A CLOSED-LOOP TWO-INDENTURE REPARABLE ITEM SYSTEM

A. Daryanto, J.C.W. van Ommeren, W.H.M. Zijm
Faculty of EBA, Tilburg University, A.Daryanto@utw.nl
Faculty of EEMCS, University of Twente, J.C.W.vanOmmeren@math.utwente.nl
Faculty of EEMCS, University of Twente, W.H.M.Zijm@math.utwente.nl

Abstract:
In this paper we study the performance of a closed-loop two-indenture reparable item system. We consider a production cell consisting of a number of identical machines. A machine may fail in which case it is removed from the production cell and replaced by a ready-for-use one, if available. The failing machine is sent to an assembly shop where it is disassembled to detect which component causes the failure. Next, a ready-for-use component is taken from stock, if available, and assembled in the machine which in turn is placed in the stock of usable machines. The malfunctioning component finally is sent to a component repair shop where it is restored, after which it is placed in the stock of ready-for-use components. Both the assembly and component repair facility are modeled as single server queues with exponentially distributed service times. If no ready-for-use machines or components are available when needed, the production cell performs worse or the assembly operation is delayed, respectively.

We show that an aggregate model of the system, slightly modified, possesses a product form solution by modelling the component repair facility as a server with state-dependent service rates (or alternatively, state-dependent blocking), of which the parameters follow from an application of Norton’s theorem for Closed Queuing Networks. This solution structure allows us to approximate several performance measures such as the probability that all machines in the production cell are working, or the expected number of working machines. These approximations for the performance measures are compared with exact results. The proposed approximation performs well. A number of extensions of the model are briefly reviewed.

Keywords: reparable item system, repair facilities, queuing-inventory models, indenture levels

1 Introduction

In this paper, we study the performance of a production cell, consisting of a number of identical working machines, each of which may fail. In cases where a high availability of machines in such a production cell is essential, it is natural to follow a repair-by-replacement policy. Hence, the failing machine is removed and immediately replaced by a ready-for-use one that is kept in stock. The broken machine is sent to an assembly shop where it is disassembled to detect which component causes the failure. Next, a ready-for-use component is taken from stock to replace the malfunctioning one, the machine is restored and placed in the stock of ready-for-use machines again. The broken component finally is sent to a second repair shop where it is restored, after which it is placed in the appropriate stock of ready-for-use components. Clearly, we need some excess inventories of both machines and components to let this system work properly. However, since these stocks are filled by the flow of repaired products from either the component repair or the assembly facility (both modeled as finite capacity queuing systems), some delay may occur, and even stocks may be temporarily depleted. In the case where no components are available, the machine restoration is delayed. In the case where no ready-for-use machines are available when needed, the production cell performs worse. See figure 1 for a graphical display of all item flows. Hence, on the one hand it is necessary to have a sufficient number of both components and machines circulating; on the other hand these items are often expensive hence there is a
natural tendency to restrict their numbers. A basic question then is whether we can quickly evaluate the performance of this two-indenture reparable item system or, more precisely, to determine the overall production cell performance, given certain failure and repair rates, as well as some target stock levels of rotating items.

This paper can be seen as a companion of the paper by Avsar and Zijm[1] where an open system with reparable items is considered (i.e. broken machines are generated by an infinite source, according to a Poisson process). In such a case, the appropriate performance measure often is the assembly fill rate (the fraction of times a ready-for-use machine is available when needed). This latter system is modelled as an open two-indenture reparable item system where assembly and repair facilities are described as open Jackson networks. In such an open network, each level receives a Poisson input. The difficulty with the closed system analysis is that this no longer holds; even if machine failure rates are exponential, the input of broken items at each level is determined by the number of working machines while in particular the input to the component repair shop depends on the assembly shop performance as well. In the body of this paper, we consider the simple case in which there is only one type of reparable components while both the component-repair facility and the assembly shop are modelled as single exponential servers. In the final section, we relax this assumption and discuss possible extensions.

