U N I V E R S I T Ä T   H A M B U R G

Integrated models
for production-inventory systems

Sonja Otten, Ruslan Krenzler, Hans Daduna

Preprint–No. 2014-01   Mai 2014

DEPARTMENT MATHEMATIK

SCHWERPUNKT MATHEMATISCHE STATISTIK
UND STOCHASTISCHE PROZESSE

BUNDESSTR. 55, D – 20146 HAMBURG

Integrated models for production-inventory systems

Sonja Otten, Ruslan Krenzler, Hans Daduna

May 23, 2014

Contents

1 Introduction

Queueing theory and inventory theory are fields of Operations Research with different methodologies to optimize e.g. production processes and inventory control. In classical Operations Research queueing theory and inventory theory are often considered as disjoint areas of research. On the other side, the emergence of complex supply chains calls for integrated queueing-inventory models, which are the focus of our present research.

We consider a supply chain consisting of production systems (servers) at several locations, each with a local inventory, and a supplier network of workstations, which manufactures items of possibly different kinds (raw material), to replenish the local inventories. Each production system manufactures according to customers' demand on a make-to-order basis. To satisfy a customer's demand a server needs exactly one unit of raw material from the associated local inventory.

Although we will throughout describe our systems in terms of production and manufacturing, there are other scenarios where our models fit, e.g. distributed retail systems where customers' demand has to be satisfied from the local inventories and delivering the goods to the customers needs a non negligible amount of time; the replenishment for the local retail stations is provided by a complex production network. Another scenario is a distributed set of repair stations where spare parts are needed to repair the brought-in items which are hold in local inventories. Production of the needed spare parts and sending them to the repair stations is again due to a complex production network.

In our standard scenario at each location the production system consists of a single server with infinite waiting room. The inventories are controlled by a base stock policy. This means that when the number of items falls below a location specific level (local base stock level), an order is sent to the supplier network. The supplier network can be of complex structure. Whenever a local inventory is depleted, the corresponding production system is not able to serve queued demand and arriving customers at this location are lost ("local lost sales").

Related literature. For a general review of single inventories, where delivering the goods to the customers needs a non negligible amount of time, we refer to Krishnamoorthy et al. [KLM11], which surveys recent literature on integrated queueing-inventory models with non-zero lead times. The simplest of our models with only one production unit with an associated inventory is closely related to the class of models under review there.

Clearly, literature on inventory theory is overwhelming, so we only point to some references closely related to our investigations. We mention that there are two extreme cases of customers' reaction in the situation that inventory is depleted when demand arrives (cf. [SPP98]):

- backordering, which means that customers are willing to wait for their demands to be fulfilled, and
- lost sales, which means that demands that occur when inventory is empty are lost.

In classic inventory models it is common to assume that excess demand is backordered ([SPP98], [Zip00, p. 40]; [Axs00]). However, studies by Gruen et al. ([GCB02]) and Verhoef and Sloot ([VS06]) that analyze customer behaviour in practice show that in many retail settings most of the original demand can be considered to be lost.

For an overview of the literature on systems with lost sales we refer to Bijvank and Vis ([BV11]). They present a classification scheme for the replenishment policies most often applied in literature and practice, and they review the proposed replenishment policies, including the base stock policy. According to [vDB13]: "Their literature review confirms that there are only a limited number of papers dealing with lost sales systems and the vast majority of these papers make simplifying assumptions to make them analytically tractable."

Because we consider queueing-inventory systems where inventories are controlled by base stock policies we mention here that Zipkin ([Zip00, p. 181]) states that the base stock policy makes sense, when the economies of scale for the supplier are negligible relative to other factors. He gives the following example: when the holding and backorder costs clearly dominate any fixed order costs, or/and when each individual unit is very valuable. Tempelmeier ([Tem05, p. 84]) additionally argued that base stock control is economically reasonable if the order quantity is limited because of technical reasons.

The idea of using queueing theory to model (pure) inventory systems that operate under a base stock policy goes back to Morse ([Mor58, p. 139]). He gives a very simple example where the concept “re-order for each item sold” is useful: items in inventory are bulky, and expensive (automobiles or TV sets[1]).

This technique has recently found interest for evaluating performance of replenishment under base stock control in more complex situations, see e.g. the work of Rubio and Wein ([RW96]) and Zazanis ([Zaz94]) on pure inventory systems under base stock policy.

Similarly to literature on inventory theory, literature on queueing theory is overwhelming. We therefore point only to the most relevant sources for our present investigation: Our production systems are classical queueing systems which can be seen as parallel queues, while the supplier network which provides the replenishment for the associated inventories is a generalized Jackson network of queues, see [Jac57]. We additionally refer to [Kel79] and [CMP99] for the routing structure inside the network which describes in our setting the schedule for the replenishment orders handled by the network. These networks with different customer types will serve in the first part of our presentation as model for multi-product inventory replenishment networks.

In our second part we assume that the produced items are exchangeable. Such situation often occurs in multi-station maintenance and repair models, see e.g. the recent paper [RBP13], and [Dad90] with more related references.

Our contribution. We develop a Markovian stochastic process description of a complex supply chain system which consists of different production units with associated local inventories combined with a supplier network which serves the replenishment orders from the inventories.

The model enables us to set optimal base stock levels by exploiting the long run and stationary distribution of the Markov process and to find explicit cost functions under stationary conditions for the global systems.

Our approach results in rather simple explicit expressions for the optimality equations because we are able to prove that the stationary distribution of the global integrated system is of so called ”product form”. This means that the joint steady state distribution of the integrated multiple production-inventory-supplier network is a (multiple) product of the stationary marginal distributions

- of the queues at the locations (production systems), and
- of the inventories at the locations, and
- of the supplier network.
- Furthermore, the joint stationary distribution for the workstations of the supplier network is a product of the stationary marginal distributions of these workstations (queues).

This product form structure of the joint stationary distribution is often characterized as the global process being ”separable”, and is interpreted as ”the components of the system decouple asymptotically and in equilibrium”. Clearly, separability is an important (but rather rare) property of complex systems. When it holds, it leads to easy to perform structural and performance analysis.

We consider two classes of systems which in terms of (pure) inventory theory can be classified as ”multi-product systems” versus ”single-product systems”. The structural difference in the models is

[1]Note, the paper is from 1958.

that in the first case each location sends out orders which can be identified on the scheduled path through the supplier network and are eventually delivered exactly to the origin location. In the second case the send-out orders are exchangeable and may be returned to locations other than the origin, depending on some dispatch rules.
In both cases we consider parallel server (queues) to serve different incoming Poissonian demand streams and come up with a product form equilibrium.

Considering the servers at the locations as devices which deliver items from the inventory to incoming demand, needing non negligible delivering time (as in the single inventory case described in [KLM11]), and with replenishment by a supplier network, our work is an extension of the recent investigations of Rubio and Wein ([RW96]) and Zazanis ([Zaz94]) on pure inventory systems under base stock policy: The time to deliver items from the inventory is zero there. So, in terms of production-inventory scenario their model is a special case of ours with zero production time.

The rest of the paper is organized as follows. In Section 2 we will describe our integrated model for production and inventory management in the multi-product case and derive the stationary distribution. In Section 2.3 we use the explicit results obtained to analyze the long time average costs with the aim to find the optimal local base stock levels. In Section 3 we analyze the single product case, where the items produced in the supplier network are exchangeable. Surprisingly, it turns out that if the supply systems consists of more than one workstation, we cannot compute the stationary distribution explicitly. However, if we aggregate the whole supplier network into a single workstation, which serves replenishment requests in parallel, we can compute the stationary distribution, and show in Section 3.2 that it is again of product form. In Section 3.3 we perform a cost analysis for this system and find out that in this case even the cost function under linear cost is separable. This means the global search for the vector of optimal base stock levels can be reduced to a set of independent optimization problems. This is similar to the cost function in the pure inventory system in [RW96].

Notations and conventions:

- $\mathbb{R}_0^+ = [0, \infty)$, $\mathbb{R}^+ = (0, \infty)$, $\mathbb{N}$=$\{1,2,3,\dots\}$, $\mathbb{N}_0 = \{0\} \cup \mathbb{N}$
- All random variables and processes occurring henceforth are defined on a common underlying probability space $(\Omega, \mathcal{F}, P)$.
- For all processes considered in this paper we can and will assume that their paths are right continuous with left limits (cadlag).
- Empty sums are 0, and empty products are 1.
- $1_{\{expression\}}$ is the indicator function which is 1 if *expression* is true and 0 otherwise.
- For better readability only the male form for customers will be used in this paper.
- $\mathbf{e}_i = (0, \dots, 0, \underbrace{1}_{i-\text{th element}}, 0, \dots, 0)$ is a vector of appropriate dimension.
- For a vector $k_m = (k_{m1}, ..., k_{m\ell})$ we denote by $\#k_m$ the length of k_m, i.e. $\#k_m = \ell$.

2 Models for production-inventory systems: Multi-product systems

2.1 The general model with supplier network

The supply chain of interest is depicted in Figure 1. We have a set of locations $\overline{J} := \{1, 2, ..., J\}$. Each of the locations consists of a production system with an attached inventory. The inventories are replenished by a supplier network, which consists of M workstations and manufactures raw material for all locations, but distinguishes between the replenishment orders from different locations. $\overline{M} := \{J+1, ..., J+M\}$ denotes the set of workstations. Each order of raw material is specified by a location $j \in \overline{J}$ and runs through the workstations according to its own deterministic route, which may depend on the location. The resulting raw material is sent back to the location which has placed the order.

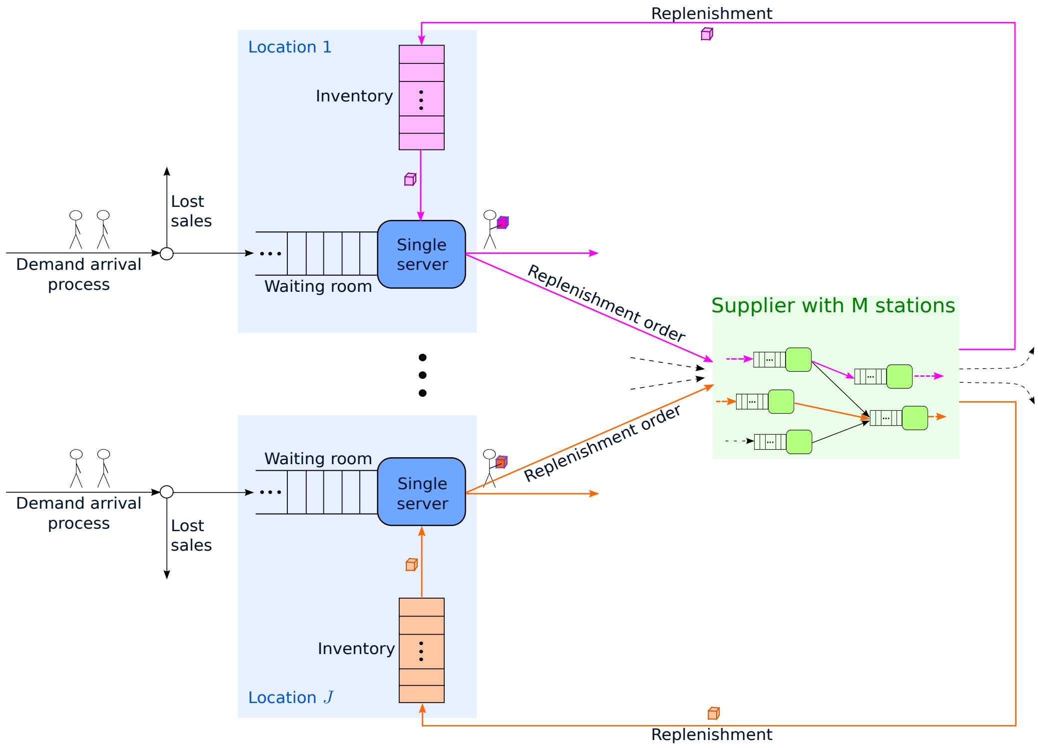

Figure 1: Production inventory system with base stock policy

Customers arriving at location j are of the same type and require the same type of service. The types arriving at the different locations are different in general.

Each production system $j \in \overline{J}$ consists of a single server with infinite waiting room under a first-come, first-served (FCFS) regime and manufactures according to customers' demand on a make-to-order basis. To satisfy a customer's demand the production system needs exactly one unit of raw material which is taken from the associated local inventory. Customers arrive one by one at the production system j according to a Poisson process with rate $\lambda_j > 0$ and require service. If the inventory is depleted at location j, arriving customers at this location decide not to join the queue and are lost ("local lost sales").

