

E-POLCA to control multi-product, multi-machine job shops

Nico Vandaele, Diederik Claerhout, Inneke Van Nieuwenhuyse
Faculty of Applied Economics, University of Antwerp, Prinsstraat 13, 2000 Antwerp, Belgium

Rony Cremmery
Dana Corporation, Ten Briele 3, 8200 Brugge, Belgium

We propose a supporting framework for the material control system Paired-cell Overlapping Loops of Cards with Authorization (POLCA), which combines the advantages of push systems and pull systems. In our load based version of the POLCA control system, we rely on a multi-product, multi-machine queueing network to determine release authorizations and allowed workloads. We report on our experiences in a metal shop, taken from Spicer Off-Highway Products Division, part of Dana Corporation. They are implementing an E-POLCA system in a paperless -cardless- environment.

1. Introduction

Intended to manage the required material flows through the available resources, the output of a material control system provides every workstation at each moment with the information what to do. Traditionally, some systems overstate the importance of material flow while other approaches stress the planning of resources too much (Vandaele and De Boeck 2003). For instance, Material Requirements Planning (MRP), a push system, computes schedules of what should be started into the production system based on external demand, bill of materials, stocks, fixed lead times and fixed lot sizes. Once an order is released, each time a workstation finishes its activities the material is pushed to the next workstation not regarding the congestion of the resources downstream. Completed with dispatching rules, which arrange the queues in front of the workstations, this control system dictates the workstations at each moment what to do. However, by applying fixed lead times, the dynamics of limited resource availability is completely ignored. On the other hand, just-in-time production (JIT), with the pull system kanban, focuses on resources. Only when the queue in front of the downstream workstation is below a certain level, which can be understood as free capacity, the upstream workstation is allowed to produce.

Furthermore, JIT requires a careful structure of the production system: implementing capacity buffers, setup reduction, cross-training the workforce, rearranging the plant layout... By using a kanban system every workstation knows at each moment what to do. Because it assumes steady flow rates, the JIT control system fails to manage the required material flow in environments which do not comply with the steadiness: typical make-to-order, customized and environments with large product variety are not well suited.

A better performing material control system for a multi-product, multi-machine job shop where highly customized products are following different routings may be developed. Even when the business environment is characterized by highly variable demand, large product variety, low volumes, short and reliable lead times. On the one hand, releases should be based on external demand, bill of materials, stocks, realistic lead-times and optimal lot sizes. We propose an aggregate MRP methodology taking into account the dynamics of limited resource availability. On the other hand, releases should be based on a WIP cap, a limitation of the work in process (WIP), to embrace the benefits of Pull systems (Hopp and Spearman 2004). We propose workload based releases founded on overlapping loops of cards to consider the upcoming available capacity downstream. In fact, we propose a load based version of the material control system Paired-cell Overlapping Loops of Cards with Authorization (POLCA) (Suri 1998) supported by an Advanced Resources Planning (ARP) system (Vandaele and Lambrecht 2003).

Findings of Jonsson and Mattsson (2003) about the fit between planning environments and manufacturing planning and control methods confirm the power of our approach. The planning environment we are dealing with bears close resemblance to their 'configure to order' environment. Conceptually, they perceived a strong match between the 'configure to order' environment and the detailed material planning method MRP. The best shop floor control method for the 'configure to order' environment seems to be Input/Output control combined with Dispatch lists. A best capacity planning method is lacking. Empirically, most frequently used in the 'configure to order' environment are: Capacity requirements planning to plan capacity; Re-order point systems or MRP for detailed material planning; Infinite capacity scheduling and Dispatch lists to control the shop floor. However, 'configure-to-order' manufacturing is the only environment with significantly less satisfied and significantly more dissatisfied users, compared to other environments. Knowing that ARP answers the capacity planning issue, an improved MRP methodology determines the releases and an extreme type of Input/Output control is generated

by the WIP-cap, our approach should be beneficial to industrial practitioners. A simulation comparing the performance of POLCA, kanban, CONWIP and MRP under selected manufacturing environments was made by Zhou, Luh and Tomastik (2000).

The remainder of the paper is organized as follows. In Section 2 we introduce the POLCA material control system. In Section 3 we satisfy the input requirements using the output of the Advanced Resources Planning system. The latter is expounded in Section 4. Section 5 describes the implementation of our approach in Spicer Off-Highway Products Division Bruges. We draw conclusions in Section 6.

2. The POLCA SYSTEM

In this section, we discuss the POLCA material control system which combines a push and a pull signal. A release on the shop floor is the materialization of a push-signal, which in our implementation is the result of an Advanced Resources Planning System (Vandaele and Lambrecht 2003). The pull-signal is the result of the knowledge of upcoming available capacity. The idea of combining push- and pull signals in a Paired-cell Overlapping Loops of Cards with Authorization (POLCA) control system traces back to Quick Response Manufacturing (QRM) principles (Suri 1998). For additional details concerning POLCA see Suri and Krishnamurthy (2003a and 2003b). In the next discussion about POLCA, we heavily rely on Suri (1998, 243-255).

The POLCA material control system requires that every workstation has a list of production orders, each with its release authorization. For each loop, which joins two successive workstations together, the number of POLCA cards has to be determined. For example, Figure 1 depicts a job shop with 8 workstations and 3 production order flows. A production order K1 follows the routing A - B - C - D, a production order K2 follows the routing E - F - C - H and a production order K3 follows the routing E - F - C - D - G. Then, for the production order flow K1, the number of A/B-, B/C- and C/D- cards has to be determined. This procedure is done for every production order flow and the number of cards represents the available resource capacity in the loop. Note that POLCA cards are not product specific. A free C/D card can be assigned to a production order K1 or K3, a free E/F card can be assigned to a production order K2 or K3 and a free F/C card can be assigned to a production order K2 or K3.

How does POLCA control the material flow? A simple rule regulates the selection of the next production order from the queue. Only when the release authorization date of the produc-

tion order has been crossed *and* a card of the next loop is available, the workstation may start its operation. In our example, only when the release authorization for a production order K1 on workstation C has been passed *and* a C/D- card is available to fix onto the production order K1, workstation C may start manufacturing the production order K1. When a workstation is allowed to start several production orders in the queue, it is supposed to follow the authorization list or to obey a second order selection rule (e.g. to minimize setup). The POLCA card is detached from the production order only when the production order leaves the loop. In our example, the C/D-card fixed onto the production order K1 becomes available only when workstation D finishes the production order K1 for its fourth operation. As a consequence, the production order K1 is accompanied in workstation C by a B/C- card and a C/D-card.

