14 91	37.03
Prio		16 85		12 81		10 53	

TABLE 6

Average shop waiting times for different norm waiting times and different ratios A_1/A_2, $L_1=2\times L_2$; utilization rate 95%, prio=absolute priority waiting time for category 1

W_1	Norm waiting time, category 1	$A_1=A_2$		$A_1=\frac{1}{2}A_2$		$A_1=\frac{1}{4}A_2$	
		Category 1	Category 2	Category 1	Category 2	Category 1	Category 2
1 0 W	122 13	121 63	59 95	124 70	61.04	123.87	60.17
0 8 W	97 70	99 77	82 06	102.38	71 82	101 40	65.61
0 6 W	73 28	77.92	104 12	80 91	82 48	79 45	71 22
0 4 W	48 85	57 46	124 87	60.11	92 85	57 27	76 44
0 2 W	24 43	38 54	144.12	41 24	102.10	37.14	81.65
0 0 W	0 00	25 93	156.73	27 51	109.13	22.24	85.19
Prio		19 89		13 76		11 32	

where

W_a = actual waiting time,

W_{fcfs} = overall FCFS waiting time,

W_{min} = the minimum waiting time that can be achieved.

These normalized waiting times were calculated using Tables 1–6 (see Appendix) and are graphically shown in Figs. 1, 2, and 3 for the first set of experiments and in Figs. 4, 5 and 6 for the second set of experiments.

Although the results differ from what we expected, these differences are rather small. For practical purposes we therefore can conclude that up till a certain value for the norm waiting time reduction (about 60%) there is approximately a one-to-one relation between the *norm* waiting time reduction and the *normalized actual* waiting time reduction. We did not use ANOVA yet so we can not state this as a general conclusion. However, because we are only interested in relative differences (percentages) and we used common random numbers, we think that our conclusion may be generalized. So for practical purposes we state that to attain a reduction of the normalized waiting time of x% ($0\leqslant x\leqslant 60$) we must decrease the norm waiting time of the fast category by x% and increase the norm waiting time for the slow category such that equation (1) holds. Using the operation due date rule then guarantees that the reduction of x% will be obtained in a controlled way (small variance in waiting time) at the order level. A second conclusion, which is

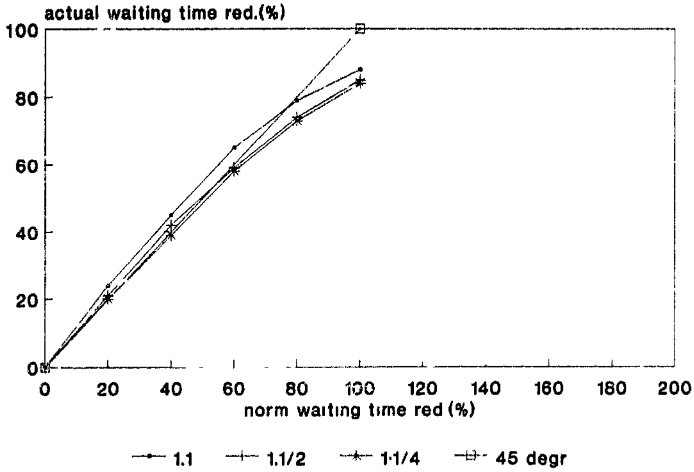

Fig. 1. Actual waiting time reduction as a function of norm waiting time reduction, $L_1=L_2$, utilization rate 84%; (■) $\Lambda_1=\Lambda_2$; (+) $\Lambda_1=\frac{1}{2}\Lambda_2$; (*) $\Lambda_1=\frac{1}{4}\Lambda_2$.

