CONFIDENTIAL. LIMITED CITCULATION. FOR REVIEW ONLY.

Performance analysis of a decoupling stock in a Make-to-Order system

E. De Cuypere K. De Turck D. Fiems

Department of Telecommunications and Information Processing, Ghent University, St-Pietersnieuwstraat 41, B-9000, Belgium (email: eline.decuypere@telin.UGent.be; koen.deturck@telin.UGent.be; dieter.fiems@telin.UGent.be)

Abstract: In a Make-to-Order system, products are only manufactured when orders are placed. As this may lead to overly long delivery times, a stock of semi-finished products can be installed to reduce production time: the so-called decoupling stock. As performance of the decoupling stock is critical to the overall performance and cost of the production system, we propose and analyse a Markovian model of the decoupling stock. In particular, we focus on a queueing model with two buffers, thereby accounting for both the decoupling stock as well as for possible backlog of orders. By means of numerical examples, we then quantify the impact of production inefficiency on delivery times and overall cost.

Keywords: performance evaluation, inventory control, production systems, decoupling point, Markov process, queueing theory.

1. INTRODUCTION

Production facilities are often characterised as Make-to-Order (MTO) or Make-to-Stock (MTS) systems. In MTS production, products are stocked in advance, while in MTO production, a product only starts to be manufactured when a customer order is received. Nowadays, as a mean to respond quickly to growing variety, shorter product life cycles and intense competition, industries are moving towards MTO production (Soman et al., 2004). Indeed, to ensure short delivery times, most production systems have a decoupling point at the inventory of their semi-finished products. In such systems, the decoupling stock is an accumulated inventory of semi-finished products waiting for an order to arrive. This allows for diminishing delivery times of customised products as only the final completion step still needs to be done. Research on the performance of the decoupling stock in a Make-to-Order system is therefore of main importance. This is the subject of the present paper.

Most research on Make-to-Order systems focusses on MTO/MTS decisions and hybrid MTO/MTS systems. In order to arrive at a MTO/MTS decision, Hoekstra et al. (1992) defined the customer order decoupling point (CODP) concept. These authors consider market, product and production related factors as well as the desired service level and associated inventory costs to locate the optimal decoupling point. Chang and Lu (2010) investigated so-called hybrid MTO/MTS systems. They studied a one-station production system dealing with two types of random demands: ordinary demand and specific demand. In this hybrid system, both types of demand arrive according to a Poisson process and production times of the workstation are exponentially distributed. Specific demand has a higher priority with respect to ordinary demand and the

performance of this system is studied by means of matrixgeometric methods. The decoupling inventory problem has also been studied as a two-stage production process in which the output of stage one, and the demand at stage two, are generated by independent stochastic processes. The stages are decoupled by storing intermediate products. Bell (1980) sets limits on the available storage capacity and the rates of flow production into and out of the decoupling inventory. He formulated a model which enables the firm to determine the optimum capacities for the storage facility, and to determine the value of an additional supply of intermediate product.

The present study of the decoupling stock also closely relates to literature on two-part assembly systems or kitting processes. For such systems, there are two queues, each storing a specific part, and production only starts when both part buffers are non-empty. In the present setting, one part-buffer corresponds to the decoupling stock, while the other corresponds to the list of backlogged orders. Also in the current setting, production only starts when both buffers are non-empty. Indeed, each delivery of a finished product requires both the order specifications and a semi-finished product and can only be satisfied if both are present. If both part-buffers have unlimited capacity, Harrison (1973) was the first to prove that, assuming no arrival control strategy, this queueing system is inherently unstable. In particular, he studied the multipleinput extension of the GI/G/1 queue in which arrivals in each stream are described by an independent renewal process and service times are independent and identically distributed. He showed that part waiting times converge to non-defective limiting distributions only if the buffer capacities are bounded. This was also demonstrated by Latouche (1981) who termed the two-part assembly system as waiting lines with paired customers. He considered a