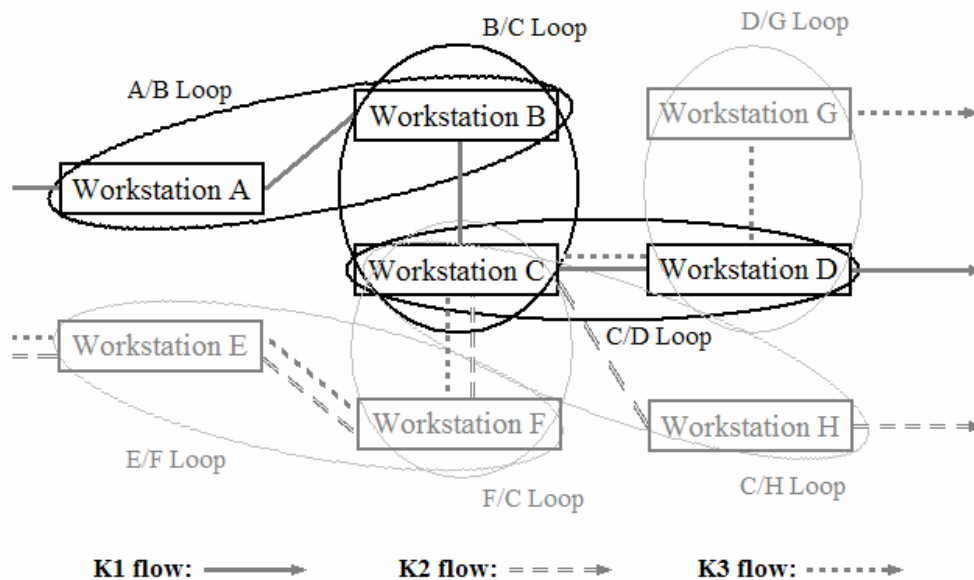


Figure 1: Job shop with 8 workstations and 3 production orders

The question remains whether the POLCA control system manages the required material flows through the resources which have the highest probability of being available? In the following two sections we will show how Advanced Resources Planning (ARP) captures the stochastic behavior of the production system in order to obtain important mid-term performance measure values to direct the mid-term planning and to guide the short term control decisions (Lambrecht et al. 1998). This approach provides the release authorizations while the instantly available capacity is controlled by a card system. In our example, a production order K1 will only be started on workstation C if there is enough capacity on workstation C *and* D. By looking

one operation ahead, the available capacity enjoys a better opportunistic allocation. Besides the improved capacity allocation, linking two successive workstations in a loop also results in a more flexible buffer allocation allowing variability pooling. For example, whether workstation C or D is the bottleneck, it will be protected by most of the C/D-cards. Linking more than two successive workstations adds complexity (because of the high number of different cards) at the expense of decreasing ‘control’ returns.

We implemented an electronic version where at each workstation a display shows the authorized production orders waiting to be processed. The authorized production orders for which a POLCA card is available are colored green, the remaining production orders red. The application is further amended by a workload based version of the original POLCA control system. We converted the allowed number of production orders in the loop, equal to the number of POLCA cards, in allowed hours of work in the loop. Only when the allowed hours of work in the loop minus the hours of work present in the loop is greater than zero, the next production order in the list will be colored green and allowed to be processed.

3. Implementing POLCA

In the previous section we discussed the principles of the POLCA control system. In this section we determine the release authorizations and the allowed workload for each loop. Both are necessary conditions to implement the POLCA control system.

3.1 Release Authorization

As described above, the POLCA control system requires that every workstation has a list of authorized production orders. Release authorizations should be based on demand and system status. To obtain realistic lead time estimations we rely on Advanced Resources Planning. Using stochastic modelling and optimization this aggregate planning explicitly recognizes the stochastic nature of manufacturing systems. We distinguish two phases (see also Vandaele and Lambrecht 2003). During the ‘Lot Sizing and Lead Time Estimation Phase’ the multi-product, multi-machine job shop is modelled as a queueing network where all parameters are a function of the lot size. By applying an optimization routine to the queueing network we obtain for each product the manufacturing lot size that minimizes the weighted average expected lead time of the job shop and the corresponding probability distribution of the lead times. Section 4 will expose

this 'Lot Sizing and Lead Time Estimation Phase'. During the 'Tuning Phase' management may consider the lead times as unacceptable and may decide to adjust capacity structure (overtime, capacity expansion), to off-load heavily loaded resources, to consider alternative routings... The queueing network provides the opportunity to conduct a large number of 'what-if' analyses. All serious capacity problems must be solved before diving into operational decisions.

Taking into account the output of the ARP system we put together production orders and create a time window for each production order. We group customer orders of product k , characterized by an order quantity and a due date, into a number of production orders of which the number of units approach the target manufacturing lot size calculated by ARP. For each production order a time window is created by subtracting the expected lead time of the production order plus safety time from the earliest due date of the customer orders grouped in the production order. The starting times of these time windows are used as release authorizations for the production orders.

The expected lead time of a production order is worked out mathematically at the end of Section 4. Safety time is defined as the difference between the desired percentile of the lead time distribution and the expected lead time. In Figure 2 we find two safety time allocation schemes. In each case we recognize 3 operations with expected waiting times (W), expected setup times (S) and expected process times (P). At the top, all safety time is pooled to benefit from variability pooling. We recognize one time window. The release authorizations of the three operations coincide. At the bottom, safety time is allocated to each operation. Each operation has its own release authorization to obtain more control over the system. Figure 2 illustrates that the same service level can be obtained with less safety time when safety time is pooled. However, the existence of assemblies, bottlenecks and material trace-ability requirements will limit the use of pooled safety time (Vandaele and De Boeck 2003). Further research should yield the optimal allocation scheme of safety time (cf *infra*).

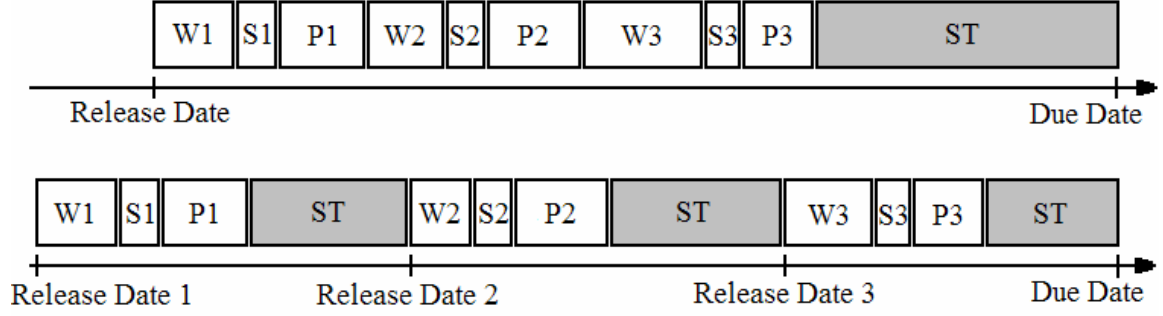


Figure 2: Safety Time Allocation Schemes

The software I-Clips, which computerizes Advanced Resources Planning (Vandaele and Lambrecht 2003), is able to provide every workstation with a list of production orders, each with its release authorization based on demand and system status.