The service time is exponentially-1 distributed and the service at location j is provided with local queue length dependent intensity. If there are $n_j > 0$ customers present and if the inventory is not depleted, the service intensity is $\mu_j(n_j) > 0$. If the server is ready to serve a customer who is at the head of the line, and the inventory is not depleted, the service immediately starts. Otherwise, the service starts at the instant of time when the next replenishment arrives at the inventory.

A served customer departs from the system immediately and the associated (consumed) raw material is removed from the inventory and an order of one unit is sent out to the supplier network at this time instant ("base stock policy"). The local base stock level b_j is the maximal size of the inventory at location j (we denote $\mathbf{b} := \big(b_j : j \in \overline{J}\big)$).

To distinguish orders from different locations, we mark each order by a "type" which for simplicity is the index of the location, where the order is triggered.

An order triggered by location j follows a type-j-dependent route for eventual replenishment, denoted by $r(j) = (r(j,1), ..., r(j, S(j)-1), r(j, S(j)))$, where $r(j,\ell) = m \in \overline{M}$ for $\ell = 2, ..., S(j)$ is the identifier of the ℓ-th workstation on the path $r(j)$, and $S(j)$ is the number of stages of the route of type j. For completeness we fix $r(j,1) := j$.

The workstation $m \in \overline{M}$ consists of a single server and a waiting room under a FCFS regime. The service times are exponentially-1 distributed and the service at workstation $m \in \overline{M}$ is provided with local queue length dependent intensity. If there are $\ell > 0$ orders present, the service intensity is $\nu_m(\ell) > 0$.

It is assumed that transmission times for orders are zero and that transportation times between the supplier network and the inventory are negligible.

Remark 1 (Non-zero transportation time). We can model independent transportations of each item (raw material or order) between supplier network and locations or between workstations by means of special (virtual) $M/G/\infty$ workstations.

To obtain a Markovian process description of the integrated queueing-inventory system, we denote by $X_j(t)$ the number of customers present at location $j \in \overline{J}$ at time $t \geq 0$ either waiting or in service (queue length). By $Y_j(t)$ we denote the contents of the inventory at location $j \in \overline{J}$ at time $t \geq 0$. By $Y_m(t)$ we denote the sequence of orders at workstation $m \in \overline{M}$ of the supplier network at time $t \geq 0$.

We denote by K_m the set of possible states at node $m \in \overline{M}$ (local state space). The state $k_m := [t_{m1}, s_{m1}; ...; t_{m\#k_m}, s_{m\#k_m}] \in K_m$ indicates that there are $\#k_m$ orders at workstation $m \in \overline{M}$, on position $p \in \{1, ..., \#k_m\}$ resides an order of type $t_{mp} \in \overline{J}$, which is on stage $s_{mp} \in \{1, ..., S(t_{mp})\}$ of his route $r(t_{mp}) = \Big(r(t_{mp}, 1), ..., r(t_{mp}, S(t_{mp}))\Big)$. More precisely (t_{m1}, s_{m1}) is the order at the head of the line, who is in service and $(t_{m\#k_m}, s_{m\#k_m})$ is the order at the tail of the line.

In order to make the reading easier, we will use a unified notation for the states of the inventories at the locations and the states of the workstations in the supplier network. Therefore, the state of the inventory at location $j \in \overline{J}$ is $k_j = [j, 1; ...; j, 1]$, since the route of type j starts in the inventory at location j (i.e. $s_{jp} = 1$ for all $p \in \{1, ..., \#k_j\}$). See Remark 3 below. The stage $s_{jp} > 1$ indicates that the item (as an order) is in the supplier network. We suppose that items of type j return to the inventory at location $r(j,1) = j$ after leaving workstation $r(j, S(j))$ of the supplier network and repeat their route through the system (cf. [Kel79, pp. 82]).

$$\mathbf{k} = [k_1, ..., k_J, k_{J+1}, ..., k_{J+M}] \in K \subseteq \prod_{j=1}^{J+M} K_j,$$

where K_j denotes the local state space at j and K denotes the feasible states composed of feasible local states.

For $\#k_j = 0$ we read

$$\left[t_{j1}, s_{j1}; ...; t_{j\#k_j}, s_{j\#k_j}\right] =: [\,].$$

We define by $Z = ((X_1(t), \dots, X_J(t), Y_1(t), \dots, Y_J(t), Y_{J+1}(t), ..., Y_{J+M}(t)) \,:\, t \geq 0)$ the joint queueing and inventory process of this system and make the usual independence assumptions. Then Z is a homogeneous strong Markov process, which we assume to be ergodic. The state space of Z is

$$E = \left\{(\mathbf{n}, \mathbf{k}) : \mathbf{n} \in \mathbb{N}_0^{\overline{J}},\, \mathbf{k} \in K\right\}.$$

We are interested in the stationary system's cost per unit time, respectively the long time average cost. These are equal because of the ergodicity of Z.

The total costs of the system are determined by specific cost values per unit of time. These costs at location $j \in \overline{J}$ are in detail: waiting costs in queue and in service per customer $c_{w,j}$, costs for providing capacity of the inventory $c_{s,j}$, holding costs per item of the inventory $c_{h,j}$, and shortage costs for lost sales $c_{ls,j}$. The holding costs per item at workstation $m \in \overline{M}$ of the supplier network are $c_{h,m}$.

Therefore, the cost function per unit of time in the respective states is

$$\begin{aligned} f :\ \mathbb{N}_0^{\overline{J}} \times K &\longrightarrow \mathbb{R}_0, \\ f\left(n_1, \dots n_J, k_1, ..., k_J, k_{J+1}, ..., k_{J+M}\right) &= \left(\sum_{j \in \bar{J}} f_j(n_j, k_j, b_j) + \sum_{m \in \overline{M}} f_m(k_m)\right) \end{aligned}$$

with the cost functions $f_j :\ \mathbb{N}_0 \times K_j \times \mathbb{N} \longrightarrow \mathbb{R}_0$ of the system state (n_j, k_j) with base stock level b_j per unit of time at location j

$$f_j(n_j, k_j, b_j) = c_{w,j} \cdot n_j + c_{s,j} \cdot b_j + c_{h,j} \cdot \#k_j + c_{ls,j} \cdot 1_{\{\#k_j = 0\}}$$

and the cost function $f_m :\ \mathbb{N}_0 \longrightarrow \mathbb{R}_0$ of the system state k_m per unit of time at workstation m of the supplier network

$$f_m(k_m) = c_{h,m} \cdot \#k_m.$$

Our aim is to analyze the long-run system behaviour and to minimize the long-run average costs.

2.2 Limiting and stationary behaviour

To apply the ergodic theorem for Markovian processes we first determine the limiting and stationary distribution of $Z = ((X_1(t), \ldots, X_J(t), Y_1(t), \ldots, Y_J(t), Y_{J+1}(t), ..., Y_{J+M}(t)) : t \geq 0)$. Our main result is

Theorem 2. *The limiting and stationary distribution of the system described above is for* $(\mathbf{n}, \mathbf{k}) \in E$

$$\lim_{t\to\infty} P\left(Z(t) = (\mathbf{n}, \mathbf{k})\right) \quad =: \quad \pi\left(\mathbf{n}, \mathbf{k}\right) = \left(\prod_{j\in\overline{J}} \pi_j(n_j)\right) \theta\left(\mathbf{k}\right), \tag{1}$$

with

$$\pi_j(n_j) = C_j^{-1} \prod_{\ell=1}^{n_j} \left(\frac{\lambda_j}{\mu_j(\ell)}\right), \qquad n_j \in \mathbb{N}_0, \tag{2}$$

$$\theta(\mathbf{k}) = C_\theta^{-1} \prod_{j\in\overline{J}} \left(\frac{1}{\lambda_j}\right)^{\#k_j} \cdot \prod_{m\in\overline{M}} \prod_{\ell=1}^{\#k_m} \left(\frac{1}{\nu_m(\ell)}\right), \qquad \mathbf{k} \in K, \tag{3}$$

and normalization constants

$$\begin{aligned} C_j &= \sum_{n_j\in\mathbb{N}_0} \prod_{\ell=1}^{n_j} \left(\frac{\lambda_j}{\mu_j(\ell)}\right), \\ C_\theta &= \sum_{\mathbf{k}\in K} \prod_{j\in\overline{J}} \left(\frac{1}{\lambda_j}\right)^{\#k_j} \cdot \prod_{m\in\overline{M}} \prod_{\ell=1}^{\#k_m} \left(\frac{1}{\nu_m(\ell)}\right). \end{aligned}$$

Remark 3. Let us recall some notation:

- It will sometimes be convenient to use the elaborate notation:

$$\begin{aligned} \mathbf{k} &= \left[\overbrace{k_1, ..., k_J}^{\substack{\text{inventories}\\ \text{at locations}}}, \overbrace{k_{J+1}, ..., k_{J+M}}^{\text{supplier network}}\right] \\ &= \Bigg(\overbrace{\underbrace{[1,1;...;1,1]}_{\#k_1}, ..., \underbrace{[J,1;...;J,1]}_{\#k_J}}^{\substack{\text{inventories}\\ \text{at locations}}}, \overbrace{\underbrace{\left[t_{(J+1)1}, s_{(J+1)1}; ...; t_{(J+1)\#k_{J+1}}, s_{(J+1)\#k_{J+1}}\right]}_{\#k_{J+1}}, ...}^{\text{supplier network}} \\ &\qquad ..., \underbrace{\left[t_{(J+M)1}, s_{(J+M)1}; ...; t_{(J+M)\#k_{J+M}}, s_{(J+M)\#k_{J+M}}\right]}_{\#k_{J+M}}\Bigg). \end{aligned}$$

- The states of the inventories at the locations $j \in \overline{J}$ are of the form $[j, 1; ...; j, 1]$ since there is only raw material of type j at location j and the route of type j starts in the inventory at location j (i.e. the stage s_{jp} is equal to 1 in all positions p at "j").

- For $\#k_j = 0$

$$\left[t_{j1}, s_{j1}; ...; t_{j\#k_j}, s_{j\#k_j}\right] \quad =: \quad [0,0].$$

Proof of Theorem 2. The stochastic queueing and inventory process Z has an infinitesimal generator $\mathbf{Q} = (q(z;\tilde{z}) : z, \tilde{z} \in E)$ with the following non-negative transition rates for $(\mathbf{n}, \mathbf{k}) \in E$:

- arrival of a customer at location $i \in \overline{J}$
  if the inventory at this location is not empty because of the lost sales rule:

$$q\left((\mathbf{n},\mathbf{k});(\mathbf{n}+\mathbf{e}_i,\mathbf{k})\right) \quad = \lambda_i \cdot 1_{\{\#k_i>0\}}, \qquad i \in \overline{J}\,,$$

- service completion of a customer at location $i \in \overline{J}$
  if there is at least one customer at i and the inventory there is not empty,
  i.e. a customer departs from location i (=station $r(i,1)$) and a unit of raw material is removed from the associated local inventory, where $\#k_i$ units are present,
  and a replenishment order is sent to the supplier network, more precisely to workstation $r(i,2) \in \overline{M}$, where $\#k_{r(i,2)}$ orders are present:

$$\begin{aligned}
q\Bigg(\Big(&\mathbf{n}, \underbrace{[1,1;...;1,1]}_{\#k_1}, ..., \underbrace{[i,1;...;i,1]}_{\#k_i>0}, ..., \underbrace{[J,1;...;J,1]}_{\#k_J}, \left[t_{(J+1)1}, s_{(J+1)1}; ...; t_{(J+1)\#k_{J+1}}, s_{(J+1)\#k_{J+1}}\right], ...\\
&..., \underbrace{\left[t_{r(i,2)1}, s_{r(i,2)1}; ...; t_{r(i,2)\#k_{r(i,2)}}, s_{r(i,2)\#k_{r(i,2)}}\right]}_{\#k_{r(i,2)}}, ..., \left[t_{(J+M)1}, s_{(J+M)1}; ...; t_{(J+M)\#k_{J+M}}, s_{(J+M)\#k_{J+M}}\right]\Big);\\
\Big(&\mathbf{n}-\mathbf{e}_i, \underbrace{[1,1;...;1,1]}_{\#k_1}, ..., \underbrace{[i,1;...;i,1]}_{\#k_i-1}, ..., \underbrace{[J,1;...;J,1]}_{\#k_J}, \left[t_{(J+1)1}, s_{(J+1)1}; ...; t_{(J+1)\#k_{J+1}}, s_{(J+1)\#k_{J+1}}\right], ...\\
&..., \underbrace{\left[t_{r(i,2)1}, s_{r(i,2)1}; ...; t_{r(i,2)\#k_{r(i,2)}}, s_{r(i,2)\#k_{r(i,2)}}; i, 2\right]}_{\#k_{r(i,2)}+1}, ..., \left[t_{(J+M)1}, s_{(J+M)1}; ...; t_{(J+M)\#k_{J+M}}, s_{(J+M)\#k_{J+M}}\right]\Big)\Bigg)\\
= {}& \mu_i(n_i) \cdot 1_{\{n_i>0\}} \cdot 1_{\{\#k_i>0\}}, \qquad i \in \overline{J}\,,
\end{aligned}$$