Fig. 1. Decoupling stock in a Make-to-Order system

system of infinite capacity queues with Poisson arrivals for both parts and exponential services. The steady state is attained, i.e. the system is stable, if the arrival rates depend on the difference between queue lengths. Bonomi (1987) extended Latouche's research by considering two exponential distributions, one for the part processing distribution, i.e. the synchronisation phase, and the other for the assembly operation distribution. Approximations for the throughput rate and average queue length were given. Lipper and Sengupta (1986) also extended the work of Latouche by considering multiple Poisson input streams arriving in buffers with a finite capacity. They derived bounds and approximations for Poisson arrivals with the same rate and an arrival is blocked and lost if the buffer is full. A more general structure in which parts are withdrawn from infinite pools and processed prior to assembly has been studied by Hopp and Simon (1989) and Som et al. (1994). Som and Wilhelm (1999) studied a two-queue system in which each part is processed according to an exponential distribution and the assembly operation times are generally distributed. They followed a matrix-geometric approach to determine numerically the marginal distributions of both kit and end-product inventory positions. Finally, assuming finite part-buffers, a two-part assembly system in a Markovian environment was studied by De Cuypere and Fiems (2011) by means of the generalized minimal residual method.

In contrast to the literature above, we here focus on a twopart system with one finite and one infinite buffer. Such an assumption comes natural in the context of the decoupling stock. Indeed, the decoupling stock needs to be sufficiently small to limit the involved costs. Hence, finite capacity is assumed. In contrast, no such assumption is imposed on the queue of backlogged orders. As in (De Cuypere and Fiems, 2011), we study the decoupling stock in a Markovian environment. This allows for studying the effect of variability in the production process on the performance of the decoupling stock.

The remainder of this paper is organised as follows. Section 2 describes the decoupled inventory system at hand. In section 3, the steady-state probabilities are derived and relevant performance measures are determined. To illustrate our approach, section 4 considers some numerical examples. Finally, conclusions are drawn in section 5.

2. MODEL DESCRIPTION

The decoupling stock is modelled as a queueing model with two queues, as depicted in Figure 1. The first queue — the product queue — has finite capacity C_p and stores the

semi-finished products prior to being processed to finished products. The second queue — the order queue — keeps track of the orders that have not yet been delivered and has infinite capacity. Arriving orders are served in accordance with a first-come-first-served queueing discipline. Each order takes a semi-finished product from the product queue and completes the product in accordance with order specifications. Note that the two queues in the model at hand are tightly coupled. Departures from the product queue are only possible when there are orders. Similarly, departures from the order queue are only possible if there are semi-finished products in the product queue.

In order to keep the analysis of this queueing model analytically tractable, it is assumed that orders arrive in accordance with a Poisson process with arrival rate λ_o and that the production times constitute a sequence of independent exponentially distributed random variables with service rate μ . In the most basic setting, we further assume that semi-finished products arrive at the decoupling stock in accordance with a Poisson process as well. Let λ_p denote the arrival rate of this process. In this case, the decoupling stock constitutes a two-dimensional continuous time Markov process $\{(n,m) \mid n \geq 0, 0 \leq m \leq C_p\}$ where the state (n,m) corresponds to n waiting orders and msemi-finished products in the decoupling stock.

To study the impact of burstiness, unreliability or inefficiency in the production process, we extend the basic model by introducing a Markovian environment with finite state space Ω which modulates the arrival rate of the semi-finished products. In this case, the decoupling stock constitutes a three-dimensional continuous time Markov process $\{(n, m, k) \mid n \geq 0, 0 \leq m \leq C_p, k \in \Omega\}$ where the state (n, m, k) corresponds to n waiting orders, m semi-finished products in the decoupling stock and modulating state k.

3. ANALYSIS

3.1 Poisson arrivals

In the basic setting, the Markov process is a homogeneous quasi-birth-and-death process (QBD), see Latouche and Ramaswami (1999). In the present setting, the so-called level or block-row number, indicates the number of waiting orders while the phase, i.e. the index within a block element, indicates the number of semi-finished products. The one-step transitions are restricted to states in the same level (from state (n, *) to state (n, *)) or in two adjacent levels (from state (n, *) to state (n+1, *) or state (n-1, *)). In particular, the transition rates between the different states of the Markov chain are summarised in the table below.