3.2 Number of Cards

In addition to the release authorizations, for each loop joining two successive workstations the workload has to be determined. By putting a cap on the allowed workload in each loop we prevent congestion of the production system.

Based on Little's Law, stating that for any given production system the work-in-process is equal to the lead time multiplied by the throughput (Hopp and Spearman 2000, 223-225), Suri proposed the following equation (Suri 1998, 255-256):

$$\text{Number of } l/m \text{ cards} = [LT(l) + LT(m)] * \frac{\text{Num}(l/m)}{D}. \quad (1)$$

$LT(l)$ and $LT(m)$ denote the average lead times (in hours) of workstations l and m over the planning horizon D . $\text{Num}(l/m)$ represents the total number of production orders that go from workstation l to workstation m during the planning horizon. Finally, D is expressed as the number of hours in the planning horizon. We specify $LT(m)$ and $\text{Num}(l/m)$ based on the output of the ARP system and convert the number of cards into allowed hours of work.

Assume k to be the product index ($k = 1 \dots K$), m the workstation index ($m = 1 \dots M$) and o the operation index for product k ($o = 1 \dots O_k$), where O_k is the number of operations for product k . Each product k is characterized by an average order quantity \overline{OQ}_k and an average order interarrival time \overline{Y}_k , expressed in hours. As far as the production characteristics are con-

cerned, the following is defined for product k and operation o , expressed in hours: \bar{T}_{ko} , the expected setup time and \bar{X}_{ko} , the expected unit processing time. In addition, define $\delta_{kom} = 1$ if operation o for product k is on workstation m (0 otherwise); $\delta_{kol \rightarrow m} = 1$ if operation o for product k is on workstation l and the next operation is on workstation m (0 otherwise); $\delta_{kol \leftrightarrow m} = 1$ if operation o for product k is on workstation l and the next operation is on workstation m or if operation o for product k is on workstation m and the previous operation was on workstation l (0 otherwise). Note that $\delta_{kol \rightarrow m} = \delta_{kol} \delta_{k(o+1)m}$ and $\delta_{kol \leftrightarrow m} = \delta_{kol} \delta_{k(o+1)m} + \delta_{kom} \delta_{k(o-1)l}$.

First, we specify $LT(m)$. In order to keep track with the focus of this section, we describe the stochastic model in more detail in the next section. To continue the line here, based on Advanced Resources Planning principles, the model provides the probability distribution of the lead time of each operation of each production order by modelling the multi-product, multi-machine job shop as a queueing network. Besides, we saw in Section 3.1 that customer orders of product k , characterized by an average order quantity \overline{OQ}_k , are grouped into optimal production orders $Q_k^* \overline{OQ}_k$. As a starting point, we propose the following weighted average lead time

$$LT(m) = \sum_{k=1}^K \sum_{o=1}^{O_k} \delta_{kom} \left[\frac{\overline{OQ}_k D}{Y_k Q_k^* \overline{OQ}_k} \right] / \left[\sum_{k=1}^K \sum_{o=1}^{O_k} \delta_{kom} \frac{\overline{OQ}_k D}{Y_k Q_k^* \overline{OQ}_k} \right] \left[E(W_{q_m}(Q^*)) + \bar{T}_{ko} + Q_k^* \overline{OQ}_k \bar{X}_{ko} \right]. \quad (2)$$

The weight between the first square brackets takes care of the relative importance of operation o of product k on workstation m . $E(W_{q_m}(Q^*))$ represents the expected waiting time at workstation m . \bar{T}_{ko} denotes the expected setup time of the production order $Q_k^* \overline{OQ}_k$ on workstation m . $Q_k^* \overline{OQ}_k \bar{X}_{ko}$ encloses the expected process time of the production order $Q_k^* \overline{OQ}_k$ on workstation m . However, by replacing $E(W_{q_m}(Q^*))$ with $S_{\%}(W_{q_m}(Q^*))$, the required percentile of the probability distribution of the time spent in the queue of workstation m , in (2), we are able to quantify the number of safety cards. Considering the WIP fluctuations in the queue, these safety cards protect the resources by increasing the buffer sizes. Our specification of $LT(m)$ becomes

$$LT(m) = \sum_{k=1}^K \sum_{o=1}^{O_k} \delta_{kom} \left[\frac{\overline{OQ}_k D}{Y_k Q_k^* \overline{OQ}_k} \right] / \left[\sum_{k=1}^K \sum_{o=1}^{O_k} \delta_{kom} \frac{\overline{OQ}_k D}{Y_k Q_k^* \overline{OQ}_k} \right] \left[S_{\%}(W_{q_m}(Q^*)) + \bar{T}_{ko} + Q_k^* \overline{OQ}_k \bar{X}_{ko} \right]. \quad (3)$$

Note that at this point each operation receives an amount of safety time, which can be operation dependent. The issue and performance of different allocation schemes is important and subject to future research. As a consequence, it will not be discussed in this paper.

Second, we specify $Num(l/m)$. This represents the total number of production orders that go from workstation l to workstation m during the planning horizon D . By using the demand for the products on the l/m loop we obtain

$$Num(l/m) = \sum_{k=1}^K \sum_{o=1}^{O_k} \delta_{kol \rightarrow m} \frac{\overline{OQ}_k D}{\overline{Y}_k Q_k^* \overline{OQ}_k}. \quad (4)$$

At this point we have determined all elements to compute the number of POLCA cards. To convert the number of cards into allowed hours of work, we multiply the number of cards by $WL(l/m)$, the average workload of the operations of the products passing the l/m loop during the planning horizon:

$$WL(l/m) = \left(\sum_{k=1}^K \sum_{o=1}^{O_k} \delta_{kol \leftrightarrow m} \frac{\overline{OQ}_k D}{\overline{Y}_k Q_k^* \overline{OQ}_k} (\overline{T}_{ko} + Q_k^* \overline{OQ}_k \overline{X}_{ko}) \right) / \left(\sum_{k=1}^K \sum_{o=1}^{O_k} \delta_{kol \rightarrow m} \frac{\overline{OQ}_k D}{\overline{Y}_k Q_k^* \overline{OQ}_k} \right) \quad (5)$$

