Resource-constrained two-echelon inventory models for repairable item systems

Faculty of Mathematical Sciences

University of Twente

University for Technical and Social Sciences

P.O. Box 217
7500 AE Enschede
The Netherlands
Phone: +31-53-4893400
Fax: +31-53-4893114
Email: memo@math.utwente.nl

Memorandum No. 1521

Resource-constrained two-echelon inventory models for repairable item systems

Z.M. Avşar[1] and W.H.M. Zijm

April 2000

ISSN 0169-2690

[1]Department of Operations Planning and Control, Faculty of Technology Management, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

Resource-Constrained Two-Echelon Inventory Models for Repairable Item Systems

Zeynep Müge Avşar and W. Henk Zijm

Department of Operations Planning and Control
Faculty of Technology Management
Eindhoven University of Technology
P.O. Box 513, 5600 MB Eindhoven, The Netherlands

January 14, 2000

Abstract

In this paper, we consider two-echelon maintenance systems with repair facilities both at a number of local service centers (called bases) and at a central location. Each repair facility may be considered to be a job shop and is modeled as a (limited capacity) open queuing network, while any transport from the central facility to the bases (and vice versa) is modeled as an ample server. At all bases as well as at the central repair facility, ready-for-use spare parts are kept in stock. Once an item in the field fails, it is returned to one of the bases and replaced by a ready-for-use item from the spare parts stock, if available. The returned failed item is either repaired at the base or shipped to and repaired at the central facility. In the case of local repair, the item is added to the local base stock as a ready-for-use item after repair. If a repair at the central facility is needed, the base orders an item from the central spare parts stock to replenish its local stock, while the failed item is added to the central stock after repair. Orders are satisfied on a first-come-first-serve basis while any requirement that cannot be satisfied immediately either at the bases or at the central facility is backlogged.

We assume that failed items are returned to the bases according to a Poisson process, and that each repair shop (at the bases as well as at the central facility) can be modeled as a Jackson network. Under these conditions, we propose a special near-product-form solution that provides an excellent approximation for the steady-state distribution of a slightly aggregated system, that permits us to calculate all relevant performance indicators (such as fill rates and stockout probabilities) at the bases as well as at the central facility, as a function of target inventory levels at each location. Errors of these performance measures are generally less than one percent, when compared with simulation results. Finally, we show how these approximations can be used to determine optimal inventory levels at both the central and local facilities.

Keywords: multi-echelon systems, repairable items, limited repair capacities

1 Introduction

Since the pathbreaking work of Sherbrooke [16], multi-echelon models for repairable items inventory control have received considerable attention. The current paper can be seen as an extension of Sherbrooke's METRIC (Multi Echelon Technique for Recoverable Item Control) models by integrating inventory systems with finite capacity repair centers (modeled as open Jackson networks) at both the local bases and a central depot. In this section, we first describe the general structure of two-echelon repairable item systems. Next, we briefly discuss some key references on multi-echelon repairable item systems, and indicate the main contributions of this paper.

Let us first outline the general structure of capacitated multi-echelon, multi-indenture repairable item models. Suppose items are placed in the field to operate, possibly as part of a large technical system. If an item fails, it is returned to a local, nearby repair facility (called a base) which at the same time ships a ready-for-use spare item, taken from a local spare parts stock, to the technical system in order to minimize the non-operating time of that system. The failed item is either repaired at the base or, in case the repair turns out to be hard and requires special engineering expertise, at a central repair shop. In the first case (local repair), upon completion of the repair the item is added to the local stock. If, however, the failed item has to be repaired at the central facility, the base immediately orders a ready-for-use item from a central stock at this repair facility, to replenish the item shipped to the field installation. Once the repair at the central facility is completed, the revised item is added to the central stock as a ready-for-use item again. Hence, as long as no condemnation occurs (i.e., each item can be repaired either at one of the bases or at the central repair shop) the total number of items (operating in the field, being in repair or in transport between depot and bases, or stocked as ready-for-use items) is constant in principle. We assume that all requirements are fulfilled on a first-come-first-serve basis and that each demand that cannot be satisfied immediately either at the bases or at the central facility is backlogged.

Models similar to the one described above have initially been considered by Sherbrooke [16] and have become known as METRIC models. As in almost all (multi-echelon) inventory models, Sherbrooke focuses on the determination of optimal order-up-to levels at both the local and the central stocking centers, and ignores any limitation on the available repair capacities at any facility. On the other hand, he considers multiple items that may operate together in complex systems (such as aircrafts, ships or production facilities) and attempts to maximize the overall system availability under a given budget constraint. Although, as mentioned already, the total number of rotating items of each type is fixed, the number of items operating in the field is assumed to be sufficiently large to allow for the field demand (due to failures) to be approximated by a Poisson process.

During the last three decades several important improvements of METRIC have been proposed. Muckstadt [15] was the first to recognize the importance of the product structure with respect to recoverable item control. He extended the existing METRIC model, which may be characterized as a two-echelon, single-indenture model, to a two-echelon, two-indenture model, which is also referred to as MOD-METRIC. Another variant of METRIC is VARI-METRIC, a two-echelon, single-indenture model developed by Slay [19]. In the core part of the analysis of the initial METRIC model, it is assumed that, for each product, the number of items in repair follows a Poisson distribution (of which the variance equals the mean). In his VARI-METRIC method, Slay derives an approximate expression for the variances of the number of items in repair. Next, for each product, he fits a negative binomial distribution on the first two moments of these items in order to obtain a more accurate approximation. Graves [10] independently developed a slightly simpler approximation for the variance of the number of items in repair. Next, he also continues with fitting a negative binomial distribution on the first two moments. Sherbrooke [17] generalized the original VARI-METRIC method and developed a two-indenture, two-echelon version of VARI-METRIC. By simulation, it has been shown that the results produced by this method are fairly accurate. An overview of METRIC type models is given in Sherbrooke [18]. Extensions to more flexible models allowing for emergency repair or emergency supply (but still assuming no resource constraints) have been studied by Verrijdt [21].

Another important line of research was initiated by Gross [11]. The main difference between the VARI-METRIC model and the models of Gross and others (see, e.g., Gross et al. [12], [13] and Albright [1]) is in the constraints of the repair process. In VARI-METRIC, it is assumed that all repair leadtimes are independent variables, which corresponds to an infinite repair capacity. In the models by Gross and others, a limited repair capacity is assumed, however at the cost of other rather restrictive assumptions such as a dedicated repair capacity and fixed repair routings. In these models, the circulation of products through repair, distribution and use is usually modeled as a closed queuing network, and hence, in the spare parts literature, these models are also known as the "closed queuing network models for spare parts management". Approximations for general single stage queuing stations at both the central and local repair centers have been proposed by Diaz and Fu [7].

For this paper, another line of research is of interest. This line starts with the classical multi-echelon inventory systems under periodic review studied initially by Clark and Scarf [6] for serial systems, and later extended to inverse aborescent or distribution structures (one central depot and multiple local warehouses) by numerous authors, see Federgruen [9], Axsäter [2], Van Houtum et al. [20] and Diks et al. [8] for reviews of the literature. Again, almost all authors assume unlimited production capacities at any facility and model all supply lead times as being either fixed or an independent random variable. Only recently, models have been proposed to integrate serial base-stock systems with limited capacities at each facility where these facilities are modeled as either open or population constrained queuing network models, see, e.g., Buzacott et al [4] and Buzacott and Shanthikumar [5]. Note, however, that in all these models items are not circulating but instead are procured from an external supplier, subsequently go through one production stage, next are stocked, subsequently go through a second downstream production stage, are stocked again, and finally leave the system to satisfy external demand. Hence, these models do not allow for distribution structures nor for the complex routings that occur in METRIC models (with either local or central repair).

In this paper, we attempt to integrate models of resource-constrained repair facilities and multi-echelon inventory models for repairable item systems. We consider the case of multiple local bases and a central repair facility, each modeled as a finite capacity open Jackson queuing network. Repair may take place either at a base or at the central facility. As in most papers, we assume that demand (due to failed items in the field) occurs at each base according to a Poisson process. The main contribution of this paper is described as follows: under a slight modification of the steady state equations of the total system, the resulting equations can be shown to have a product form solution. Based on the latter solution, several performance measures (such as fill rates, stockout probabilities and expected stockouts) can be calculated both at the bases and at the central facility, as a function of target inventory levels. Numerical experiments show that the calculated performance measures deviate in general less than one percent from those determined through simulation. This in turn allows the use of the modified models to determine optimal base stock inventory levels, e.g., in order to achieve target service levels, taking into account all repair constraints.

We conclude with an outline of the remaining part of this paper. In the next section, we consider a very simple two-echelon system with one base next to the central repair facility. This model mainly serves to explain the essential elements of the modification discussed above. We present a proof for the product form solution of the modified system and present numerical results to show the accuracy of the approximation. Next, we turn to more general repairable item network structures and more general repair facilities in Section 3, present proofs for the main results for the modified systems and again discuss numerical results. In Section 4, we show how to use the modified models for the purpose of optimizing the target inventory levels of ready-for-use items at both the central repair facility and the bases. In Section 5, we summarize our results and discuss a number of extensions that are currently being investigated.

2 Analysis of a simple two-echelon system with central repair

In this section, we first discuss a highly simplified repairable item system, to explain how a slight modification turns this system into a near-product form network that can be completely analyzed. The results of this analysis are shown to serve as excellent approximations for key performance measures of the original system.

Consider the system displayed in Figure 1, consisting of a single base and a central repair facility. Failed items that arrive at the base are always shipped to the central repair facility, hence no repair at the base occurs. Both the central repair facility and the base hold a number of ready-for-use items in stock. Each failed item generates a demand for a new item at the base, while at the same time the base orders an item from the stock at the central repair facility to update its local stock again. Each demand not immediately fulfilled at either the base or the depot is backordered. The central repair facility is modeled as an exponential server with repair rate μ_0 while transport of an item from the depot to the base stock is modeled as an ample server with exponential service rate μ_1. Transport of a failed item from the base to the depot is not modeled explicitly. In the next section, the single server repair center will be replaced by a product form network, including possible ample servers, hence then transport from base to depot is easily included in the network structure.

We assume that the system operates according to a base-stock policy. Let S_0 and S_1 be the specified target spare part inventory levels at the central repair facility and at the base, respectively. The stock location of the central repair facility is often referred to as the depot. The number of items to be repaired or being in repair at the central repair shop is denoted by the random variable N while the number of items in transport from the depot to the base is denoted by M. The number of ready-for-use items (i.e., items already repaired, now being stocked) at the depot (base) is denoted by $\bar{N}$ ($\bar{M}$). The number of items backordered in case there is no spare part in stock is denoted by K_0 for the depot and by K_1 for the base. Note that each failure of an item in the field results in a request both at the depot and at the base, due to the fact that any demand fulfillment at the base leads to a replenishment order from the base to the depot at the same time. We assume that items fail according to a Poisson process with rate λ where naturally $(\lambda/\mu_0) < 1$ in order for the system to be ergodic.

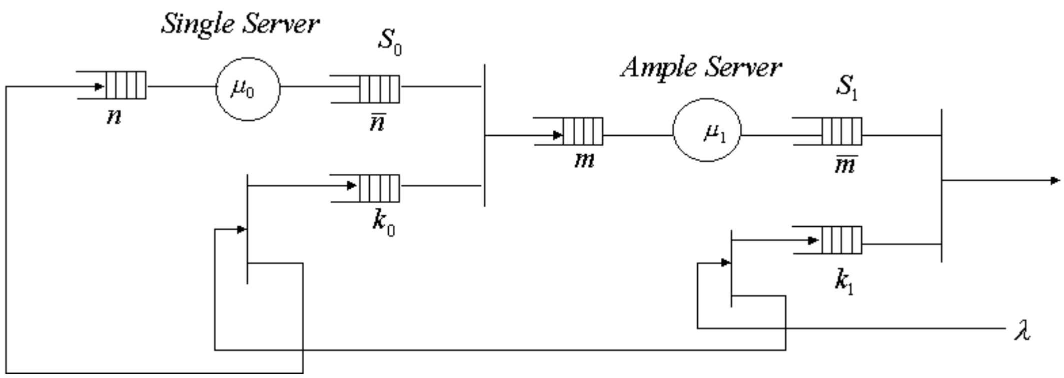

Figure 1: A single item, two-echelon model

For $N = n$, $\bar{N} = \bar{n}$, $M = m$, $\bar{M} = \bar{m}$, $K_0 = k_0$ and $K_1 = k_1$, the following equations hold as a result of the operating inventory control policies:

$$n + \bar{n} - k_0 = S_0,$$

$$m + k_0 + \bar{m} - k_1 = S_1.$$

Since any request is satisfied immediately if there are available spare parts at the depot (base), the request and the spare part are merged just upon the arrival of the request. Also, if requests are being backordered at the depot (at the base), then just after the repair completion of a part (arrival of a part shipped from the depot to the base) it is merged with the longest waiting request. So, the values of both $\bar{n}$ ($\bar{m}$) and k_0 (k_1) at the upper (lower) echelon can never be positive at the same time. More precisely,

$$If \;\; n \leq S_0, \;\; then \;\; \bar{n} = S_0 - n \;\; and \;\; k_0 = 0;$$

$$If \;\; n > S_0, \;\; then \;\; \bar{n} = 0 \;\; and \;\; k_0 = n - S_0;$$

$$If \;\; m + k_0 \leq S_1, \;\; then \;\; \bar{m} = S_1 - (m + k_0) \;\; and \;\; k_1 = 0;$$

$$If \;\; m + k_0 > S_1, \;\; then \;\; \bar{m} = 0 \;\; and \;\; k_1 = (m + k_0) - S_1.$$

From these relations, it follows immediately that n and m completely determine the state of the system, including the values of $\bar{n}$, k_0, $\bar{m}$ and k_1. Thus, this repairable item system can be modeled as a continuous time Markov chain with state description (n, m). The corresponding transition diagram is displayed in Figure 2. $P(N = n, M = m)$ is the steady-state probability of being in state (n, m). Note that for any $n > S_0$ the system behaves as an open tandem queuing system, since in this case a backlog occurs at the depot, causing each completed item at the central repair facility to be transferred immediately to the base. The more difficult part of the transition diagram (similar to that of a fork-join queue system since indeed here one demand generates both an additional repair request and the start of a transport activity) arises for $n < S_0$. Fortunately, that part of the state space can be very naturally aggregated since the states with $0 \leq n \leq S_0$ are precisely those states with no backlog at the depot, i.e., with $k_0 = 0$, while any $k_0 > 0$ corresponds to the set of states with $n = (S_0 + k_0)$.