The paper by Avsar and Zijm[1] is an extension to Sherbrooke’s METRIC (Multi Echelon Technique for Recoverable Item Control) models [9] by integrating inventory systems with finite capacity repair centers. The METRIC model was extended by Muckstadt[6] to the so-called MODMETRIC model which assumes that the service facilities are uncapacitated and the failures of the items follow a Poisson process. This makes it possible to use Palm’s theorem to analyze such systems. The performance analysis of capacitated two-indenture models for open repairable item systems is studied in Zijm and Avsar[10]. The notion indenture refers to the product structure of an assembly, consisting of multiple components. The study of multi-indenture problems, is conducted by Scudder and Hausman[5], Gupta and Albright[3], Rustenburg[7], and Rustenburg et al.[8], who approximate the steady state behavior under less restrictive assumptions than the MODMETRIC model.

The organization of this paper is as follows: In section 2, we model the reparable item system and show that it basically is identical to a two-dimensional finite Markov chain. In practical situations the state space is large. Since the analysis of large Markov chain is, in general, intractable, we need approximations. We propose an approximation based on a modification of the model which is obtained by replacing the component-repair facility by a special server with state-dependent service rates (alternatively, this server may be seen as a server that sometimes
blocks and reroutes customers). Section 4 is devoted to finding an approximation for the probability that a request for a ready-for-use component cannot be fulfilled although so far there is no backlog of components for the assembly operation. This probability plays a crucial role in the analysis; we estimate it by using the main idea of Norton’s theorem. The quality of the approximation in terms of the production cell performance is presented in Section 5 by comparing the results based on the approximation with exact results. Generalizations of the model to more complicated product structures (multiple component types) and more complex repair and assembly shops are discussed in Section 6.

2 The Reparable Item System

In this section we show that, by making some simplifying assumptions, we can model the repairable item system described above as a two-dimensional finite Markov chain. The steady state behavior of this Markov chain can, in principle, be found. However, since the number of states increases rapidly with the inventory sizes and the number of machines, the analysis becomes infeasible. In addition, when generalizing to more complex systems, an exact analysis is deemed to fail.

Suppose that in the production cell a maximum of $M$ machines are ideally working properly. Each machine may fail at a Poisson rate $\lambda$, which is basically the component failure rate. Assume that in addition to the working machines we have a base stock of $S_2$ ready-for-use machines (note that in general the stock level is lower than $S_2$ due to the fact that some machines are either restored or waiting for restore in the assembly shop, or even waiting for components before sent to the assembly shop). Further, we have a base stock of $S_1$ ready-for-use components (again, the real stock may be lower due to the fact that some components are in repair or waiting to be repaired). When a machine fails, it is, ideally, immediately replaced by a ready-for-use machine, taken from stock. The broken machine is disassembled at the assembly room where the failing component is removed immediately (in a neglectable time). Next, a ready-for-use component is matched with the disassembled machine, after which the machine can be restored (assembled) again. If no component is available, some further delay is incurred. The failing component in turn is repaired at the component-repair facility. Both the assembly and component repair shops are modeled as single exponential servers with rates $\mu_2$ and $\mu_1$ respectively. As indicated already in figure 1, the matching of a request and a ready-for-use machine from stock is modeled as a synchronization queue, and similar for the component requests and stocks. The performance measure that we are interested in is either the probability that all machines in the production cell are working properly, or the expected number of working machines in the cell.
As a first step toward the analysis we note that the synchronization queue between the assembly room and the production cell can be seen as a normal queue where machines are waiting to be moved into the production cell. This is possible only if there is place, that is, if a machine in the production cell has failed. This gives us the model in figure 2 (similar to the classical single stage machine-repairman problem). So only one synchronization queue is left, at the component level.

Initially, we assume without loss of generality that all stocks are completely filled, the repair and assembly facilities are empty and \( M \) machines are working properly. Hence, the total number of machines in the system is \( S_2 + M \) and the total number of components equals \( S_1 + S_2 + M \) (including those assembled in machines). Let \( N_1 \) and \( N_2 \) be the number of items being processed or waiting in the queue at the component-repair facility and assembly shop, respectively, and let \( N_3 \) be the number of good machines either in the production cell or in the queue. Let \( K_1 \) be the number of backordered components. Denote the number of machines presently working in the production cell by \( M \). Note that \( M = \min(M, N_3) \). See figure 3 for a graphical representation of the model.