- service completion of an order at workstation $m \in \overline{M}$
  if there is at least one order,
  i.e. an order of type t_{m1} on stage s_{m1} of its route is removed from workstation m, where $\#k_m$ orders are present, and is sent to the next stage $s_{m1}+1$ of his route, i.e. either

- if $s_{m1} < S(t_{m1})$, to workstation $r(t_{m1}, s_{m1}+1) \in \overline{M}$ (where $\#k_{r(t_{m1},s_{m1}+1)}$ orders are already present):

$$q\Bigg(\Big(\mathbf{n}, \underbrace{[1,1;...;1,1]}_{\#k_1}, ..., \underbrace{[J,1;...;J,1]}_{\#k_J}, \big[t_{(J+1)1}, s_{(J+1)1}; ...; t_{(J+1)\#k_{J+1}}, s_{(J+1)\#k_{J+1}}\big], ..., \underbrace{[t_{m1}, s_{m1}; ...; t_{m\#k_m}, s_{m\#k_m}]}_{\#k_m>0}, ...$$
$$..., \underbrace{\Big[t_{r(t_{m1},s_{m1}+1)1}, s_{r(t_{m1},s_{m1}+1)1}; ...; t_{r(t_{m1},s_{m1}+1)\#k_{r(t_{m1},s_{m1}+1)}}, s_{r(t_{m1},s_{m1}+1)\#k_{r(t_{m1},s_{m1}+1)}}\Big]}_{\#k_{r(t_{m1},s_{m1}+1)}}, ...$$
$$..., \big[t_{(J+M)1}, s_{(J+M)1}; ...; t_{(J+M)\#k_{J+M}}, s_{(J+M)\#k_{J+M}}\big]\Big);$$
$$\Big(\mathbf{n}, \underbrace{[1,1;...;1,1]}_{\#k_1}, ..., \underbrace{[J,1;...;J,1]}_{\#k_J}, \big[t_{(J+1)1}, s_{(J+1)1}; ...; t_{(J+1)\#k_{J+1}}, s_{(J+1)\#k_{J+1}}\big], ..., \underbrace{[t_{m2}, s_{m2}; ...; t_{m\#k_m}, s_{m\#k_m}]}_{\#k_m-1}, ...$$
$$..., \underbrace{\Big[t_{r(t_{m1},s_{m1}+1)1}, s_{r(t_{m1},s_{m1}+1)1}; ...; t_{r(t_{m1},s_{m1}+1)\#k_{r(t_{m1},s_{m1}+1)}}, s_{r(t_{m1},s_{m1}+1)\#k_{r(t_{m1},s_{m1}+1)}}; t_{m1}, s_{m1}+1\Big]}_{\#k_{r(t_{m1},s_{m1}+1)}+1}, ...$$
$$..., \big[t_{(J+M)1}, s_{(J+M)1}; ...; t_{(J+M)\#k_{J+M}}, s_{(J+M)\#k_{J+M}}\big]\Big)\Bigg)$$
$$= \nu(\#k_m) \cdot 1_{\{\#k_m>0\}} \cdot 1_{\{s_{m1}<S(t_{m1})\}}, \qquad m \in \overline{M},$$

- or if $s_{m1} = S(t_{m1})$, to the inventory at location $t_{m_1} \in \overline{J}$ (where $\#k_{t_{m1}}$ units of raw material are already present):

$$q\Bigg(\Big(\mathbf{n}, \underbrace{[1,1;...;1,1]}_{\#k_1}, ..., \underbrace{[t_{m1},1;...;t_{m1},1]}_{\#k_{t_{m1}}}, ..., \underbrace{[J,1;...;J,1]}_{\#k_J}, \big[t_{(J+1)1}, s_{(J+1)1}; ...; t_{(J+1)\#k_{J+1}}, s_{(J+1)\#k_{J+1}}\big], ...$$
$$..., \underbrace{[t_{m1}, s_{m1}; ...; t_{m\#k_m}, s_{m\#k_m}]}_{\#k_m>0}, ..., \big[t_{(J+M)1}, s_{(J+M)1}; ...; t_{(J+M)\#k_{J+M}}, s_{(J+M)\#k_{J+M}}\big]\Big);$$
$$\Big(\mathbf{n}, \underbrace{[1,1;...;1,1]}_{\#k_1}, ..., \underbrace{[t_{m1},1;...;t_{m1},1;t_{m1},1]}_{\#k_{t_{m1}}+1}, ..., \underbrace{[J,1;...;J,1]}_{\#k_J}, \big[t_{(J+1)1}, s_{(J+1)1}; ...; t_{(J+1)\#k_{J+1}}, s_{(J+1)\#k_{J+1}}\big], ...$$
$$..., \underbrace{[t_{m2}, s_{m2}; ...; t_{m\#k_m}, s_{m\#k_m}]}_{\#k_m-1}, ..., \big[t_{(J+M)1}, s_{(J+M)1}; ...; t_{(J+M)\#k_{J+M}}, s_{(J+M)\#k_{J+M}}\big]\Big)\Bigg)$$
$$= \nu(\#k_m) \cdot 1_{\{\#k_m>0\}} \cdot 1_{\{s_{m1}=S(t_{m1})\}}, \qquad m \in \overline{M}.$$

Furthermore, $q(z, \tilde{z}) = 0$ for any other pair $z \neq \tilde{z}$, and

$$q(z; z) = -\sum_{\tilde{z}\in E, z\neq\tilde{z}} q(z; \tilde{z}), \qquad \forall z \in E.$$

The global balance equations $\pi\mathbf{Q} = \mathbf{0}$ of the stochastic queueing and inventory process Z represents for each state $(\mathbf{n}, \mathbf{k}) \in E$ the property "flux out of the state is equal to the flux into that state". Therefore, the global balance equations are given as follows

flux out of the state

$$(\mathbf{n},\mathbf{k}) = \Big(\mathbf{n}, \underbrace{[1,1;...;1,1]}_{\#k_1}, ..., \underbrace{[J,1;...;J,1]}_{\#k_J}, \left[t_{(J+1)1}, s_{(J+1)1}; ...; t_{(J+1)\#k_{J+1}}, s_{(J+1)\#k_{J+1}}\right], ...$$
$$..., \left[t_{(J+M)1}, s_{(J+M)1}; ...; t_{(J+M)\#k_{J+M}}, s_{(J+M)\#k_{J+M}}\right]\Big)$$

through:

- an arrival of a customer at location $i \in \overline{J}$
  if the inventory at this location is not empty (i.e. $\#k_i > 0$) because of the lost sales rule,
- a service completion of a customer at location $i \in \overline{J}$
  if there is at least one customer (i.e. $n_i > 0$)
  and the inventory at this location is not empty (i.e. $\#k_i > 0$),
- a completion of an order at workstation $\ell \in \overline{M}$ of the supplier network
  if there is at least one order at this workstation (i.e. $\#k_\ell > 0$):

$$\pi\Big(\mathbf{n}, \underbrace{[1,1;...;1,1]}_{\#k_1}, ..., \underbrace{[J,1;...;J,1]}_{\#k_J}, \left[t_{(J+1)1}, s_{(J+1)1}; ...; t_{(J+1)\#k_{J+1}}, s_{(J+1)\#k_{J+1}}\right], ...$$
$$..., \left[t_{(J+M)1}, s_{(J+M)1}; ...; t_{(J+M)\#k_{J+M}}, s_{(J+M)\#k_{J+M}}\right]\Big)$$
$$\cdot\Big(\underline{\sum_{i\in\overline{J}} \lambda_i \cdot 1_{\{\#k_i>0\}}} + \underline{\sum_{i\in\overline{J}} \mu_i(n_i) \cdot 1_{\{n_i>0\}} \cdot 1_{\{\#k_i>0\}}} + \underline{\sum_{\ell\in\overline{M}} \nu_\ell(\#k_\ell) \cdot 1_{\{\#k_\ell>0\}}}\Big)$$

$=$ flux into the state $(\mathbf{n},\mathbf{k})$ through:

- an arrival of a customer at location $i \in \overline{J}$
  if there is at least one customer at location i (i.e. $n_i > 0$)
  and the inventory at this location is not empty (i.e. $\#k_i > 0$):

$$\sum_{i\in\overline{J}} \pi\Big(\mathbf{n}-\mathbf{e}_i, \underbrace{[1,1;...;1,1]}_{\#k_1}, ..., \underbrace{[J,1;...;J,1]}_{\#k_J}, \left[t_{(J+1)1}, s_{(J+1)1}; ...; t_{(J+1)\#k_{J+1}}, s_{(J+1)\#k_{J+1}}\right], ...$$
$$..., \left[t_{(J+M)1}, s_{(J+M)1}; ...; t_{(J+M)\#k_{J+M}}, s_{(J+M)\#k_{J+M}}\right]\Big)$$
$$\cdot\lambda_i \cdot 1_{\{\#k_i>0\}} \cdot 1_{\{n_i>0\}}$$

- a service completion of a customer at location $t_{\ell\#k_\ell} = r(t_{\ell\#k_\ell}, 1) \in \overline{J}$
  if there is at least one order at workstation ℓ (i.e. $\#k_\ell > 0$)
  and the order at the tail of the queue at workstation ℓ is in stage 2 of his route (i.e. $s_{\ell\#k_\ell} = 2$)
  (i.e. a customer departs from location $t_{\ell\#k_\ell}$
  and an item is removed from the associated local inventory there,
  and an order is sent to workstation $r(t_{\ell\#k_\ell}, 2) = \ell$)
  (note that $\{s_{\ell\#k_\ell} = 2\}$ implies $\left\{\#k_{t_{\ell\#k_\ell}} < b_{t_{\ell\#k_\ell}}\right\}$ to hold):

$$+\sum_{\ell\in\overline{M}}\pi\Big(\mathbf{n}+\mathbf{e}_{t_{\ell\#k_\ell}},\underbrace{[1,1;...;1,1]}_{\#k_1},...,\underbrace{[t_{\ell\#k_\ell},1;t_{\ell\#k_\ell},1;...;t_{\ell\#k_\ell},1]}_{\#k_{t_{\ell\#k_\ell}}+1},...,\underbrace{[J,1;...;J,1]}_{\#k_J},$$
$$\big[t_{(J+1)1},s_{(J+1)1};...;t_{(J+1)\#k_{J+1}},s_{(J+1)\#k_{J+1}}\big],...,\underbrace{[t_{\ell1},s_{\ell1};...;t_{\ell(\#k_\ell-1)},s_{\ell(\#k_\ell-1)}]}_{\#k_\ell-1},...$$
$$...,\big[t_{(J+M)1},s_{(J+M)1};...;t_{(J+M)\#k_{J+M}},s_{(J+M)\#k_{J+M}}\big]\Big)$$
$$\cdot\mu_{t_{\ell\#k_\ell}}(n_{t_{\ell\#k_\ell}}+1)\cdot 1_{\{s_{\ell\#k_\ell}=2\}}\cdot 1_{\{\#k_\ell>0\}}$$