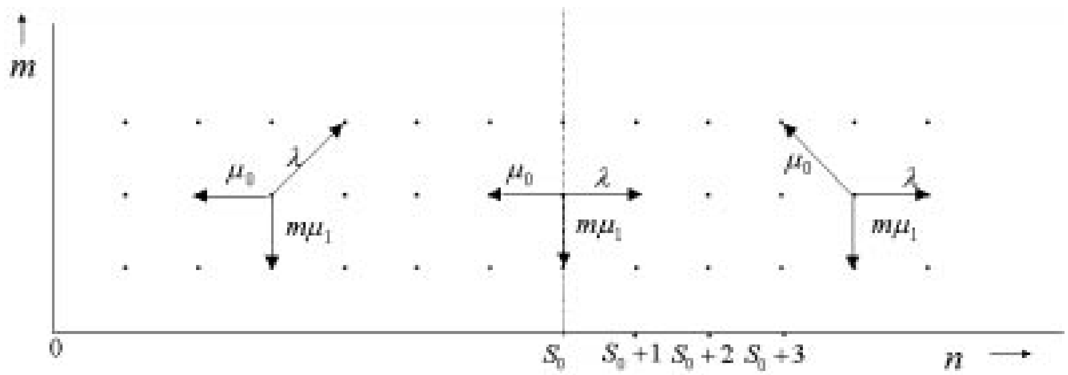

Figure 2: Transition diagram for state (n, m)

Therefore, a natural aggregation is a description of the system through the states (k_0, m). Denote the steady-state probabilities of this new model by $\tilde{P}$. For simplification of the notation, P_{nm} and $\tilde{P}_{k_0 m}$ will be used for $P(N = n, M = m)$ and $P(K_0 = k_0, M = m)$, respectively. Then, for any m,

$$\tilde{P}_{0m} = \sum_{n \leq S_0} P_{nm},$$

$$\tilde{P}_{k_0 m} = P_{S_0 + k_0, m} \qquad for \ \ k_0 > 0.$$

Let $q(m)$ be the steady-state probability that an arriving request for an item at the depot has to wait, given that it finds no other waiting requests in front of it. Using the fact that Poisson arrivals see time averages, we have

$$q(m) = \frac{P_{S_0 m}}{\sum_{n \leq S_0} P_{nm}} \qquad for \ every \ m,$$

while the transition diagram corresponding to the alternative state space description is displayed by Figure 3.

Lemma 1. *The model with state description (k_0, m) and transition rates as denoted in Figure 3 is an aggregate formulation of the one with state description (n, m) and transition rates as denoted in Figure 2. Balance equations of the aggregate model for any m are*

$$(\lambda + m\mu_1)\, \tilde{P}_{0m} = \lambda(1 - q(m-1))\, I_{\{m>0\}} \tilde{P}_{0,m-1} + \mu_0\, I_{\{m>0\}} \tilde{P}_{1,m-1} + (m+1)\mu_1\, \tilde{P}_{0,m+1} \quad for \ k_0 = 0, \ (1)$$

$$(\lambda + \mu_0 + m\mu_1)\ \tilde{P}_{1m} \ = \ \lambda q(m)\tilde{P}_{0m} \ + \ \mu_0 I_{\{m>0\}}\tilde{P}_{2,m-1} \ + \ (m+1)\mu_1 \tilde{P}_{1,m+1} \quad for\ k_0 = 1, \tag{2}$$

$$(\lambda + \mu_0 + m\mu_1)\ \tilde{P}_{k_0 m} \ = \ \lambda \tilde{P}_{k_0-1,m} \ + \ \mu_0 I_{\{m>0\}}\tilde{P}_{k_0+1,m-1} \ + \ (m+1)\mu_1 \tilde{P}_{k_0,m+1} \quad for\ k_0 > 1. \tag{3}$$

Proof: The first part of the lemma (aggregation) follows immediately from the above discussion and the fact that any demand arrival at the depot in case no backlog exists, leads to a backlog with probability $q(m)$, and to an immediate shipment of an item from the depot stock to the base with probability $1-q(m)$. To prove the second part of the lemma, we start with the balance equations for the model with state (n, m). For any m, these equations are

$$(\lambda + \mu_0 I_{\{n>0\}} + m\mu_1)P_{nm} \ = \ \lambda I_{\{n>0\}} I_{\{m>0\}} P_{n-1,m-1} + \mu_0 P_{n+1,m} + (m+1)\mu_1 P_{n,m+1} \quad for\ \ n < S_0, \tag{4}$$

$$(\lambda + \mu_0 I_{\{S_0>0\}} + m\mu_1)P_{S_0 m} = \lambda I_{\{S_0>0\}} I_{\{m>0\}} P_{S_0-1,m-1} + \mu_0 I_{\{m>0\}} P_{S_0+1,m-1} + (m+1)\mu_1 P_{S_0,m+1} \quad for\ \ n = S_0, \tag{5}$$

$$(\lambda + \mu_0 + m\mu_1)P_{nm} \ = \ \lambda I_{\{n>0\}} P_{n-1,m} \ + \ \mu_0 I_{\{m>0\}} P_{n+1,m-1} \ + \ (m+1)\mu_1 P_{n,m+1} \quad for\ \ n > S_0. \tag{6}$$

Balance equations of the model with state (k_0, m) would be obtained as follows: For each $k_0 \geq 1$, i.e., $n = S_0 + k_0 \geq (S_0 + 1)$, and any m, the balance equations will be (6). Summation of the balance equations in (4) and (5) over all $0 \leq n \leq S_0$ results in the balance equation for $k_0 = 0$ and any m. Below, we give the details.

For $k_0 > 1$, since $\tilde{P}_{k_0 m} = P_{S_0+k_0,m}$, equation (6) is written as in (3).

For $k_0 = 1$, (6) becomes

$$(\lambda + \mu_0 + m\mu_1)\ \tilde{P}_{1m} = \lambda P_{S_0 m} \ + \ \mu_0 I_{\{m>0\}}\tilde{P}_{2,m-1} \ + \ (m+1)\mu_1 \tilde{P}_{1,m+1}.$$

Rewriting the first term on the right hand side in terms of $q(m)$ yields (2).

For $k_0 = 0$, summation of (4) and (5) gives

$$\sum_{n \leq S_0} (\lambda + \mu_0 I_{\{n>0\}} + m\mu_1)P_{nm} \ = \ \sum_{n \leq S_0} \lambda I_{\{n>0\}} I_{\{m>0\}} P_{n-1,m-1} \ + \ \sum_{n \leq (S_0-1)} \mu_0 P_{n+1,m}$$
$$+ \ \mu_0 I_{\{m>0\}} P_{S_0+1,m-1} \ + \ \sum_{n \leq S_0} (m+1)\mu_1 P_{n,m+1}.$$

Taking $n' = n - 1$, we obtain

$$(\lambda + m\mu_1)\ \tilde{P}_{0m} \ + \ \sum_{n' \leq (S_0-1)} \mu_0 P_{n'+1,m} \ = \ \sum_{n' \leq (S_0-1)} \lambda I_{\{m>0\}} P_{n',m-1} \ + \ \sum_{n \leq (S_0-1)} \mu_0 P_{n+1,m}$$
$$+ \ \mu_0 I_{\{m>0\}} \tilde{P}_{1,m-1} \ + \ (m+1)\mu_1 \tilde{P}_{0,m+1}.$$

The second term on the left hand side and the second term on the right hand side are cancelled. Since

$$\sum_{n' \leq (S_0-1)} P_{n',m-1} \ = \ \tilde{P}_{0,m-1} - P_{S_0,m-1}$$

we may rewrite the first term on the right hand side as $\lambda(1 - q(m-1))I_{\{m>0\}}\tilde{P}_{0,m-1}$ to obtain (1). $\square$

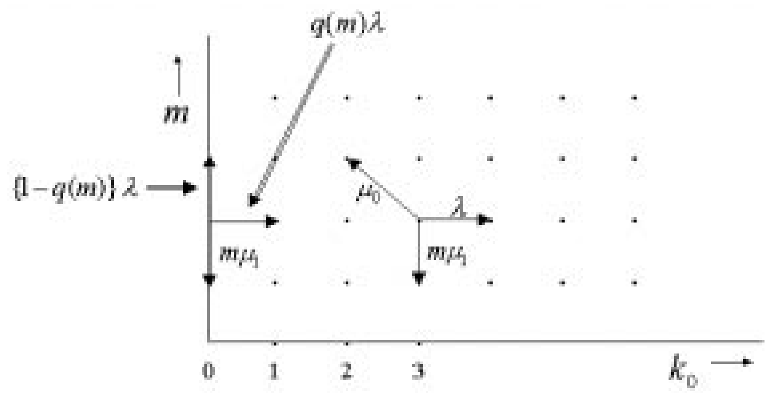

Figure 3: Transition diagram for state (k_0, m)

The difficulty in the description of the aggregate model is of course in the dependence of $q(m)$ on m. Basically, the modification to be discussed below comes down to ignoring this dependence. Let $P_0(N = n)$ and $P_1(M = m)$ denote the marginal probability distributions of the states at the central repair facility and the base, respectively. Let further $\rho_0 = (\lambda/\mu_0)$ and $\rho_1 = (\lambda/\mu_1)$. Then,

$$P_0(N = n) \; = \; (1 - \rho_0)\rho_0^n,$$

$$P_1(M = m) \; = \; \frac{\rho_1^m}{m!}\, e^{-\rho_1}.$$

The modification of the aggregate model with state description (k_0, m) proposed in this article is based on the ignorance of the dependence of $q(m)$ on m, meaning that in the balance equations of the states (k_0, m) the conditional probabilities $q(m)$ are all replaced by

$$q \; = \; \frac{P_0(N = S_0)}{P_0(N \le S_0)} \; = \; \frac{(1 - \rho_0)\rho_0^{S_0}}{(1 - \rho_0^{S_0+1})}$$

Note that q is the expected probability that a request arriving at the depot has to be backordered when this request does not see any backordered item at the depot at the time of its arrival, while not observing the value of m, i.e.,

$$q \; = \; \sum_{m=0}^{\infty} q(m) P(M = m | N \le S_0).$$

Hence, q can be seen as a weighted average of the values $q(m)$. The following result is essential for the analysis in this paper.

Theorem 1. *For the modified aggregate model, the steady-state distribution is given by*

$$\check{P}(K_0 = k_0, M = m) \; = \; \begin{cases} (1 - \rho_0^{S_0+1})\; P_1(M = m) & \text{for } k_0 = 0, \\ (1 - \rho_0)\rho_0^{S_0+k_0}\; P_1(M = m) & \text{for } k_0 > 0. \end{cases}$$

Proof: The result follows immediately by substituting the given distribution in the balance equations of Lemma 1, in which $q(m)$ is replaced by q. □

As noted earlier, for $k_0 > 0$ the system behaves as a tandem queuing system and hence the product form seems natural, given the one-to-one correspondence between k_0 and n in this case. The states with $k_0 = 0$ represent the aggregation and hence satisfy different transition rates. The solution given in the above theorem will be referred to as a near-product form solution, for obvious reasons. The reader may note that for $S_0 = 0$ the modified model is equivalent to the original model, since in this case $q(m) = q = 1$ for all m.