By using (1), (3) and (4), we obtain the number of l/m cards. Multiplying this number by (5) yields the allowed workload in the l/m loop:

$$\text{Number of } l/m \text{ cards} = \left(\sum_{k=1}^K \sum_{o=1}^{O_k} XY \right) Z \quad \text{Allowed workload on } l/m \text{ loop} = \left(\sum_{k=1}^K \sum_{o=1}^{O_k} XY \right) W$$

$$\text{with } X = \delta_{kol} \left(\frac{1}{\overline{Y}_k Q_k^*} / \sum_{k=1}^K \sum_{o=1}^{O_k} \frac{\delta_{kol}}{\overline{Y}_k Q_k^*} \right) + \delta_{kom} \left(\frac{1}{\overline{Y}_k Q_k^*} / \sum_{k=1}^K \sum_{o=1}^{O_k} \frac{\delta_{kom}}{\overline{Y}_k Q_k^*} \right)$$

$$Y = \delta_{kol} S_{\%} (W_{q_l} (Q^*)) + \delta_{kom} S_{\%} (W_{q_m} (Q^*)) + \overline{T}_{ko} + Q_k^* \overline{OQ}_k \overline{X}_{ko}$$

$$Z = \sum_{k=1}^K \sum_{o=1}^{O_k} (\delta_{kol} \delta_{k(o+1)m} / \overline{Y}_k Q_k^*)$$

$$W = \sum_{k=1}^K \sum_{o=1}^{O_k} ((\delta_{kol} \delta_{k(o+1)m} + \delta_{kom} \delta_{k(o-1)l}) (\overline{T}_{ko} + Q_k^* \overline{OQ}_k \overline{X}_{ko}) / \overline{Y}_k Q_k^*).$$

Note that the duration of the planning horizon does not affect the number of cards or allowed workload on the loops because the constant D vanishes.

Once I-Clips has provided every workstation with a list of release authorizations and we have determined the allowed workload for each loop by the above mentioned formulas, the POLCA

control system will manage the required material flows through the available capacity by signaling every workstation at each moment what to do.

4. Advanced Resources Planning, a necessary prerequisite

By now it should be clear that Advanced Resources Planning (ARP), a high-level tuning and planning tool recognizing the stochastic nature of manufacturing systems (Vandaele and De Boeck 2003), is playing an important part in the implementation of the POLCA control system. ARP captures the stochastic behavior of the production capacity to obtain the release authorizations and the allowed workloads in the loops. In Section 3 we mentioned that ARP starts with the modeling of the multi-product, multi-machine job shop as a queueing network in order to obtain realistic lot sizes, lead times... In this fourth section we focus our attention on this modeling. We refer to Lambrecht, Ivens and Vandaele (1998) for further details.

Assume k to be the product index ($k = 1 \dots K$), m the workstation index ($m = 1 \dots M$) and o the operation index for product k ($o = 1 \dots O_k$), where O_k is the number of operations for product k . Each product k is characterized by an average order quantity \overline{OQ}_k , an average order interarrival time \overline{Y}_k , the variance of the order interarrival time $s_{Y_k}^2$, the squared coefficient of variation (SCV) of the order interarrival time $c_{Y_k}^2$ and the arrival rate $\lambda_k = 1/\overline{Y}_k$.

As far as the production characteristics are concerned, the following are defined for product k and operation o , expressed in hours: T_{ko} , the setup time random variable; X_{ko} , the unit processing time random variable; \overline{T}_{ko} , the expected setup time; \overline{X}_{ko} , the expected unit processing time; μ_{ko} , the unit processing rate ($= 1/\overline{X}_{ko}$); $s_{T_{ko}}^2$, the variance of the setup time; $s_{X_{ko}}^2$, the variance of the unit processing time; $c_{T_{ko}}^2$, the SCV of the setup time; $c_{X_{ko}}^2$, the SCV of the unit processing time. In addition, let us define $\delta_{kom} = 1$ if operation o for product k is on workstation m and 0 otherwise. At this point all the input parameters are given.

In the proposed queueing network each workstation is modeled as a multi-product lot sizing model with queueing delays. The multiple arrival processes of the k products are superposed into one aggregate arrival process. All characteristics of the aggregate arrival process and the aggregate production process are functions of the lot sizes Q_k . Note that the lot size is expressed

as a multiplier Q_k of the average order quantity \overline{OQ}_k . For each workstation m we have to obtain: l_m , the aggregate batch arrival rate; ca_m^2 , the SCV of the aggregate batch interarrival time; $ca_m'^2$, the SCV of the external aggregate batch interarrival time; μ_m , the aggregate batch processing rate; cs_m^2 , the SCV of the aggregate batch processing time; ρ_m' , the adapted traffic intensity.

The aggregate arrival process at workstation m is characterized by the average and the SCV of the aggregate batch interarrival times. Note that the batch arrival rate of product k at the first workstation of its routing equals $\lambda_{b_k} = \lambda_k / Q_k$ which is a result of grouping the order quantities into a manufacturing batch of size $Q_k \overline{OQ}_k$ (expressed in units). The aggregate batch arrival rate of product k at workstation m equals $l_{mk} = \sum_{o=1}^{O_k} \lambda_{b_k} \delta_{kom}$. Then the aggregate batch arrival rate at workstation m equals $l_m = \sum_{k=1}^K \sum_{o=1}^{O_k} \lambda_{b_k} \delta_{kom}$ which includes both the internal and the external batch arrivals at workstation m . The external aggregate batch arrival rate at workstation m equals $l_m' = \sum_{k=1}^K \lambda_{b_k} \delta_{k1m}$.

We now turn to the production process at workstation m . The aggregate batch processing time on workstation m equals

$$1/\mu_m = \sum_{k=1}^K \frac{l_{mk}}{l_m} \sum_{o=1}^{O_k} \frac{\lambda_{b_k} \delta_{kom}}{l_{mk}} (\bar{T}_{ko} + Q_k \overline{OQ}_k \bar{X}_{ko})$$

where l_{mk}/l_m is the probability that a randomly picked product in front of workstation m is of product type k . The expression for $1/\mu_m$ is a weighted average over product batch processing times, which are in turn weighted averages of the operations on workstation m for the same product.