- a transition of an order of type $t_{\ell\#k_\ell}$ from workstation $r(t_{\ell\#k_\ell},s_{\ell\#k_\ell}-1)$ to the next workstation of the supplier network if there is at least one order at workstation ℓ (i.e. $\#k_\ell>0$) and the order in the tail of the queue at workstation ℓ is not in stage 2 of his route (i.e. $s_{\ell\#k_\ell}>2$) (i.e. an order of type $t_{\ell\#k_\ell}$ is removed from workstation $r(t_{\ell\#k_\ell},s_{\ell\#k_\ell}-1)$ and is sent to workstation $\ell=r(t_{\ell\#k_\ell},s_{\ell\#k_\ell})$):

$$+\sum_{\ell\in\overline{M}}\pi\Big(\mathbf{n},\underbrace{[1,1;...;1,1]}_{\#k_1},...,\underbrace{[J,1;...;J,1]}_{\#k_J},\big[t_{(J+1)1},s_{(J+1)1};...;t_{(J+1)\#k_{J+1}},s_{(J+1)\#k_{J+1}}\big],...$$
$$...,\underbrace{\Big[t_{\ell\#k_\ell},s_{\ell\#k_\ell}-1;t_{r(t_{\ell\#k_\ell},s_{\ell\#k_\ell}-1)1},s_{r(t_{\ell\#k_\ell},s_{\ell\#k_\ell}-1)1};...;t_{r(t_{\ell\#k_\ell},s_{\ell\#k_\ell}-1)\#k_{r(t_{\ell\#k_\ell},s_{\ell\#k_\ell}-1)}},s_{r(t_{\ell\#k_\ell},s_{\ell\#k_\ell}-1)\#k_{r(t_{\ell\#k_\ell},s_{\ell\#k_\ell}-1)}}\Big]}_{\#k_{r(t_{\ell\#k_\ell},s_{\ell\#k_\ell}-1)}+1},...$$
$$...,\underbrace{[t_{\ell1},s_{\ell1};...;t_{\ell(\#k_\ell-1)},s_{\ell(\#k_\ell-1)}]}_{\#k_{t_{\ell\#k_\ell}}-1},...,\big[t_{(J+M)1},s_{(J+M)1};...;t_{(J+M)\#k_{J+M}},s_{(J+M)\#k_{J+M}}\big]\Big)$$
$$\cdot\nu_{r(t_{\ell\#k_\ell},s_{\ell\#k_\ell}-1)}(\#k_{r(t_{\ell\#k_\ell},s_{\ell\#k_\ell}-1)}+1)\cdot 1_{\{\#k_\ell>0\}}\cdot 1_{\{s_{\ell\#k_\ell}>2\}}$$

- a replenishment of the inventory at location $i\in\overline{J}$ if there is at least one unit of raw material at location i (i.e. $\#k_i>0$) (i.e. an order of type i is removed from workstation $r(i,S(i))$ and is sent to the inventory at location i):

$$+\sum_{i\in\overline{J}}\pi\Big(\mathbf{n},\underbrace{[1,1;...;1,1]}_{\#k_1},...,\underbrace{[i,1;...;i,1]}_{\#k_i-1},...,\underbrace{[J,1;...;J,1]}_{\#k_J},\big[t_{(J+1)1},s_{(J+1)1};...;t_{(J+1)\#k_{J+1}},s_{(J+1)\#k_{J+1}}\big],...$$
$$...,\underbrace{\Big[i,S(i);t_{r(i,S(i))1},s_{r(i,S(i))1};......;t_{r(i,S(i))\#k_{r(i,S(i))}},s_{r(i,S(i)-1)\#k_{r(i,S(i))}}\Big]}_{\#k_{r(i,S(i))}+1},...$$
$$...,\big[t_{(J+M)1},s_{(J+M)1};...;t_{(J+M)\#k_{J+M}},s_{(J+M)\#k_{J+M}}\big]\Big)$$
$$\cdot\nu_{r(i,S(i))}(\#k_{r(i,S(i))}+1)\cdot 1_{\{\#k_i>0\}}\,.$$

Substitution of (1) into the global balance equations directly leads to

$$
\begin{aligned}
&\left(\prod_{j\in\overline{J}}\pi_j(n_j)\right)\theta\Big(\underbrace{[1,1;...;1,1]}_{\#k_1},...,\underbrace{[J,1;...;J,1]}_{\#k_J},\left[t_{(J+1)1},s_{(J+1)1};...;t_{(J+1)\#k_{J+1}},s_{(J+1)\#k_{J+1}}\right],...\\
&\qquad ...,\left[t_{(J+M)1},s_{(J+M)1};...;t_{(J+M)\#k_{J+M}},s_{(J+M)\#k_{J+M}}\right]\Big)\\
&\cdot\Big(\sum_{i\in\overline{J}}\lambda_i\cdot 1_{\{\#k_i>0\}}+\sum_{i\in\overline{J}}\mu_i(n_i)\cdot 1_{\{n_i>0\}}\cdot 1_{\{\#k_i>0\}}+\sum_{\ell\in\overline{M}}\nu_\ell(\#k_\ell)\cdot 1_{\{\#k_\ell>0\}}\Big)\\
=&\sum_{i\in\overline{J}}\left(\prod_{j\in\overline{J}\setminus\{i\}}\pi_j(n_j)\right)\pi_i(n_i-1)\theta\Big(\underbrace{[1,1;...;1,1]}_{\#k_1},...,\underbrace{[J,1;...;J,1]}_{\#k_J},\left[t_{(J+1)1},s_{(J+1)1};...;t_{(J+1)\#k_{J+1}},s_{(J+1)\#k_{J+1}}\right],...\\
&\qquad ...,\left[t_{(J+M)1},s_{(J+M)1};...;t_{(J+M)\#k_{J+M}},s_{(J+M)\#k_{J+M}}\right]\Big)\\
&\cdot\lambda_i\cdot 1_{\{n_i>0\}}\cdot 1_{\{\#k_i>0\}}\\
&+\sum_{\ell\in\overline{M}}\left(\prod_{j\in\overline{J}\setminus\{t_{\ell\#k_\ell}\}}\pi_j(n_j)\right)\pi_{t_{\ell\#k_\ell}}(n_{t_{\ell\#k_\ell}}+1)\theta\Big(\underbrace{[1,1;...;1,1]}_{\#k_1},...,\underbrace{[t_{\ell\#k_\ell},1;t_{\ell\#k_\ell},1;...;t_{\ell\#k_\ell},1]}_{\#k_{t_{\ell\#k_\ell}}+1},...,\underbrace{[J,1;...;J,1]}_{\#k_J},\\
&\qquad\left[t_{(J+1)1},s_{(J+1)1};...;t_{(J+1)\#k_{J+1}},s_{(J+1)\#k_{J+1}}\right],...,\underbrace{[t_{\ell1},s_{\ell1};...;t_{\ell(\#k_\ell-1)},s_{\ell(\#k_\ell-1)}]}_{\#k_\ell-1},...\\
&\qquad ...,\left[t_{(J+M)1},s_{(J+M)1};...;t_{(J+M)\#k_{J+M}},s_{(J+M)\#k_{J+M}}\right]\Big)\\
&\cdot\mu_{t_{\ell\#k_\ell}}(n_{t_{\ell\#k_\ell}}+1)\cdot 1_{\{\#k_\ell>0\}}\cdot 1_{\{s_{\ell\#k_\ell}=2\}}\\
&+\sum_{\ell\in\overline{M}}\left(\prod_{j\in\overline{J}}\pi_j(n_j)\right)\theta\Big(\underbrace{[1,1;...;1,1]}_{\#k_1},...,\underbrace{[J,1;...;J,1]}_{\#k_J},\left[t_{(J+1)1},s_{(J+1)1};...;t_{(J+1)\#k_{J+1}},s_{(J+1)\#k_{J+1}}\right],...\\
&\qquad ...,\underbrace{\Big[t_{\ell\#k_\ell},s_{\ell\#k_\ell}-1;t_{r(t_{\ell\#k_\ell},s_{\ell\#k_\ell}-1)1},s_{r(t_{\ell\#k_\ell},s_{\ell\#k_\ell}-1)1};...;t_{r(t_{\ell\#k_\ell},s_{\ell\#k_\ell}-1)\#k_{r(t_{\ell\#k_\ell},s_{\ell\#k_\ell}-1)}},s_{r(t_{\ell\#k_\ell},s_{\ell\#k_\ell}-1)\#k_{r(t_{\ell\#k_\ell},s_{\ell\#k_\ell}-1)}}\Big]}_{\#k_{r(t_{\ell\#k_\ell},s_{\ell\#k_\ell}-1)}+1},...\\
&\qquad ...,\underbrace{[t_{\ell1},s_{\ell1};...;t_{\ell(\#k_\ell-1)},s_{\ell(\#k_\ell-1)}]}_{\#k_{t_{\ell\#k_\ell}}-1},...,\left[t_{(J+M)1},s_{(J+M)1};...;t_{(J+M)\#k_{J+M}},s_{(J+M)\#k_{J+M}}\right]\Big)\\
&\cdot\nu_{r(t_{\ell\#k_\ell},s_{\ell\#k_\ell}-1)}(\#k_{r(t_{\ell\#k_\ell},s_{\ell\#k_\ell}-1)}+1)\cdot 1_{\{\#k_\ell>0\}}\cdot 1_{\{s_{\ell\#k_\ell}>2\}}\\
&+\sum_{i\in\overline{J}}\left(\prod_{j\in\overline{J}}\pi_j(n_j)\right)\theta\Big(\underbrace{[1,1;...;1,1]}_{\#k_1},...,\underbrace{[i,1;...;i,1]}_{\#k_i-1},...,\underbrace{[J,1;...;J,1]}_{\#k_J},\left[t_{(J+1)1},s_{(J+1)1};...;t_{(J+1)\#k_{J+1}},s_{(J+1)\#k_{J+1}}\right],...\\
&\qquad\underbrace{\Big[i,S(i);t_{r(i,S(i))1},s_{r(i,S(i))1};......;t_{r(i,S(i))\#k_{r(i,S(i))}},s_{r(i,S(i)-1)\#k_{r(i,S(i))}}\Big]}_{\#k_{r(i,S(i))}+1},...\\
&\qquad ...,\left[t_{(J+M)1},s_{(J+M)1};...;t_{(J+M)\#k_{J+M}},s_{(J+M)\#k_{J+M}}\right]\Big)\\
&\cdot\nu_{r(i,S(i))}(\#k_{r(i,S(i))}+1)\cdot 1_{\{\#k_i>0\}}.
\end{aligned}
$$

Substitution of (2) we obtain

$$\left(\prod_{j\in\overline{J}}\pi_j(n_j)\right)\theta\Big(\underbrace{[1,1;...;1,1]}_{\#k_1},...,\underbrace{[J,1;...;J,1]}_{\#k_J},\left[t_{(J+1)1},s_{(J+1)1};...;t_{(J+1)\#k_{J+1}},s_{(J+1)\#k_{J+1}}\right],...$$
$$,...,\left[t_{(J+M)1},s_{(J+M)1};...;t_{(J+M)\#k_{J+M}},s_{(J+M)\#k_{J+M}}\right]\Big)$$
$$\cdot\Big(\sum_{i\in\overline{J}}\lambda_i\cdot 1_{\{\#k_i>0\}}+\sum_{i\in\overline{J}}\mu_i(n_i)\cdot 1_{\{n_i>0\}}\cdot 1_{\{\#k_i>0\}}+\sum_{\ell\in\overline{M}}\nu_\ell(\#k_\ell)\cdot 1_{\{\#k_\ell>0\}}\Big)$$

$$=\sum_{i\in\overline{J}}\left(\prod_{j\in\overline{J}}\pi_j(n_j)\right)\theta\Big(\underbrace{[1,1;...;1,1]}_{\#k_1},...,\underbrace{[J,1;...;J,1]}_{\#k_J},\left[t_{(J+1)1},s_{(J+1)1};...;t_{(J+1)\#k_{J+1}},s_{(J+1)\#k_{J+1}}\right],...$$
$$...,\left[t_{(J+M)1},s_{(J+M)1};...;t_{(J+M)\#k_{J+M}},s_{(J+M)\#k_{J+M}}\right]\Big)$$
$$\cdot\mu_i(n_i)\cdot 1_{\{n_i>0\}}\cdot 1_{\{\#k_i>0\}}$$

$$+\sum_{\ell\in\overline{M}}\left(\prod_{j\in\overline{J}}\pi_j(n_j)\right)\theta\Big(\underbrace{[1,1;...;1,1]}_{\#k_1},...,\underbrace{[t_{\ell\#k_\ell},1,t_{\ell\#k_\ell},1;...;t_{\ell\#k_\ell},1]}_{\#k_{t_{\ell\#k_\ell}}+1},...,\underbrace{[J,1;...;J,1]}_{\#k_J},$$
$$\left[t_{(J+1)1},s_{(J+1)1};...;t_{(J+1)\#k_{J+1}},s_{(J+1)\#k_{J+1}}\right],...,\underbrace{\left[t_{\ell1},s_{\ell1};...;t_{\ell(\#k_\ell-1)},s_{\ell(\#k_\ell-1)}\right]}_{\#k_\ell-1},...$$
$$...,\left[t_{(J+M)1},s_{(J+M)1};...;t_{(J+M)\#k_{J+M}},s_{(J+M)\#k_{J+M}}\right]\Big)$$
$$\cdot\lambda_{t_{\ell\#k_\ell}}\cdot 1_{\{\#k_\ell>0\}}\cdot 1_{\{s_{\ell\#k_\ell}=2\}}$$