Basically, the modification suggested above is similar to approximations suggested by Buzacott et al. [4] and Buzacott and Shanthikumar [5]. However, by viewing the approximation in terms of the conditional probabilities $q(m)$ and q, and by explicitly considering the balance equations of the aggregate system with states (k_0, m), the near-product form immediately follows, which in turn eliminates the need to consider the marginal state space of the base explicitly. The above references decompose the system completely, and therefore require an explicit description of the arrival process of items at the second stage (the base) which is definitely *not* a Poisson process (see, e.g., the derivation in the first appendix of Buzacott et al. [4]). The derivation of the near-product form solution eliminates the need for this decomposition.

Performance measures typically considered for repairable item systems are the stockout probability, the fill rate and the expected stockout at the base. The fill rate is defined as the probability that any demand arriving at the base can be fulfilled immediately. To obtain these performance measures, the distribution of the random variable K_1 must be computed. This distribution satisfies

$$
\begin{aligned}
Pr(K_1 = 0) &= \sum_{k_0+m \leq S_1} \tilde{P}_{k_0 m}, \\
Pr(K_1 = k_1) &= \sum_{k_0+m = S_1+k_1} \tilde{P}_{k_0 m} \qquad for \;\; k_1 > 0.
\end{aligned}
$$

Then, in terms of this distribution the performance measures at the base are

Stockout Probability (SP) : $Pr(K_1 > 0)$,

Fill Rate (FR) : $Pr(K_0 + M < S_1) = Pr(K_1 = 0) - Pr(K_0 + M = S_1)$,

Expected Stockout (ES) : $E(K_1)$.

In order to assess the performance of the proposed approximation, the measures regarding K_1 listed above are computed both by simulation and by using the analytical near product form solution of the modified system. The simulation results are based on an experimental setting with 15 replications and a simulation time of 10^6 for each replication. In Tables 1, 2 and 3, averages of the measures obtained by the replications and confidence intervals are given for a confidence level of 0.95. Service rates μ_0 and μ_1 are 10. Table 1 is for the failure rate λ being 8, Table 2 is for $\lambda = 9.5$ and Table 3 is for $\lambda = 5$. The results clearly show that the approximate solution performs excellently. A comparison of the results based on the solution of the modified model with the simulation results for different target stock levels shows errors less than 1 percent. Hence, it seems natural to investigate a similar modification for more complex repairable item systems, to see whether similar analytical results can be obtained from which equally accurate approximations can be derived. That will be the topic of the next section.

3 General capacitated two-echelon repairable item systems

In this section, we investigate complex but more realistic two-echelon repairable item systems, derive analytical results for a slightly modified aggregate system and study the performance of the solution of this modified system as an approximation for the exact performance measures. The basic characteristics of the systems studied in this section are listed below.

Table 1: Performance Measures for $\rho_0 = \rho_1 = 8/10$

S_0	S_1	SP_{sim}	SP_{app}	FR_{sim}	FR_{app}	ES_{sim}	ES_{app}
0	3	0.49996 (0.49973,0.50019)	0.49987	0.37741 (0.37730,0.37753)	0.37744	2.50451 (2.49982,2.50920)	2.50095
0	4	0.40034 (0.40002,0.40065)	0.40018	0.50004 (0.49981,0.50027)	0.50013	2.00455 (2.00003,2.00907)	2.00109
1	3	0.40355 (0.40326,0.40385)	0.40171	0.48996 (0.48975,0.49018)	0.49246	2.00918 (2.00462,2.01373)	2.00290
0	5	0.32034 (0.31999,0.32069)	0.32018	0.59966 (0.59935,0.59998)	0.59982	1.60421 (1.59992,1.60850)	1.60091
1	4	0.32138 (0.32103,0.32174)	0.32042	0.59645 (0.59615,0.59674)	0.59829	1.60562 (1.60129,1.60996)	1.60119
2	3	0.32557 (0.32524,0.32589)	0.32318	0.58155 (0.58127,0.58183)	0.58449	1.61139 (1.60730,1.61549)	1.60446
0	6	0.25633 (0.25599,0.25668)	0.25615	0.67966 (0.67931,0.68001)	0.67982	1.28387 (1.27986,1.28788)	1.28073
1	5	0.25664 (0.25629,0.25699)	0.25618	0.67862 (0.67826,0.67897)	0.67958	1.28424 (1.28019,1.28829)	1.28077
2	4	0.25783 (0.25750,0.25817)	0.25662	0.67443 (0.67411,0.67476)	0.67682	1.28583 (1.28198,1.28967)	1.28128
3	3	0.26256 (0.26225,0.26287)	0.26036	0.65558 (0.65529,0.65587)	0.65811	1.29213 (1.28819,1.29608)	1.28571
0	7	0.20512 (0.20477,0.20547)	0.20492	0.74367 (0.74332,0.74401)	0.74385	1.02754 (1.02381,1.03127)	1.02459
1	6	0.20521 (0.20486,0.20556)	0.20492	0.74336 (0.74301,0.74371)	0.74382	1.02760 (1.02383,1.03136)	1.02459
2	5	0.20553 (0.20520,0.20586)	0.20498	0.74217 (0.74183,0.74250)	0.74338	1.02799 (1.02441,1.03158)	1.02466
3	4	0.20671 (0.20640,0.20701)	0.20558	0.73744 (0.73713,0.73775)	0.73964	1.02958 (1.02587,1.03328)	1.02535
4	3	0.21203 (0.21171,0.21236)	0.21011	0.71490 (0.71460,0.71520)	0.71700	1.03673 (1.03308,1.04037)	1.03071
0	0	0.91012 (0.91007,0.91017)	0.91013	0	0	4.80357 (4.79877,4.80837)	4.80000
10	10	0.01138 (0.01126,0.01150)	0.01127	0.98579 (0.98566,0.98592)	0.98592	0.05718 (0.05574,0.05863)	0.05633
5	5	0.10549 (0.10517,0.10582)	0.10504	0.86700 (0.86666,0.86733)	0.86792	0.52748 (0.52450,0.53046)	0.52472
8	2	0.14490 (0.14461,0.14520)	0.14392	0.71250 (0.71224,0.71275)	0.71230	0.57690 (0.57401,0.57980)	0.57241
2	8	0.10509 (0.10477,0.10541)	0.10492	0.86865 (0.86833,0.86897)	0.86885	0.52687 (0.52399,0.52975)	0.52459

Table 2: Performance Measures for $\rho_0 = \rho_1 = 9.5/10$

S_0	S_1	SP_{sim}	SP_{app}	FR_{sim}	FR_{app}	ES_{sim}	ES_{app}
0	4	0.81349 (0.81319,0.81378)	0.81339	0.14390 (0.14370,0.14411)	0.14392	16.27863 (16.17366,16.38359)	16.26181
0	9	0.62979 (0.62904,0.63054)	0.62940	0.33718 (0.33652,0.33783)	0.33744	12.59754 (12.49388,12.70119)	12.58164
2	7	0.62983 (0.62905,0.63061)	0.62940	0.33708 (0.33639,0.33776)	0.33743	12.59683 (12.49285,12.70081)	16.58164
4	5	0.63020 (0.62947,0.63094)	0.62948	0.33599 (0.33533,0.33664)	0.33691	12.59774 (12.49277,12.70270)	12.58173
5	4	0.63104 (0.63028,0.63181)	0.63005	0.33270 (0.33202,0.33337)	0.33394	12.59836 (12.49373,12.70298)	12.58240
0	20	0.35868 (0.35717,0.36020)	0.35799	0.62243 (0.62099,0.62388)	0.62313	7.16242 (7.06803,7.25681)	7.15376
5	15	0.35869 (0.35719,0.36019)	0.35799	0.62243 (0.62098,0.62388)	0.62313	7.16133 (7.06639,7.25627)	7.15376
10	10	0.35870 (0.35722,0.36018)	0.35799	0.62242 (0.62101,0.62384)	0.62313	7.16251 (7.06806,7.25696)	7.15376
16	4	0.36065 (0.35916,0.36214)	0.35963	0.61298 (0.61157,0.61439)	0.61422	7.16435 (7.06914,7.25956)	7.15569
20	30	0.07650 (0.07498,0.07802)	0.07681	0.91943 (0.91789,0.92101)	0.91911	1.53359 (1.48697,1.58021)	1.53122

Table 3: Performance Measures for $\rho_0 = \rho_1 = 5/10$

S_0	S_1	SP_{sim}	SP_{app}	FR_{sim}	FR_{app}	ES_{sim}	ES_{app}
0	2	0.20392 (0.20380,0.20405)	0.20393	0.60652 (0.60643,0.60662)	0.60653	0.40977 (0.40919,0.41036)	0.40980
2	4	0.01335 (0.01327,0.01342)	0.01301	0.97183 (0.97173,0.97193)	0.97298	0.02636 (0.02614,0.02658)	0.02590
2	3	0.02817 (0.02807,0.02827)	0.02702	0.93564 (0.93553,0.93575)	0.93823	0.05453 (0.05422,0.05485)	0.05292
3	3	0.01510 (0.01503,0.01517)	0.01439	0.96048 (0.96039,0.96057)	0.96192	0.02842 (0.02819,0.02865)	0.02743
1	4	0.02623 (0.02612,0.02634)	0.02584	0.94638 (0.94626,0.94651)	0.94770	0.05211 (0.05179,0.05244)	0.05161

- *Repair units both at the depot and at each base:* Failed items arrive at each base according to a (possibly base-dependent) Poisson rate. The base is responsible for replacing the failed item with a ready-for-use spare part. A failed item is either repaired at a local base repair shop, after which it is stocked at the base as a ready-for-use item again, or sent to the central repair facility. In the latter case, at the same time a ready-for-use item is requested by the base from the depot to replenish its local base stock, while the item repaired at the central repair facility is placed in the depot upon completion.

- *Multiple bases:* The complete system consists of several bases and one central repair facility with a depot of ready-for-use spare parts. Requests from the bases for items from the depot stock are fulfilled on a first-come-first-serve basis.

- *Product form repair networks:* Each repair shop is modeled as an open Jackson network. Failed items that are returned from a base to the central repair facility follow a (probabilistic) routing in the central repair facility, with given process time distributions at each node being independent of the base that shipped them (see Section 5 for a relaxation of this assumption). However, requests from the different bases for new items to replenish their local stocks are distinguished in order to make sure that each shipped item arrives at the correct base. Transport times from the central depot to the bases may be base-dependent.

The most complicated maintenance system for which the approximation is investigated in this article is obtained by combining all the extensions above. Such a system with two bases is displayed in Figure 4.

