MPE — Volume 1, pp. 225-243
Reprints available directly from the publisher
Photocopying permitted by license only

SYSTEM-THEORETIC ANALYSIS OF DUE-TIME PERFORMANCE IN PRODUCTION SYSTEMS*

DAVID JACOBS and SEMYON M. MEERKOV

Department of Electrical Engineering and Computer Science, The University of Michigan, Ann Arbor, MI 48109-2122

(Received 5 January 1995)

Along with the average production rate, the due-time performance is an important characteristic of manufacturing systems. Unlike the production rate, the due-time performance has received relatively little attention in the literature, especially in the context of large volume production. This paper is devoted to this topic. Specifically, the notion of due-time performance is formalized as the probability that the number of parts produced during the shipping period reaches the required shipment size. This performance index is analyzed for both lean and mass manufacturing environments. In particular, it is shown that, to achieve a high due-time performance in a lean environment, the production system should be scheduled for a sufficiently small fraction of its average production rate. In mass production, due-time performance arbitrarily close to one can be achieved for any scheduling practice, up to the average production rate.

AMS Nos.: 90B30, 60J20, 93A30
KEYWORDS: Manufacturing; performance measures; due-time performance index; lean and mass production systems; stochastic analysis; system-theoretic properties

1. INTRODUCTION AND PROBLEM FORMULATION

1.1. Manufacturing Considerations

The likelihood to meet the shipping schedule is an important feature of manufacturing systems with unreliable machines. Indeed, modern automotive component plants are often expected to produce the required number of specified units (e.g., car seats with specified trim, engines and transmissions with given characteristics) with four hours notice, referred to as the broadcast. Analogously, automotive assembly plants are often expected to build cars for a six-hour shipping schedule with specified car options produced. Due to the extremely large nomenclature of all items, these demands cannot be efficiently satisfied by maintaining finished-goods inventories, and therefore must be satisfied by the production system. Thus the due-time performance is of importance.

The likelihood to meet the shipping schedule can be made large by maintaining a large stock of finished goods. This is referred to as mass production. However, large finished-goods buffers lead to many negative consequences (see, for instance [1]–[3]). To avoid this, buffers must be small. This is called lean production. The question arises: How should the production be organized so that the likelihood to meet the

*This work has been supported by the National Science Foundation under grant No. DDM-9105086.

shipping requirements by an unreliable production system is high even in lean manufacturing? This is the question addressed in this paper.

1.2. Systems Considerations

The likelihood to meet the shipping schedule can be conceptualized as a functional defined on the trajectories of a stochastic system. We refer to this functional as the due-time performance measure (DTP). In this paper, we address the following questions concerning this functional:

(i) Given a production system, how can DTP be defined and evaluated?
(ii) What are system-theoretic properties of this functional in both lean and mass manufacturing?
(iii) How should the parameters of a production system be chosen so that DTP is as high as desired?

1.3. Problem Formulation

Let T and R be the shipping interval and the shipment size, respectively.

DEFINITION 1.1. The *due-time performance index* is defined as

$$\mathrm{DTP} = \mathrm{Prob}\,\{n(t) = R, t \leq T\}$$

where $n(t)$ is the number of parts produced during the period $[0, t]$ in the steady-state operation of an ergodic production system.

Remark 1.1. Let ρ be the average production rate. Then, obviously,

$$R_{max} = \rho T.$$

All other shipments can be characterized by a parameter L, referred to as the *load factor*:

$$R = LR_{max} = L\rho T, \qquad L \in [0, 1].$$

Thus,

$$\mathrm{DTP} = \mathrm{Prob}\,\{n(t) = L\rho T, t \leq T\}.$$

The values of DTP parameterized by L will be considered throughout this paper.

Remark 1.2. When T is large, the manufacturing is refered to as mass production; when T is small, the production is lean.

Problem 1.1. Given a manufacturing system, analyze DTP, both qualitatively and quantitatively, as a function of L, ρ, and T.

Problem 1.2. Given a manufacturing system and parameters ρ, T, and ϵ, $0 < \epsilon < 1$, determine the largest $L^* \in [0, 1]$ so DTP $\geq 1 - \epsilon$ for all $L \leq L^*$.

Problems 1.1 and 1.2 are addressed in sections 2 and 3, respectively. The conclusions are found in section 4. The proofs are given in the Appendices A-C.

2. ANALYSIS

2.1. The Model

As it is obvious from its definition, the calculation of DTP requires the knowledge of the probability distribution of $n(t)$. In general, the calculation of this distribution is a formidable task. Therefore, we analyze here a very simple production-shipment system for which such a calculation can be carried out, and then, following Gershwin's suggestion [4], investigate the possibility of DTP evaluation using the Gaussian approximation.

Production Subsystem:

(i) The production system has two states: "up" (operational) and "down" (non operational). When up, the production system is producing parts at the rate c parts/min. When down, the production rate is zero.
(ii) The production system's up time is a random variable, $T_{up} \in [0, \infty)$, distributed exponentially with parameter p; thus, the average up time is $\overline{T}_{up} = 1/p$ min.
(iii) The production system's down time, $T_{down} \in [0, \infty)$, is an exponentially distributed random variable with parameter r; thus, the average down time is $\overline{T}_{down} = 1/r$ min.

Shipment Subsystem:

(iv) Parts produced are immediately prepared for shipment.
(v) Each shipment is of the size R parts and the shipping period is T min, $T \in (0, \infty)$.

Remark 2.1. The two state production systems are quite typical in large-volume production, where the machines are either operational and, therefore, produce at a fixed rate, or broken down, starved, or blocked, and, therefore, do not produce at all.

Remark 2.2. The assumption that the up and down times are distributed exponentially is introduced for analytical convenience. Real systems may or, most often, may not satisfy this assumption. We believe, however, that the insight developed here is of significance even when the distributions of T_{up} and T_{down} are different. We hope also that the results are robust enough to tolerate perturbations to exponential distributions.