Along the same lines, the SCV of the aggregate batch processing time are obtained (Lambrecht et al. 1998):

$$cs_m^2 = \left[\sum_{k=1}^K \frac{l_{mk}}{l_m} \sum_{o=1}^{O_k} \frac{\lambda_{b_k} \delta_{kom}}{l_{mk}} \left[\bar{T}_{ko} + Q_k \overline{OQ}_k \bar{X}_{ko} \right]^2 \right] \mu_m^2 - 1 + \sum_{k=1}^K \frac{l_{mk}}{l_m} \sum_{o=1}^{O_k} \frac{\lambda_{b_k} \delta_{kom}}{l_{mk}} \frac{\left[s_{T_{ko}}^2 + Q_k \overline{OQ}_k s_{X_{ko}}^2 \right]}{\left[\bar{T}_{ko} + Q_k \overline{OQ}_k \bar{X}_{ko} \right]^2}$$

The adapted traffic intensity for workstation m , which includes both the utilization due to setups and the utilization due to processing, becomes

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

At this point, only ca_m^2 , the SCV of the aggregate batch interarrival time and $ca_m'^2$, the SCV of the external aggregate batch interarrival time remain to be determined. Solving the following set of linear equations yields the M unknowns ca_m^2 , $m = 1, \dots, M$ (Lambrecht et al. 1998):

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

In this set of linear equations, $f_{0n} = l'_n / \sum_{m=1}^M l'_m$ represents the proportion of batches from outside

and going to workstation n ; $f_{mn} = (1/l'_m) \sum_{k=1}^K \sum_{l=1}^{O_k-1} \lambda_{b_k} \delta_{kom} \delta_{k(o+1)n}$ represents the proportion of

batches leaving workstation m and going to workstation n ; $f_{m0} = (1/l'_m) \sum_{k=1}^K \lambda_{b_k} \delta_{kO_k m}$ represents

the proportion of batches leaving workstation m and going outside. To obtain $ca_m'^2$ the following approximation is used:

$$ca_m'^2 \approx \frac{1}{3} + \frac{2}{3} \sum_{k=1}^K \frac{\lambda_{b_k} \delta_{k1m} c_{Y_k}^2}{l'_m Q_k} \quad \text{if } \sum_{k=1}^K \delta_{k1m} \geq 2$$

$$ca_m'^2 \approx c_{Y_k}^2 / Q_k \quad \text{if } \sum_{k=1}^K \delta_{k1m} = 1.$$

Finally the weighted average lead time of workstation m can be stated as:

$$E(W_{M_m}) = E(W_{q_m}) + \sum_{k=1}^K \frac{\sum_{o=1}^{O_k} \lambda_k \overline{OQ}_k \delta_{kom}}{\sum_{k=1}^K \sum_{o=1}^{O_k} \lambda_k \overline{OQ}_k \delta_{kom}} \left(\sum_{o=1}^{O_k} \frac{\lambda_k \overline{OQ}_k \delta_{kom}}{\sum_{o=1}^{O_k} \lambda_k \overline{OQ}_k \delta_{kom}} [T_{ko} + Q_k \overline{OQ}_k \bar{X}_{ko}] \right) \quad (6)$$

$$\text{with } E(W_{q_m}) = \frac{\rho_m'^2 (ca_m^2 + cs_m^2)}{2l'_m (1 - \rho_m')} \exp \left\{ \frac{-2(1 - \rho_m') (1 - ca_m^2)^2}{3\rho_m' (ca_m^2 + cs_m^2)} \right\} \quad \text{if } ca_m^2 \leq 1$$

$$E(W_{q_m}) = \frac{\rho_m'^2 (ca_m^2 + cs_m^2)}{2l'_m (1 - \rho_m')} \quad \text{if } ca_m^2 > 1.$$

This is the weighted average over the products visiting workstation m , which on their turn are weighted averages over the operations on workstation m for product k . The weight

$\sum_{o=1}^{O_k} \lambda_k \overline{OQ}_k \delta_{kom} / \sum_{k=1}^K \sum_{o=1}^{O_k} \lambda_k \overline{OQ}_k \delta_{kom}$, independent from the manufacturing lot size multiplier,

measures the relative importance of product k for workstation m .

If we add the weighted average lead time of each workstation and take into account the average waiting time of finished batches until their due date, the objective function for the total job shop becomes

$$E(W) = \sum_{m=1}^M E(W_{q_m}) + \sum_{k=1}^K \frac{\lambda_k \overline{OQ}_k}{\sum_{k=1}^K \lambda_k \overline{OQ}_k} \frac{(Q_k \overline{OQ}_k - 1) \overline{Y}_k}{2 \overline{OQ}_k} \quad (7)$$

$$+ \sum_{m=1}^M \sum_{k=1}^K \frac{\sum_{o=1}^{O_k} \lambda_k \overline{OQ}_k \delta_{kom}}{\sum_{k=1}^K \sum_{o=1}^{O_k} \lambda_k \overline{OQ}_k \delta_{kom}} \left(\sum_{o=1}^{O_k} \frac{\lambda_k \overline{OQ}_k \delta_{kom}}{\sum_{o=1}^{O_k} \lambda_k \overline{OQ}_k \delta_{kom}} \left[T_{ko} + Q_k \overline{OQ}_k \overline{X}_{ko} \right] \right).$$

The second sum measures the average waiting time of finished batches until their due date. The weight $\lambda_k \overline{OQ}_k / \sum_{k=1}^K \lambda_k \overline{OQ}_k$ takes care of the relative importance of product k for the total job shop.

By using a dedicated optimization routine, the non-linear objective function of the total job shop, equation (7), is minimized taking into account a set of simultaneous, non-linear constraints (Vandaele 1996). We obtain the vector Q^* , containing the optimal multiplier Q_k^* for each product.

Now, we are able to complete our exposition started in Section 3. First, we determine the release authorizations. In this paper safety time is pooled at the end of the routing. Further research should yield the influence of other safety time allocation schemes on the queuing network. Consequentially, the fixation of the time windows requires the probability distributions of the total lead time of the production orders. Assume QL_{kl} to be the lot size of production order L_{kl} , where l is the production order index for product k ($l = 1 \dots S_k$) and S_k represents the number of production orders for product k . The expected total lead time of production order L_{kl} is given by

$$E(W_k) = \sum_{o=1}^{O_k} \left(\sum_{m=1}^M E(W_{q_m}(Q^*)) \delta_{kom} + \overline{T}_{ko} + QL_{kl} \overline{X}_{ko} \right) \quad (8)$$

The variance of the total lead time of production order L_{kl} is approximated by

$$V(W_k) = \sum_{o=1}^{O_k} V(W_{q_m}(Q^*)) \delta_{kom} + \sum_{o=1}^{O_k} s_{T_{ko}}^2 + \sum_{o=1}^{O_k} Q_{L_{ko}} s_{X_{ko}}^2 \quad (9)$$

in which the term $V(W_{q_m})$ is given in Vandaele and Lambrecht (2003). If a distribution is postulated, the total lead time of product k is fully characterized.