$$+\sum_{\ell\in\overline{M}}\left(\prod_{j\in\overline{J}}\pi_j(n_j)\right)\theta\Big(\underbrace{[1,1;...;1,1]}_{\#k_1},...,\underbrace{[J,1;...;J,1]}_{\#k_J},\left[t_{(J+1)1},s_{(J+1)1};...;t_{(J+1)\#k_{J+1}},s_{(J+1)\#k_{J+1}}\right],...$$
$$...,\underbrace{\Big[t_{\ell\#k_\ell},s_{\ell\#k_\ell}-1;t_{r(t_{\ell\#k_\ell},s_{\ell\#k_\ell}-1)1},s_{r(t_{\ell\#k_\ell},s_{\ell\#k_\ell}-1)1};...;t_{r(t_{\ell\#k_\ell},s_{\ell\#k_\ell}-1)\#k_{r(t_{\ell\#k_\ell},s_{\ell\#k_\ell}-1)}},s_{r(t_{\ell\#k_\ell},s_{\ell\#k_\ell}-1)\#k_{r(t_{\ell\#k_\ell},s_{\ell\#k_\ell}-1)}}\Big]}_{\#k_{r(t_{\ell\#k_\ell},s_{\ell\#k_\ell}-1)}+1},...$$
$$...,\underbrace{\left[t_{\ell1},s_{\ell1};...;t_{\ell(\#k_\ell-1)},s_{\ell(\#k_\ell-1)}\right]}_{\#k_{t_{\ell\#k_\ell}}-1},...,\left[t_{(J+M)1},s_{(J+M)1};...;t_{(J+M)\#k_{J+M}},s_{(J+M)\#k_{J+M}}\right]\Big)$$
$$\cdot\nu_{r(t_{\ell\#k_\ell},s_{\ell\#k_\ell}-1)}(\#k_{r(t_{\ell\#k_\ell},s_{\ell\#k_\ell}-1)}+1)\cdot 1_{\{\#k_\ell>0\}}\cdot 1_{\{s_{\ell\#k_\ell}>2\}}$$

$$+\sum_{i\in\overline{J}}\left(\prod_{j\in\overline{J}}\pi_j(n_j)\right)\theta\Big(\underbrace{[1,1;...;1,1]}_{\#k_1},...,\underbrace{[i,1;...;i,1]}_{\#k_i-1},...,\underbrace{[J,1;...;J,1]}_{\#k_J},\left[t_{(J+1)1},s_{(J+1)1};...;t_{(J+1)\#k_{J+1}},s_{(J+1)\#k_{J+1}}\right],...$$
$$\underbrace{\left[i,S(i);t_{r(i,S(i))1},s_{r(i,S(i))1};...;t_{r(i,S(i))\#k_{r(i,S(i))}},s_{r(i,S(i)-1)\#k_{r(i,S(i))}}\right]}_{\#k_{r(i,S(i))}+1},...$$
$$...,\left[t_{(J+M)1},s_{(J+M)1};...;t_{(J+M)\#k_{J+M}},s_{(J+M)\#k_{J+M}}\right]\Big)$$
$$\cdot\nu_{r(i,S(i))}(\#k_{r(i,S(i))}+1)\cdot 1_{\{\#k_i>0\}}.$$

Cancelling $\left(\prod_{j\in\overline{J}}\pi_j(n_j)\right)$ and the sums with the terms $\mu_i(n_i)\cdot 1_{\{n_i>0\}}\cdot 1_{\{\#k_i>0\}}$ on both sides of the equation, and substitution of (3) leads to

$$\begin{aligned}
&\theta\Big(\underbrace{[1,1;...;1,1]}_{\#k_1},...,\underbrace{[J,1;...;J,1]}_{\#k_J},\left[t_{(J+1)1},s_{(J+1)1};...;t_{(J+1)\#k_{J+1}},s_{(J+1)\#k_{J+1}}\right],...\\
&\qquad\qquad ...,\left[t_{(J+M)1},s_{(J+M)1};...;t_{(J+M)\#k_{J+M}},s_{(J+M)\#k_{J+M}}\right]\Big)\\
&\cdot\Big(\sum_{i\in\overline{J}}\lambda_i\cdot 1_{\{\#k_i>0\}}+\sum_{\ell\in\overline{M}}\nu_\ell(\#k_\ell)\cdot 1_{\{\#k_\ell>0\}}\Big)\\
=\;&\sum_{\ell\in\overline{M}}\theta\Big(\underbrace{[1,1;...;1,1]}_{\#k_1},...,\underbrace{[J,1;...;J,1]}_{\#k_J},\left[t_{(J+1)1},s_{(J+1)1};...;t_{(J+1)\#k_{J+1}},s_{(J+1)\#k_{J+1}}\right],...\\
&\qquad\qquad ...,\left[t_{(J+M)1},s_{(J+M)1};...;t_{(J+M)\#k_{J+M}},s_{(J+M)\#k_{J+M}}\right]\Big)\\
&\quad\cdot\left(\frac{1}{\lambda_{t_{\ell\#k_\ell}}}\right)\cdot\nu_\ell(\#k_\ell)\cdot\lambda_{t_{\ell\#k_\ell}}\cdot 1_{\{\#k_\ell>0\}}\cdot 1_{\{s_{\ell\#k_\ell}=2\}}\\
&+\sum_{\ell\in\overline{M}}\theta\Big(\underbrace{[1,1;...;1,1]}_{\#k_1},...,\underbrace{[J,1;...;J,1]}_{\#k_J},\left[t_{(J+1)1},s_{(J+1)1};...;t_{(J+1)\#k_{J+1}},s_{(J+1)\#k_{J+1}}\right],...\\
&\qquad\qquad ...,\left[t_{(J+M)1},s_{(J+M)1};...;t_{(J+M)\#k_{J+M}},s_{(J+M)\#k_{J+M}}\right]\Big)\\
&\quad\cdot\left(\frac{1}{\nu_{r(t_{\ell\#k_\ell},s_{\ell\#k_\ell}-1)}(\#k_{r(t_{\ell\#k_\ell},s_{\ell\#k_\ell}-1)}+1)}\right)\cdot\nu_\ell(\#k_\ell)\\
&\quad\cdot\nu_{r(t_{\ell\#k_\ell},s_{\ell\#k_\ell}-1)}(\#k_{r(t_{\ell\#k_\ell},s_{\ell\#k_\ell}-1)}+1)\cdot 1_{\{\#k_\ell>0\}}\cdot 1_{\{s_{\ell\#k_\ell}>2\}}\\
&+\sum_{i\in\overline{J}}\theta\Big(\underbrace{[1,1;...;1,1]}_{\#k_1},...,\underbrace{[J,1;...;J,1]}_{\#k_J},\left[t_{(J+1)1},s_{(J+1)1};...;t_{(J+1)\#k_{J+1}},s_{(J+1)\#k_{J+1}}\right],...\\
&\qquad\qquad ...,\left[t_{(J+M)1},s_{(J+M)1};...;t_{(J+M)\#k_{J+M}},s_{(J+M)\#k_{J+M}}\right]\Big)\\
&\quad\cdot\lambda_i\cdot\left(\frac{1}{\nu_{r(i,S(i))}(\#k_{r(i,S(i))}+1)}\right)\\
&\quad\cdot\nu_{r(i,S(i))}(\#k_{r(i,S(i))}+1)\cdot 1_{\{\#k_i>0\}}.
\end{aligned}\tag{4}$$

Cancelling, this yields

$$\begin{aligned}
&\Big(\sum_{i\in\overline{J}}\lambda_i\cdot 1_{\{\#k_i>0\}}+\sum_{\ell\in\overline{M}}\nu_\ell(\#k_\ell)\cdot 1_{\{\#k_\ell>0\}}\cdot\left(1_{\{s_{\ell\#k_\ell}=2\}}+1_{\{s_{\ell\#k_\ell}>2\}}\right)\Big)\\
=\;&\sum_{\ell\in\overline{M}}\nu_\ell(\#k_\ell)\cdot 1_{\{\#k_\ell>0\}}\cdot 1_{\{s_{\ell\#k_\ell}=2\}}+\sum_{\ell\in\overline{M}}\nu_\ell(\#k_\ell)\cdot 1_{\{\#k_\ell>0\}}\cdot 1_{\{s_{\ell\#k_\ell}>2\}}\\
&+\sum_{i\in\overline{J}}\lambda_i\cdot 1_{\{\#k_i>0\}}.
\end{aligned}$$

□

Remark 4. $\theta(\mathbf{k}) = \theta(k_1, \ldots, k_J, k_{J+1}, ..., k_{J+M})$ is obtained as a stochastic solution of (4), which resembles the global balance equations of a Kelly network with $J+M$ nodes and $\sum_{j\in\overline{J}} b_j$ customers and exponentially distributed service times with rate λ_j at node $j \in \{1, ..., J\}$ and with rate ν_m at node $m \in \{J+1, ..., J+M\}$ and with deterministic, type dependent routing.

2.3 Average costs

We will analyze average long term costs of the system as a function of the base stock levels $\mathbf{b} = (b_1, ..., b_J)$.

Lemma 5. *Optimal solutions for the problem described in Section 2.1 are the set*

$$\arg\min\left(\overline{g}(\mathbf{b})\right)$$

with

$$\overline{g}(\mathbf{b}) \quad := \quad \sum_{j\in\overline{J}} c_{s,j}\cdot b_j + \sum_{\mathbf{k}\in K}\left(\sum_{j\in\overline{J}} c_{ls,j}\cdot 1_{\{\#k_j=0\}} + \sum_{j\in\overline{J}\cup\overline{M}} c_{h,j}\cdot\#k_j\right)\theta(\mathbf{k}).$$

Proof. The asymptotic average costs for an ergodic system can be calculated as

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T f(Z(\omega,t))dt \quad = \quad \sum_{(\mathbf{n},\mathbf{k})} f(\mathbf{n},\mathbf{k})\,\pi(\mathbf{n},\mathbf{k}) =: \overline{f}(\mathbf{b}) \qquad P-a.s.$$

Using product form properties of the system we get

$$\begin{aligned}
&\overline{f}(\mathbf{b})\\
=\ & \sum_{(\mathbf{n},\mathbf{k})}\left(\sum_{j\in\overline{J}} f_j(n_j,k_j,b_j) + \sum_{m\in\overline{M}} f_m(k_m)\right)\left(\prod_{\ell\in\overline{J}}\pi_\ell(n_\ell)\right)\theta(\mathbf{k})\\
=\ & \sum_{(\mathbf{n},\mathbf{k})}\left(\sum_{j\in\overline{J}}(c_{w,j}\cdot n_j + c_{s,j}\cdot b_j + c_{h,j}\cdot\#k_j + c_{ls,j}\cdot 1_{\{\#k_j=0\}}) + \sum_{m\in\overline{M}} c_{h,m}\cdot\#k_m\right)\left(\prod_{\ell\in\overline{J}}\pi_\ell(n_\ell)\right)\theta\Big(\mathbf{k}\Big)\\
=\ & \sum_{\mathbf{n}}\sum_{\mathbf{k}}\left(\sum_{j\in\overline{J}}(c_{w,j}\cdot n_j)\right)\left(\prod_{\ell\in\overline{J}}\pi_\ell(n_\ell)\right)\theta(\mathbf{k})\\
&+\sum_{\mathbf{n}}\sum_{\mathbf{k}}\left(\sum_{j\in\overline{J}}(c_{s,j}\cdot b_j + c_{h,j}\cdot\#k_j + c_{ls,j}\cdot 1_{\{\#k_j=0\}}) + \sum_{m\in\overline{M}} c_{h,m}\cdot\#k_m\right)\left(\prod_{\ell\in\overline{J}}\pi_\ell(n_\ell)\right)\theta(\mathbf{k})\\
=\ & \underbrace{\sum_{\mathbf{k}}\theta(\mathbf{k})}_{=1}\sum_{\mathbf{n}}\sum_{j\in\overline{J}} c_{w,j}\cdot n_j\left(\prod_{\ell\in\overline{J}}\pi_\ell(n_\ell)\right)\\
&+\underbrace{\sum_{\mathbf{n}}\left(\prod_{\ell\in\overline{J}}\pi_\ell(n_\ell)\right)}_{=1}\sum_{\mathbf{k}}\left(\sum_{j\in\overline{J}}(c_{s,j}\cdot b_j + c_{h,j}\cdot\#k_j + c_{ls,j}\cdot 1_{\{\#k_j=0\}}) + \sum_{m\in\overline{M}} c_{h,m}\cdot\#k_m\right)\theta(\mathbf{k}).
\end{aligned}$$