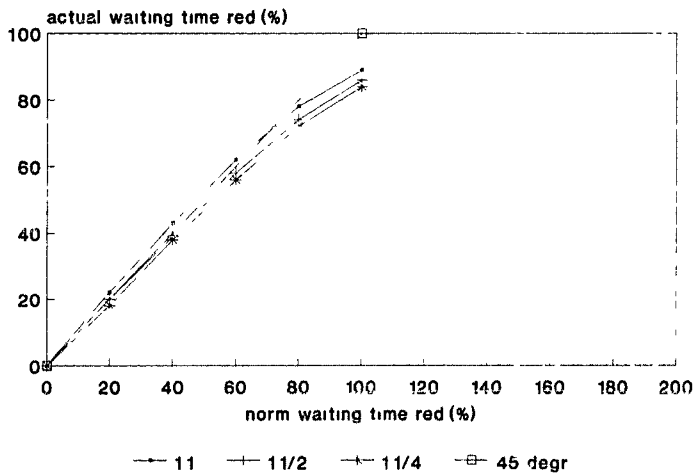

Fig 2 Actual waiting time reduction as a function of norm waiting time reduction, $L_1=L_2$, utilization rate 89% (■) $\Lambda_1=\Lambda_2$, (+) $\Lambda_1=\frac{1}{2}\Lambda_2$, (*) $\Lambda_1=\frac{1}{4}\Lambda_2$.

implicit included in the previous statement, is that utilization rate and ratio of relative arrival rates hardly have any influence. The very small influence of the utilization rate which may be present is important for real life situations: *the higher the utilization rate the "stronger" the one-to-one relation is.*

From the results of the second set of experiments we also can conclude that there is a more or less one-to-one relationship between the

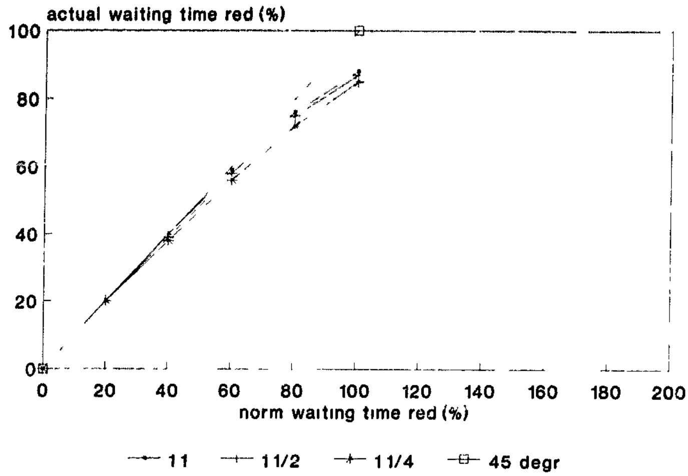

Fig 3 Actual waiting time reduction as a function of norm waiting time reduction, $L_1 = L_2$, utilization rate 95% (■) $A_1 = A_2$, (+) $A_1 = \frac{1}{2}A_2$, (∗) $A_1 = \frac{1}{4}A_2$

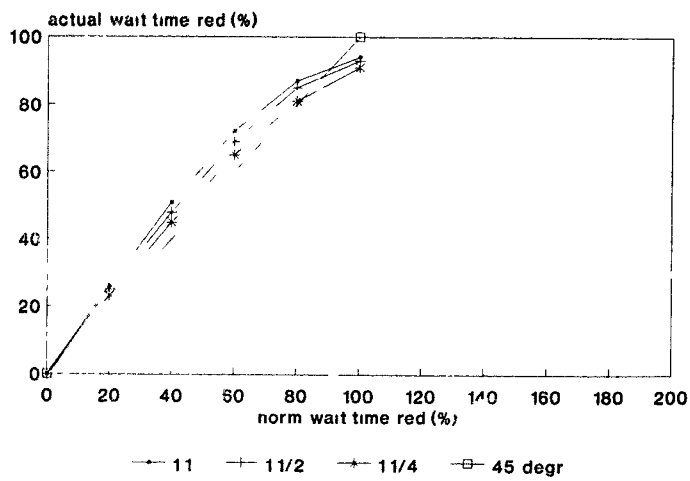

Fig 4 Actual waiting time reduction as a function of norm waiting time reduction; $L_1 = 2 \times L_2$; utilization rate 85%. (■) $A_1 = A_2$, (+) $A_1 = \frac{1}{2}A_2$, (∗) $A_1 = \frac{1}{4}A_2$