From	То	Rate	
(n,m)	(n, m+1)	λ_p	for $m \ge C_p - 1$
(n,m)	(n+1,m)	λ_o	
(n,m)	(n-1, m-1)	μ	for $n \ge 1, m \ge 1$

We then find that the generator matrix of the Markov chain has the following block matrix representation,

$$\mathbf{Q} = \begin{bmatrix} \mathbf{L}'_{p} \ \mathbf{L}_{o} \ \mathbf{0} \ 0 \ \cdots \\ \mathbf{M} \ \mathbf{L}_{p} \ \mathbf{L}_{o} \ 0 \ \cdots \\ \mathbf{0} \ \mathbf{M} \ \mathbf{L}_{p} \ \mathbf{L}_{o} \ \cdots \\ \mathbf{0} \ \mathbf{0} \ \mathbf{M} \ \mathbf{L}_{p} \ \cdots \\ \vdots \ \vdots \ \vdots \ \cdots \end{bmatrix} .$$
(1)

The blocks are given by, $\mathbf{L}_o = \lambda_0 \mathbf{I}, \ \mathbf{L}_p = \mathbf{L}'_p - \mu \hat{\mathbf{I}}$, with

$$\mathbf{L}_{p}^{\prime} = \begin{bmatrix} -(\lambda_{p} + \lambda_{o}) & \lambda_{p} & 0 & \cdots & 0 \\ 0 & -(\lambda_{p} + \lambda_{o}) & \lambda_{p} & \cdots & 0 \\ 0 & 0 & -(\lambda_{p} + \lambda_{o}) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -(\lambda_{p} + \lambda_{o}) \end{bmatrix}$$

and,

$$\mathbf{M} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ \mu & 0 & \cdots & 0 & 0 \\ 0 & \mu & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \mu & 0 \end{bmatrix} ,$$

and where \mathbf{I} and $\mathbf{\ddot{I}}$ denote the identity matrix and the diagonal matrix with first diagonal element equal to zero and the others equal to one, respectively.

3.2 Interrupted Poisson arrivals

To model the burstiness in the arrival process of the semifinished products, we replace the Poisson processes by a Markovian arrival process (Buchholz et al., 2010; Fiems et al., 2011). We here limit the presentation to interrupted Poisson processes, but the methodology easily extends to general Markovian arrival processes. In either case, the QBD structure of the Markov chain is retained, the phase now describing both the content of the decoupling stock and the state of the Markovian environment.

For the interrupted Poisson process, the Markovian environment has two states: the active state in which there are arrivals in accordance with a Poisson process with rate λ_p and the inactive state in which there are no arrivals. Let α and β denote the rate from the inactive to the active state and vice versa, respectively. Retaining orders arriving according to a Poisson process and exponentially distributed kitting times, we have the following transition rates:

From	To	Rate	
(n, m, 0)	(n,m,1)	α	
(n, m, 1)	(n,m,0)	β	
(n,m,k)	(n, m+1, 1)	λ_p	for $m \ge C_p - 1$
(n,m,k)	(n+1, m, k)	λ_o	*
(n,m,k)	(n-1, m-1, k)	μ	for $n \ge 1, m \ge 1$
		-	

As mentioned above, for state (n, m, k), the level is still defined by the state value n, indicating the number of waiting orders, while the phases are defined by m (the number of semi-finished products) and k (the state of the Markovian environment). It is assumed that when the state k equals one, the process is in an active state while when it equals zero the process is inactive. Clearly, the generator matrix of the Markov model still has the block matrix representation (1). The blocks are now defined as follows. We have $\mathbf{L}_o = \lambda \mathbf{I}$, $\mathbf{L}_p = \mathbf{L}'_p - \mu \tilde{\mathbf{I}}$, with,

$$\mathbf{L}_{p}^{\prime} = \begin{bmatrix} -(\alpha + \lambda_{o}) & \alpha & 0 & \cdots & 0 \\ \beta & -(\beta + \lambda_{p} + \lambda_{o}) & \lambda_{p} & \cdots & 0 \\ 0 & 0 & -(\alpha + \lambda_{o}) \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots -(\beta + \lambda_{p} + \lambda_{o}) \end{bmatrix}$$

and,
$$\mathbf{M} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 \\ \mu & 0 & \cdots & 0 & 0 & 0 \\ 0 & \mu & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \mu & 0 & 0 \end{bmatrix},$$

and where $\hat{\mathbf{I}}$ denotes the diagonal matrix whose two first diagonal entries equal zero while the others equal 1.