Let X_j, $j \in \overline{J}$, denote random variables which are distributed according to π_j. Using

$$\begin{aligned}
\sum_{\mathbf{n}} \sum_{j\in\bar{J}} c_{w,j} \cdot n_j \left(\prod_{\ell\in\overline{J}} \pi_\ell(n_\ell) \right) &= \sum_{j\in\bar{J}} \sum_{n_1=0}^{\infty} \cdots \sum_{n_J=0}^{\infty} c_{w,j} \cdot n_j \left(\prod_{\ell\in\overline{J}} \pi_\ell(n_\ell) \right) \\
&= \sum_{j\in\bar{J}} \left(c_{w,j} \sum_{n_j=0}^{\infty} n_j \cdot \pi_j(n_j) \underbrace{\sum_{(n_i)_{i\neq j}} \left(\prod_{\ell\in\overline{J}\setminus\{j\}} \pi_\ell(n_\ell) \right)}_{=1} \right) \\
&= \sum_{j\in\bar{J}} c_{w,j} \sum_{n_j=0}^{\infty} n_j \cdot \pi_j(n_j) \\
&= \sum_{j\in\bar{J}} c_{w,j} \cdot E_{\pi_j}(X_j),
\end{aligned}$$

we get for the asymptotic average costs

$$\begin{aligned}
\overline{f}(\mathbf{b}) &= \sum_{\mathbf{k}} \left(\sum_{j\in\bar{J}} (c_{s,j} \cdot b_j + c_{h,j} \cdot \#k_j + c_{ls,j} \cdot 1_{\{\#k_j=0\}}) + \sum_{m\in\overline{M}} c_{h,m} \cdot \#k_m \right) \theta(\mathbf{k}) \\
&\quad + \underbrace{\sum_{j\in\bar{J}} c_{w,j} E_{\pi_j}(X_j)}_{\text{independet of } b_j} .
\end{aligned}$$

$$\Longrightarrow \arg\min\left(\overline{f}(\mathbf{b})\right) = \arg\min\left(\overline{g}(\mathbf{b})\right),$$

where

$$\begin{aligned}
\overline{g}(\mathbf{b}) &:= \sum_{\mathbf{k}\in K} \left(\sum_{j\in\bar{J}} (c_{s,j} \cdot b_j + c_{h,j} \cdot \#k_j + c_{ls,j} \cdot 1_{\{\#k_j=0\}}) + \sum_{m\in\overline{M}} c_{h,m} \cdot \#k_m \right) \theta(\mathbf{k}) \\
&= \sum_{\mathbf{k}\in K} \left(\sum_{j\in\bar{J}} (c_{s,j} \cdot b_j + c_{ls,j} \cdot 1_{\{\#k_j=0\}}) + \sum_{j\in\overline{J}\cup\overline{M}} c_{h,j} \cdot \#k_j \right) \theta(\mathbf{k}) \\
&= \sum_{j\in\bar{J}} c_{s,j} \cdot b_j + \sum_{\mathbf{k}\in K} \left(\sum_{j\in\bar{J}} c_{ls,j} \cdot 1_{\{\#k_j=0\}} + \sum_{j\in\overline{J}\cup\overline{M}} c_{h,j} \cdot \#k_j \right) \theta(\mathbf{k}).
\end{aligned}$$

□

Remark 6 (Multi-server). Since we analyzed production systems and workstations with single server, where the service is provided with local queue length dependent intensity, our result covers especially the case of multi-server queues.

If we have a multi-server with $m_j \in \mathbb{N}$ or $m_j = \infty$ servers at location resp. workstation $j \in \overline{J} \cup \overline{M}$, the service intensities are given by

$$\mu_j(n_j) = \begin{cases} n_j \cdot \mu_j & \text{if } n_j < m_j, \\ m_j \cdot \mu_j & \text{if } n_j \geq m_j. \end{cases}$$

3 Reduced model with aggregated supplier network and indistinguishable orders

In the previous section, we considered the multi-product case, i.e. the situation, where orders from different locations are different and can be uniquely identified on their scheduled path through the supplier network and are delivered to exactly the station where the order was generated. We derived there the stationary distribution of the global supply chain process in explicit product form.

In this section, we consider the case where the items (raw material) produced in the supplier network according to the orders from the locations are exchangeable, i.e. we have a single-product case. Indistinguishable orders and item types, which are exchangeable, have been investigated in connection with maintenance and repair, see the recent paper [RBP13]. If an item required according to some order is produced, it is sent with some prespecified probability p_j to location $j \in \overline{J}$. If the inventory at this location is full, i.e. the base stock level is reached already, the item is not sent to the selected location and is hold for an additional random time in the queue of the supplier. If this time expires, the produced item is sent out anew. Surprisingly, this simpler network structure results in a Markovian stochastic process model which resists to solving the equilibrium balance equations explicitly in smooth form.

However, if we aggregate the supplier network into a single node (workstation) where orders are served in parallel, we can solve the global balance equation of the describing Markov process and obtain a stationary distribution which is of product form again.

3.1 Model with single workstation at the supplier

The supply chain of interest is depicted in Figure 2. We have a set of locations $\overline{J} := \{1, 2, ..., J\}$. Each of the locations consists of a production system with an attached inventory. The inventories are replenished by a single supplier, who is referred to as workstation $J+1$ and manufactures raw material for all locations. The items of raw material are indistinguishable (exchangeable). The resulting raw material is sent with probability p_j to location $j \in \overline{J}$ if the inventory is not full at this location. Otherwise, it is hold for an additional random time in the queue of the supplier. If this time expires, the produced item is sent out anew.

Each production system $j \in \overline{J}$ consists of a single server with infinite waiting room under a first-come, first-served (FCFS) regime and manufactures according to customers' demand on a make-to-order basis. To satisfy a customer's demand the production system needs exactly one unit of raw material which is taken from the associated local inventory. Customers arrive one by one at the production system j according to a Poisson process with rate $\lambda_j > 0$ and require service. If the inventory is depleted at location j, arriving customers at this location decide not to join the queue and are lost ("local lost sales").

The service time is exponentially-1 distributed and the service at location j is provided with local queue length dependent intensity. If there are $n_j > 0$ customers present and if the inventory is not depleted, the service intensity is $\mu_j(n_j) > 0$. If the server is ready to serve a customer who is at the head of the line, and the inventory is not depleted, the service immediately starts. Otherwise, the service starts at the instant of time when the next replenishment arrives at the inventory.

A served customer departs from the system immediately and the associated consumed raw material is removed from the inventory and an order of one unit is placed at this time instant ("base stock policy"). The local base stock level b_j is the maximal size of the inventory at location j (we denote $\mathbf{b} := \big(b_j : j \in \overline{J}\big)$).

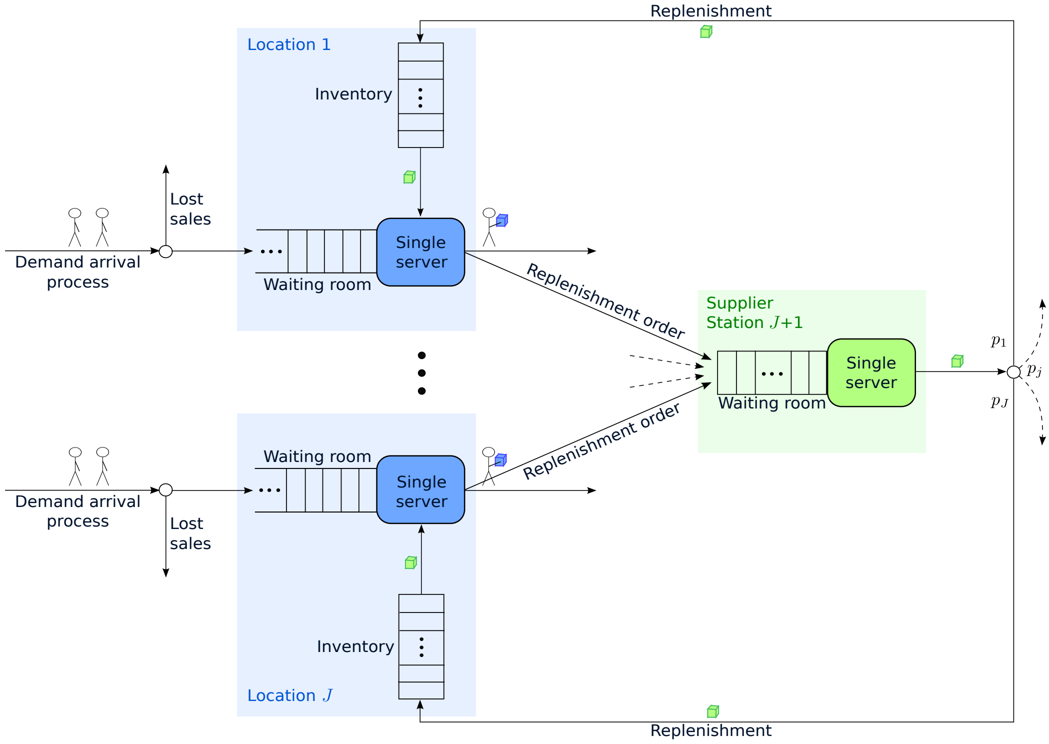

Figure 2: Production inventory system with base stock policy

The supplier (which is referred to as workstation $J+1$) consists of a single server and a waiting room under a FCFS regime. The waiting room is finite and has size $\sum_{j\in\overline{J}} b_j - 1$. The service time of the supplier is exponentially distributed with parameter ν. A finished item of raw material departs from the supplier immediately and is sent with probability $p_j > 0$ to location j, where $\sum_{j\in\overline{J}} p_j = 1$. If the inventory at this location is full, the item is not sent to the selected location and is hold for an additional random time in the queue of the supplier. If this time expires, the produced item is sent out anew.

It is assumed that transmission times for orders are zero and that transportation times between the supplier and the inventory are negligible.

Remark 7. Our strategy in case of a full inventory resembles the following two rerouting mechanism to handle breakdowns for nodes in networks of queues or buffers of finite size which are full:

- Blocking principle "repetitive service - random destination" (RS-RD): A customer being served at node i chooses his next destination node j according to the routing table. If he cannot join node j the customer stays at node i to obtain another service. When his additional service expires, the customer selects his destination node anew according to his routing instruction.

- "Skipping" principle: If a customer selects for his jump's destination a node, where the jump is not allowed, he only performs an imaginary jump to that node, spending no time there, but jumps onto a next node immediately according to the routing matrix.

There are various strategies in practice to handle breakdowns. The skipping principle was introduced by Schassberger [Sch84] and later on it was used e.g. in [DS96]. The RS-RD principle occurred e.g. in [Kle76, Section 5.11] and [Lig85, Proposition 5.10]. Sauer and Daduna ([SD04]) discussed both principles and they gave a short survey about the most prominent strategies to model in case of blocking.

It has to be noted that our system does not correspond to the usual networks with breakdown and repair mechanism since our service intensity of the blocked node (i.e. the production system at the location with the full inventory) is not set to 0. This is also the case e.g. in [vdGdKAR12].

To obtain a Markovian process description of the integrated queueing-inventory system, we denote by $X_j(t)$ the number of customers present at location $j \in \overline{J}$ at time $t \geq 0$ either waiting or in service (queue length). By $Y_j(t)$ we denote the size of the inventory at location $j \in \overline{J}$ at time $t \geq 0$. By $Y_{J+1}(t)$ we denote the number of replenishment orders at the supplier at time $t \geq 0$ either waiting or in service (queue length).

We define by $Z = ((X_1(t), \ldots, X_J(t), Y_1(t), \ldots, Y_J(t), Y_{J+1}(t)) : t \geq 0)$ the joint queueing and inventory process of this system and make the usual independence assumptions. Then Z is a homogeneous strong Markov process, which we assume to be ergodic. The state space of Z is

$$E = \left\{ (\mathbf{n}, \mathbf{k}) : \mathbf{n} \in \mathbb{N}_0^{\overline{J}}, \mathbf{k} \in K \right\}$$

with

$$K := \{(k_1, ..., k_J, \underbrace{\sum_{j=1}^{J} (b_j - k_j))}_{=:k_{J+1}} \, | 0 \leq k_j \leq b_j, \; j = 1, ..., J\} .$$

Remark 8. It has to be noted, that in the previous model k_j, $j \in \overline{J} \cup \{J+1\}$, was given by a complex structure with types and stages in Kelly's [Kel79] terminology. In the present model additionally to the queue lengths at the locations we only need the information about the size of the inventories and the number of orders at the supplier.