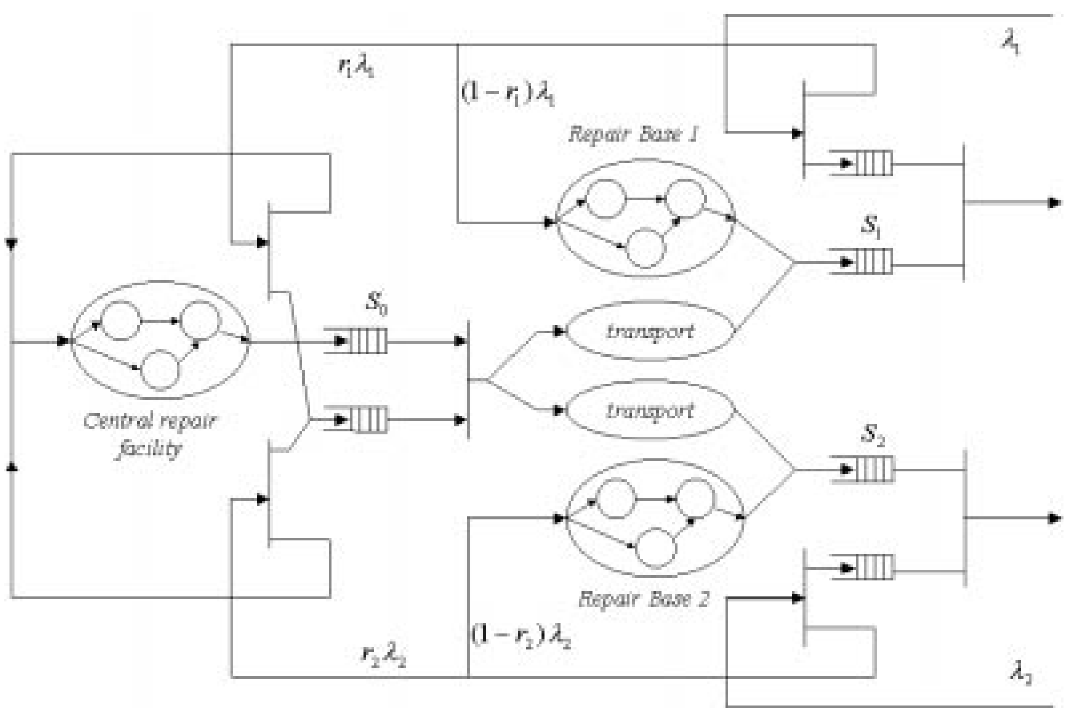

Figure 4: A two-echelon, single indenture model

3.1 Modification and analysis of general repairable item systems

In this subsection, an extension of the analysis of modified aggregate systems is obtained and the existence of near-product form steady-state distributions is proven for the complex systems introduced above. First, we have to revise the notations introduced in the previous section. Let L be the number of bases and let M_l be the random variable describing the number of items in transfer from the depot to base l $(l = 1, 2, ..., L)$. The target spare part inventory level is denoted by S_0 at the depot and by S_l at base l, the random variable denoting the number of requests not filled (backlogs) at that base is K_l while $\bar{M}_l$ is the random variable denoting the number of available spare parts that are ready-for-use, for $l = 1, 2, ..., L$. Suppose there are J nodes within the central repair network including an ample server node that represents transportation from the bases to the depot. The number of items being served or waiting to be served at node j of this network is denoted by random variable N_j for $j = 1, ..., J$. The states of the central repair network are represented by vectors $\boldsymbol{n} = (n_1, n_2, ..., n_J)$, where n_j is the number of items being in repair at node j $(j = 1, ..., J)$ while $\bar{n}$ denotes the number of ready-for-use spare parts stocked at the depot. Let Z_l be the number of nodes at the repair network of base l, then similarly the states of the local repair network of base l are represented by vectors $\boldsymbol{h}_l = (h_{l1}, h_{l2}, ..., h_{lZ_l})$, for $l = 1, 2, ..., L$. Also, let $\boldsymbol{h} = (\boldsymbol{h}_1, \boldsymbol{h}_2, ..., \boldsymbol{h}_L)$. The vector $\boldsymbol{m} = (m_1, m_2, ..., m_L)$ describes the number of spare parts being transported from the depot to each base by base-specific ample servers, while $\bar{m}_l$ denotes the number of ready-for-use spare parts at base l $(l = 1, 2, ..., L)$. Finally, let k_{0l} denote the number of requests from base l $(l = 1, 2, ..., L)$, waiting to be fulfilled at the central depot, and let $\boldsymbol{k}_0 = (k_{01}, k_{02}, ..., k_{0L})$. For reasons to be explained below, we also define $k_0 = \sum_{l=1}^{L} k_{0l}$ as the total number of backordered requests at the central depot. Then, the state definition for the complete repairable item system is denoted by $(\boldsymbol{k}_0, \boldsymbol{n}, \boldsymbol{m}, \boldsymbol{h})$, while furthermore the following relations have to be satisfied:

$$If \;\; |\boldsymbol{n}| \le S_0, \;\; then \;\; \bar{n} = S_0 - |\boldsymbol{n}| \;\; and \;\; k_0 = 0;$$

$$If \;\; |\boldsymbol{n}| > S_0, \;\; then \;\; \bar{n} = 0 \;\; and \;\; k_0 = |\boldsymbol{n}| - S_0;$$

$$For \;\; l = 1, 2, ..., L,$$

$$If \;\; |\boldsymbol{h}_l| + m_l + k_{0l} \le S_l, \;\; then \;\; \bar{m}_l = S_l - (|\boldsymbol{h}_l| + m_l + k_{0l}) \;\; and \;\; k_l = 0;$$

$$If \;\; |\boldsymbol{h}_l| + m_l + k_{0l} > S_l, \;\; then \;\; \bar{m}_l = 0 \;\; and \;\; k_l = (|\boldsymbol{h}_l| + m_l + k_{0l}) - S_l.$$

The total failure rate is $\lambda = \sum_{l=1}^{L} \lambda_l$ with λ_l being the failure rate at base l, $l = 1, 2, ...L$. For each node j (z) of the central network (base l), the exponential service rate is denoted by μ_{0j} (μ_{lz}). Defining c_{0j} as the number of parallel servers at node j and $\mu_{0j}(n_j)$ being the service rate when there are n_j items at the node, we have

$$\mu_{0j}(n_j) = \begin{cases} n_j \mu_{0j} & if \;\; n_j < c_{0j}, \\ c_{0j} \mu_{0j} & if \;\; n_j \ge c_{0j}. \end{cases}$$

For each base l, c_{lz} and $\mu_{lz}(h_{lz})$ are defined similarly. Let r_l denote the probability that a failed item arriving at base l has to be sent to the central repair facility, hence $(1 - r_l)$ denotes the probability that the item can be repaired at the local base repair shop. Then, $\lambda_{CR} = \sum_{l=1}^{L} r_l \lambda_l$ denotes the total arrival rate of failed items (and hence also of replenishment requests from the bases) at the central repair shop while $(1 - r_l)\lambda_l$ is the arrival rate of items to be repaired at local repair shop l, for $l = 1, 2, ..., L$.

Let p_{ij} be the routing probability for an item being repaired at the central network to go from node i to node j. Denote the routing probability that an arriving item is sent to node j by p_{0j} and the routing probability of leaving the network after completion of the service at node j by p_{j0}. The total arrival rate $\hat{\lambda}_j$

into each node j of the central network is computed from

$$\hat{\lambda}_j = \left(\sum_{l=1}^{L} r_l \lambda_l\right) p_{0j} \; + \; \sum_{i=1}^{J} \hat{\lambda}_i \, p_{ij} \qquad for \;\; j = 1, ..., J. \tag{7}$$

Clearly the system is ergodic if $\hat{\lambda}_j < c_{0j}\mu_{0j}$ for $j = 1, ..., J$. The routing matrices and the effective arrival rates for the networks at the bases are defined similarly and similar ergodicity conditions hold.

Before continuing, we will slightly simplify the state space description by making an early aggregation step. Note that for a full system description the vector $\boldsymbol{k}_0 = (k_{01}, k_{02}, ..., k_{0L})$ is needed in order to determine the number of ready-for-use items and the backlogs at each base l. However, the system structure does not change at all if, as soon as a shipment from the central depot to a local base is initiated, we assume that with probability $r_l\lambda_l/\lambda_{CR}$ the item is shipped to base l. Note that physically all requests can be seen as standing in one queue in the order of arrival, since they are handled on a first-come-first serve basis. Hence, if we know the total number k_0 of backordered requests at the depot, the probability distribution of the values k_{0l} is easily determined by the binomial formula

$$P(K_{0l} = k_{0l}) = \binom{k_0}{k_{0l}} (r_l\lambda_l/\lambda_{CR})^{k_{0l}} (1 - (r_l\lambda_l/\lambda_{CR}))^{k_0 - k_{0l}},$$

while the joint probability distribution of the vectors $\boldsymbol{k}_0$ is given by the well-known multinomial distribution

$$P(K_{01} = k_{01}, ..., K_{0L} = k_{0L}) = \frac{k_0!}{\left(\prod_{l=1}^{L} k_{0l}!\right)} \left(\prod_{l=1}^{L} (r_l\lambda_l/\lambda_{CR})^{k_{0l}}\right).$$

Since k_0 is completely determined by the cardinality of the vector $\boldsymbol{n}$ (see the above state relations), it follows that we may restrict ourselves to the state description $(\boldsymbol{n}, \boldsymbol{m}, \boldsymbol{h})$ without loosing any generality, since the routing probabilities $(r_l\lambda_l/\lambda_{CR})$ can be used to derive the more detailed probability distributions on the initial state space described by $(\boldsymbol{k}_0, \boldsymbol{n}, \boldsymbol{m}, \boldsymbol{h})$, at least if some sort of product form solution arises again.

Now, we first concentrate on a special parameter setting for the above system in which $r_l = 1$ for all l, i.e., we ignore for the moment the local base repair shops and assume that all failed items have to be repaired at the central facility. Hence, for the moment we assume $\lambda_{CR} = \lambda$, i.e., we concentrate on the two-echelon system consisting of the central repair facility, its associated depot stock and possible requests from the bases, and the transport nodes to all bases. For this system, we first show that a modification similar to the one made in Section 2 leads to a set of balance equations with a near-product form solution. Next, the results are easily extended to the system with possible repairs at the local base shops as well.

For the system with $r_l = 1$ for all l, the state description becomes $(\boldsymbol{n}, \boldsymbol{m})$. Analogous to the development in section 2, we define an aggregate model with states described by $(k_0, \boldsymbol{m})$ as follows:

$$(0, \boldsymbol{m}) \qquad for \;\; |\boldsymbol{n}| \le S_0,$$

$$(k_0, \boldsymbol{m}) \quad for \;\; |\boldsymbol{n}| = S_0 + k_0, \;\; k_0 > 0.$$

The relation between the two models is investigated below. For any $\boldsymbol{m}$,

$$\tilde{P}_{0\boldsymbol{m}} \;=\; \sum_{|\boldsymbol{n}| \le S_0} P_{\boldsymbol{n}\boldsymbol{m}},$$

$$\tilde{P}_{k_0\boldsymbol{m}} \;=\; \sum_{|\boldsymbol{n}| = S_0 + k_0} P_{\boldsymbol{n}\boldsymbol{m}} \qquad for \;\; k_0 > 0.$$

Note that now also for $k_0 > 0$ the state $(k_0, \boldsymbol{m})$ is an aggregate state. Using notations similar to those introduced in Section 2 for the steady-state probabilities, we define $q(\boldsymbol{m})$ by:

$$q(\boldsymbol{m}) \;=\; \frac{\sum_{|\boldsymbol{n}| = S_0} P_{\boldsymbol{n}\boldsymbol{m}}}{\sum_{|\boldsymbol{n}| \le S_0} P_{\boldsymbol{n}\boldsymbol{m}}} \qquad for\ every\ \boldsymbol{m}.$$

Again, $q(\boldsymbol{m})$ denotes the probability that a request arriving at the central depot for a ready-for-use item finds no other requests waiting to be fulfilled while nevertheless the depot stock is depleted. In addition, let

$$\hat{\mu}(k_0, \boldsymbol{m}) \;=\; \sum_{|\boldsymbol{n}|=S_0+k_0} \left(\sum_{j=1}^{J} p_{j0}\, \mu_{0j}(n_j)\, I_{\{n_j>0\}} \right) \frac{P_{\boldsymbol{nm}}}{\sum_{|\boldsymbol{n}|=S_0+k_0} P_{\boldsymbol{nm}}} \qquad \text{for every } \; k_0 > 0 \; \text{ and } \; \boldsymbol{m}.$$

The rate $\hat{\mu}(k_0, \boldsymbol{m})$ is the conditional departure rate from the depot given that there are $k_0 > 0$ items being backordered. Define $\boldsymbol{e}_j$ as a J-dimensional vector with all of its entries being zero except the j^{th} entry which is 1 and $\boldsymbol{f}_l$ as a similar L-dimensional vector.

Lemma 2. *The model with state description* $(k_0, \boldsymbol{m})$ *is an aggregate formulation of the one with state description* $(\boldsymbol{n}, \boldsymbol{m})$*. The balance equations of the aggregate model for any* $\boldsymbol{m}$ *are*

$$\begin{aligned}
&\lambda \tilde{P}_{0\boldsymbol{m}} \;+\; \left(\sum_{l=1}^{L} m_l \mu_l \right) \tilde{P}_{0\boldsymbol{m}} \\
&= \sum_{l=1}^{L} \lambda_l (1 - q(\boldsymbol{m} - \boldsymbol{f}_l)) \tilde{P}_{0,\boldsymbol{m}-\boldsymbol{f}_l} I_{\{m_l>0\}} \;+\; \sum_{l=1}^{L} \hat{\mu}(1, \boldsymbol{m} - \boldsymbol{f}_l) \tilde{P}_{1,\boldsymbol{m}-\boldsymbol{f}_l} I_{\{m_l>0\}} \left(\frac{\lambda_l}{\lambda} \right) \\
&\quad + \sum_{l=1}^{L} (m_l + 1) \mu_l \; \tilde{P}_{0,\boldsymbol{m}+\boldsymbol{f}_l} \qquad \text{for } \; k_0 = 0,
\end{aligned} \tag{8}$$

$$\begin{aligned}
&\lambda \tilde{P}_{1\boldsymbol{m}} \;+\; \left(\sum_{l=1}^{L} m_l \mu_l \right) \tilde{P}_{1\boldsymbol{m}} \;+\; \hat{\mu}(1, \boldsymbol{m}) \tilde{P}_{1\boldsymbol{m}} \\
&= \lambda q(\boldsymbol{m}) \tilde{P}_{0\boldsymbol{m}} \;+\; \sum_{l=1}^{L} \hat{\mu}(2, \boldsymbol{m} - \boldsymbol{f}_l) \tilde{P}_{2,\boldsymbol{m}-\boldsymbol{f}_l} I_{\{m_l>0\}} \left(\frac{\lambda_l}{\lambda} \right) \;+\; \sum_{l=1}^{L} (m_l + 1) \mu_l \; \tilde{P}_{1,\boldsymbol{m}+\boldsymbol{f}_l} \;\; \text{for } \; k_0 = 1,
\end{aligned} \tag{9}$$

$$\begin{aligned}
&\lambda \tilde{P}_{k_0\boldsymbol{m}} \;+\; \left(\sum_{l=1}^{L} m_l \mu_l \right) \tilde{P}_{k_0\boldsymbol{m}} \;+\; \hat{\mu}(k_0, \boldsymbol{m}) \tilde{P}_{k_0\boldsymbol{m}} \\
&= \lambda \tilde{P}_{k_0-1,\boldsymbol{m}} \;+\; \sum_{l=1}^{L} \hat{\mu}(k_0 + 1, \boldsymbol{m} - \boldsymbol{f}_l) \tilde{P}_{k_0+1,\boldsymbol{m}-\boldsymbol{f}_l} I_{\{m_l>0\}} \left(\frac{\lambda_l}{\lambda} \right) \\
&\quad + \sum_{l=1}^{2} (m_l + 1) \mu_l \; \tilde{P}_{k_0,\boldsymbol{m}+\boldsymbol{f}_l} \qquad \text{for } \; k_0 > 1.
\end{aligned} \tag{10}$$