scheduled waiting time reduction and the normalized actual reduction. This relationship is the strongest in case we have a utilization rate of 95%. For lower utilization rates we get more reduction of the waiting time as scheduled. Besides that the relationship seems somewhat dependent on the ratio of the relative arrival rates. Compared with the first set of experiments, we only changed the mean routing length, which in this case is influenced by the ratio of the relative arrival rates.

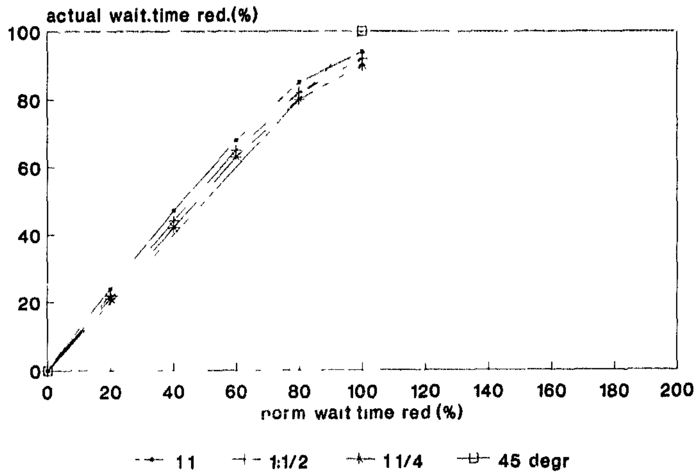

Fig 5. Actual waiting time reduction as a function of norm waiting time reduction, $L_1 = 2\times L_2$, utilization rate 90%. (■) $A_1 = A_2$; (+) $A_1 = \frac{1}{2}A_2$, (*) $A_1 = \frac{1}{4}A_2$

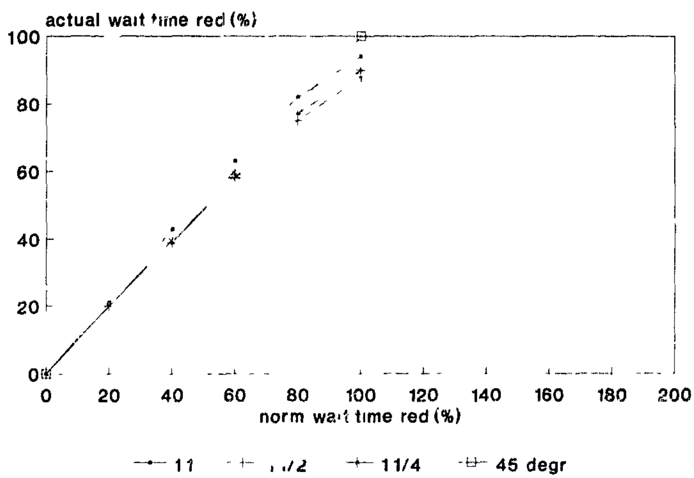

Fig 6 Actual waiting time reduction as a function of norm waiting time reduction, $L_1 = 2\times L_2$, utilization rate 95% (■) $A_1 = A_2$, (+) $A_1 = \frac{1}{2}A_2$, (*) $A_1 = \frac{1}{4}A_2$

Therefore we did some experiments with different mean routing lengths, which holds for both product categories (as in the first set of experiments). The outcome of these experiments confirmed our observations from Table A2: the longer the mean routing length, the more ("normalized") reduction than has been scheduled (via the norm waiting times) we get. Further research is needed to explain and to "control" this phenomenon.

5. Conclusions

The results from the simulation study described in this paper show that it is possible to create in a simple way different *predictable* flow rates for different categories of work orders in production departments. The method requires that the average work order flow rate is under control, which has been achieved via Input/Output planning.