The total costs of the system are determined by specific cost values per unit of time. These costs at location $j \in \overline{J}$ are in detail: waiting costs in queue and in service per customer $c_{w,j}$, costs for providing capacity of the inventory $c_{s,j}$, holding costs per item of the inventory $c_{h,j}$, and shortage costs for lost sales $c_{ls,j}$. The holding costs per item at the supplier are $c_{h,J+1}$.

Therefore, the cost function per unit of time in the respective states is

$$\begin{aligned} f: \mathbb{N}_0^{\overline{J}} \times K &\longrightarrow \mathbb{R}_0, \\ f(n_1, \ldots n_J, k_1, ..., k_J, k_{J+1}) &= \left(\sum_{j \in \bar{J}} f_j(n_j, k_j, b_j) + f_{J+1}(k_{J+1}) \right) \end{aligned}$$

with the cost functions $f_j : \mathbb{N}_0^2 \times \mathbb{N} \longrightarrow \mathbb{R}_0$ at location j of the local system state (n_j, k_j) with base stock level b_j per unit of time

$$f_j(n_j, k_j, b_j) = c_{w,j} \cdot n_j + c_{s,j} \cdot b_j + c_{h,j} \cdot k_j + c_{ls,j} \cdot 1_{\{k_j = 0\}}$$

and the cost function $f_{J+1} : \mathbb{N}_0 \longrightarrow \mathbb{R}_0$ of the system state $k_{J+1} = \sum_j (b_j - k_j)$ per unit of time at the supplier

$$f_{J+1}(k_{J+1}) = c_{h,J+1} \cdot k_{J+1}.$$

Our aim is to analyze the long-run system behaviour and to minimize the long-run average costs.

3.2 Limiting and stationary behaviour

To apply the ergodic theorem for Markovian processes we first determine the limiting and stationary distribution of $Z = ((X_1(t), \ldots, X_J(t), Y_1(t), \ldots, Y_J(t), Y_{J+1}(t)) : t \geq 0)$. Our main result is

Theorem 9. *The limiting and stationary distribution of the system described above is*

$$\lim_{t \to \infty} P(Z(t) = (\mathbf{n}, \mathbf{k})) =: \pi(\mathbf{n}, \mathbf{k}) = \left(\prod_{j \in \overline{J}} \pi_j(n_j) \right) \theta(\mathbf{k}), \tag{5}$$

with

$$\pi_j(n_j) = C_j^{-1} \prod_{\ell=1}^{n_j} \left(\frac{\lambda_j}{\mu_j(\ell)} \right), \qquad n_j \in \mathbb{N}_0, \tag{6}$$

$$\theta(\mathbf{k}) = \theta(k_1, \ldots, k_J, k_{J+1}) = C_\theta^{-1} \prod_{j \in \overline{J}} \left(\frac{\nu p_j}{\lambda_j} \right)^{k_j}, \qquad \mathbf{k} \in K, \tag{7}$$

and normalization constants

$$C_j = \textstyle\sum_{n_j \in \mathbb{N}} \prod_{\ell=1}^{n_j} \left(\frac{\lambda_j}{\mu_j(\ell)} \right) \qquad \text{and} \qquad C_\theta = \displaystyle\sum_{\mathbf{k} \in K} \prod_{j \in \overline{J}} \left(\frac{\nu p_j}{\lambda_j} \right)^{k_j}.$$

Proof. The stochastic queueing and inventory process Z has an infinitesimal generator $\mathbf{Q} = (q(z, \tilde{z}) : z, \tilde{z} \in E)$ with the following strictly positive transition rates for $(\mathbf{n}, \mathbf{k}) \in E$:

$$\begin{aligned} q((\mathbf{n}, \mathbf{k}); (\mathbf{n} + \mathbf{e}_i, \mathbf{k})) &= \lambda_i \cdot 1_{\{k_i > 0\}}, && i \in \overline{J}, \\ q((\mathbf{n}, \mathbf{k}); (\mathbf{n} - \mathbf{e}_i, \mathbf{k} - \mathbf{e}_i + \mathbf{e}_{J+1})) &= \mu_i(n_i) \cdot 1_{\{n_i > 0\}} \cdot 1_{\{k_i > 0\}}, && i \in \overline{J}, \\ q((\mathbf{n}, \mathbf{k}); (\mathbf{n}, \mathbf{k} + \mathbf{e}_i - \mathbf{e}_{J+1})) &= \nu p_i \cdot 1_{\{k_i < b_i\}}, && i \in \overline{J}. \end{aligned}$$

Furthermore,

$$q\,(z;z) = -\sum_{\tilde{z}\in E, z\neq\tilde{z}} q\,(z;\tilde{z})\,.$$

It has to be noted that from $k_i < b_i$ follows that $k_{J+1} > 0$. Therefore, the global balance equations $\pi\mathbf{Q} = \mathbf{0}$ of stochastic queueing and inventory process Z are given as follows:

$$\begin{aligned}
&\pi\,(\mathbf{n},\mathbf{k})\left(\sum_{i\in\overline{J}}\lambda_i\cdot 1_{\{k_i>0\}} + \sum_{i\in\overline{J}}\mu_i(n_i)\cdot 1_{\{n_i>0\}}\cdot 1_{\{k_i>0\}} + \sum_{i\in\overline{J}}\nu p_i\cdot 1_{\{k_i<b_i\}}\right)\\
= \;&\sum_{i\in\overline{J}}\pi\,(\mathbf{n}-\mathbf{e}_i,\mathbf{k})\,\lambda_i\cdot 1_{\{n_i>0\}}\cdot 1_{\{k_i>0\}}\\
&+\sum_{i\in\overline{J}}\pi\,(\mathbf{n}+\mathbf{e}_i,\mathbf{k}+\mathbf{e}_i-\mathbf{e}_{J+1})\,\mu_i(n_i+1)\cdot 1_{\{k_i<b_i\}}\\
&+\sum_{i\in\overline{J}}\pi\,(\mathbf{n},\mathbf{k}-\mathbf{e}_i+\mathbf{e}_{J+1})\,\nu p_i\cdot 1_{\{k_i>0\}}\,.
\end{aligned}$$

It has to be shown that the distribution (5) satisfies these global balance equations. Substitution of (5) into the global balance equations directly leads to

$$\begin{aligned}
&\left(\prod_{j\in\overline{J}}\pi_j(n_j)\right)\theta\,(\mathbf{k})\left(\sum_{i\in\overline{J}}\lambda_i\cdot 1_{\{k_i>0\}} + \sum_{i\in\overline{J}}\mu_i(n_i)\cdot 1_{\{n_i>0\}}\cdot 1_{\{k_i>0\}} + \sum_{i\in\overline{J}}\nu p_i\cdot 1_{\{k_i<b_i\}}\right)\\
= \;&\sum_{i\in\overline{J}}\left(\prod_{j\in\overline{J}\setminus\{i\}}\pi_j(n_j)\right)\pi_i(n_i-1)\theta\,(\mathbf{k})\,\lambda_i\cdot 1_{\{n_i>0\}}\cdot 1_{\{k_i>0\}}\\
&+\sum_{i\in\overline{J}}\left(\prod_{j\in\overline{J}\setminus\{i\}}\pi_j(n_j)\right)\pi_i(n_i+1)\theta\,(\mathbf{k}+\mathbf{e}_i-\mathbf{e}_{J+1})\,\mu_i(n_i+1)\cdot 1_{\{k_i<b_i\}}\\
&+\sum_{i\in\overline{J}}\left(\prod_{j\in\overline{J}}\pi_j(n_j)\right)\theta\,(\mathbf{k}-\mathbf{e}_i+\mathbf{e}_{J+1})\,\nu p_i\cdot 1_{\{k_i>0\}}
\end{aligned}$$

$$\begin{aligned}
&\overset{(6)}{\Leftrightarrow} && \left(\prod_{j\in\overline{J}}\pi_j(n_j)\right)\theta(\mathbf{k})\left(\sum_{i\in\overline{J}}\lambda_i\cdot 1_{\{k_i>0\}}+\sum_{i\in\overline{J}}\mu_i(n_i)\cdot 1_{\{n_i>0\}}\cdot 1_{\{k_i>0\}}+\sum_{i\in\overline{J}}\nu p_i\cdot 1_{\{k_i<b_i\}}\right)\\
&= && \sum_{i\in\overline{J}}\left(\prod_{j\in\overline{J}}\pi_j(n_j)\right)\theta(\mathbf{k})\,\mu_i(n_i)\cdot 1_{\{n_i>0\}}\cdot 1_{\{k_i>0\}}\\
& && +\sum_{i\in\overline{J}}\left(\prod_{j\in\overline{J}}\pi_j(n_j)\right)\theta(\mathbf{k}+\mathbf{e}_i-\mathbf{e}_{J+1})\,\lambda_i\cdot 1_{\{k_i<b_i\}}\\
& && +\sum_{i\in\overline{J}}\left(\prod_{j\in\overline{J}}\pi_j(n_j)\right)\theta(\mathbf{k}-\mathbf{e}_i+\mathbf{e}_{J+1})\,\nu p_i\cdot 1_{\{k_i>0\}}\\
&\Leftrightarrow && \theta(\mathbf{k})\left(\sum_{i\in\overline{J}}\lambda_i\cdot 1_{\{k_i>0\}}+\sum_{i\in\overline{J}}\nu p_i\cdot 1_{\{k_i<b_i\}}\right)\\
&= && \sum_{i\in\overline{J}}\theta(\mathbf{k}+\mathbf{e}_i-\mathbf{e}_{J+1})\,\lambda_i\cdot 1_{\{k_i<b_i\}}\\
& && +\sum_{i\in\overline{J}}\theta(\mathbf{k}-\mathbf{e}_i+\mathbf{e}_{J+1})\,\nu p_i\cdot 1_{\{k_i>0\}} \qquad (8)\\
&\overset{(7)}{\Leftrightarrow} && \theta(\mathbf{k})\left(\sum_{i\in\overline{J}}\lambda_i\cdot 1_{\{k_i>0\}}+\sum_{i\in\overline{J}}\nu p_i\cdot 1_{\{k_i<b_i\}}\right)\\
&= && \sum_{i\in\overline{J}}\theta(\mathbf{k})\,\nu p_i\cdot 1_{\{k_i<b_i\}}+\sum_{i\in\overline{J}}\theta(\mathbf{k})\,\lambda_i\cdot 1_{\{k_i>0\}}\,.
\end{aligned}$$

□

Remark 10. $\theta(\mathbf{k})=\theta(k_1,\dots,k_J,k_{J+1})$ is obtained as a stochastic solution of (8), which resembles the global balance equations of an artificial non-standard Gordon-Newell network with $J+1$ nodes and $\sum_{j\in\overline{J}}b_j$ customers, exponentially distributed service times with rate λ_j for $k_j\leq b_j$ and "∞" otherwise at node $j\in\{1,...,J\}$ and with rate ν at node $J+1$.

3.3 Average costs

We will analyze average long term costs of the system as a function of the base stock levels $\mathbf{b}=(b_1,...,b_J)$.

Lemma 11. *Optimal solutions for the problem described in 3.1 are the set*

$$\arg\min\left(\overline{g}(\mathbf{b})\right)$$

with

$$\begin{aligned}\overline{g}(\mathbf{b}) \quad &:= \quad \sum_{j\in\bar{J}} c_{s,j}\cdot b_j + \sum_{\mathbf{k}\in K}\left(\sum_{j\in\bar{J}} c_{ls,j}\cdot 1_{\{k_j=0\}} + \sum_{j\in\overline{J}\cup\{J+1\}} c_{h,j}\cdot k_j\right)\theta(\mathbf{k}) \\ &= \quad \sum_{j\in\bar{J}}(c_{s,j}+c_{h,J+1})\cdot b_j + \sum_{\mathbf{k}\in K}\left(\sum_{j\in\bar{J}}\left(c_{ls,j}\cdot 1_{\{k_j=0\}} + (c_{h,j}-c_{h,J+1})\cdot k_j\right)\right)\theta(\mathbf{k}).\end{aligned}$$

The proof is the same as the proof of Lemma 5 with $k_{J+1} = \sum_j (b_j - k_j)$.