As a result of the operating inventory control policies we immediately observe that:

\[
K_1 = (S_1 - N_1)^+ \tag{1}
\]

\[
K_1 + N_2 + N_3 = M + S_2 \tag{2}
\]

The system can be described by \( N_1 \) and \( N_2 \) since these quantities induce the values of \( K_1 \) and \( N_3 \) by the relations (1) and (2). Hence we can model this system as a continuous time Markov chain with states \((n_1, n_2)\). The transition diagram is shown in figure 4.

Let \( P(n_1, n_2) = P(N_1 = n_1, N_2 = n_2) \) be the steady state probability of being in state \((n_1, n_2)\). Note that in regions I and II we find transitions similar to those in a fork-join queue (which in general are hard to analyze). In regions III and IV on the other hand we find transitions similar to those in a tandem queue (which is obvious since \( n_1 > S_1 \) indicates a backlog hence any repaired component is immediately transferred to the assembly shop). This can also be seen from relation (1) which shows that \( K_1 = k_1 > 0 \) is equivalent to \( N_1 = S_1 + k_1 \) and that \( K_1 = 0 \) is equivalent to \( N_1 \leq S_1 \). Together with relation (2), we find that we only need to study the behavior of \( K_1 \) and \( N_2 \). In the next section we derive an approximation for the probabilities \( P(K_1 = k_1, N_2 = n_2) \).

The performance measures we are interested in are the probability that all machines in the cell
are working properly, defined by

\[ A = P(M = M) = P(N_3 \geq M) = \sum_{n_1=0}^{S_1 + S_2 \min(S_2, S_1 + S_2 - n_1)} \sum_{n_2=0} P(n_1, n_2) \]

and the expected number of working machines in the production cell

\[ E[M] = E(M - (N_2 + (N_1 - S_1)^+ - S_2)^+) = E(M - (N_2 + K_1 - S_2)^+) \]

Note that \( N_2 + (N_1 - S_1)^+ = N_2 + K \) in the latter formula denotes the number of machines that are either waiting to be or being assembled \( (N_2) \) or still waiting for a spare part \( (K_1 = (N_1 - S_1)^+) \). Hence, \( (N_2 + K_1 - S_2)^+ \) is the backlog of machines (unfilled requests of machines needed in the production cell).

3 Modification of the model

Note that from the transition diagram displayed in the previous section we may immediately deduce the global balance equations and hence in principle solve the system, i.e. determine the steady state probabilities \( P(n_1, n_2) \) for the \((N_1, N_2)\) process. Due to the transition on the left side of the state space (see figure 4), it is in general difficult to find explicit expression for these steady state probabilities. We already observed however that it suffices to study the \((K_1, N_2)\) process. Recall that \( k_1 = 0 \) if and only if \( n_1 \leq S_1 \) while \( k_1 = k > 0 \) if and only if \( n_1 = S_1 + k \), hence, apart from the rates out of the states \((0, n_2)\), we observe transition rates for the states \((k_1, n_2)\) equal to those in the regions III and IV of the \((n_1, n_2)\) state space. In these regions, we recognize the transitions of a closed tandem queue with two single server stations and one multiple server station. Hence, if we aggregate the states \((n_1, n_2)\) with \( n_1 = 0, \ldots, S_1 \) into one state \((0, n_2)\) and adjust the rates at the \( n_2 \)-axis \((k_1 = 0)\), we obtain the same state space and transition rates as in a special closed queueing network with a product form for the steady state probabilities. Of course this will not lead to an exact solution but to an approximation.