Remark 2.3. In reality, the parts flow is discrete. Assumption (i), however, postulates a continuous flow. Although the discrete part model also can be analyzed, the continuity assumption makes the calculations more transparent. Moreover, we believe that this assumption does not cause an extreme violence to the system. Indeed, in most automotive manufacturing processes, the cycle time, that is, the time necessary to accomplish a machining or an assembly operation, is of the order of magnitude of seconds. The machines' up and down times are, typically, hours. Therefore, during the up time, a "continuous" stream of parts is observed. This, coupled with the fact that in large-volume production thousands of parts are "flowing" simultaneously through the manufacturing facility, makes it impossible and, moreover, unnecessary to keep track of each particular part and, thus, justifies the rate-based assumption (i). Note that rate-based models have been considered in other studies as well ([5]–[8]).

Remark 2.4. Assumption (iv) implies that no finished goods buffers are available. Although this may be the case in some situations, many plants still have some finished-goods inventories. In this case, the results derived here can be viewed as a lower bound on attainable DTP.

Remark 2.5. The maximum shipment size, $R_{\max}$, is constrained by the production system parameters, c, p, r, and the shipping period, T, as follows: Due to assumptions (i)–(iii), it is easy to show that the average production rate, ρ, is:

$$\rho = c \frac{r}{r+p} \quad \text{[parts/min]}.$$

Let

$$P_{up}(t) = \text{Prob}\,\{\text{production system is up at time } t\},$$

$$P_{down}(t) = \text{Prob}\,\{\text{production system is down at time } t\}.$$

Due to the ergodicity of the production system (i)–(iii), the following limits exist:

$$P_{up} = \lim_{t \to \infty} P_{up}(t), \tag{1}$$

$$P_{down} = \lim_{t \to \infty} P_{down}(t).$$

Therefore, Definition 1.1 is applicable and DTP is calculated next.

2.2. Exact Calculation

THEOREM 2.1 *Under assumptions (i)–(v),*

$$\text{DTP}\,(c, p, r, T, L) = \left[\frac{r}{r+p}\right] e^{-[pF+r(1-F)]T} \sum_{i=0}^{\infty} \sum_{j=0}^{i} \frac{(pFT)^j}{j!} \frac{[r(1-F)T]^i}{i!} \tag{2}$$

$$+ \left[\frac{p}{r+p}\right]\left[1 - e^{-[pF+r(1-F)]T} \sum_{j=0}^{\infty} \sum_{i=0}^{j} \frac{(pFT)^j}{j!} \frac{[r(1-F)T]^i}{i!}\right],$$

where $F = L\,(r/(r+p))$, and the infinite sums converge.

Proof. See Appendix A.

The behavior of DTP as a function of L is illustrated in Figure 1 for the production system defined by parameters $c = 10$, $p = 0.15$, and $r = 0.30$. These parameters correspond to a machining department of an automotive component plant which we have studied. The number of terms used to approximate the infinite sums in (2) was chosen so that the inclusion of additional terms resulted in practically no change of the curves. For the sake of mathematical completeness, Figure 1 uses the load factor, L, defined on $[0, (p + r)/r]$ rather than on $[0, 1]$.

As it follows from this figure, DTP is close to 1 for all $L \in (0, 1)$ if the shipping period T is much larger than the production system's "up–down" period, $T_{u\text{-}d}$, defined as $\overline{T}_{up} + \overline{T}_{down}$. For the example of Figure 1, $T_{u\text{-}d} = 10$ min. Thus, if the production is "sufficiently mass" in the sense that

$$T >> \overline{T}_{up} + \overline{T}_{down}$$

the scheduling of the production for almost the average production rate, that is, for L close to one, would still result in a good DTP. However, when

$$T \approx \overline{T}_{up} + \overline{T}_{down},$$

the load factor must be sufficiently small in order to meet the shipping schedule with large probability. The choice of such a load factor is addressed in section 3.

2.3. An Approximation

As it has been mentioned before, exact calculation of DTP is possible only for very simple manufacturing systems. On the other hand, the behavior of $n(t)$ is defined by many "independent" events and, following Gershwin's suggestion [4], may be characterized by a Gaussian distribution. Since system (i)–(v) admits an exact analysis, it is of interest to compare it with a Gaussian approximation. The variance of $n(T)$ necessary for this evaluation is given below:

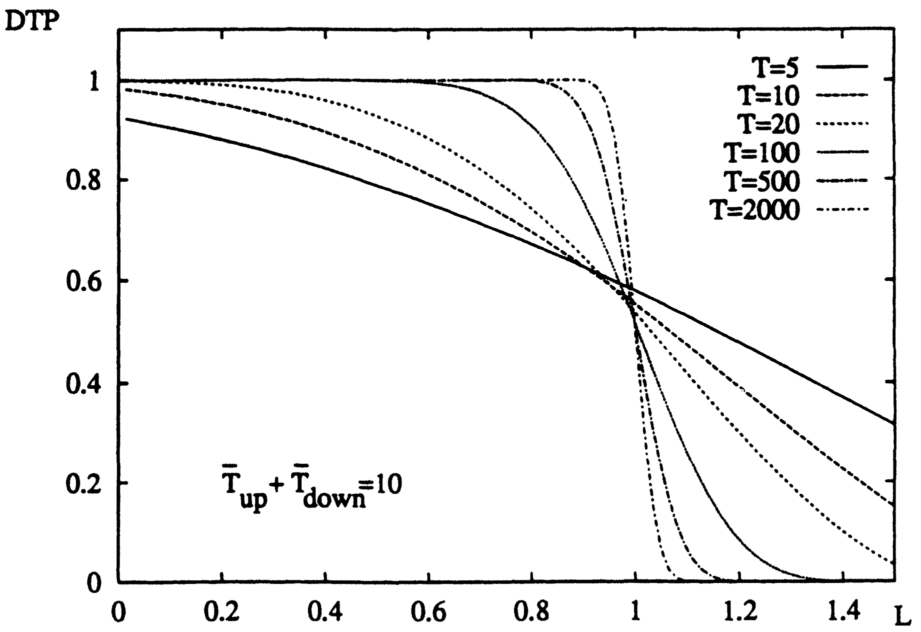

Figure 1 DTP as a function of L.

THEOREM 2.2 *Under assumptions (i)–(v),*

$$\mathrm{E}\{n(T)\} = c\frac{r}{r+p}T \tag{3}$$

$$\mathrm{Var}\{n(T)\} = \frac{2c^2rpT}{(r+p)^3} - \frac{2c^2rp}{(r+p)^4}\left[1 - e^{-(r+p)T}\right]$$

Proof. See Appendix B.