Second, we determine $S_{\%}(W_{q_m}(Q^*))$, the required percentile of the probability distribution of the time spent in the queue of workstation m , in order to calculate the allowed workload in the loops. In equation (6) we have defined $E(W_{q_m})$. Besides, $V(W_{q_m})$ is given in Vandaele and Lambrecht (2003). If we postulate a lognormal distribution for the time spent in the queue of workstation m (Vandaele 1996):

$$S_{\%}(W_{q_m}(Q^*)) = \exp\{\beta + z_{\%} \gamma\}$$

with $z_{\%}$ the required percentile of the standard normal distribution,

$$\beta = \ln \left(E(W_q(Q^*)) / \sqrt{\frac{V(W_q(Q^*))}{E(W_q(Q^*))^2} + 1} \right) \text{ and } \gamma^2 = \ln \left(\frac{V(W_q(Q^*))}{E(W_q(Q^*))^2} + 1 \right).$$

Note that we assume the same service level for all products processed on workstation m . One could easily substitute the required percentile by a weighted average percentile taking into account the different service levels and the relative importance of the products on workstation m .

At this point, the theoretical modeling for determining the prerequisites for the POLCA control system has been developed. In Section 5 we report on the industrial implementation of E-POLCA at Spicer Off-Highway, Bruges (Belgium).

5. E-POLCA implementation at Spicer Off-Highway Products Division Bruges

A few years ago, Spicer Off-Highway Products Division Bruges, which produces power shift transmissions for off-highway vehicles, realized substantial operational improvements by implementing the software I-CLIPS which computerizes the mentioned ARP system (Vandaele et al. 2000). The output of the ARP system was essential to improve detailed scheduling of the job shop. At the moment, the metal working company is implementing our workload based version of the POLCA control system with a view to improve job shop control and make job shop

scheduling redundant. In this section we place the POLCA control system within the existing framework at the company.

We first describe the production environment. The production process of power shift transmissions is roughly a four-step process. First, raw steel parts are processed in the soft steel shop. Second, the steel parts are hardened through a heat-treatment. Third, the hardened parts are processed in the hard steel shop. Finally, the steel parts are assembled into housings that are shipped from another production unit. The workstations of the steel shops contain one or more machines. Initially, the workstations were arranged in a job shop layout. Currently, job shop layout and cellular layout are alternating. The material handling at the steel shops is fully automated. First, all workstations of the steel shops are arranged around two high-stacker cranes. Second, the steel shops are equipped with an 'Automatic Storage and Retrieval System' which stores the work-in-process and is accessible from the assembly lines. Third, an 'Automotive Guided Vehicle' moves the parts from the soft steel shop to heat-treatment and brings them back to the hard steel shop.

Secondly, the production planning and scheduling procedure can be described as follows. Producing a wide variety of highly customized products Spicer Off-Highway Products Division Bruges opted for an assemble-to-order environment. The company starts the production planning with a forecasting process to estimate the component requirements. Every four weeks, the demand for final products and spare parts for the next 48 weeks are forecasted taking into account the received orders, customer forecasts, the economic outlook and business cycles. The process provides a sales plan for final products in which 10% has to be delivered a particular day, 50% a particular week and 40% a particular month. Further, a load leveling process transforms the sales plan into a master production schedule (MPS). After two weeks, this MPS is adjusted for significant deviations. Elaborating the MPS by using the bill of materials, demand for spare parts, stock data... yields the requirements for components for the next 48 weeks.

Taking into account the demand for components, the available resources, the calendar, the bill of materials and the production routings for the next 16 weeks, I-CLIPS provides weekly for each component the optimal lot size and the corresponding lead time distribution. The user interface of I-CLIPS can be found in Vandaele and Lambrecht (2003). The output of I-CLIPS, adjusted weekly to the daily changes of the MPS, is used to group component requirements into a number of production orders of which the number of units approach the target manufacturing lot

size. For some components the requirements for several weeks are grouped into a production order. Because Spicer Off-Highway Products Division Bruges opted for pooled safety time in the steel shops and heat treatment department, cf. the top part of Figure 2, equation (8) and (9), the I-CLIPS output and the postulation of a lognormal distribution are sufficient to set the release authorization of a production order on the first workstation of its routing. Once this release authorization has been crossed, the production order is authorized on all workstations.

The assembly lines are controlled by final assembly schedules. On a daily basis, a linear and mixed integer programming solver optimizes the sequence of operations at each assembly line for the next 20 days considering the master production schedule, the availability of resources, the availability of components and the priorities. Intended to gear the component production and final assembly to one other, the company implemented a queue management system. With a view to timely delivery at the assembly lines, reduce setups... the system adjusts daily the sequence of the jobs in the queue of the workstations of the steel shops and heat treatment department. In fact, group leaders responsible for a number of workstations and having all required information at their disposal, are ordering the production orders in the queue of their workstations by applying some priority rules and going along with unforeseen conditions.

Finally, we place the POLCA control system within the current framework. Spicer Off-Highway Products Division Bruges is implementing the POLCA control system in the steel shops and heat treatment department to support the queue management system by considering the real time available capacity downstream. All requirements to implement the POLCA control system are fulfilled. First, at each workstation a display produces the output of the queue management system, an ordered list of production orders in the queue of the workstation. With pooled safety time, every production order released on the shop floor is authorized. Second, I-CLIPS provides the data required to calculate the allowed workload for each loop. Currently, the calculations described in Section 3 are executed in Microsoft Access. After testing and refining the application, the allowed workload for each loop will be calculated weekly by a procedure in the AS/400 system of the company.

To illustrate the real-life implementation we discuss the input and output of the Microsoft Access application. At the time of writing this paper, the company was producing 583 different components requiring 3,967 operations on 123 different workstations. We discerned 343 POLCA loops. The input of the database encloses 5 tables linked to I-CLIPS output files. In

Figure 3 the demand table shows that the component with I-CLIPS identification 580 faces an average order quantity of 42 units. The average inter arrival time is 14 days with a squared coefficient of variation of 0.5. The manufacturing process of the component consists of 7 operations and the optimal manufacturing lot size calculated by I-CLIPS is 255 units. The available capacity is visualized in Figure 4. There is one workstation with I-CLIPS identification 115, or Dana identification 99W03, which has an availability of 66.5% on the ‘24/7’ time scale. This percentage summarizes the overall availability of the workstation. It contains shift patterns, calendar information, downtime, breaks, meetings...