By the observation that we only need to study the \((K_1, N_2)\) process and the relation \( K_1 = (N_1 - S_1)^+ \), it is natural to aggregate the states \( \{(n_1, n_2) | n_1 = 0, \ldots, S_1\} \) into one state \((0, n_2)\) for \( n_2 = 0, \ldots, M + S_2 \). The systems behavior can be described through the states \((k_1, n_2)\). To
use this approach, we need that for any $n_2$

$$P(K_1 = 0, N_2 = n_2) = \sum_{n_1 = 0}^{S_1} P(n_1, n_2)$$

$$P(K_1 = k_1, N_2 = n_2) = P(S_1 + k_1, N_2 = n_2), \quad k_1 > 0.$$  

This can be accomplished by defining the right transition rates for the states at the boundary $k_1 = 0$. However, in order to find these rates, we would need the steady state probabilities of the original system. In the following, we sketch an alternative way of adjusting the rates at the boundary to obtain a product form approximation.

Consider in the original system the states with $n_1 = 0, \ldots, S_1 - 1$. In all these states the rate at which $n_2$ is increased does not depend on $n_1$ and is equal to $(M - (S_2 - n_2)\lambda$. When $n_1 = S_1$ the rate of increase of $n_2$ is zero. So, in the aggregated states, this up going rate is almost fixed. Define

$$q(n_2) = P(N_1 = S|N_1 \leq S_1, N_2 = n_2) \quad \text{for} \quad n_2 = 0, \ldots, M + S_2.$$  

We can interpret $q(n_2)$ as the probability that a request for a component has to wait, given it finds no other waiting requests, while there are $n_2$ machines in the assembly room. It is clear that the rate of increase of $n_2$ in the aggregated state $(0, n_2)$ is $(1 - q(n_2))(M - (n_2 - S_2)\lambda$. Since finding $q(n_2)$ is difficult, we focus on the conditional probability $q$ defined by

$$q = \sum_{n_2} q(n_2)P(N_2 = n_2|N_1 \leq S) = P(N_1 = S|N_1 \leq S),$$  

which can easily be approximated (see the next section) and replace $q(n_2)$ in the rate of of increase by $q$. From now on we consider the transformed model with transition rates given in figure 5 where $q$ is given by formula (4). Denote the steady state probabilities for this system by $P(k, n_2)$.

**Lemma 1** The steady state probability for the model with state description $(k_1, n_2)$ and transition rates as denoted in figure 5 (for arbitrary $q$) has a product form.

**Proof.** Consider a closed queueing network with $M + S_2$ identical jobs circulating between two stations with a single exponential server operating with rate $\mu_1$ and $\mu_2$, respectively and
a station with $M$ exponential servers operating with rate $\lambda$. Jobs flow cyclically from the first single server station to the second single server station to the multiserver station. If a job arrives at the first station and finds no other jobs there, it is with probability $1 - q$ sent to the second station; otherwise it is taken into service (with rate $\mu_1$). The state space and transition diagram are identical to the ones in the transformed model. The lemma now follows immediately from checking the balance equations. Alternatively, the closed queueing network just introduced can be seen as a special case of a Jackson network with jump-over blocking which is known to have a product form solution (see Boucherie and van Dijk[2]).

Above, we used the transition rates first of all to identify an alternative system with a known product form solution. In the following we give another interpretation of this transformation. Let us look both at the original model in figure 3 and at the alternative model in figure 6. We see that the component-repair facility with synchronization queue is replaced by a typical server. This server has, with probability $1 - q$, an infinite service rate for jobs which find the server idle. By looking at the open system with the typical server, and conditioning on the fact that there are exactly $M + S_2$ jobs in the network, we find the following expression for $\hat{P}(k_1, n_2, n_3)$, the steady state distribution of the TCQN:

$$
\hat{P}(k_1, n_2, n_3) = \begin{cases} 
\hat{D} q \left( \frac{1}{\mu_1} \right)^{k_1} \left( \frac{1}{\mu_2} \right)^{n_2} \left( \frac{1}{\mu_3} \right)^{n_3}, & \text{for } k_1 > 0 \text{ and } n_3 < M, \\
\hat{D} q \left( \frac{1}{\mu_1} \right)^{k_1} \left( \frac{1}{\mu_2} \right)^{n_2} \frac{1}{M!M^{n_3-M}}, & \text{for } k_1 > 0 \text{ and } n_3 \geq M, \\
\hat{D} \left( \frac{1}{\mu_2} \right)^{n_2} \frac{1}{n_3!}, & \text{for } k_1 = 0 \text{ and } n_3 < M, \\
\hat{D} \left( \frac{1}{\mu_2} \right)^{n_2} \frac{1}{M!M^{n_3-M}}, & \text{for } k_1 = 0 \text{ and } n_3 \geq M,
\end{cases}
$$

(5)

with $k_1 + n_2 + n_3 = M + S_2$ where $\hat{D}$ is the normalization constant. This immediately gives us the following lemma.