Given this expectation and variance, an estimate of DTP can be calculated from

$$\mathrm{DTP}_{est} = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{L\rho T}^{\infty} e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2} dx, \tag{4}$$

where $\mu = \mathrm{E}\{n(T)\}$ and $\sigma^2 = \mathrm{Var}\{n(T)\}$ are the expectation and variance of $n(T)$, respectively.

At present, we do not have an analytical bound on the accuracy of the estimate (3)–(4). Therefore, a numerical investigation has been carried out by comparing DTP values derived using (2) and (3)–(4). The results are illustrated in Figure 2 for the same production system as in Figure 1. As it follows from this figure, approximation (3)–(4) is quite good for T's which are more than five times greater than T_{u-d}. For $T >> T_{u-d}$, the two curves are practically indistinguishable. Thus, when the production is not "very" lean, the Gaussian approximation of DTP can be utilized. However, in a very lean environment, this approximation may lead to errors as high as 5% and more.

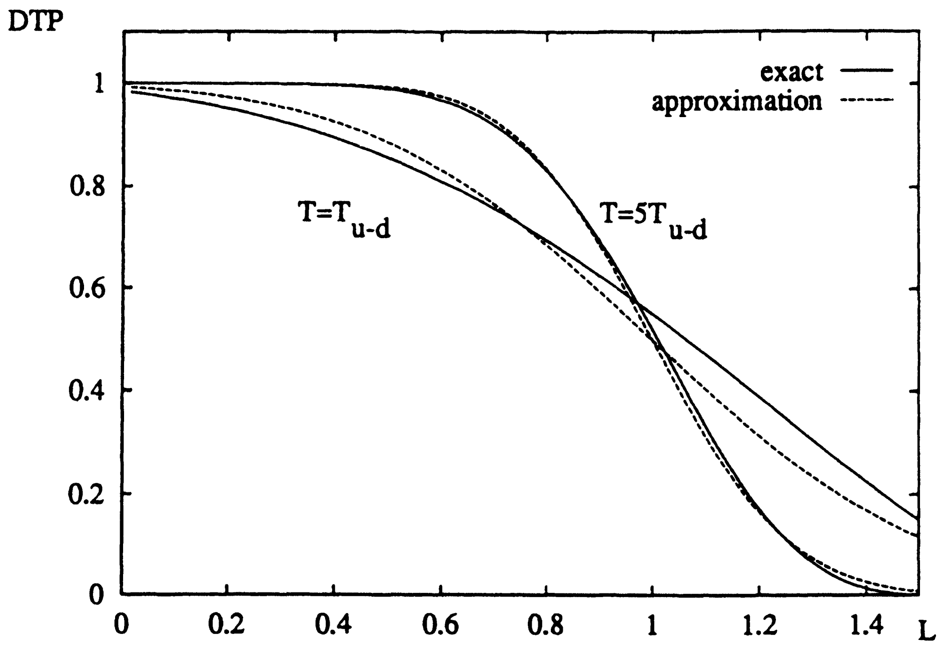

Figure 2 Exact and approximate DTP calculations.

3. DESIGN

3.1. System-Theoretic Properties

In this section we describe the monotonicity properties of DTP with respect to the production–shipment system's parameters and analyze DTP limits when the shipping period and the load factor tend to their minimum and maximum values.

THEOREM 3.1 *Under assumptions (i)–(v)*

$$\lim_{T\to 0} \mathrm{DTP}\,(L, T) = \frac{r}{r+p}, \qquad \forall\, L \in \left[0, \frac{1}{\rho}\right]$$

$$\lim_{T\to\infty} \mathrm{DTP}\,(L, T) = \begin{cases} 1 & \text{for } 0 < L < 1 \\ 0 & \text{for } L > 1. \end{cases}$$

Proof. See Appendix C.

Thus, in an "extremely" lean environment, DTP is independent of L and approaches the normalized (by c) average production rate. Since this value can be quite small, Theorem 3.1 shows that the "extremely" lean operation of an unreliable system cannot produce acceptable DTP. To improve the performance, either p must be decreased or r must be increased. Since $\partial DTP(L, 0)/\partial r = p/(r + p)^2$ and $\partial DTP(L, 0)/\partial p = -r/(r + p)^2$, the largest DTP improvement in an extremely lean system occurs when r is increased, if $p > r$, or p is decreased, if $p < r$. When $p = r$ equal improvements take place when either p or r are changed appropriately. Note that a decrease in p amounts to improvement in the preventative maintenance; an increase in r amounts to improvements in the skilled trades (repair) operations.

THEOREM 3.2 *Under assumptions (i)–(v),*

$$\lim_{L\to 0} \mathrm{DTP}\,(L, T) = 1 - \frac{p}{r+p}\, e^{-rT}.$$

Proof. Follows immediately from (2) with $F = 0$.

Thus, for any shipping period $T \in (0, \infty)$, the DTP is less than 1 for any load factor, even for L as small as desired. Also, no production–shipment system (i)–(v) can ensure DTP greater than indicated in Theorem 3.2.

THEOREM 3.3 *Under assumptions (i)–(v), DTP = DTP(p, r, L, T) is monotonically decreasing in p and L, and monotonically increasing in r.*

Proof. See Appendix C.

This result is quite expected, however monotonicity with respect to T, which also could have been expected, does not take place. This can be explained as follows: Consider a system which is up most of the time ($r/r + p \approx 1$), and which is operating with a load factor L slightly smaller than one. If the shipping period T is increased from zero to infinity, the DTP passes through three regimes. For $T \approx 0$, from Theorem 3.1 it follows that DTP $\approx r/(r + p) \approx 1$. As T increases, the satisfaction of the schedule requires that the system be up at the start of the shipping period and remain up for most of it. Therefore, DTP decreases. Indeed, from approximation (3)–(4) it follows that DTP decreases to approximately $\frac{1}{2}$. However, in the limit $T \to \infty$, from Theorem 3.1 DTP $\to$ 1, and hence DTP must increase for large values of T.

The qualitative properties, specified by Theorems 3.1–3.3, can be used as a guide for the choice of L so that DTP reaches a desired value. This is described next.