PartDescription	AvgDemand	IAT	SCVIAT	NbrOperations	LotSize
583	22	7	0,0585	1	60
582	43	7	0,0622	7	200
581	23	14	0,5	1	60
580	42	14	0,5	7	255
579	68	7	0,4722	1	96
578	58	7	0,4982	7	200
577	58	7	0,4982	7	200
576	16	7	0,5	1	60
575	13	7	0,5	7	100
574	13	7	0,5	7	100
573	22	7	0,5	1	60
572	18	7	0,5	9	150
571	18	7	0,5	6	163
570	45	7	0	1	73
569	38	7	0	7	200
568	38	7	0	6	200
567	68	7	0	1	129
566	62	7	0	7	250

Figure 3: Input Demand

MachineId	MachineDescription	Availability	NbrMachines
12399W42		0,665	1
12299W36		0,665	1
12199W25		0,665	1
12099W20		0,665	1
11999W09		0,665	1
11899W08		0,665	1
11799W07		0,665	1
11699W04		0,665	1
11599W03		0,665	1
11499W01		0,665	1
11399W00		0,665	1
11297BBB		0,218	1
11193KOG		0,218	1
11092475		0,218	1
10992326		0,218	1
10892A80		0,114	1
10790964		0,114	1
1069090W		0,437	1

Figure 4: Input Resources

The table depicted in Figure 5 reveals the routing of the components. The 7th operation on component 580 consists of operation 132 and is performed by the workstation with I-CLIPS identification 75. Process- and setup times are recorded in Figure 6. To manufacture 1 component 580, operation 132 performed by workstation 75 takes on average 0.07 hours with a variance of 0.0002 squared hours. The average setup time of this operation is 4 hours with a variance of 0.25 squared hours. Only the required percentile of the probability distribution of the waiting time is yet missing. Spicer Off-Highway products Division Bruges opted for the 85-percentile. Using the average waiting times and variances calculated by I-CLIPS and postulating lognormal distributions, a spreadsheet calculates the required percentile for each workstation. In Figure 7, we find that the workstation with Dana identification 08892, or I-CLIPS identification 12, knows an average waiting time of 14.36 hours with a variance of 694.1 squared hours and an 85-percentile of 24.2 hours.

ProductId	Order	Operation	MachinelD
583	1	10	93
582	7	130	75
582	6	125	89
582	5	120	91
582	4	110	104
582	3	100	101
582	2	10	17
582	1	1	96
581	1	10	93
580	7	132	75
580	6	125	89
580	5	120	91
580	4	110	104
580	3	100	101
580	2	10	17
580	1	1	96
579	1	10	93
578	7	130	75

Figure 5: Input Routings

PartDescription	Operation	MachinelD	Setup	VarSetup	Process	VarProcess
583	10	93	0,5	0	0,0004	0
582	130	75	2,7875	0,777	0,075	0,0002
582	125	89	1,3	0,01	0,025	0
582	120	91	1,34	0,1795	0,0604	0,0001
582	110	104	0,2	0,004	0,005	0
582	100	101	0,01	0	0,01	0
582	10	17	4,1211	1,6983	0,0691	0,0002
582	1	96	0,2	0	0,03	0
581	10	93	0,55	0,0302	0,0004	0
580	132	75	4	0,25	0,07	0,0002
580	125	89	1,514	0,0546	0,025	0
580	120	91	1,545	0,2387	0,06	0,0001
580	110	104	0,2	0,004	0,0046	0
580	100	101	0,01	0	0,0084	0
580	10	17	7,75	0	0,0667	0,0002
580	1	96	0,3931	0,0154	0,0225	0
579	10	93	0,55	0,0302	0,0004	0
578	130	75	2,1373	0,4568	0,07	0,0002

Figure 6: Input Process- and Setup Times

Using these linked tables and some queries, the Process- and Setup Times table, see Figure 6, is extended with a NextOper column and a NextMach column, showing the I-CLIPS identification of the next operation and next workstation. Finally, some queries calculate the allowed workload on each loop. In Figure 8 we find that the allowed workload on the loop joining workstation 04X15 to workstation 05374 comes to 97.8 hours.

Machine	Wait1	MachinelD	VarWait	Wait
02427	138,08	1	29385,921928	235,67
02998	56,47	2	7160,550042	96,51
03290	67,62	3	7885,603832	115,63
04X15	42,87	4	2954,57303	73,23
05374	26,79	5	1364,41195	45,84
06652	56,91	6	6977,873306	97,31
08A16	0,08	7	1,254341	0,06
08251	0,001	8	0,001	0,00
08264	0,49	9	10,642192	0,55
08426	4,09	10	184,61322	0,00
08477	2,51	11	60,508534	3,79
08892	14,36	12	694,112732	24,19
08952	11,88	13	954,08077	18,80
10633	76,59	14	11999,978362	131,02
10945	76,26	15	12108,217788	130,44
11X80	0,35	16	8,960997	0,35
11966	78,75	17	7797,707155	9,82
12774	66,81	18	9065,020437	114,30

Figure 7: Input Waiting Times

FIRST	LAST	Hours
02427	02427	4,1374
02427	23950	5,4413
03290	05374	29,7510
03290	16786	20,7494
03290	17996	64,4712
03290	21760	332,9061
03290	27530	9,8543
03290	27661	11,5166
04X15	05374	97,7960
04X15	16786	50,3457
04X15	16A17	15,9680
04X15	17996	99,0960
04X15	27530	25,7350
04X15	27661	16,9538
05374	05374	25,4360
05374	10633	48,8217
05374	17996	19,7016
05374	18725	69,1560

Figure 8: Output E-POLCA

To simplify the application, all tables, queries, macros... are hidden with the exception of 1 form containing 7 buttons. Clicking the first 4 buttons reproduces the Demand table (see Figure 3), the Resource table (see Figure 4), the Routings table (extended Process- and Setup Times table) and the Waiting Time table (see Figure 7). The fifth button refreshes the input, the sixth button reproduces the output of the application (see Figure 8) and the last button closes the program.

At each workstation, the output of the E-POLCA control system will be coloured-wise visualized on a networked pc. Figure 9 shows a black and white version of one of the screens help-

ing workstation 04X15 to select the next production order. We retrieved information about the 04X15/05374 loop. At the left side, we find the output of the queue management system for workstation 04X15 with some routing information. The first production order comes from workstation 90600 and will go to workstation 05374 after workstation 04X15 finishes operations. Besides, we find the sum of the setup- and process times on these workstations. At the right side, we find the same information for the second workstation of the loop we are interested in. All relevant information about the selected loop is highlighted in grey. At the bottom we recover the allowed workload on the loop, the workload currently on the loop and the difference between these amounts. The allowed workload on the 04X15/05374 loop is 97.8 hours (see also Figure 8), the workload currently on the loop is 74.1 hours and as a result the free capacity comes to 23.7 hours. Because the available capacity (23.7 hours) minus the workload of the first three 04X15/05374 production orders (22 hours) is still positive, workstation 04X15 is allowed to start the first four 04X15/05374 production orders. In Figure 9, the green colour is replaced by a bold font style. Obvious, workstation 04X15 may retrieve similar information for other loops.