3.3 Stability conditions

The above defined Markov processes are stable iff,

$$\lambda_o < \lambda_n$$

where λ_0 and ρ_0 equal the order arrival rate and the achievable throughput of the product queue, respectively.

Concerning the two-dimensional Markov process (with Poisson arrivals), the throughput of the product queue equals the outgoing rate of a $M/M/1/C_p$ queue and can explicitly be calculated as,

$$\lambda_n = (1 - q_0)\mu.$$

with the steady-state zero probability,

$$q_0 = \begin{cases} \frac{1-\gamma}{1-\gamma^{Cp+1}} & \text{if } \gamma \neq 1\\ \frac{1}{C_p+1} & \text{if } \gamma = 1 \end{cases}$$

where $\gamma = \frac{\lambda_p}{\mu}$.

To calculate the achievable throughput of the product queue for the three-dimensional Markov process (with interrupted Poisson arrivals), we consider a two-dimensional Markov process with state (m, k), m being the number of semi-finished products in the queue and k the modulating state. Here, the Markov process constitutes a homogeneous finite QBD with block size equal to two. The generator matrix of the Markov chain has the following block matrix representation,

$$\mathbf{Q} = \begin{bmatrix} \mathbf{P}_{00} \ \mathbf{A}_{o} \ \mathbf{0} \ \cdots \ \mathbf{0} \\ \mathbf{A}_{2} \ \mathbf{A}_{1} \ \mathbf{A}_{0} \ \cdots \ \mathbf{0} \\ \mathbf{0} \ \mathbf{A}_{2} \ \mathbf{A}_{1} \ \cdots \ \mathbf{0} \\ \vdots \ \vdots \ \vdots \ \ddots \ \mathbf{0} \\ \mathbf{0} \ \mathbf{0} \ \mathbf{0} \ \cdots \ \mathbf{P}_{MM} \end{bmatrix}$$

with,

$$\begin{aligned} \mathbf{A}_{0} &= \begin{bmatrix} 0 & 0 \\ 0 & \lambda_{p} \end{bmatrix}, \\ \mathbf{A}_{1} &= \begin{bmatrix} -(\alpha + \mu) & \alpha \\ \beta & -(\beta + \mu + \lambda_{p}) \end{bmatrix}, \quad \mathbf{A}_{2} &= \begin{bmatrix} \mu & 0 \\ 0 & \mu \end{bmatrix}, \\ \mathbf{P}_{MM} &= \begin{bmatrix} -(\alpha + \mu) & \alpha \\ \beta & -(\beta + \mu) \end{bmatrix}, \quad \mathbf{P}_{00} &= \begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix}. \end{aligned}$$

The stationary distribution for homogeneous finite QBD's can easily be solved by linear level reduction, see Latouche

and Ramaswami (1999). This then allows to calculate the maximal arrival rate λ_n of the orders. Having established the modelling assumptions and settled our notation, we now derive the steady-state probabilities and determine relevant performance measures for the decoupled inventory system.

3.4 Steady-state probabilities

As previously mentioned, we first consider a Markov process on the two-dimensional state space $\{(n,m) \mid n \geq 0, 0 \leq m \leq C_p\}$ and refer by level k to the set of states $\{(k,0), (k,1), \ldots, (k,C_p)\}$. A well-known method for finding the stationary distribution of QBD processes is the matrix-geometric method. With $\pi(n,m)$ the stationary probability of the process being in state (n,m), and using the vector notation $\pi_k = (\pi(k,0), \pi(k,1), \ldots, \pi(k,C_p))$, the probability vectors can be expressed as,