3.2. Existence of Solution

THEOREM 3.4 *Under assumptions (i)–(vi), the design problem 1.2 has a unique solution if* $\epsilon > (p/(r+p))e^{-rT}$. *This solution is unique.*

Proof. The existence follows directly from Theorem 3.2. The uniqueness is a consequence of the monotonicity of DTP with respect to L (Theorem 3.3). Indeed, since $DTP(L)$ is monotonically decreasing for $L \in [0, (r+p)/r]$, for any given c, p, r and T, equation

$$\left[\frac{r}{r+p}\right] e^{-[pF+r(1-F)]T} \sum_{i=0}^{\infty} \sum_{j=0}^{i} \frac{(pFT)^j}{j!} \frac{[r(1-F)T]^i}{i!}$$

$$+ \left[\frac{p}{r+p}\right]\left[1 - e^{-[pF+r(1-F)]T} \sum_{j=0}^{\infty} \sum_{i=0}^{j} \frac{(pFT)^j}{j!} \frac{[r(1-F)T]^i}{i!}\right] = 1 - \epsilon$$

has a unique solution, $F^*(\epsilon)$, for all $\epsilon \in [(p/(r+p))\, e^{-rT}, 1]$. Then $L^*(\epsilon)$ can be calculated then from

$$L^*(\epsilon) = F^*(\epsilon)(1 + \frac{p}{r}).$$

The behavior of $L^*(\epsilon, c, p, r, T)$ is illustrated in Figure 3 for the same production system as in Figures 1 and 2. As it follows from this figure, when ϵ is, for instance, 0.1, the DTP requirement is satisfied for L's as high as 0.95, if the production is mass, and as low as 0.1, if the production is lean. Thus, to achieve high DTP in a lean operation, the shipment should be scheduled for a sufficiently small fraction of the system capacity. This conclusion is in agreement with the report [9] where it is indicated that the Japanese semiconductor manufacturers plan to load their production systems, by the end of this century, only up to 10%–15% of their capacity, in order to rapidly satisfy their customers' demands. Analogous practices have been reported by the best U.S. semiconductor manufacturers [10].

4. CONCLUSION

In this paper, a formalization of the due-time performance measure for large-volume production systems has been introduced and analyzed. Both quantitative and qualitative properties have been investigated. It has been shown that no perfect DTP can be provided by an unreliable system. However, DTP arbitrarily close to perfect can be attained in a mass-production environment. No such possibility exists in a lean mode. Nevertheless,

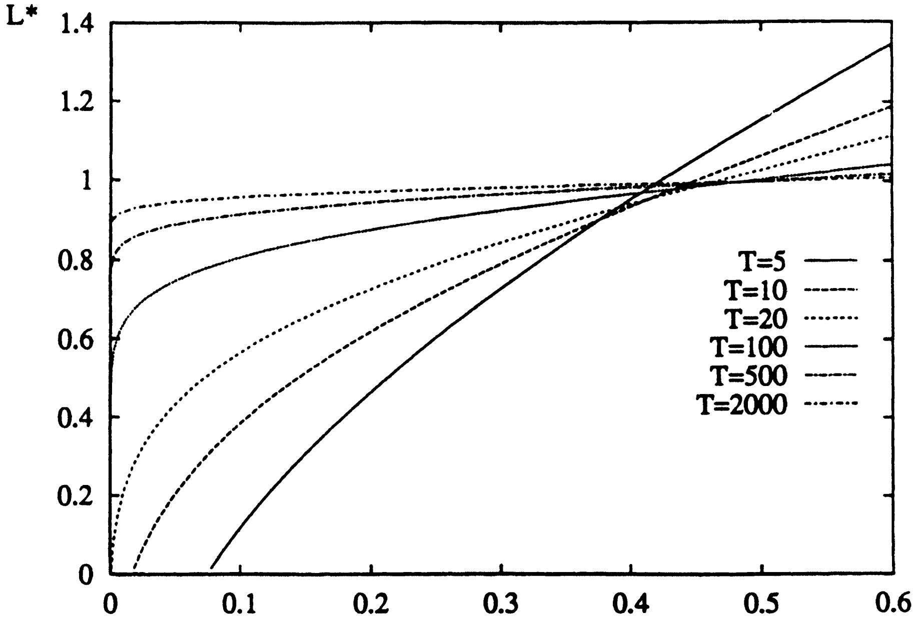

Figure 3 L^* as a function of ϵ.

even in lean production, a sufficiently high DTP (commensurable with the system reliability as stated in Theorem 3.4) can be attained by choosing an appropriate load factor.

These conclusions seem to explain the following observation: During the last 5–10 years, many of the U.S. automotive manufacturers have been transferred from mass to lean production. Almost universally, it has been observed that the due-time performance measure has deteriorated dramatically. In view of the results described here this is not surprising, since, again almost universally, this change has not been accompanied by a change in scheduling practices, and the newly introduced lean systems have been scheduled for shipments defined by the average production rate. In this situation, nothing else but a dramatic drop in DTP is possible.

Lean production cannot be viewed as a tool of producing as much as mass production; rather, it is a tool of producing less but in a more efficient manner, as articulated, for example, in [3].

APPENDIX A. PROOFS FOR SECTION 2.2

Proof of Theorem 2.1 First, we calculate P_{up} and P_{down}, defined in (1). As it follows from assumptions (i)–(iii),

$$P_{up}(t + \Delta t) = P_{up}(t)e^{-p\Delta t} + P_{down}(t)\,(1 - e^{-r\Delta t}) + O(\Delta t^2)$$
$$= P_{up}(t)\,(1 - p\Delta t) + P_{down}(t)\,(r\Delta t) + O(\Delta t^2).$$

In the limit as $t \to \infty$,

$$P_{up} = P_{up}(1 - p\Delta t) + P_{down}(r\Delta t) + O(\Delta t^2),$$
$$p\Delta t P_{up} = r\Delta t P_{down} + O(\Delta t^2).$$

Since $P_{up} + P_{down} = 1$, in the limit $\Delta t \to 0$, we obtain

$$P_{up} = \frac{r}{r+p} \tag{5}$$
$$P_{down} = \frac{p}{r+p}.$$

Next, we consider system (i)–(v) in the steady state defined by (1), (5) and calculate the DTP as a sum of two terms: the first one, DTP_{up}, is calculated under the assumption that at $t = 0$ the production system is up; the second, DTP_{down}, is calculated under the assumption that it is down at $t = 0$. Then,

$$\text{DTP} = \left[\frac{r}{r+p}\right] \text{DTP}_{up} + \left[\frac{p}{r+p}\right] \text{DTP}_{down}. \tag{6}$$