Proof: First, consider the balance equations for the model with states $(\boldsymbol{n}, \boldsymbol{m})$. For any $\boldsymbol{m}$, these equations are

$$\Big(\lambda + \sum_{j=1}^{J} \mu_{0j}(n_j) I_{\{n_j>0\}} + \sum_{l=1}^{L} m_l \mu_l \Big) P_{\boldsymbol{n}\boldsymbol{m}}$$

$$= \sum_{j=1}^{J} \sum_{l=1}^{L} \lambda_l \, P_{\boldsymbol{n}-\boldsymbol{e}_j,\boldsymbol{m}-\boldsymbol{f}_l} I_{\{m_l>0\}} I_{\{n_j>0\}} p_{0j} \; + \; \sum_{j=1}^{J} \mu_{0j}(n_j+1) \, P_{\boldsymbol{n}+\boldsymbol{e}_j,\boldsymbol{m}} \, p_{j0},$$

$$+ \sum_{j=1}^{J} \sum_{i=1}^{J} \mu_{0i}(n_i+1) \, P_{\boldsymbol{n}+\boldsymbol{e}_i-\boldsymbol{e}_j,\boldsymbol{m}} \, I_{\{n_j>0\}} p_{ij} \; + \; \sum_{l=1}^{L} (m_l+1)\mu_l \, P_{\boldsymbol{n},\boldsymbol{m}+\boldsymbol{f}_l} \qquad \textit{for } \boldsymbol{n}, \; |\boldsymbol{n}| < S_0, \quad (11)$$

$$\Big(\lambda + \sum_{j=1}^{J} \mu_{0j}(n_j) I_{\{n_j>0\}} + \sum_{l=1}^{L} m_l \mu_l \Big) P_{\boldsymbol{n}\boldsymbol{m}}$$

$$= \sum_{j=1}^{J} \sum_{l=1}^{L} \lambda_l \, P_{\boldsymbol{n}-\boldsymbol{e}_j,\boldsymbol{m}-\boldsymbol{f}_l} I_{\{m_l>0\}} I_{\{n_j>0\}} p_{0j} \; + \; \sum_{j=1}^{J} \mu_{0j}(n_j+1) \sum_{l=1}^{L} P_{\boldsymbol{n}+\boldsymbol{e}_j,\boldsymbol{m}-\boldsymbol{f}_l} I_{\{m_l>0\}} p_{j0} \left(\frac{\lambda_l}{\lambda} \right)$$

$$+ \sum_{j=1}^{J} \sum_{i=1}^{J} \mu_{0i}(n_i+1) \, P_{\boldsymbol{n}+\boldsymbol{e}_i-\boldsymbol{e}_j,\boldsymbol{m}} \, I_{\{n_j>0\}} p_{ij} \; + \; \sum_{l=1}^{L} (m_l+1)\mu_l \, P_{\boldsymbol{n},\boldsymbol{m}+\boldsymbol{f}_l} \qquad \textit{for } \boldsymbol{n}, \; |\boldsymbol{n}| = S_0, \quad (12)$$

$$\Big(\lambda + \sum_{j=1}^{J} \mu_{0j}(n_j) I_{\{n_j>0\}} \; + \; \sum_{l=1}^{L} m_l \mu_l \Big) P_{\boldsymbol{n}\boldsymbol{m}}$$

$$= \sum_{j=1}^{J} \lambda \, P_{\boldsymbol{n}-\boldsymbol{e}_j,\boldsymbol{m}} \, I_{\{n_j>0\}} p_{0j} \; + \; \sum_{j=1}^{J} \mu_{0j}(n_j+1) \sum_{l=1}^{L} P_{\boldsymbol{n}+\boldsymbol{e}_j,\boldsymbol{m}-\boldsymbol{f}_l} I_{\{m_l>0\}} p_{j0} \left(\frac{\lambda_l}{\lambda} \right)$$

$$+ \sum_{j=1}^{J} \sum_{i=1}^{J} \mu_{0i}(n_i+1) \, P_{\boldsymbol{n}+\boldsymbol{e}_i-\boldsymbol{e}_j,\boldsymbol{m}} \, I_{\{n_j>0\}} p_{ij} \; + \; \sum_{l=1}^{L} (m_l+1)\mu_l \, P_{\boldsymbol{n},\boldsymbol{m}+\boldsymbol{f}_l}$$

$$\textit{for } \boldsymbol{n}, \; |\boldsymbol{n}| = S_0 + k_0, \; k_0 \geq 1. \quad (13)$$

Next, we turn to the aggregate model with states $(k_0, \boldsymbol{m})$. Summation of the balance equations in (11) and (12) over all $\boldsymbol{n}$ with $|\boldsymbol{n}| \leq S_0$ gives the balance equation for the states $(0, \boldsymbol{m})$. Summation of (13) over all $\boldsymbol{n}$ with $|\boldsymbol{n}| = S_0 + k_0$ leads to the balance equations for all states $(k_0, \boldsymbol{m})$ with $k_0 \geq 1$. Below, we present the detailed derivation.

For $k_0 > 1$,

$$\lambda \tilde{P}_{k_0 \boldsymbol{m}} \; + \; \left(\sum_{l=1}^{L} m_l \mu_l \right) \tilde{P}_{k_0 \boldsymbol{m}} \; + \; \sum_{j=1}^{J} \; \sum_{|\boldsymbol{n}|=S_0+k_0} \mu_{0j}(n_j) P_{\boldsymbol{n}\boldsymbol{m}} I_{\{n_j>0\}}$$

$$= \sum_{j=1}^{J} \lambda p_{0j} \sum_{|\boldsymbol{n}'|=S_0+k_0-1} P_{\boldsymbol{n}'\boldsymbol{m}} \; + \; \sum_{j=1}^{J} p_{j0} \sum_{|\boldsymbol{n}''|=S_0+k_0+1} \mu_{0j}(n''_j) \sum_{l=1}^{L} P_{\boldsymbol{n}'',\boldsymbol{m}-\boldsymbol{f}_l} I_{\{m_l>0\}} I_{\{n''_j>0\}} \left(\frac{\lambda_l}{\lambda} \right)$$

$$+ \sum_{j=1}^{J} \sum_{i=1}^{J} p_{ij} \sum_{|\boldsymbol{n}'''|=S_0+k_0} \mu_{0i}(n'''_i) \, P_{\boldsymbol{n}'''\boldsymbol{m}} I_{\{n'''_i>0\}} \; + \; \sum_{l=1}^{L} (m_l+1)\mu_l \, \tilde{P}_{k_0,\boldsymbol{m}+\boldsymbol{f}_l},$$

where $\boldsymbol{n}' = \boldsymbol{n} - \boldsymbol{e}_j$, $\boldsymbol{n}'' = \boldsymbol{n} + \boldsymbol{e}_j$ and $\boldsymbol{n}''' = \boldsymbol{n} + \boldsymbol{e}_i - \boldsymbol{e}_j$. The first term on the right hand side can be rewritten as $\lambda\tilde{P}_{k_0-1,\boldsymbol{m}}$ because $\sum_{j=1}^{J} p_{0j} = 1$. Next, we note the following relation:

$$\sum_{j=1}^{J} A_j = \sum_{j=1}^{J}\left(p_{j0}A_j \ + \ \sum_{i=1}^{J} p_{ij}A_i\right),$$

for any series $\{A_j;\ j = 1, ..., J\}$. Applying this relation for the third terms both on the left and the right hand sides, and rearranging some other terms yield (10).

For $k_0 = 1$, going through the same steps as in the case of $k_0 > 1$, we find (9).

For $k_0 = 0$, proceeding in the same way and using relations similar to the ones above, we find

$$\lambda\tilde{P}_{0\boldsymbol{m}} \ + \ \left(\sum_{l=1}^{L} m_l\mu_l\right)\tilde{P}_{0\boldsymbol{m}} \ + \ \sum_{j=1}^{J} p_{j0} \sum_{|\boldsymbol{n}|\le S_0} \mu_{0j}(n_j)I_{\{n_j>0\}}P_{\boldsymbol{nm}}$$

$$= \sum_{j=1}^{J} p_{0j}\sum_{l=1}^{L}\lambda_l \sum_{|\boldsymbol{n}'|\le S_0-1} P_{\boldsymbol{n}',\boldsymbol{m}-\boldsymbol{f}_l}I_{\{m_l>0\}} \ + \ \sum_{j=1}^{J} p_{j0}\sum_{|\boldsymbol{n}''|\le S_0} \mu_{0j}(n''_j)I_{\{n''_j>0\}}P_{\boldsymbol{n}''\boldsymbol{m}}$$

$$+ \sum_{j=1}^{J} p_{j0}\sum_{|\boldsymbol{n}''|=S_0+1} \mu_{0j}(n''_j)I_{\{n''_j>0\}}\sum_{l=1}^{L} P_{\boldsymbol{n}'',\boldsymbol{m}-\boldsymbol{f}_l}I_{\{m_l>0\}}\left(\frac{\lambda_l}{\lambda}\right) \ + \ \sum_{l=1}^{L}(m_l+1)\mu_l\ \tilde{P}_{0,\boldsymbol{m}+\boldsymbol{f}_l}.$$

The third term on the left hand side and the second one on the right hand side are cancelled. With further rearrangements, the balance equations in (8) are obtained. □

Note that, since the transportation nodes are ample exponential servers, we have $P_l(M_l = m_l) = \frac{\rho_l^{m_l}}{m_l!}\,e^{-\rho_l}$, where $\rho_l = (\lambda_l/\mu_l)$ for $l = 1, 2, ...L$. P_0 is expressed in terms of the usual product form solutions of the nodes, P_{0j}'s, within the network representing the central repair facility, i.e.,

$$P_0(\boldsymbol{N} = \boldsymbol{n}) \ = \ \prod_{j=1}^{J} P_{0j}(N_j = n_j),$$

with $P_{0j}(N_j = n_j) = \frac{\alpha_j(n_j)}{\sum_{u=0}^{\infty}\alpha_j(u)}$ for $n_j = 0, 1, ...$, where $\alpha_j(0) = 1$ and $\alpha_j(u) = \frac{(\hat{\lambda}_j)^u}{\prod_{v=1}^{u}\mu_{0j}(v)}$ for $u = 1, 2, ...$ Now, similar to the approach in the preceding section, the modification of the aggregate model with state description $(k_0, \boldsymbol{m})$ is based on the ignorance of the dependence of $q(\boldsymbol{m})$ on $\boldsymbol{m}$. Hence, in the balance equations of the states $(k_0, \boldsymbol{m})$ the conditional probabilities $q(\boldsymbol{m})$ are all replaced by

$$q \ = \ \frac{\sum_{|\boldsymbol{n}|=S_0} P_0(\boldsymbol{N} = \boldsymbol{n})}{\sum_{|\boldsymbol{n}|\le S_0} P_0(\boldsymbol{N} = \boldsymbol{n})}$$

Note that the approximation only concerns the probabilities $q(\boldsymbol{m})$ but not the transition rates $\hat{\mu}(k_0, \boldsymbol{m})$. This may be a bit surprising since the definition of $\hat{\mu}(k_0, \boldsymbol{m})$ does involve the detailed probabilities $P_{\boldsymbol{nm}}$. However, note that $\hat{\mu}(k_0, \boldsymbol{m})$ has only been defined for $k_0 > 0$. In fact, in the proof of the next theorem we show that in the detailed model all states $(\boldsymbol{n}, \boldsymbol{m})$ with $k_0 > 0$ satisfy a product form solution, once the values $q(\boldsymbol{m})$ have been replaced by q.