$$\boldsymbol{\pi}_k = \boldsymbol{\pi}_0 \mathbf{R}^k. \tag{2}$$

where $\boldsymbol{\pi}_0 = (\pi(0,0), \pi(0,1), \dots, \pi(0,C_p))$ and the socalled rate matrix R is the minimal non-negative solution of the non-linear matrix equation $\mathbf{R}^2\mathbf{M} + \mathbf{RL_p} + \mathbf{L_0} =$ **0**. Here, we compute the rate matrix by implementing the iterative algorithm in the book of Latouche and Ramaswami (1999). Note that this method can also be applied for the Markov process on the three-dimensional state space $\{(n,m,s) \mid n \geq 0, 0 \leq m \leq C_p, k = 0, 1\}$ where k denotes the state of production/delivery efficiency of the semi-finished products, as the phase set k is defined in the finite state space Ω (see section 2).

3.5 Performance measures

Once the steady state probabilities have been determined numerically, we can calculate a number of interesting performance measures for the decoupled inventory system. For ease of notation, we introduce the values $\pi(n,m) \doteq \sum_{k \in \Omega} \pi(n,m,k)$ for the interrupted Poisson process. For both Poisson arrivals and IPP, we can then define the marginal distributions of the number of semi-finished products $\pi^{(p)}(m) = \sum_n \pi(n,m)$ and of the number of orders $\pi^{(0)}(n) = \sum_m \pi(n,m)$ in their respective queues.

Note that as the queue of backlogged orders is infinite, the throughput of the decoupled inventory system η equals the order arrival rate λ_o and the effective load of the system ρ_{eff} equals $\frac{\lambda_o}{\mu}$. In addition, we have the following performance measures.

• The mean semi-finished product queue and the order backlog content: $E Q_p$ and $E Q_o$ respectively,

$$E Q_p = \sum_{m}^{C_p} \pi^{(p)}(m)m, \quad E Q_o = \sum_{n}^{\infty} \pi^{(o)}(n)n.$$

• The variance of the semi-finished product queue and the order backlog content: $\operatorname{Var} Q_p$ and $\operatorname{Var} Q_o$ respectively,

Var
$$Q_p = \sum_{m}^{C_p} \pi^{(p)}(m) m^2 - (E Q_p)^2$$

Var
$$Q_o = \sum_{n=1}^{\infty} \pi^{(o)}(n) n^2 - (E Q_o)^2$$

• The average waiting time W (calculated based on Little's theorem) the average time an order is waiting before being launched into production:

٦

$$\mathbf{W} = \frac{\mathbf{E} Q_o}{\lambda_o}$$

• The average lead time LT (calculated based on Little's theorem) is the average amount of time between the placement of an order and the completion to a finished product:

$$LT = W + \frac{1}{\mu}$$

• As the product queue has finite capacity, production prior to the decoupling stock may be blocked. This happens when there is a product arrival and the queue is full. Hence, blocking corresponds to the loss probability in the product buffer. The loss probability is most easily expressed in terms of the throughput. We have,

$$b_p = \frac{\lambda_p - \eta}{\lambda_p} = \frac{\lambda_p - \lambda_o}{\lambda_p}$$

In most manufacturing systems, there is a trade-off between achieving a high service level and minimising the inventory costs. Therefore, it is of main importance to understand the effects of the system parameters on both the average lead time (characterising the service level) and the average stock level (determining the inventory costs). Table 1 indicates the relation between the parameters of a decoupled inventory system and the two performance measures $E Q_p$ and LT. Intuitively, a higher arrival rate and stock capacity for the semi-finished products lead to a higher average stock and a lower average lead time while an increase in the order arrival rate decreases the average stock and increases the average lead time. Finally, a decrease in completion time on average reduces both the average stock of semi-finished products and the average lead time.

Table 1. Effects of the system parameters on $\mathbf{E} Q_p$ and \mathbf{LT}

	$\lambda_p(\nearrow)$	$C_p(\nearrow)$	$\lambda_o(\nearrow)$	$\mu(\nearrow)$
EQ_p	7	7	\searrow	\searrow
LT	\searrow	\searrow	7	\searrow

Obviously, inefficiency in the production process impacts negatively the performance of the decoupled inventory system. For example, a higher order arrival rate λ_o leads to a much higher average lead time for an unreliable production process. We now illustrate our approach by means of some numerical examples.