To calculate these two terms, we need additional variables. As it follows from assumptions (i)–(iii), during its operation the production system changes its states between "up" and "down." Let $U(j)$ be the duration of the jth uptime, and let $D(j)$ be the duration of the jth downtime. Due to assumptions (ii) and (iii), $U(j)$ and $D(j)$ are distributed exponentially with parameters p and r, respectively. Define

$$D_j = D(1) + \cdots + D(j)$$
$$U_j = U(1) + \cdots + U(j),\ j \geq 1.$$

The cumulative distributions of U_j and D_j are calculated in [11] to be:

$$\text{Prob}\,\{U_j \leq x\} = 1 - e^{-px}\left\{ 1 + px + \ldots + \frac{(px)^{j-1}}{(j-1)!} \right\}$$

$$\text{Prob}\,\{D_j \leq x\} = 1 - e^{-rx}\left\{ 1 + rx + \ldots + \frac{(rx)^{j-1}}{(j-1)!} \right\} \tag{7}$$

Finally, let $V(j)$, $j \geq 1$, denote the event that the parts required for the shipment are completed during the jth uptime and that they are completed in time for shipment, given that the system was up at time $t = 0$. Then $\text{Prob}\{V(j)\} = \text{Prob}\{n(t) = L\rho T,\ t \leq T$, and t occurs during the jth uptime $\mid$ system is up at $t = 0\}$.

Event $V(j)$, $j \geq 2$, occurs if and only if

$$(\alpha) \quad U_{j-1} < FT$$

$$(\beta) \quad U_j \geq FT$$

$$(\gamma) \quad D_{j-1} \leq (1 - F)\, T,$$

where $F = L(r/(r + p))$, and F can be interpreted as the fraction of time the system must be up for $t \in [0, T]$ to satisfy the shipping schedule. Condition (α) implies that the required production was not completed during the $(j - 1)th$ uptime; condition (β) implies that the production is completed by the end of the jth uptime; conditions (β) and (γ) together imply that the production is completed at time $t \leq T$. Based on (7), the joint probability of $U_{j-1} < FT$ and $U_j \geq FT$ can be calculated to be:

$$\text{Prob}\{U_{j-1} < FT \text{ and } U_j \geq FT\} = \frac{(pFT)^{j-1}}{(j-1)!}\, e^{-pFT}.$$

Therefore,

$$\text{Prob}\,\{V(j)\} = \text{Prob}\{U_{j-1} < FT \text{ and } U_j \geq FT\} \times \text{Prob}\,\{D_{j-1} \leq (1 - F)T \mid U_{j-1} < FT \text{ and } U_j \geq FT\} = e^{-pFT}\left[\frac{(pFT^{j-1}}{(j-1)!}\right]\left[1 - e^{-r(1-F)T}\left\{1 + r\,(1 - F)T + \cdots + \frac{(r(1 - F)T)^{j-2}}{(j-2)!}\right\}\right], \quad j \geq 2. \tag{8}$$

The probability of $V(1)$ is the probability that $U_1 \geq FT$, and is, therefore, given by

$$\text{Prob}\{V(1)\} = 1 - \text{Prob}\{U(1) < FT\} = e^{-pFT}. \tag{9}$$

Using (8) and (9), DTP_{up} can now be calculated as the probability that the parts required for shipment are completed during any uptime before the shipping time T, that is,

$$\begin{aligned}\text{DTP}_{up} &= \sum_{j=1}^{\infty}\text{Prob}\,\{V(j)\} \\ &= e^{-pFT} + e^{-pFT}\,[\,pFT] - e^{-pFT}[\,pFT]e^{-r(1-F)T} + \cdots \\ &\quad + e^{-pFT}\left[\frac{(pFT)^j}{j!}\right] - e^{-pFT}\left[\frac{(pFT)^j}{j!}\right]e^{-r(1-F)T}(1 + r(1 - F)T + \cdots \\ &\quad + \frac{(r(1 - F)T)^{j-1}}{(j-1)!}) + \cdots.\end{aligned}$$

This can be re-arranged as

$$\mathrm{DTP}_{up} = e^{-[pF+r(1-F)]T} \sum_{i=0}^{\infty} \sum_{j=0}^{i} \frac{(pFT)^j}{j!} \frac{(r(1-F)T)^i}{i!} \tag{10}$$

Similar analysis applied to DTP_{down} yields:

$$\mathrm{DTP}_{down} = 1 - e^{-[pF+r(1-F)]T} \left[\sum_{j=0}^{\infty} \sum_{i=0}^{i} \frac{(pFT)^j}{j!} \frac{(r(1-F)T)^i}{i!} \right] \tag{11}$$

Then, the expression for DTP, given in Theorem 2.1, is obtained directly from (6), (10), and (11).

The convergence of the infinite sums follows immediately from the ratio convergence test applied to the inner summations in (10) and (11).

APPENDIX B. PROOFS FOR SECTION 2.3

Proof of Theorem 2.2. Taking into account that

$$\mathrm{Prob}\,\{n(t) = ct \text{ and the system is up at time } t\} = \frac{r}{r+p} e^{-pt} \tag{12}$$

$$\mathrm{Prob}\,\{n(t) = 0 \text{ and the system is down at time } t\} = \frac{p}{r+p} e^{-rt}$$

introduce functions $\pi_{up}(n, t)$ and $\pi_{down}(n, t)$, $0 < n < ct$, $t > 0$, as follows:

$$\mathrm{Prob}\,\{n(t) \leq n, \text{ and the system is up at time t}\} = \int_0^n \pi_{up}(s, t) ds \tag{13}$$

$$\mathrm{Prob}\{n(t) \leq n, \text{ and the system is down at time t}\} = \int_0^n \pi_{down}(s, t) ds + \frac{p}{r+p} e^{-rt}.$$

The density functions $\pi_{up}(n, t)$ and $\pi_{down}(n, t)$ are smooth. Indeed, for $n \in (0, ct)$, from (13)

$$\int_0^n \pi_{up}(s, t)\, ds = 1 - \mathrm{DTP}_{up}(L, t), \tag{14}$$

where $L = n(r + p)/crt$. Since, as it follows from (10), the right-hand side of (14) is a smooth function of L (note that $F = L\, r/(r + p)$,

$$\pi_{up}(n, t) = -\frac{\partial \mathrm{DTP}_{up}(L, t)}{\partial L} \left[\frac{r+p}{ctr} \right]$$

$$\pi_{down}(n, t) = -\frac{\partial \mathrm{DTP}_{down}(L, t)}{\partial L} \left[\frac{r+p}{ctr} \right].$$

Therefore, $\pi_{up}(n, t)$ and $\pi_{down}(n, t)$ are smooth for $n \in (0, ct)$.