Theorem 2. *For the modified aggregate model, the steady-state distribution is given by*

$$\check{P}(K_0 = 0, \boldsymbol{M} = \boldsymbol{m}) \ = \ \left(\sum_{|\boldsymbol{n}|\le S_0} P_0(\boldsymbol{N} = \boldsymbol{n})\right)\left(\prod_{l=1}^{L} P_l(M_l = m_l)\right),$$

$$\check{P}(K_0 = k_0, \boldsymbol{M} = \boldsymbol{m}) \ = \ \left(\sum_{|\boldsymbol{n}|=S_0+k_0} P_0(\boldsymbol{N} = \boldsymbol{n})\right)\left(\prod_{l=1}^{L} P_l(M_l = m_l)\right) \qquad \textit{for} \ \ k_0 > 0.$$

Proof: In order to give the proof for $k_0 > 1$, we return to the detailed system with states $(\boldsymbol{n}, \boldsymbol{m})$. Since in this system, upon repair completion of a part at the central facility, that part is immediately transferred to one of the bases if $k_0 > 0$, it seems natural to invoke in (13) the product form solution

$$\left(\prod_{j=1}^{J} P_{0j}(N_j = n_j)\right)\left(\prod_{l=1}^{L} P_l(M_l = m_l)\right)$$

for all states $(\boldsymbol{n}, \boldsymbol{m})$ with $|\boldsymbol{n}| > S_0+1$. Then, since $\tilde{P}_{k_0\boldsymbol{m}} = \sum_{|\boldsymbol{n}|=S_0+k_0} P_{\boldsymbol{n}\boldsymbol{m}}$, by summing up the equations in (13) we immediately obtain (10) for $k_0 > 1$ with the steady-state solutions given by the theorem.

For $k_0 = 0$, the proof is given with the use of the local balance equations

$$\lambda \tilde{P}_{0\boldsymbol{m}} = \sum_{l=1}^{L} (m_l+1)\mu_l \, \tilde{P}_{0,\boldsymbol{m}+\boldsymbol{f}_l}, \tag{14}$$

$$m_l \mu_l \tilde{P}_{0\boldsymbol{m}} = \lambda_l(1-q(\boldsymbol{m}-\boldsymbol{f}_l))\tilde{P}_{0,\boldsymbol{m}-\boldsymbol{f}_l} I_{\{m_l>0\}} + \hat{\mu}(1,\boldsymbol{m}-\boldsymbol{f}_l)\tilde{P}_{1,\boldsymbol{m}-\boldsymbol{f}_l} I_{\{m_l>0\}}\left(\frac{\lambda_l}{\lambda}\right) \quad \textit{for } l = 1,...,L. \tag{15}$$

Equation (14) balances the total arrivals into the system against the total departures from the system. Equation (15) balances the transitions out of state $(0, \boldsymbol{m})$ due to completions of transportations to base l against the transitions into state $(0, \boldsymbol{m})$ due to the start of transportations to node l. In order to see that these equations hold for the modified aggregate model, replace the $q(\boldsymbol{m}-\boldsymbol{f}_l)$'s and the $\tilde{P}$'s in (8) with q and the $\check{P}$'s given in the theorem, respectively. After cancellation of the common terms obtained through the use of $P_l(M_l = m_l) = P_l(M_l = m_l-1)\frac{\rho_l}{m_l}$ and $\alpha_j(n_j+1) = \alpha_j(n_j)\left(\frac{\hat{\lambda}_j}{\mu_{0j}(n_j+1)}\right)$ for $n_j \geq 0$, equation (14)becomes $\lambda = \sum_{l=1}^{L} \lambda_l$ and the equation in (15) for any base l can be rewritten as

$$\lambda_l P_0(|\boldsymbol{N}| \leq S_0) P_l(M_l = m_l-1) I_{\{m_l>0\}} \left(\prod_{d=1,d\neq l}^{L} P_d(M_d = m_d)\right)$$

$$= \lambda_l\left(P_0(|\boldsymbol{N}| \leq S_0) - P_0(|\boldsymbol{N}| = S_0)\right) P_l(M_l = m_l-1) I_{\{m_l>0\}} \left(\prod_{d=1,d\neq l}^{L} P_d(M_d = m_d)\right)$$

$$+ \sum_{j=1}^{J} p_{j0} \sum_{|\boldsymbol{n}|=S_0} \mu_{0j}(n_j+1)\left(P_0(\boldsymbol{N}=\boldsymbol{n})\frac{\hat{\lambda}_j}{\mu_{0j}(n_j+1)}\right)\left(\frac{\lambda_l}{\lambda}\right) P_l(M_l = m_l-1) I_{\{m_l>0\}} \left(\prod_{d=1,d\neq l}^{L} P_d(M_d = m_d)\right).$$

Since $r_l = 1$ for all l, a summation of the equations in (7) over j yields $\sum_{j=1}^{J} p_{j0}\hat{\lambda}_j = \lambda$. From this, it is immediately observed that the above equation holds .

For $k_0 = 1$, consider the local balance equations

$$\lambda \tilde{P}_{1\boldsymbol{m}} = \sum_{l=1}^{L} (m_l+1)\mu_l \, \tilde{P}_{1,\boldsymbol{m}+\boldsymbol{f}_l}, \tag{16}$$

$$\left(\sum_{l=1}^{L} m_l \mu_l\right)\tilde{P}_{1\boldsymbol{m}} + \hat{\mu}(1,\boldsymbol{m})\tilde{P}_{1\boldsymbol{m}} = \lambda q(\boldsymbol{m})\tilde{P}_{0\boldsymbol{m}} + \sum_{l=1}^{L} \hat{\mu}(2,\boldsymbol{m}-\boldsymbol{f}_l)\tilde{P}_{2,\boldsymbol{m}-\boldsymbol{f}_l} I_{\{m_l>0\}}\left(\frac{\lambda_l}{\lambda}\right). \tag{17}$$

Note that equation (17) is in fact the summation of L local balance equations for the L bases. To observe that the equations (16) and (17) hold for the modified aggregate model, in (9) substitute q and $\check{P}$'s for $q(\boldsymbol{m})$'s and $\tilde{P}$'s, respectively. Proceeding as in the case of $k_0 = 0$, equation (16) is reduced to $\lambda = \sum_{l=1}^{L} \lambda_l$

and equation (17) becomes

$$\sum_{l=1}^{L} \lambda_l P_0(|\boldsymbol{N}| = S_0+1) P_l(M_l = m_l - 1) I_{\{m_l>0\}} \left(\prod_{d=1,d\neq l}^{L} P_d(M_d = m_d) \right)$$

$$+ \sum_{j=1}^{J} p_{j0} \sum_{|\boldsymbol{n}|=S_0+1} \mu_{0j}(n_j) I_{\{n_j>0\}} P_{\boldsymbol{nm}}$$

$$= \lambda P_0(|\boldsymbol{N}| = S_0) \prod_{l=1}^{L} P_l(M_l = m_l) \; + \; \sum_{j=1}^{J} p_{j0} \sum_{|\boldsymbol{n}|=S_0+1} \mu_{0j}(n_j+1) \sum_{l=1}^{L} P_{\boldsymbol{n}+\boldsymbol{e}_j, \boldsymbol{m}-\boldsymbol{f}_l} I_{\{m_l>0\}} \left(\frac{\lambda_l}{\lambda} \right).$$

Then, the result follows by referring to

$$P_{\boldsymbol{nm}} \; = \; \left(P_0(\boldsymbol{N} = \boldsymbol{n} - \boldsymbol{e}_j) \frac{\hat{\lambda}_j}{\mu_{0j}(n_j)} \right) \left(\prod_{l=1}^{L} P_l(M_l = m_l) \right)$$

for any $\boldsymbol{n}$ with $|\boldsymbol{n}| = S_0 + 1$ on the left hand side, and, for $l = 1, 2, ..., L$,

$$P_{\boldsymbol{n}+\boldsymbol{e}_j, \boldsymbol{m}-\boldsymbol{f}_l} = \left(P_0(\boldsymbol{N} = \boldsymbol{n}) \frac{\hat{\lambda}_j}{\mu_{0j}(n_j+1)} \right) P_l(M_l = m_l - 1) I_{\{m_l>0\}} \left(\prod_{d=1,d\neq l}^{L} P_d(M_d = m_d) \right),$$

for any $\boldsymbol{n}$ with $|\boldsymbol{n}| = S_0 + 1$ on the right hand side, using $\sum_{j=1}^{J} p_{j0} \hat{\lambda}_j = \lambda$. □

Recall that Theorem 2 has been proven for the special situation that $r_l = 1$ for all l. However, it is easy to see that the result can be generalized to the situation for arbitrary $r_l, l = 1, 2, ..., L$. First of all, the result and the proof of Theorem 2 does not change at all if we replace λ_l by $r_l \lambda_l$ and λ by $\lambda_{CR} = \sum_{l=1}^{L} r_l \lambda_l$. Second, it is well-known that splitting a Poisson process into two streams with probabilities r_l and $1 - r_l$, respectively, results in two independent Poisson processes. Thirdly, merging L independent Poisson processes (as occurs at the central repair facility) again leads to a Poisson process (basically, we used this fact already in the preceding proofs without explicitly mentioning it). Finally, since each base repair shop is modeled as a Jackson network and hence has a product form steady-state solution on its own, and since the modified aggregate system (consisting of the central repair facility followed by the base transportation nodes) also has a product form solution (Theorem 2), and all arrival processes are Poisson, the system as a whole (aggregated and modified in exactly the same way as before with all conditional probabilities $q(\boldsymbol{m})$ replaced by q) also has a product form solution (in which the initial product form solutions arise as marginal probabilities). This is summarized in the following theorem.

Theorem 3. *Consider the general capacitated repairable item system with one central repair facility and L local bases, as introduced in the beginning of this section. Suppose that each local repair shop and the central repair facility can be modeled as a Jackson network, while the transportation nodes are modeled as ample exponential servers. Then, the modified aggregate system that is obtained by considering only the backlogs at the central repair facility and replacing the conditional probabilities $q(\boldsymbol{m})$ by q for all $\boldsymbol{m}$, has a product form steady-state distribution. In particular,*

$$P(K_0 = 0, \boldsymbol{M} = \boldsymbol{m}, \boldsymbol{H} = \boldsymbol{h}) = \left(\sum_{|\boldsymbol{n}| \leq S_0} P_0(\boldsymbol{N} = \boldsymbol{n}) \right) \left(\prod_{l=1}^{L} P_l(M_l = m_l) \right) \left(\prod_{l=1}^{L} P_l(\boldsymbol{H}_l = \boldsymbol{h}_l) \right)$$

$$P(K_0 = k_0, \boldsymbol{M} = \boldsymbol{m}, \boldsymbol{H} = \boldsymbol{h}) = \left(\sum_{|\boldsymbol{n}| = S_0+k_0} P_0(\boldsymbol{N} = \boldsymbol{n}) \right) \left(\prod_{l=1}^{L} P_l(M_l = m_l) \right) \left(\prod_{l=1}^{L} P_l(\boldsymbol{H}_l = \boldsymbol{h}_l) \right) \; \textit{for } k_0 > 0,$$

where $P_0(\boldsymbol{N} = \boldsymbol{n})$ *is the product form solution of the central repair shop with arrival rate* $\lambda_{CR} = \sum_{l=1}^{L} r_l \lambda_l$*,* $P_l(M_l = m_l)$ *denotes the steady-state probability of transportation node* l*, and* $P_l(\boldsymbol{H}_l = \boldsymbol{h}_l)$ *is the product form solution of local repair shop* l*.*

Since Theorem 3 holds, we can now exploit the multinomial expression to obtain the detailed probabilities $P(\boldsymbol{K}_0 = \boldsymbol{k}_0, \boldsymbol{M} = \boldsymbol{m}, \boldsymbol{H} = \boldsymbol{h})$. By using appropriate convolutions and the balance relations between the various state variables discussed in the beginning of this section, all relevant performance measures are now easily calculated. Note, however, that all these values are only exact for the modified aggregate system. In the next subsection, we show that they also provide excellent approximations for the performance measures (obtained through simulation) of the original system.

3.2 Numerical Experiments

In order to assess the performance of the approximation for a complicated maintenance system, numerical experiments are performed for a two-base system as displayed in Figure 4 where the repair units at the depot and at the bases are all single exponential servers. In Table 4, simulation results and approximate values are given for the case $r_1 = r_2 = 0.5$ and the exponential service rates for local and central repairs and for the transportation nodes all being equal to 10. This system will be referred as the symmetric system. Simulation runs are as designed for the simple maintenance system studied in Section 2. One other system for which the experiments are performed is of the same type with the repair rate at the bases being higher, 15, to make the system more realistic, $r_1 = 0.7$, $r_2 = 0.1$ and $\lambda_1 = 10$, $\lambda_2 = 8$, and other parameters being the same as before. Unequal r and λ values make this system asymmetric. The results are given by Table 5.