4. NUMERICAL EXAMPLES

In order to quantify the impact of inefficiency in the production process on the performance of a decoupled inventory system, we compare the decoupling stock with Poisson arrivals to corresponding decoupled inventory system with interrupted Poisson arrivals. The arrival interruptions account for inefficiency in the production process.

Preprint submitted to 14th IFAC Symposium on Information Control Problems in Manufacturing. Received November 18, 2011. We then use the following parameters to characterise the interrupted Poisson process,

$$\sigma = \frac{\beta}{\alpha + \beta} \,, \quad \kappa = \frac{1}{\alpha} + \frac{1}{\beta} \,, \quad \rho = \lambda_p \sigma \,.$$

Note that σ is the fraction of time that the interrupted Poisson process is active, the absolute time parameter κ is the average duration of an active and an inactive period, and ρ is the arrival load of the semi-finished products.



Fig. 2. Production inefficiency and a low buffer capacity result in a higher average lead time.

Figure 2 shows the average lead time versus the arrival load of the semi-finished products with buffer capacity C_p equal to 10 and 15 for Poisson arrivals as well as for interrupted Poisson arrivals. The order arrival rate λ_o equals 0.6 and service completion times are exponentially distributed with service rate μ equal to one for all curves. In addition, we set $\sigma = 0.8$ and $\kappa = 10$ for the interrupted Poisson processes. As expected, the average lead time decreases as the arrival load of the semi-finished products increases. Moreover, if more buffer capacity is available, it will also be used: the average lead time decreases for increasing values of C_p . Comparing interrupted Poisson and Poisson processes, burtiness in the production process has a negative impact on performance – there is on average more time required to deliver an order.

The mean number of waiting orders in a decoupled inventory system with buffer capacity equal to 10 and 15 for Poisson arrivals as well as for interrupted Poisson arrivals is depicted in Figure 3. As in the preceding figure, the interrupted Poisson process is characterised by $\sigma = 0.8$ and $\kappa = 10$. Also, the order arrival rate and the service rate are the same ($\lambda_o = 0.6$ and $\mu = 1$). As expected, the mean number of waiting orders decreases as the arrival load of semi-finished products increases. Moreover, to reduce the average number of waiting orders, more buffer capacity of semi-finished products is required for the case of the interrupted Poisson process than for the case of the Poisson process.

Figure 4 refers to the maximum value for λ_n such that the queue of the waiting orders remains stable, i.e. the number of waiting orders doesn't increase to an infinite number. The arrival load of the semi-finished products ρ equals 0.8 and the service rate μ equals 1. For the interrupted Poisson

5.0 $C_n = 10 \& Poisson$ $C_{-} = 15 \& Poisson$ 4.5 $C_p = 10 \& IPP$ $C_n = 15 \& IPP$ 4.0 3.5 3.0 2.5 2.0 15L 0.70 0.72 0.74 0.76 ρ

Fig. 3. Production inefficiency and a low buffer capacity result in a higher mean number of waiting orders.



Fig. 4. The limit values for λ_n are lower for an interrupted Poisson process than for a Poisson process.

process, we consider the same parameter values as in the previous figures ($\sigma = 0.8$ and $\kappa = 10$). As the figure shows, the maximum value for λ_n increases as the buffer capacity of the semi-finished products C_p increases. Indeed, the loss probability decreases as the capacity increases such that more orders per time unit can be satisfied. Furthermore, the limit value for λ_n is smaller for the interrupted Poisson process than for the Poisson process. Obviously, burstiness yields larger periods without arrivals during which the queue size of the waiting orders increases.