We have shown that the system can be "forced" to realize these flow rates, and therefore different throughput times, by using operation due date sequencing as a priority rule. Important for application in practice is the observation that the validity of the balance equation (1) is especially strong for high utilization rates.

Further we observed in our experiments that, when using different flow rates, there always is a difference between the normative flow rates and the actual flow rates (due to the minimum waiting time which can be achieved). The general observation is that the fast category always lags behind schedule and the slow category will always be ahead of schedule. This is important for scheduling in practice: when scheduled flow times of the orders differ from the average flow times in the system, there always will be a *systematic* difference between scheduled due date and actual due date. This difference however can be calculated and therefore can be accounted for when predicting external due dates. Further research is needed to explain and control the earlier mentioned "routing length" effect and to test the implementation of the principles outlined in this paper in practice.

Appendix

We give an example of the calculations used to derive the values listed in Tables A1 and A2. For a shop utilization rate of 84% and $A_1=A_2$, the total (shop) waiting time if all orders have the same norm waiting time is 17.75 (see Table 1). The total waiting time if all orders of product category 1 always have priority over orders of product category 2 is 5.48. Therefore, the greatest possible reduction in waiting time that can be obtained is

$$17.75-5.48=12.27$$

Using a norm waiting time for product category 1 orders of 60% of the average waiting time, the total actual waiting time for product category 1 orders is 12.25 time units (see Table 1). The actual reduction in total waiting time expressed as a fraction of the greatest possible reduction is thus

$$\frac{(17.75-12.25)}{(17.75-5.48)}=0.448=\sim 0.45$$

TABLE A1

Reduction in waiting time as a percentage of the greatest possible reduction, for all simulated situations in case $L_1=L_2$, (A) $A_1=A_2$, (B) $A_1=\frac{1}{2}A_2$, (C) $A_1=\frac{1}{4}A_2$

Reduction in norm waiting time (%)	Utilization rate (%)								
	84%			89%			95%		
	A	B	C	A	B	C	A	B	C
0%	0	0	0	0	0	0	0	0	0
20%	24	21	20	22	20	18	20	20	20
40%	45	42	39	43	39	38	40	39	38
60%	65	59	58	62	58	56	59	58	56
80%	79	74	73	78	74	72	76	75	72
100%	88	85	84	89	86	84	88	87	85

TABLE A2

Reduction in waiting time as a percentage of the greatest possible reduction, for all simulated situations in case $L_1=2\times L_2$, (a) $A_1=A_2$, (B) $A_1=\frac{1}{2}A_2$, (C) $A_1=\frac{1}{4}A_2$

Reduction in norm waiting time (%)	Utilization rate (%)								
	85%			90%			95%		
	A	B	C	A	B	C	A	B	C
0%	0	0	0	0	0	0	0	0	0
20%	26	25	23	24	22	21	21	20	20
40%	51	48	45	47	44	42	43	39	39
60%	72	69	65	68	65	63	63	58	59
80%	87	85	81	85	82	80	82	75	77
100%	94	93	91	94	92	90	94	88	90

References

1 Baker, K., 1974. Introduction to Sequencing and Scheduling. Wiley, New York

2 Conway, R W , Maxwell, W L. and Miller, L W., 1967. Theory of Scheduling, Addison-Wesley, Reading, MA.
3 Bechte, W , 1982. Controlling manufacturing lead times and work in process inventory by means of load oriented order release. In: APICS Conf Proc
4 Bertrand, J.W M. and Wortmann, J.C , 1981. Production Control and Information Systems for Component Manufacturing Shops Elsevier, Amsterdam
5 Kingsman, B.G. and Tatsiopoulos, I.P., 1983. Lead time management. European J. Oper. Res., 14: 351–358.
6 Kanet, J.J. an Hayya, J.C., 1982. Priority dispatching with operation due dates in job shops. J. Oper. Management, 2(3): 161–175.