All experiments carried out suggest that the approximation works very well for various sets of system parameters. Errors are generally less than one percent, with the largest error occurring in the determination of the expected backlogs at the bases.

4 Optimization

In the preceding section, we have observed that the performance measures as calculated for the modified model serve as an excellent approximation for those of the original model. This suggests that the modified model can be used to determine an optimal investment strategy, given that a target service level has to be attained, or an optimal allocation of stock between the depot and the local bases, given a certain budget restriction. As an example, we present in this section an approach to the first problem. Hence, we are interested in the minimum number of SKU's (stock keeping units or spare parts), and the optimal allocation of this number between depot and bases, such that a target fill rate is met. The realistic assumption here is that the cost per unit of stock to be allocated is the same for the depot and the bases.

Before answering this question we first have to define what is an overall fill rate. We have chosen this fill rate as a weighted average of the fill rates at the bases, i.e., we define the overall fill rate FR_b by

$$FR_b(\boldsymbol{S}) = \frac{\sum_{l=1}^{L} \lambda_l FR_l(\boldsymbol{S})}{\lambda},$$

where FR_l is the fill rate at base l, $l = 1, 2, ...L$, and $\boldsymbol{S} = (S_0, S_1, S_2, ..., S_L)$. Note that in this way FR_b is indeed the expected total proportion of requests that are fulfilled immediately. Next, we define a greedy, step-wise approach to find the minimum number of SKU's and the optimal allocation such that for some given $\alpha, 0 < \alpha < 1$, we have $FR_b(\boldsymbol{S}) \geq \alpha$.

The greedy approach works as follows: First, we determine the minimum order-up-to levels at the bases, $\widehat{S}_1, \widehat{S}_2, ..., \widehat{S}_L$, say, such that, at each base, the service level is at least α, while assuming that the central depot is never out of stock (equivalent to $S_0 = \infty$). It is clear that in an optimal allocation in which the stock

Table 4: Performance Measures for the symmetric system

S_0	S_1	S_2	SP_{sim}	SP_{app}	FR_{sim}	FR_{app}	ES_{sim}	ES_{app}
0	10	10	base 1 0.02578	0.02572	0.96137	0.96137	0.07713	0.07626
			(0.02558,0.02598)		(0.96113,0.96161)		(0.07630,0.07796)	
			base 2 0.02573	0.02572	0.96146	0.96137	0.07709	0.07626
			(0.02550,0.02595)		(0.96118,0.96175)		(0.07616,0.07803)	
2	9	9	base 1 0.02479	0.02472	0.96282	0.96283	0.07411	0.07322
			(0.02458,0.02500)		(0.96257,0.96307)		(0.07327,0.07495)	
			base 2 0.02474	0.02472	0.96292	0.96283	0.07408	0.07322
			(0.02451,0.02497)		(0.96264,0.96319)		(0.07314,0.07503)	
4	8	8	base 1 0.02395	0.02388	0.96392	0.96394	0.07138	0.07049
			(0.02376,0.02415)		(0.96366,0.96417)		(0.07056,0.07221)	
			base 2 0.02391	0.02388	0.96401	0.96394	0.07138	0.07049
			(0.02368,0.02413)		(0.96373,0.96429)		(0.07045,0.07231)	
6	7	7	base 1 0.02350	0.02342	0.96408	0.96412	0.06936	0.06847
			(0.02330,0.02369)		(0.96383,0.96434)		(0.06856,0.07017)	
			base 2 0.02347	0.02342	0.96415	0.96412	0.06937	0.06847
			(0.02326,0.02368)		(0.96389,0.96441)		(0.06846,0.07028)	
8	6	6	base 1 0.02404	0.02396	0.96184	0.96187	0.06905	0.06813
			(0.02384,0.02424)		(0.96158,0.96209)		(0.06824,0.06985)	
			base 2 0.02400	0.02396	0.96191	0.96187	0.06907	0.06813
			(0.02379,0.02422)		(0.96164,0.96218)		(0.06817,0.06997)	
10	5	5	base 1 0.02710	0.02700	0.95333	0.95338	0.07299	0.07204
			(0.02689,0.02730)		(0.95306,0.95359)		(0.07217,0.07381)	
			base 2 0.02706	0.02700	0.95339	0.95338	0.07298	0.07204
			(0.02684,0.02727)		(0.95312,0.95367)		(0.07209,0.07386)	
12	4	4	base 1 0.03657	0.03647	0.92888	0.92895	0.08774	0.08670
			(0.03637,0.03678)		(0.92862,0.92915)		(0.08693,0.08854)	
			base 2 0.03655	0.03647	0.92894	0.92895	0.08772	0.08670
			(0.03633,0.03676)		(0.92867,0.92921)		(0.08684,0.08860)	
14	3	3	base 1 0.06226	0.06214	0.86496	0.86498	0.12957	0.12846
			(0.06205,0.06247)		(0.86471,0.86521)		(0.12874,0.13040)	
			base 2 0.06222	0.06214	0.86499	0.86498	0.12949	0.12846
			(0.06200,0.06243)		(0.86473,0.86524)		(0.12862,0.13036)	
16	2	2	base 1 0.12792	0.12779	0.71043	0.71043	0.23886	0.23759
			(0.12771,0.12812)		(0.71024,0.71063)		(0.23799,0.23972)	
			base 2 0.12786	0.12779	0.71049	0.71043	0.23871	0.23759
			(0.12767,0.12805)		(0.71027,0.71071)		(0.23780,0.23961)	
18	1	1	base 1 0.28470	0.28462	0.39640	0.39629	0.50696	0.50573
			(0.28454,0.28486)		(0.39627,0.39654)		(0.50617,0.50775)	
			base 2 0.28468	0.28462	0.39639	0.39629	0.50678	0.50573
			(0.28449,0.28486)		(0.39624,0.39655)		(0.50592,0.50764)	
20	0	0	base 1 0.60148	0.60150	0	0	1.09479	1.09360
			(0.60136,0.60160)				(1.09407,1.09551)	
			base 2 0.60152	0.60150	0	0	1.09466	1.09360
			(0.60138,0.60166)				(1.09385,1.09547)	
5	10	5	base 1 0.00844	0.00842	0.98728	0.98721	0.02529	0.02446
			(0.00833,0.00856)		(0.98714,0.98743)		(0.02480,0.02578)	
			base 2 0.06731	0.06727	0.89591	0.89599	0.19760	0.19665
			(0.06698,0.06765)		(0.89553,0.89630)		(0.19600,0.19921)	

Table 5: Performance Measures for the asymmetric system

S_0	S_1	S_2	SP_{sim}	SP_{app}	FR_{sim}	FR_{app}	ES_{sim}	ES_{app}
0	10	10	base 1 0.08329	0.08335	0.89051	0.89042	0.34775	0.34771
			(0.08288,0.08371)		(0.89008,0.89095)		(0.34437,0.35113)	
			base 2 0.00061	0.00061	0.99873	0.99870	0.00117	0.00117
			(0.00060,0.00062)		(0.99871,0.99875)		(0.00114,0.00120)	
0	13	5	base 1 0.03667	0.03670	0.95180	0.95173	0.15290	0.15271
			(0.03635,0.03700)		(0.95145,0.95215)		(0.15050,0.15530)	
			base 2 0.02361	0.02362	0.95142	0.95135	0.04562	0.04562
			(0.02351,0.02372)		(0.95129,0.95156)		(0.04535,0.04590)	
2	11	5	base 1 0.03854	0.03857	0.94934	0.94927	0.16070	0.16053
			(0.03821,0.03887)		(0.94899,0.94969)		(0.15825,0.16314)	
			base 2 0.01967	0.01967	0.95940	0.95935	0.03796	0.03794
			(0.01957,0.01976)		(0.95928,0.95953)		(0.03771,0.03820)	
4	9	5	base 1 0.04051	0.04053	0.94674	0.94669	0.16892	0.16874
			(0.04018,0.04084)		(0.94638,0.94709)		(0.16640,0.17144)	
			base 2 0.01719	0.01719	0.96442	0.96437	0.03314	0.03312
			(0.01710,0.01728)		(0.96429,0.96455)		(0.03291,0.03337)	
6	7	5	base 1 0.04263	0.04263	0.94380	0.94383	0.17768	0.17741
			(0.04229,0.04297)		(0.94344,0.94417)		(0.17511,0.18025)	
			base 2 0.01569	0.01568	0.96747	0.96744	0.03024	0.03019
			(0.01561,0.01577)		(0.96735,0.96758)		(0.03002,0.03045)	
8	5	5	base 1 0.04570	0.04558	0.93708	0.93738	0.18791	0.18746
			(0.04536,0.04603)		(0.93672,0.93744)		(0.18532,0.19051)	
			base 2 0.01477	0.01476	0.96933	0.96930	0.02844	0.02840
			(0.01469,0.01485)		(0.96923,0.96944)		(0.02824,0.02865)	
10	3	5	base 1 0.06503	0.06465	0.87030	0.87063	0.21975	0.21876
			(0.06469,0.06536)		(0.86995,0.87065)		(0.21709,0.22240)	
			base 2 0.01421	0.01420	0.97046	0.97044	0.02734	0.02732
			(0.01414,0.01428)		(0.97036,0.97056)		(0.02714,0.02755)	
3	7	7	base 1 0.08979	0.08981	0.88183	0.88187	0.37475	0.37468
			(0.08938,0.09021)		(0.88140,0.88226)		(0.37128,0.37821)	
			base 2 0.00424	0.00424	0.99119	0.99116	0.00816	0.00815
			(0.00420,0.00428)		(0.99113,0.99124)		(0.00807,0.00826)	

position of the depot is taken into account, these values represent a lower bound for the base order-up-to levels. Next, we initialize a step-wise improvement procedure on $S_0 = 0$ and $S_l = \widehat{S}_l$, for $l = 1, 2, ..., L$. At each next step of the greedy approach, one more item is assigned to either of the bases or to the depot, based on the maximum improvement on the fill rate $FR_b(\boldsymbol{S})$. We stop as soon as a fill rate α is reached.

As an example, we have performed a number of numerical experiments for the same repairable item system as studied in Section 3.2, where we have chosen $\alpha = 0.95$. In Tables 6 and 7, the steps of the greedy approach are displayed for the symmetric and asymmetric systems, respectively. At each step, one more item is assigned to either of the bases or to the depot based on the improvements implied at the current (S_0, S_1, S_2) by each assignment. For the symmetric system, the minimum resource to be utilized for FR_b to be at 0.95 is $S_0 + S_1 + S_2 = 6 + 7 + 6 = 19$. If a budget constraint allows for an investment of maximally 20 spare parts to be allocated to the depot and the bases, then the allocation obtained by the greedy approach is $S_0 = 6$, $S_1 = 7$, $S_2 = 7$. Note that, as observed in Table 2, this allocation gives the best fill rate for the bases among all possible symmetric allocations, i.e., $S_1 = S_2$, of 20 spare part stocks.

Although a formal proof cannot be given here, extensive numerical experiments suggest that the fill rate defined above behaves as a multi-dimensional concave function as soon as $S_l \geq \widehat{S}_l$, for $l = 1, 2, ..., L$. Therefore, we believe that the resulting allocations are generally close to optimal. Apart from this, the optimization shows the power of the approximations based on the modified system analyzed in this paper. Without these approximations, extensive simulation studies would be needed to determine optimal allocations, which seems to be rather impractical for realistic systems.

5 Summary and possible extensions

In this paper, we have analyzed a fairly general two-echelon repairable item system with limited repair capacities at both a central repair facility and a number of local bases. Both the bases and the central facility are able to keep a number of ready-for-use items in stock. Items that have failed in the field are returned to the closest base and are replaced immediately by a ready-for-use item, if available. The returned item is either repaired at the base, in which case it is put in the local stock after repair, or sent to the central repair facility, in which case a ready-for-use item is immediately shipped to the base to update its stock level. Each repair shop (bases and central facility) can be modeled as an open Jackson queuing network. We have shown that a slight modification of the overall model reveals a product form solution which allows for a relatively easy calculation of several performance measures. Numerical experiments indicate that these functions excellently approximate the performance measures of the original system, given the order-up-to levels of the stocks at both the depot and the bases. Based upon this result, a greedy approach is defined to determine optimal order-up-to levels, in order to meet a given target fill rate.

The approach presented here can be extended easily for more complicated models. Note that the only approximations that are made in the paper concern the replacement of the conditional probabilities $q(m)$ by q for all m. We have presented the analysis under the assumption that all nodes (in either the base repair shop or the central facility, or the transportation nodes) show exponential behavior. In fact, all results remain valid as long as the repair networks are product form networks, i.e., each open BCMP network will do. In particular, this allows for general (non-exponential) transportation servers since in any BCMP network an ample server may have a general service time distribution without destroying the product form property.