Due to assumptions (ii) and (iii), $\pi_{up}(n, t)$ and $\pi_{down}(n, t)$ obey the following evolution equations:

$$\pi_{up}(n, t+\Delta t) = \pi_{up}(n - c\Delta t, t)e^{-p\Delta t} + \int_t^{t+\Delta t} \pi_{down}(n - c(t+\Delta t - \tau), t) r e^{-r(\tau - t)}\, d\tau + O(\Delta t^2)$$

$$\pi_{down}(n, t+\Delta t) = \pi_{down}(n, t)e^{-r\Delta t} + \int_t^{t+\Delta t} \pi_{up}(n - c(\tau - t), t) p e^{-p(\tau - t)}\, d\tau + O(\Delta t^2)$$

$$c\Delta t < n < ct.$$

With the accuracy of $O(\Delta t^2)$, these equations can be rewritten as

$$\pi_{up}(n, t+\Delta t) = (1 - p\Delta t)\pi_{up}(n - c\Delta t, t) + (r\Delta t)\,\pi_{down}(n, t) + O(\Delta t^2)$$

$$\pi_{down}(n, t+\Delta t) = (p\Delta t)\pi_{up}(n - c\Delta t, t) + (1 - r\Delta t)\,\pi_{down}(n, t) + O(\Delta t^2)$$

$$c\Delta t < n < ct. \tag{15}$$

Taking into account (12), introduce now the following expectations:

$$E_{up}(t) = E_{up}\{n(t)\} = \int_0^{ct} n\, \pi_{up}(n, t) dn + \frac{ctr}{r+P} e^{-pt}$$

$$E_{down}(t) = E_{down}\{n(t)\} = \int_0^{ct} n \pi_{down}(n, t) dn \tag{16}$$

and

$$S(t) = E\{n^2(t)\} = \int_0^{ct} n^2[\pi_{up}(n,t) + \pi_{down}(n, t)]\, dn + \frac{c^2t^2r}{r+p} e^{-pt}. \tag{17}$$

Then

$$E(t) = E\{n(t)\} = E_{up}(t) + E_{down}(t) \tag{18}$$

and

$$V(t) = \mathrm{Var}\{n(t)\} = S(t) - (E(t))^2. \tag{19}$$

Next, using the evolution equations, we derive and solve the differential equations for $E_{up}(t)$, $E_{down}(t)$, and $S(t)$, and thus obtain $\mathrm{Var}\{n(T)\}$. From (15) and (16) we write:

$$
\begin{aligned}
E_{up}(t+\Delta t) &= \int_0^{c(t+\Delta t)} n\pi_{up}(n, t+\Delta t)dn + \frac{c(t+\Delta t)r}{r+p} e^{-p(t+\Delta t)} \\
&= \int_0^{c\Delta t} n\pi_{up}(n, t+\Delta t)dn + \int_{c\Delta t}^{ct} n\,[(1-p\Delta\, t)\,\pi_{up}(n - c\Delta\, t, t) \\
&\qquad + (r\Delta t)\pi_{down}(n, t)]\, dn \\
&\quad + \int_{ct}^{c(t+\Delta t)} n\pi_{up}(n, t+\Delta t)\, dn + \frac{c(t+\Delta t)r}{r+p} e^{-p(t+\Delta t)} + O\,(\Delta\, t^2).
\end{aligned} \tag{20}
$$

Evaluating each of the terms in the right hand side of (20), we obtain:

$$
\left| \int_0^{c\Delta t} n\pi_{up}(n, t+\Delta t)\, dn \right| \leq (c\,\Delta\, t) \max_{n\in[0,c\Delta t]} n\pi_{up}(n, t+\Delta t) \backsim O(\Delta\, t^2) \tag{21}
$$

Introducing $n_1 = n - c\Delta t$, for the second integral we have:

$$
\begin{aligned}
&\int_{c\Delta t}^{ct} n[(1-p\Delta t)\,\pi_{up}(n - c\Delta t, t) + (r\Delta t)\pi_{down}(n, t)]dn \\
&= (1-p\Delta t) \int_0^{c(t-\Delta t)} (n_1 + c\Delta t)\,\pi_{up}(n_1, t)dn_1 + (r\,\Delta t) \int_{c\Delta t}^{ct} n\pi_{down}(n, t)\, dn \\
&= (1-p\Delta t) \int_0^{c(t-\Delta t)} n_1\,\pi_{up}(n_1, t)dn_1 + (c\,\Delta\, t) \int_0^{c(t-\Delta t)} \pi_{up}(n_1, t)\, dn_1 \\
&\quad + (r\,\Delta t) \int_{c\Delta t}^{ct} n\pi_{down}(n, t)\, dn + O(\Delta t^2) \\
&= (1-p\Delta t) \left[\int_0^{ct} n_1\,\pi_{up}(n_1, t)dn_1 - \int_{c(t-\Delta t)}^{ct} n_1\,\pi_{up}(n_1, t)\, dn_1 \right] \\
&\quad + (c\Delta t) \left[\int_0^{ct} \pi_{up}(n_1, t)dn_1 - \int_{c(t-\Delta t)}^{ct} \pi_{up}(n_1, t)\, dn_1 \right] \\
&\quad + (r\Delta t) \left[\int_0^{ct} n\,\pi_{dowm}(n, t)dn - \int_0^{c\Delta t} n\pi_{down}(n, t)\, dn \right] + O(\Delta t^2) \\
&= (1-p\Delta t) \int_0^{ct} n_1\,\pi_{up}(n_1, t)dn_1 - \int_{c(t-\Delta t)}^{ct} n_1\,\pi_{up}(n_1, t)\, dn_1 \\
&\quad + (c\Delta t) \int_0^{ct} \pi_{up}(n_1, t)dn_1 + (r\Delta t) \int_0^{ct} n\,\pi_{down}(n, t)\, dn + O(\Delta t^2) \\
&= (1-p\Delta t) \left(E_{up}(t) - \frac{ctr}{r+p} e^{-pt} \right) - \int_{c(t-\Delta t)}^{ct} n_1\,\pi_{up}(n_1, t)dn_1 \\
&\quad + (c\,\Delta t) \left(P_{up} - \frac{r}{r+p} e^{-pt} \right) + (r\Delta t)\, E_{down}(t) + O(\Delta t^2),
\end{aligned} \tag{22}
$$