**Lemma 2** The steady state distribution for the aggregate model is given by

$$
\tilde{P}(k_1, n_2) = \begin{cases} 
\frac{D q}{(M+S_2-k_1-n_2)!} \left( \frac{1}{\mu_1} \right)^{k_1} \left( \frac{1}{\mu_2} \right)^{n_2}, & \text{for } k_1 > 0 \text{ and } k_1 + n_2 > S_2, \\
\frac{D q}{M!M^{n_3-M}} \left( \frac{1}{\mu_1} \right)^{k_1} \left( \frac{1}{\mu_2} \right)^{n_2}, & \text{for } k_1 > 0 \text{ and } k_1 + n_2 \leq S_2, \\
\frac{D}{(M+S_2-n_2)!} \left( \frac{1}{\mu_2} \right)^{n_2}, & \text{for } k_1 = 0 \text{ and } n_2 > S_2, \\
\frac{D}{M!M^{n_3-M}} \left( \frac{1}{\mu_2} \right)^{n_2}, & \text{for } k_1 = 0 \text{ and } n_2 \leq S_2
\end{cases}
$$

where $D$ is the normalization constant.
The previous lemma gives us an explicit expression for the steady state probabilities. In practice it can be difficult to find the normalization constant $D$. However, we are dealing with a product form network and so we can use the MVA (Marginal Value Analysis, see eg. [4]) to find the Availability ($A$) and the expected number of working machines $EM$.

So far, we have presented results that hold true for any value of $q$. In the reparable item system, $q$ can be interpreted as the conditional probability that a machine who arrives at the central repair facility does not have to wait, given that there are no machines waiting (see formula (4)). However, we did not use this interpretation. The previous lemmas hold for any $q$ with $0 \leq q \leq 1$. So we have some freedom in choosing $q$. In the next section we find a $q$ which gives good results and has a meaningful interpretation.

4 Applying Norton’s Theorem to approximate $q$

Although we have stated in the previous section that the product form does not depend on $q$, it is still needed to find a $q$ that gives a good approximation for the performance measures. In this section, we use the basic idea of Norton’s theorem (see [4] for an overview) to find an approximation for $q$ that gives good results. This basic idea is that we can analyze a product form network by replacing sub-networks by state dependent servers. Norton’s theorem then states that the joint distribution for the numbers of customers in the sub-networks and the queue lengths at the replacing state dependent servers are the same.

To use this idea, first recall the original model as shown in figure 3. We want to find $q$, the conditional probability that when a machine fails no spare parts are in stock (see formula (4)). We take the assembly room and the production cell apart and replace them by a state dependent server. In order to find the service rates we short-circuit the original network by setting the service rate at the component-repair facility to infinity. This gives us a new, evidently smaller, network as shown in figure 7. This sub-network is, in the original model, replaced by a server with state dependent service rates. The service rate of the new server with $m$ jobs present is equal to the throughput of the network in figure 7 with $m$ jobs present. Let $TH(m)$ be this state dependent service rate. At the end of the section we describe how to find these rates.

The evolution $N_1 = n_1$, the number of components in queue at station 1, can now be described as a birth-death process. The transition diagram is shown in figure 8. Note that this is only an approximation; in case $S_1 = 0$ we would have a product form network and the results would be exact. From the transition diagram we see that

$$P(N_1 = n_1)TH(M + S_2 - (n_1 - S_1)\uparrow) = P(N_1 = n_1 - 1)\mu_1$$

(6)
for \( n_1 = 1, \cdots, M + S_1 + S_2 \).

\[ \text{Figure 8: Transition diagram for } n_1. \]

In principal, we can now find an approximation for the distribution of \( N_1 \). However, by the definition of \( q \) (see formula (4)), we only need to study the behavior for \( N_1 \leq S_1 \). For these states, the service rate of the state dependent server is equal to \( TH(M+S_2) \). Let \( \gamma = TH(M+S_2)/\mu_1 \).