One other extension to be mentioned here is for multi-echelon systems. Aggregate models of such systems would include conditional probabilities, namely q's, to be defined for each echelon as a function of the state of the downstream echelons. The approximation is then based upon the assumption of independence of the q values at each echelon from the states of the downstream echelons. The exact analysis of this modified model is similar to the one performed in this article for two-echelon systems.

If we skip the local repair shops and instead replace each transportation node between the central facility

Table 6: Greedy approach for the symmetric system

(S_0, S_1, S_2)	FR_b	(S_0+1, S_1, S_2)	(S_0, S_1+1, S_2)	(S_0, S_1, S_2+1)
(0,4,4)	0.60025	FR_1=0.66884 FR_2=0.66884 FR_b=0.66884	FR_1=0.72229 FR_2=0.60025 FR_b=0.66127	FR_1=0.60025 FR_2=0.72229 FR_b=0.66127
(1,4,4)	0.66884	FR_1=0.72530 FR_2=0.72530 FR_b=0.72530	FR_1=0.77325 FR_2=0.66884 FR_b=0.72105	FR_1=0.66884 FR_2=0.77325 FR_b=0.72105
(2,4,4)	0.72530	FR_1=0.77085 FR_2=0.77085 FR_b=0.77085	FR_1=0.81468 FR_2=0.72530 FR_b=0.76999	FR_1=0.72530 FR_2=0.81468 FR_b=0.76999
(3,4,4)	0.77085	FR_1=0.80736 FR_2=0.80736 FR_b=0.80736	FR_1=0.84798 FR_2=0.77085 FR_b=0.80942	FR_1=0.77085 FR_2=0.84798 FR_b=0.80942
(3,5,4)	0.80942	FR_1=0.87465 FR_2=0.80736 FR_b=0.84101	FR_1=0.89896 FR_2=0.77085 FR_b=0.83491	FR_1=0.84798 FR_2=0.84798 FR_b=0.84798
(3,5,5)	0.84798	FR_1=0.87465 FR_2=0.87465 FR_b=0.87465	FR_1=0.89896 FR_2=0.84798 FR_b=0.87347	FR_1=0.84798 FR_2=0.89896 FR_b=0.87347
(4,5,5)	0.87465	FR_1=0.89599 FR_2=0.89599 FR_b=0.89599	FR_1=0.91767 FR_2=0.87465 FR_b=0.89616	FR_1=0.87465 FR_2=0.91767 FR_b=0.89616
(4,6,5)	0.89616	FR_1=0.93264 FR_2=0.89599 FR_b=0.91432	FR_1=0.94561 FR_2=0.87465 FR_b=0.91013	FR_1=0.91767 FR_2=0.91767 FR_b=0.91767
(4,6,6)	0.91767	FR_1=0.93264 FR_2=0.93264 FR_b=0.93264	FR_1=0.94561 FR_2=0.91767 FR_b=0.93164	FR_1=0.91767 FR_2=0.94561 FR_b=0.93164
(5,6,6)	0.93264	FR_1=0.94462 FR_2=0.94462 FR_b=0.94462	FR_1=0.95589 FR_2=0.93264 FR_b=0.94427	FR_1=0.93264 FR_2=0.95589 FR_b=0.94427
(6,6,6)	0.94462	FR_1=0.95421 FR_2=0.95421 FR_b=0.95421	FR_1=0.96412 FR_2=0.94462 FR_b=0.95437	FR_1=0.94462 FR_2=0.96412 FR_b=0.95437
(6,7,6)	0.95437	FR_1=0.97070 FR_2=0.95421 FR_b=0.96246	FR_1=0.97649 FR_2=0.94462 FR_b=0.96056	FR_1=0.96412 FR_2=0.96412 FR_b=0.96412
(6,7,7)	0.96412	FR_1=0.97070 FR_2=0.97649 FR_b=0.97649	FR_1=0.97070 FR_2=0.96412 FR_b=0.96412	FR_1=0.97070 FR_2=0.97031 FR_b=0.97031

Table 7: Greedy approach for the asymmetric system

(S_0, S_1, S_2)	FR_b	(S_0+1, S_1, S_2)	(S_0, S_1+1, S_2)	(S_0, S_1, S_2+1)
(0,3,5)	0.59206	FR_1=0.43103 FR_2=0.95577 FR_b=0.66425	FR_1=0.44960 FR_2=0.95135 FR_b=0.67260	FR_1=0.30462 FR_2=0.97635 FR_b=0.60317
(0,4,5)	0.67260	FR_1=0.56225 FR_2=0.95577 FR_b=0.73715	FR_1=0.57423 FR_2=0.95135 FR_b=0.74184	FR_1=0.44960 FR_2=0.97635 FR_b=0.68371
(0,5,5)	0.74184	FR_1=0.66554 FR_2=0.95577 FR_b=0.79453	FR_1=0.67403 FR_2=0.95135 FR_b=0.79728	FR_1=0.57423 FR_2=0.97635 FR_b=0.75295
(0,6,5)	0.79728	FR_1=0.74515 FR_2=0.95577 FR_b=0.83876	FR_1=0.75146 FR_2=0.95135 FR_b=0.84030	FR_1=0.67403 FR_2=0.97635 FR_b=0.80839
(0,7,5)	0.84030	FR_1=0.80600 FR_2=0.95577 FR_b=0.87256	FR_1=0.81077 FR_2=0.95135 FR_b=0.87325	FR_1=0.75146 FR_2=0.97635 FR_b=0.85141
(0,8,5)	0.87325	FR_1=0.85237 FR_2=0.95577 FR_b=0.89833	FR_1=0.85599 FR_2=0.95135 FR_b=0.89837	FR_1=0.81077 FR_2=0.97635 FR_b=0.88436
(0,9,5)	0.89837	FR_1=0.88767 FR_2=0.95577 FR_b=0.91794	FR_1=0.89042 FR_2=0.95135 FR_b=0.91750	FR_1=0.85599 FR_2=0.97635 FR_b=0.90948
(1,9,5)	0.91794	FR_1=0.91238 FR_2=0.95935 FR_b=0.93326	FR_1=0.91453 FR_2=0.95577 FR_b=0.93286	FR_1=0.88767 FR_2=0.97853 FR_b=0.92805
(2,9,5)	0.93326	FR_1=0.93165 FR_2=0.96217 FR_b=0.94521	FR_1=0.93333 FR_2=0.95935 FR_b=0.94489	FR_1=0.91238 FR_2=0.98030 FR_b=0.94257
(3,9,5)	0.94521	FR_1=0.94669 FR_2=0.96437 FR_b=0.95455	FR_1=0.94800 FR_2=0.96217 FR_b=0.95430	FR_1=0.93165 FR_2=0.98169 FR_b=0.95389

and a base by a product form network, then again the complete analysis can be repeated, with similar results. Again, by replacing $q(m)'s$ by q, the whole system turns out to be a product form network. The resulting model is the full equivalent of a two-echelon production-inventory system with limited production capacities at each site. Hence, such systems can be completely analyzed as long as the demand process is Poisson and no batching in production or shipping is allowed. Buzacott et al. [4] analyze serial systems (hence without the inverse aborescent or distribution structure). A further extension concerns the modeling of systems where in each stage only a limited number of products can be in production simultaneously, leading to so-called generalized Kanban systems (see, e.g., Buzacott [3], Buzacott and Shanthikumar [5] or Di Mascolo et al. [14]). However, as soon as the amount of work-in-process in a stage is restricted, the replacement of $q(m)'s$ by q no longer leads to product form solutions.

Clearly, although computational complexities increase, the extension of the models to multiple products presents no essential theoretical problems, as long as the repair shops can be modeled as product form networks. This allows in particular for base-dependent transportation times to the central repair facilities. Note that in the analysis in Section 3 this transportation time has been included in the central repair shop model, making these transportation times indistinguishable between bases. A more interesting application however concerns field operating systems that are built from several items where the complete system is down as soon as one item fails. The extension of the framework set up in this paper then leads to a complete generalization of METRIC models for capacitated systems.

In the same way, one may study multiple indenture levels, i.e, the case where each item is an assembly that may fail due to the failure of precisely one subassembly which may be replaced. This leads to questions on how many subassemblies and how many assemblies to store at either the central depot or each local base. In addition, one may study the impact of different repair policies. For instance, depending on the actual work load, one may decide to replace a complete assembly instead of only the component that caused the failure, to get the field system back to operation as soon as possible. Clearly, however, this leads to a shift of work because eventually the complete assembly has to be revised while in addition an assembly is obviously a more expensive SKU than a single component.

Finally, the impact of different levels of criticality of items in a system's operation may lead to different priority rules in the repair shops. This leads to the study of priority systems in a multi-echelon network. So far, no results are known to us on these systems.

In conclusion, we believe that the current analytical framework provides a powerful tool to assess the performance of fairly general capacitated multi-echelon repairable item systems, and subsequently to optimize these systems, while various extensions seem to be possible. However, the analysis of the models will be limited due to computational complexities and hence further numerical approximations may be needed (e.g. two moment approximations instead of a full characterization of probability distributions for stockouts, number of items in stock, etc.). These will be the topic of future research.

References

[1] S.C. Albright. An Approximation to the stationary distribution of a multi-echelon repairable item inventory system with finite sources and repair channels. *Naval Research Logistics*, 36:179–195, 1989.

[2] S. Axsater. *Continuous Review Policies for Multi-Level Inventory Systems with Stochastic Demand*, volume 4. S. Graves, A. Rinnooy Kan and P. Zipkin: Logistics of Production and Inventory, Handbook of Operations Research, Elsevier-North-Holland, Amsterdam, The Netherlands, 1993.

[3] J.A. Buzacott. Generalized Kanban/MRP systems. Technical Report,Department of Management Science, University of Waterloo, 1989.

[4] J.A. Buzacott, S.M. Price, and J.G. Shantikumar. *Service Level in Multistage MRP and Base Stock Controlled Production Systems.* G. Fandel, Th. Gulledge and A. Jones (eds.): New Directions for Operations Research in Manufacturing, Springer, Berlin, 1992.

[5] J.A. Buzacott and J.G. Shantikumar. *Stochastic Models of Manufacturing Systems.* Prentice-Hall, EnglewoodCliffs, NJ, 1993.

[6] A.J. Clark and H. Scarf. Optimal policies for a multi-echelon inventory problem. *Management Science*, 6:475–490, 1960.

[7] A. Diaz and M.C. Fu. Models for multi-echelon repairable item inventory systems with limited repair capacity. *European Journal of Operational Research*, 97:480–492, 1997.

[8] E.B. Diks, A.G. de Kok, and A.G. Lagodimos. Multi-echelon systems: A service measure perspective. *European Journal of Operational Research*, 95:241–263, 1996.

[9] A. Federgruen. *Centralized Planning Models for Multi-Echelon Inventory Systems under uncertainty*, volume 4. S. Graves, A. Rinnooy Kan and P. Zipkin: Logistics of Production and Inventory, Handbook of Operations Research, Elsevier-North-Holland, Amsterdam, The Netherlands, 1993.

[10] S.C. Graves. A multi-echelon inventory model for a repairable item with one-for-one replenishment. *Management Science*, 31:1247–1256, 1985.

[11] D. Gross. On the ample service assumption of palm's theorem in inventory modeling. *Management Science*, 28:1065–1079, 1982.

[12] D. Gross, L.C. Kioussin, and D.R. Miller. A network decomposition approach for approximate steady state behavior of Markovian multi-echelon repariable item inventory systems. *Management Science*, 33:1453–1468, 1987.

[13] D. Gross, D.R. Miller, and R.M. Soland. A closed queuing network model for multi-echelon repairable item provisioning. *IIE transactions*, 15:344–352, 1983.

[14] M. Di Mascolo, Y. Frein, and Y. Dallery. An analytical method for performance evaluation of Kanban-controlled production systems. *Operations Research*, 44 (1):50–64, 1996.

[15] J. Muckstadt. A model for multi-item, multi-echelon, multi-indenture inventory system. *Management Science*, 20:472–481, 1973.

[16] C.C. Sherbrooke. METRIC: A multi-echelon technique for recoverable item control. *Operations Research*, 16:122–141, 1968.

[17] C.C. Sherbrooke. VARI-METRIC: Improved approximations for multi-indenture, multi-echelon availability models. *Operations Research*, 34:311–319, 1986.

[18] C.C. Sherbrooke. *Optimal Inventory Modeling of Systems: Multi-Echelon Techniques.* Wiley, New York, 1992.

[19] F.M. Slay. VARI-METRIC: An approach to modelling multi-echelon resupply when the demand process is poisson with a gamma prior. Report AF301-3, Logistics Management Institute, Washington D.C., 1984.

[20] G.J. van Houtum, K. Inderfurth, and W.H.M. Zijm. Materials coordination in stochastic multi-echelon systems. *European Journal of Operational Research*, 95:1–23, 1996.

[21] J.H.C.M. Verrijdt. *Design and Control of Service Part Distribution Systems.* PhD thesis, Eindhoven University of Technology, Faculty of Technology Management, Eindhoven, 1997.