Finally, Figure 5 represents the trade-off between the maximum probability to have the lead time higher or equal to 30 (left side) and the average stock of the semifinished products (right side) versus the buffer capacity for Poisson arrivals as well as for interrupted Poisson arrivals. Note that we calculated the lead time distribution by using the one-sided chebyshev's inequality. Under the same parameter assumptions of Figure 4 and considering that $\lambda_n = 0.7$, we notice that the the maximum probability to have the lead time higher or equal to 30 decreases and the average stock increases as the inventory capacity increases, as expected. The maximum probability to have the lead time higher or equal to 30 is higher but the average stock of

Mean number of unsatisfied orders



Fig. 5. There is a trade-off between the lead time and the average stock of the semi-finished products.

the semi-finished products is lower for interrupted Poisson arrivals than for Poisson arrivals. Indeed, the results in Figure 6 for $C_p = 25$ show that the probability to have an empty buffer is higher for an interrupted Poisson process than for a poisson process. Also, the larger the average duration of an active and an inactive period for an IPP (determined by κ), the higher the zero probability.



Fig. 6. The zero probability is the highest for an interrupted Poisson process with $\kappa = 50$.

5. CONCLUSION

In this paper, we investigate the impact of inefficiency in a Make-to-Order system (with a decoupling stock) in a Markovian setting. Here, decoupling means that the completion of the semi-finished product is only possible when there is an order. These orders are backlogged and can be satisfied only when the semi-finished products are available. Therefore, the studied Make-to-Order system is modelled as a homogeneous birth-and-death process (QBD) and solved with matrix analytic methods.

As our numerical examples show, system performance is highly sensitive to the arrival process parameters. Inefficiency in the production process causes on average a longer lead time and a higher number of waiting orders, even with a higher stock capacity for the semi-finished products. Still, it is of main importance to determine the total cost of a decoupled inventory system. This will remain the focus of future work.

REFERENCES

- Bell, P. (1980). A decoupling inventory problem with storage capacity constraints. *Operations Research*, 28, 476–488.
- Bonomi, F. (1987). An approximate analysis for a class of assembly-like queues. Queueing Systems Theory and Applications, 289–309.
- Buchholz, P., Kemper, P., and Kriege, J. (2010). Multiclass Markovian arrival processes and their parameter fitting. *Performance Evaluation*, 67(11), 1092–1106.
- Chang, K. and Lu, Y. (2010). Queueing analysis on a single-station make-to-stock/make-to-order inventoryproduction system. *Applied Mathematical Modelling*, 34, 978–991.
- De Cuypere, E. and Fiems, D. (2011). Performance evaluation of a kitting process. In Proceedings of the 17th International Conference on analytical and stochastic modelling techniques and applications, Lecture Notes in Computer Science, volume 6751, 175–188. Venice, Italy.
- Fiems, D., Steyaert, B., and Bruneel, H. (2011). approach genetic to Markovian charac-А H.264 terisation of scalable video. Multi-URL media Tools andApplications, 1 - 22.http://dx.doi.org/10.1007/s11042-010-0713-x.
- Harrison, J. (1973). Assembly-like queues. Journal Of Applied Probability, 10(2), 354–367.
- Hoekstra, S., Romme, J., and Argelo, S. (1992). Integral logistic structures : developing customer-oriented goods flow. McGraw-Hill.
- Hopp, W.J. and Simon, J.T. (1989). Bounds and heuristics for assembly-like queues. *Queueing Systems*, 4, 137 – 156.
- Latouche, G. (1981). Queues with paired customers. Journal of Applied Probability, 18, 684–696.
- Latouche, G. and Ramaswami, V. (1999). Introduction to Matrix Analytic Methods in Stochastic Modeling. SIAM.
- Lipper, E. and Sengupta, B. (1986). Assembly-like queues with finite capacity: bounds, asymptotics and aproximations. *Queueing Systems: Theory and Applications*, 18, 684.
- Som, P. and Wilhelm, W. (1999). Analysis of stochastic assembly with GI-distributed assembly time. *INFORMS Journal on Computing*, 11, 104 – 116.
- Som, P., Wilhelm, W., and Disney, R. (1994). Kitting process in a stochastic assembly system. Queueing Systems, 17, 471 – 490.
- Soman, C., van Donk, D., and Gaalman, G. (2004). Combined make-to-order and make-to-stock in a food production system. *International Journal of Production Economics*, 90, 223 – 235.