where the last expression was obtained using (12), (13), and (16). For the last integral in (20), using the smoothness of $\pi_{up}(n, t)$, $n<ct$, we write:

$$\int_{ct}^{c(t+\Delta t)} n\pi_{up}(n,t+\Delta t)\,dn = \int_{c(t-\Delta t)}^{ct} (n_1 + c\Delta t)\pi_{up}(n_1 + c\Delta t, t+\Delta t)dn_1$$
$$= \int_{c(t-\Delta t)}^{ct} n_1\pi_{up}(n_1, t)dn_1 + O(\Delta t^2). \tag{23}$$

Substituting (21)–(23) back into (20), we obtain:

$$E_{up}(t+\Delta t) = (1 - p\Delta t)\left(E_{up}(t) - \frac{ctr}{r+p}e^{-pt}\right) + (c\Delta t)\left(p_{up} - \frac{r}{r+p}e^{-pt}\right)$$
$$+ (r\Delta t)\,E_{down}(t) + \frac{c(t+\Delta t)\,r}{r+p}e^{-p(t+\Delta t)} + O(\Delta t^2)$$
$$= (1 - p\Delta t)\,E_{up}(t) + (r\Delta t)\,E_{down}(t) + (c\Delta t)\,P_{up} + O(\Delta t^2).$$

Using (5) and rearranging,

$$\frac{E_{up}(t+\Delta t) - E_{up}(t)}{\Delta t} = -pE_{up}(t) + rE_{down}(t) + \frac{cr}{r+p} + O(\Delta t),$$

and, therefore, in the limit $\Delta t \to 0$, we obtain

$$\frac{dE_{up}(t)}{dt} = -pE_{up}(t) + rE_{down}(t) + \frac{cr}{r+p}, \quad E_{up}(0) = 0, \tag{24}$$

where the initial condition follows from (16).

Similar arguments applied to $E_{down}(t)$ and $S(t)$ yield

$$\frac{dE_{down}(t)}{dt} = pE_{up}(t) - rE_{down}(t), \quad E_{down}(0) = 0 \tag{25}$$

$$\frac{dS(t)}{dt} = 2cE_{up}(t), \qquad S(0) = 0. \tag{26}$$

The unique solution of (24) and (25) is:

$$E_{up}(t) = \left[\frac{cr^2}{(r+p)^2}\right]t + \left[\frac{crp}{(r+p)^3}\right](1 - e^{-(r+p)t})$$
$$E_{down}(t) = \left[\frac{crp}{(r+p)^2}\right]t - \left[\frac{crp}{(r+p)^3}\right](1 - e^{-(r+p)t}). \tag{27}$$

Then, the solution of (26) is

$$S(t) = \left[\frac{c^2r^2}{(r+p)^2}\right] t^2 + \left[\frac{2c^2rp}{(r+p)^3}\right] t - \left[\frac{2c^2rp}{(r+p)^4}\right] (1 - e^{-(r+p)t}). \tag{28}$$

The statement of Theorem 2.2 follows directly from (18), (19), (27), and (28).

APPENDIX C. PROOFS FOR SECTION 3.1

Proof of Theorem 3.1. The first statement of the theorem follows immediately from (2) with $T = 0$.

To prove the second statement, we use the inequalities [11]:

$$\text{Prob}\{X < x\} \le \frac{\sigma^2}{\sigma^2 + (x-\mu)^2}, \quad x < \mu \tag{29}$$

$$\text{Prob}\{X > x\} \le \frac{\sigma^2}{\sigma^2 + (x-\mu)^2}, \quad x > \mu \tag{30}$$

where X is a random variable with mean μ and variance σ^2. Then, for $0 < L < 1$, from (29), and Theorem 2.2,

$$\text{DTP}(L, T) = \text{Prob}\{n(T) \ge Lc(\frac{r}{r+p})T\} = 1 - \text{Prob}\{n(T) < Lc(\frac{r}{r+p})T\}$$

$$\ge 1 - \frac{\dfrac{2c^2rpT}{(r+p)^3} - \dfrac{2c^2rp}{(r+p)^4}(1 - e^{-(r+p)T})}{\dfrac{2c^2rpT}{(r+p)^3} - \dfrac{2c^2rp}{(r+p)^4}(1 - e^{-(r+p)T}) + \left[(L-1)(\dfrac{r}{r+p})cT\right]^2}.$$

Therefore, $\lim_{T\to\infty} \text{DTP}(L, T) = 1$. The result for $L > 1$ is derived similarly using (30).

Proof of Theorem 3.3. Define functions $g_1(x, y)$, $g_2(x, y)$, $g_{up}(x, y)$, and $g_{down}(x, y)$ as follows:

$$g_1(x, y) = \sum_{i=0}^{\infty} \sum_{j=0}^{i} \frac{x^j}{j!}\frac{y^i}{i!}$$

$$g_2(x, y) = \sum_{j=1}^{\infty} \sum_{i=0}^{j-1} \frac{x^j}{j!}\frac{y^i}{i!}$$

$$g_{up}(x, y) = \frac{g_1(x, y)}{g_1(x, y) + g_2(x, y)}$$

$$g_{down}(x, y) = 1 - \frac{g_1(y, x)}{g_1(x, y) + g_2(x, y)}. \tag{31}$$

Observe that

$$g_1(x, y) + g_2(x, y) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{x^j}{j!} \frac{y^i}{i!} = \sum_{i=0}^{\infty} \frac{y^i}{i!} \sum_{j=0}^{\infty} \frac{x^j}{j!} = e^{x+y} \tag{32}$$

and, from (31), (32), (10), and (11),

$$\begin{aligned} \mathrm{DTP}_{up} &= g_{up}\,(pFT, r(1 - F)T) \\ \mathrm{DTP}_{down} &= g_{down}\,(pFT, r(1 - F)T). \end{aligned} \tag{33}$$