Although the locations and their describing processes are obviously strongly correlated because of the common replenishment mechanisms, it can be shown that this optimization problem is separable in the sense that we can split the global optimization problem into a set of independent local optimization problems.

Let $(Y_1, ..., Y_J, Y_{J+1})$ denote random variables which are distributed according to θ.

Theorem 12. *The optimal base stock level* $\mathbf{b} = (b_1, ..., b_J)$ *is determined as*

$$b_j \in \arg\min(\overline{g}_j) \quad \forall j \in \overline{J}$$

with

$$\overline{g}_j(b_j) := b_j\cdot(c_{s,j}+c_{h,J+1}) + P(Y_j=0)\cdot c_{ls,j} + E(Y_j)\cdot(c_{h,j}-c_{h,J+1}),$$

where $P(Y_j=0) = \left(\sum_{k_j=0}^{b_j}\left(\frac{\nu p_j}{\lambda_j}\right)^{k_j}\right)^{-1}$ *and* $E(Y_j) = \sum_{k_j=0}^{b_j} k_j\left(\frac{\nu p_j}{\lambda_j}\right)^{k_j}\left(\sum_{k_j=0}^{b_j} k_j\left(\frac{\nu p_j}{\lambda_j}\right)^{k_j}\right)^{-1}$.

Proof. For $\mathbf{k} \in E = \left\{(\mathbf{n},\mathbf{k}) : \mathbf{n}\in\mathbb{N}_0^{\overline{J}},\, \mathbf{k}\in K\right\}$ with

$$K := \{(k_1, ..., k_J, \underbrace{\sum_{j=1}^{J}(b_j-k_j)}_{=:k_{J+1}})\,|0\le k_j\le b_j,\ j\in\overline{J}\}$$

we have

$$\theta(\mathbf{k}) = \theta(k_1,\ldots,k_J,k_{J+1}) = C_\theta^{-1}\prod_{j\in\overline{J}}\left(\frac{\nu p_j}{\lambda_j}\right)^{k_j}.$$

We transform K and θ by an isomorphism $(k_1,\ldots,k_J,k_{J+1}) \leftrightarrow (k_1,\ldots,k_J)$ and obtain as image of K:

$$K_- := \{(k_1, ..., k_J)|0\le k_j\le b_j,\ j\in\overline{J}\} = \prod_{j=1}^{J}\{0,1,...,b_j\}.$$

Then θ is transformed to the image measure θ_-: With $\mathbf{k}_- := (k_1, ..., k_J) \leftrightarrow (k_1,\ldots,k_J,\sum_{j=1}^{J}(b_j-k_j))$ we obtain

$$\theta_-(\mathbf{k}_-) = \theta_-(k_1,...,k_J) = C_\theta^{-1}\prod_{j=1}^{J}\left(\frac{\nu p_j}{\lambda_j}\right)^{k_j} = \theta(K).$$

The reason for introducing K_- and θ_- is that K_- is a product space. Therefore, K is isomorph to such a product space and θ_- is a product measure on K_-. Hence, the normalization constant factorizes as well:

$$\theta_-(\mathbf{k}_-) = \theta_-(k_1, ..., k_J) = \prod_{j=1}^{J} C_{\theta,j}^{-1} \left(\frac{\nu p_j}{\lambda_j}\right)^{k_j}, \qquad k_j \in \{1, ..., b_j\},\ j \in \overline{J}.$$

Then for $\overline{g}(\mathbf{b})$ holds

$$\begin{aligned}
\overline{g}(\mathbf{b}) &:= \sum_{j\in\overline{J}} (c_{s,j} + c_{h,J+1}) \cdot b_j + \sum_{\mathbf{k}\in K} \left(\sum_{j\in\overline{J}} \left(c_{ls,j} \cdot 1_{\{k_j=0\}} + (c_{h,j} - c_{h,J+1}) \cdot k_j \right) \right) \theta(\mathbf{k}) \\
&= \sum_{j\in\overline{J}} (c_{s,j} + c_{h,J+1}) \cdot b_j + \sum_{\mathbf{k}_-\in K_-} \left(\sum_{j\in\overline{J}} \left(c_{ls,j} \cdot 1_{\{k_j=0\}} + (c_{h,j} - c_{h,J+1}) \cdot k_j \right) \right) \theta_-(\mathbf{k}_-).
\end{aligned}$$

Using the product form, we get

$$\begin{aligned}
\overline{g}(\mathbf{b}) &= \sum_{j\in\overline{J}} (c_{s,j} + c_{h,J+1}) \cdot b_j + \sum_{j\in\overline{J}} \sum_{\mathbf{k}_-\in K_-} \left(\theta_-(\mathbf{k}_-) \left(c_{ls,j} \cdot 1_{\{k_j=0\}} + (c_{h,j} - c_{h,J+1}) \cdot k_j \right) \right) \\
&= \sum_{j\in\overline{J}} (c_{s,j} + c_{h,J+1}) \cdot b_j + \sum_{j\in\overline{J}} \sum_{\mathbf{k}_-\in K_-} \left(\prod_{j=1}^{J} C_{\theta,j}^{-1} \left(\frac{\nu p_j}{\lambda_j}\right)^{k_j} \left(c_{ls,j} \cdot 1_{\{k_j=0\}} + (c_{h,j} - c_{h,J+1}) \cdot k_j \right) \right) \\
&= \sum_{j\in\overline{J}} (c_{s,j} + c_{h,J+1}) \cdot b_j \\
&\quad + \sum_{j\in\overline{J}} \left(\sum_{k_j=0}^{b_j} C_{\theta,j}^{-1} \left(\frac{\nu p_j}{\lambda_j}\right)^{k_j} \left(c_{ls,j} \cdot 1_{\{k_j=0\}} + (c_{h,j} - c_{h,J+1}) \cdot k_j \right) . \right. \\
&\qquad\qquad \left. \underbrace{\cdot \prod_{\substack{i=1,\\ i\neq j}}^{J} \sum_{k_i=0}^{b_i} C_{\theta,i}^{-1} \left(\frac{\nu p_i}{\lambda_i}\right)^{k_i}}_{=1} \right)
\end{aligned}$$

Set $(Y_1, ..., Y_J) \sim \theta_-$, then

$$\overline{g}(\mathbf{b}) = \sum_{j\in\bar{J}} (c_{s,j} + c_{h,J+1}) \cdot b_j + \sum_{j\in\bar{J}} \left(P(Y_j = 0) \cdot c_{ls,j} + (c_{h,j} - c_{h,J+1}) \cdot E(Y_j) \right),$$

where Y_j distributed according to a truncated geometric distribution. It follows

$$\overline{g}(\mathbf{b}) = \sum_{j\in\bar{J}} \left(b_j \cdot (c_{s,j} + c_{h,J+1}) + P(Y_j = 0) \cdot c_{ls,j} + E(Y_j) \cdot (c_{h,j} - c_{h,J+1}) \right),$$

where $P(Y_j = 0) = \left(\sum_{k_j=0}^{b_j} \left(\frac{\nu p_j}{\lambda_j}\right)^{k_j} \right)^{-1}$ and $E(Y_j) = \sum_{k_j=0}^{b_j} k_j \left(\frac{\nu p_j}{\lambda_j}\right)^{k_j} \left(\sum_{k_j=0}^{b_j} k_j \left(\frac{\nu p_j}{\lambda_j}\right)^{k_j} \right)^{-1}$.

To find the/a global minimum it suffices to minimize each summand

$$\overline{g}_j(b_j) := b_j \cdot (c_{s,j} + c_{h,J+1}) + P(Y_j = 0) \cdot c_{ls,j} + E(Y_j) \cdot (c_{h,j} - c_{h,J+1}), \qquad j \in \overline{J},$$

separately, since $\overline{g}_j(b_j) > 0$ for all $j \in \overline{J}$.

Therefore, the optimal base stock level $\mathbf{b} = (b_1, ..., b_J)$ is determined as

$$b_j \in \arg\min(\overline{g}_j) \quad \forall j \in \overline{J} \ .$$

□

Remark 13. For the model described in Section 2.1 we cannot show that the optimization problem is separable in the sense that we can split the global optimization problem into a set of independent local optimization problems.

References

[Axs00] S. Axsäter. *Inventory Control*. Kluwer Academic Publishers, Norwell, Massachusetts, 2000.

[BV11] M. Bijvank and F.A. Vis. Lost-sales inventory theory: A review. *European Journal of Operations Research*, 215:1–13, 2011.

[CMP99] X. Chao, M. Miyazawa, and M. Pinedo. *Queueing Networks – Customers, Signals, and Product Form Solutions*. Wiley, Chichester, 1999.

[Dad90] H. Daduna. Exchangeable items in repair systems: Delay times. Technical Report 2, 1990.

[DS96] H. Daduna and R. Szekli. A queueing theoretical proof of increasing property of Polya frequency functions. *Statistics and Probability Letters*, 26:233–242, 1996.

[GCB02] T.W. Gruen, D. Corsten, and S. Bharadwaj. Retail out-of-stocks: A worldwide examination of extent causes and consumer responses. *Grocery Manufacutres of America*, 2002.

[Jac57] J.R. Jackson. Networks of waiting lines. *Operations Research*, 5:518–521, 1957.

[Kel79] F. P. Kelly. *Reversibility and Stochastic Networks*. John Wiley and Sons, Chichester – New York – Brisbane – Toronto, 1979.

[Kle76] L. Kleinrock. *Queueing Theory*, volume II. John Wiley and Sons, New York, 1976.

[KLM11] A. Krishnamoorthy, B. Lakshmy, and R. Manikandan. A survey on inventory models with positive service time. *OPSEARCH*, 48(2):153–169, 2011.

[Lig85] T.M. Liggett. *Interacting Particle Systems*, volume 276 of *Grundlehren der mathematischen Wissenschaften*. Springer, Berlin, 1985.

[Mor58] P.M. Morse. *Queues, Inventories and Maintenance.* Wiley, New York, 1958.

[RBP13] R. Ravid, O.J. Boxma, and D. Perry. Repair systems with exchangeable items and the longest queue mechanism. *Queueing Syst.*, pages 295–316, 2013.

[RW96] R. Rubio and L.M. Wein. Setting base stock levels using product-form queueing networks. *Operations Research*, 42:259–268, 1996.

[Sch84] R. Schassberger. Decomposable stochastic networks: Some observations. In F. Baccelli and G. Fayolle, editors, *Modelling and Performance Evaluation Methodology*, volume 60 of *Lecture Notes in Control and Information Sciences*, chapter IV, pages 137 – 150. Springer, Berlin, 1984. Proceedings of the International Seminar, Paris, France, January 24-26, 1983.

[SD04] C. Sauer and H. Daduna. Degradable networks with general up and down time distributions. In P. Buchholz, R. Lehnert, and M. Pioro, editors, *Proceedings of the 12th GI/ITG Conference on MMB together with 3rd PGTS*, pages 185–194, Berlin, 2004. VDE-Verlag.

[SPP98] E.A. Silver, D. Pyke, and R. Peterson. *Inventory Management and Production Planning and Scheduling.* John Wiley and Sons, Inc., Chichester – New York – Brisbane – Toronto – Singapore, 3 edition, 1998.

[Tem05] H. Tempelmeier. *Bestandsmanagement in Supply Chains.* Books on Demand, 2005.

[vDB13] K.H. van Donselaar and R.A.C.M. Broekmeulen. Determination of safety stocks in a lost sales inventory system with periodic review, positive lead-time, lot-sizing and a target fill rate. *Int. J. Production Economics*, 143:440 – 448, 2013.

[vdGdKAR12] J.P. van der Gaast, M.B.M. de Koster, I.J.B.F. Adan, and J.A.C. Resing. Modeling and performance analysis of sequential zone picking systems. http://alexandria.tue.nl/repository/books/751517.pdf, November 2012.

[VS06] P. Verhoef and L.M. Sloot. Out-of-stock: Reactions, antecedents, management solutions, and a future perspective. In *Krafft, M., Mantrala, M.K. (Eds.), Retailing in the 21st Century: Current and Future Trends*, pages 239–253. Springer Publishers, Philadelphia, Pennsylvania, USA, 2006.

[Zaz94] M. Zazanis. Push and pull systems with external demands. In *Proceedings of the 32nd Allerton Conference on Communication, Control, and Computing*, Allerton, Illinois, 1994.

[Zip00] P.H. Zipkin. *Foundations of inventory management.* Jeffrey J.Shelstad, 2000.