MACHINE 04X15						MACHINE 05374					
Previous		Current		Next		Previous		Current		Next	
Machine	Load	Machine	Load	Machine	Load	Machine	Load	Machine	Load	Machine	Load
90600	2,7	04X15	2,6	05374	3,1	04X15	4,1	05374	2,6	08952	2,3
47250	3,1	04X15	2,9	99W03	3,0	04X15	2,9	05374	3,7	89423	3,6
24E79	4,8	04X15	5,1	05374	4,2	40X30	2,8	05374	2,8	08264	4,1
89423	3,1	04X15	3,4	05374	3,6	90600	3,1	05374	3,1	99W03	2,9
40X30	2,9	04X15	4,2	16A17	5,3	03290	4,0	05374	3,5	8989W	3,4
99W03	5,6	04X15	3,2	40X30	3,0	04X15	5,3	05374	5,2	27661	3,6
89423	4,1	04X15	2,1	05374	2,7	99W42	2,9	05374	4,1	16A17	4,1
08952	3,5	04X15	4,0	99W42	3,4	04X15	4,7	05374	4,3	89423	2,9
47250	3,0	04X15	3,8	47706	2,0	04X15	5,1	05374	4,7	99W42	3,5
41898	4,1	04X15	3,4	99W03	2,8	03290	3,9	05374	3,9	24E79	3,0
90600	3,7	04X15	3,1	05374	2,9	04X15	2,6	05374	5,1	47706	4,1
19IOZ	4,1	04X15	2,9	47706	3,8	04X15	3,3	05374	4,9	47250	3,9
08952	2,6	04X15	2,7	99W03	6,2	04X15	4,9	05374	3,6	99W42	4,0
03290	3,3	04X15	3,1	05374	3,4	19IOZ	2,6	05374	4,0	47706	3,5
89423	3,5	04X15	4,2	05374	5,1	8989W	3,4	05374	3,4	08952	2,9
27661	4,2	04X15	3,6	16A17	1,9	04X15	3,2	05374	3,9	47250	2,7
08426	3,6	04X15	3,5	8989W	5,1	16A17	2,9	05374	2,9	40X30	3,9
03290	4,0	04X15	3,8	05374	2,7						
08426	2,9	04X15	3,1	05374	2,2						
27661	3,7	04X15	1,9	47706	3,6						
LOOP 04X15 / 05374					97,8	WAIT 05374					74,1
						OPEN 04X15					23,7

Figure 9: E-POLCA Display

We conclude this section with some additional remarks, which are of less importance for the theory, but which are relevant in practice and became prevalent during the implementation.

First, the company is creating a procedure to ensure the continuation of the overlapping loops of cards when production routings leave the shop floor. Second, it was decided that a production order may leave the mapped production routing when another machine, with free capacity, can finish the operation sooner. A procedure to bring the production order back on routing was implemented. Third, when started production orders are deferred because of changed customer requirements, adjusted priorities... the concerning workload should be removed from the loops. Fourth, a temporarily wild card can be introduced if a manufacturing order is taken aside for some reason (e.g. quality problem); otherwise the card would never return. If the problem is solved, the wild card should be removed again at the earliest opportunity. Finally, small amounts of scrap are set off by safety margins in the sales program. Rework is executed immediately or planned in a next production order.

6. Conclusions

In this paper we proposed a supporting framework for the new material control system Paired-cell Overlapping Loops of Cards with Authorization (POLCA). We recommended Advanced Resources Planning (ARP), a high-level tuning and planning tool recognizing the stochastic nature of manufacturing systems, to determine the release authorizations and the allowed workload in each loop. We worked out the required computations and reported on our experiences in a metal shop which is implementing a load based version of the POLCA control system. Our efforts resulted in a system that manages the required material flows through the available resources which have the highest probability of being available. Our experiences in Spicer Off-Highway Products Division Bruges lead us to believe that the approach will be valuable for industrial practice. Further research should yield the optimal allocation of safety time and evaluate the performance of our approach.

References

- Hopp, W.J., M.L. Spearman. 2000. *Factory Physics*. Irwin/ McGraw-Hill, New York.
- Hopp, W.J., M.L. Spearman. 2004. To Pull or Not to Pull: What Is the Question? *Manufacturing & Service Operations Management* **6(2)** 133-148.

- Jonsson, P., S-A. Mattsson. 2003. The implications of fit between planning environments and manufacturing planning and control methods. *International Journal of Operations & Production Management* **23(8)** 872-900.
- Lambrecht, M.R., L.I. Ivens, N.J. Vandaele. 1998. ACLIPS: A Capacity and Lead Time Integrated Procedure for Scheduling. *Management Science* **44(11)** 1548-1561.
- Suri, R. 1998. *Quick Response Manufacturing, A Companywide Approach to Reducing Lead Times*, Productivity Press, Portland, OR.
- Suri, R., A. Krishnamurthy. 2003a. How to Plan and Implement POLCA: A Material Control System for High-Variety or Custom-Engineered Products. Technical Report May 2003, Center for Quick Response manufacturing, University of Wisconsin, Madison, WI.
- Suri, R., A. Krishnamurthy. 2003b. Planning and Implementing POLCA: A Card-based Control System for High Variety or Custom Engineered Products. *Proceedings of FAIM 2003: The International Conference on Flexible Automation and Intelligent Manufacturing*, Tampa, FL, June 2003.
- Vandaele, N.J. 1996. The impact of lot sizing on queueing delays: multi-product, multi-machine models. Ph.D. Thesis, Department of Applied Economics, Katholieke Universiteit Leuven, Belgium.
- Vandaele, N., M. Lambrecht, N. De Schuyter, R. Cremmery. 2000. Improved Lead Time Performance at Spicer Off-Highway. *Interfaces* **30(1)** 83-95.
- Vandaele, N., L. De Boeck. 2003. Advanced Resource Planning. *Robotics and Computer Integrated Manufacturing* **19** 211-218.
- Vandaele, N.J., M.R. Lambrecht. 2003. Reflections on the use of stochastic manufacturing models for planning decisions. J.G. Shanthikumar, D.D. Yao, W.H.M. Zijm. *Stochastic Modeling and Optimization of Manufacturing Systems and Supply Chains*. Kluwer Academic Publishers, Dordrecht. 53-85.
- Zhou, X., P.B. Luh, R.N. Tomastik. 2000. The Performance of a New Material Control and Replenishment System: A Simulation and Comparative Study. *Proceedings of the Quick Response Manufacturing 2000 Conference*, Madison, WI., June 2000, Society of Manufacturing Engineers, Dearborn, MI. 807-826.