Also, from (31) with $x > 0$, $y > 0$, we obtain:

$$\begin{aligned} \frac{\partial g_{up}\,(x, y)}{\partial x} &= -e^{-(x+y)} \sum_{i=0}^{\infty} \frac{(xy)^i}{(i!)^2} < 0 \\ \frac{\partial g_{up}\,(x, y)}{\partial y} &= -e^{-(x+y)} \sum_{i=1}^{\infty} \frac{x^i}{(i!)} \frac{y^{i-1}}{(i-1)!} > 0 \\ \frac{\partial g_{down}\,(x, y)}{\partial x} &= -e^{-(x+y)} \sum_{i=1}^{\infty} \frac{x^{i-1}}{(i-1)!} \frac{y^i}{i!} < 0 \\ \frac{\partial g_{down}\,(x, y)}{\partial y} &= e^{-(x+y)} \sum_{i=0}^{\infty} \frac{(xy)^i}{(i!)^2} > 0 \end{aligned} \tag{34}$$

and

$$\begin{aligned} g_{down}(x, y) - g_{up}(x, y) &= 1 - \frac{g_1(y, x)}{g_1\,(x, y) + g_2\,(x, y)} - \frac{g_1\,(x, y)}{g_1\,(x, y) + g_2\,(x, y)} \\ &= 1 - e^{-(x+y)} \left[\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{x^j}{j!} \frac{y^i}{i!} + \sum_{i=0}^{\infty} \frac{(xy)^i}{(i!)^2} \right] \\ &= 1 - e^{-(x+y)} \sum_{i=0}^{\infty} \frac{(xy)^i}{(i!)^2} < 0. \end{aligned} \tag{35}$$

Using (6), (34), and (35),

$$\frac{\partial \mathrm{DTP}\,(c, p, r, T, L)}{\partial p}$$

$$= \frac{\partial \left[\left(\dfrac{r}{r+p} \right) g_{up}\,(p\mathrm{FT}, r(1 - \mathrm{F})\mathrm{T}) + \left(\dfrac{p}{r+p} \right) g_{down}\,(p\mathrm{FT}, r(1 - \mathrm{F})\mathrm{T}) \right]}{\partial p}$$

$$= \left(\frac{r}{r+p}\right) \frac{\partial g_{up}(pFT, r(1-F)T)}{\partial x} FT$$

$$+ \left(\frac{p}{r+p}\right) \frac{\partial g_{down}(pFT, r(1-F)T)}{\partial x} FT$$

$$+ \left(\frac{r}{(r+p)^2}\right) [g_{down}(pFT, r(1-F)T) - g_{up}(pFT, r(1-F)T)] < 0;$$

$$\frac{\partial \text{DTP}\ (c, p, r, T, L)}{\partial r}$$

$$= \frac{\partial \left[\left(\frac{r}{r+p}\right) g_{up}\ (p\text{FT}, r(1-\text{F})\text{T}) + \left(\frac{p}{r+p}\right) g_{down}\ (p\text{FT}, r(1-\text{F})\text{T})\right]}{\partial r}$$

$$= \left(\frac{r}{r+p}\right) \frac{\partial g_{up}(pFT, r(1-F)T)}{\partial y} (1-F)T$$

$$+ \left(\frac{p}{r+p}\right) \frac{\partial g_{down}(pFT, r(1-F)T)}{\partial y} (1-F)T$$

$$+ \left(\frac{p}{(r+p)^2}\right) [g_{up}(pFT, r(1-F)T) - g_{down}(pFT, r(1-F)T)] > 0,$$

and, therefore, DTP(c, p, r, T, L) is a monotonically decreasing function of p and a monotonically increasing function of r. Additionally, since $F = L\ (r/(\text{r} + p)$, using (6), (34), and

$$\frac{\partial \text{DTP}\ (p, r, T, L)}{\partial F} = \left(\frac{r}{r+p}\right) \left[\frac{\partial\ g_{up}\ (pFT, r(1-F)T)}{\partial x} pT - \frac{\partial\ g_{up}\ (pFT, r(1-F)T)}{\partial y} rT\right]$$

$$+ \left(\frac{p}{r+p}\right) \left[\frac{\partial\ g_{down}\ (pFT, r(1-F)T)}{\partial x} pT - \frac{\partial\ g_{down}\ (pFT, r(1-F)T)}{\partial y} rT\right] < 0,$$

we conclude that DTP(p, r, L, T) is a monotonically decreasing function of L.

References

1. Demming, E., *Out of the crisis*, Massachusetts Institute of Technology, Cambridge, MA, 1986.
2. Goldratt, E. and J. Cox, *The Goal*, North River; Press, New York, 1986.
3. Womack, J., D. Jones, and D. Roos, *The machine that changed the world*, Rawson Associates, New York, 1990.
4. Gershwin, S. B., Variance of output of a tandem production system. *Proceedings of the Second International Conference on Queueing Networks with Finite Capacity*, Elsevier Science Publishers, 1993.
5. Sevastyanov, B. A., Influence of storage bin capacity on the average standstill time of a production line. *Theory of Probability with Applications*, *7*, (1962), 429–438.
6. Wijngaard, J., The effect of interstage buffer storage on the output of two unreliable production units in series, with different production rates. *AIIE Transactions*, ***11***, (1979), 42–47.
7. Dallery, Y., R. David, and X. Xie, Approximate analysis of transfer lines with unreliable machines and finite buffers. *IEEE Transactions on Automatic Control*, ***34***, (1989), 943–953.
8. Chen, H., and D. Yao, A fluid model for systems with random disruptions. *Operations Research*, ***40***, (1992), 239–247.
9. *Manufacturing 21 report: The future of Japanese manufacturing.* Translated from *Communications of the Operations Research Society of Japan*, ***34***, by Association for Manufacturing Excellence, Wheeling, Illinois 1990.
10. Borrus, M., and R. Leachman, The competitive semiconductor manufacturing survey: Results of the pilot phase. *Competitive semiconductor manufacturing report no. 01*, University of California, Berkeley, 1992.
11. Allen, A., *Probability, statistics, and queuing theory with computer science applications*, Academic Press, New York, 1978.