From formula (6) we see that

\[ q = \frac{P(N_1 = S_1)}{P(N_1 \leq S)} = \gamma S_1 \frac{1 - \gamma}{1 - \gamma S_1 + 1} \]  

(7)

It remains to find the throughput of the short-circuited network in figure 7 with \( M + S_2 \) jobs present. It is easily seen that \( P(N_3 = n) \min(n, M) \lambda = P(N_3 = n - 1) \mu_2 \) for \( n = 1, \cdots, M + S_2 \) and that the throughput satisfies

\[ TH(M + S_2) = \sum_{n=1}^{M+S_2} P(N_3 = n) \min(n, M) \lambda = P(N_2 > 0) \mu_2. \]

The numerical results presented in the next section show that the approximation for \( q \), given in formula (7) performs very well.

In the remark below lemma 2 we already mentioned the possibility to use an MVA scheme to analyze the TCQN (see figure 6). For this system, we can apply Norton’s theorem with the same sub-network (state dependent server) as introduced in this section. By doing so, we do not have to compute the normalizing constant \( D \).

5 Performance of the approximation

In this section two performance measures of the system are studied. We consider the probability \( A \) that all machines in the production cell are working properly, as well as the expected number of working machines in the cell \( (E \overline{M}) \), defined by

\[ A = P(K_1 + N_2 \leq S_2) = P(N_3 \geq M) \]  

(8)

\[ E \overline{M} = E[M - (K_1 + N_2 - S_2)^+] \]  

(9)

Note that \( A \) and \( E \overline{M} \) seem to depend only on the base stock level \( (S_2) \), and not on \( S_1 \). Of course this is not true, since \( A \) and \( E \overline{M} \) are implicitly influenced by the quantity \( q \) which depends on \( S_1 \).

To test how accurate the proposed approximations are, the performance measures derived by the product-form solution of the modified and aggregated model, are compared with exact results. The evaluation of the performance measures is checked by trying several settings for \( M, S_1, S_2, \lambda/\mu_1 \) and \( \lambda/\mu_2 \). Here we take a system with \( \mu_1 = 2\lambda \) and \( \mu_2 = \lambda \). In the table, we do not give the value of \( E \overline{M} \) but of the expected shortage of machines in the production cell \( (M - E \overline{M}) \).
The numbers show that in these systems, the approximation gives an error of at most 5%. In all other cases that we tested, we got similar results.

6 Summary and Possible Extensions

In this paper we have analyzed a closed two indenture repairable item system with a fixed number of items circulating in the network. The system consists of a production cell with machines failing with rate $\lambda$, a component-repair facility and a machine repair facility (an assembly room). Both these repair facilities are able to keep a number of ready-to-use items in stock. Machines with rate $\mu_i$ are sent to the inventory.

The exact analysis of a Markov chain model for this system with many machines or with large inventories is difficult to handle. Therefore, we aggregate a number of states and adjust some rates to obtain a product form solution. We can view this new system, related to the Markov chain with aggregated states as a so-called Typical Closed Queueing Network (TCQN). The notion ”typical" comes from modelling the central repair facility together with the synchronization station, as a typical server with state dependent blocking. Numerical results show that the proposed approximation performs well for the availability ($A$) and the expected number of working machines ($\mu - \bar{M}$).

The extension of the above model to a system in which both the component repair and the machine assembly shops are BCMP networks (instead of single exponential servers) seems to be possible. In fact, similar to results in Avsar and Zijm[1] it is possible to show that the system with state space $(k_1, \bar{w}_2)$ (where $k_1$ denotes again the backlog at the component shop, and $\bar{w}_2$ represents the vector of machines at the various nodes in the machine assembly shop) has a product form solution. Also the case with multiple failing component types can in principle be handled, but requires a multi-class version of Norton’s theorem. These extensions, as well as further generalizations to multi-echelon, multi-indenture closed loop reparable item systems, will be worked out in a subsequent paper.
References


