Submitted to *Operations Research*
manuscript (Please, provide the mansucript number!)

Modeling and performance analysis of sequential zone picking systems

J.P. van der Gaast, R.B.M. de Koster
Department of Management of Technology & Innovation, Erasmus Universiteit Rotterdam, The Netherlands, jgaast@rsm.nl, rkoster@rsm.nl

I.J.B.F. Adan
Department of Mechanical Engineering, Technische Universiteit Eindhoven, The Netherlands, i.j.b.f.adan@tue.nl

J.A.C. Resing
Department of Mathematics and Computer Science, Technische Universiteit Eindhoven, The Netherlands, j.a.c.resing@tue.nl

This paper develops an analytical model to study sequential zone picking systems, which is a very popular method of order picking. The various elements of the system, like the conveyor and the pick zones, are modeled as a network of queues with multiple order classes and capacity constraints on subnetworks. The stationary distribution of the network is highly intractable, due to the dynamic block-and-recirculate protocol. We develop an approximative, iterative method, using jump-over blocking, to accurately assess key performance statistics such as throughput and recirculation. Multi-class jump-over networks admit a product-form stationary distribution, and can be efficiently evaluated by Mean Value Analysis (MVA) and use of Norton's theorem. The method can be used to support the design of complex zone picking systems, in terms of number of segments, number and length of the zones, buffer capacities, and storage allocation of products to zones, in order to meet prespecified performance targets. Comparison of the approximation results to simulation show that for a wide range of parameters the mean relative error in the system throughput is typically less than 1%. The approximation is also used to evaluate a real-life zone picking system of a large wholesaler supplying non-food items.

Subject classifications: Warehousing; Queues: algorithms, approximations.
Area of review: Transportation.

1. Introduction

Order picking, the process of picking products to fill customer orders, is the most labor-intensive and costly activity in warehouses due to its high contribution (about 55%) to the total operating cost (Drury 1988). Recent trends in distribution and manufacturing, like e-commerce, have increased the importance of efficient order picking even more (Le-Duc and de Koster 2007). The focus of

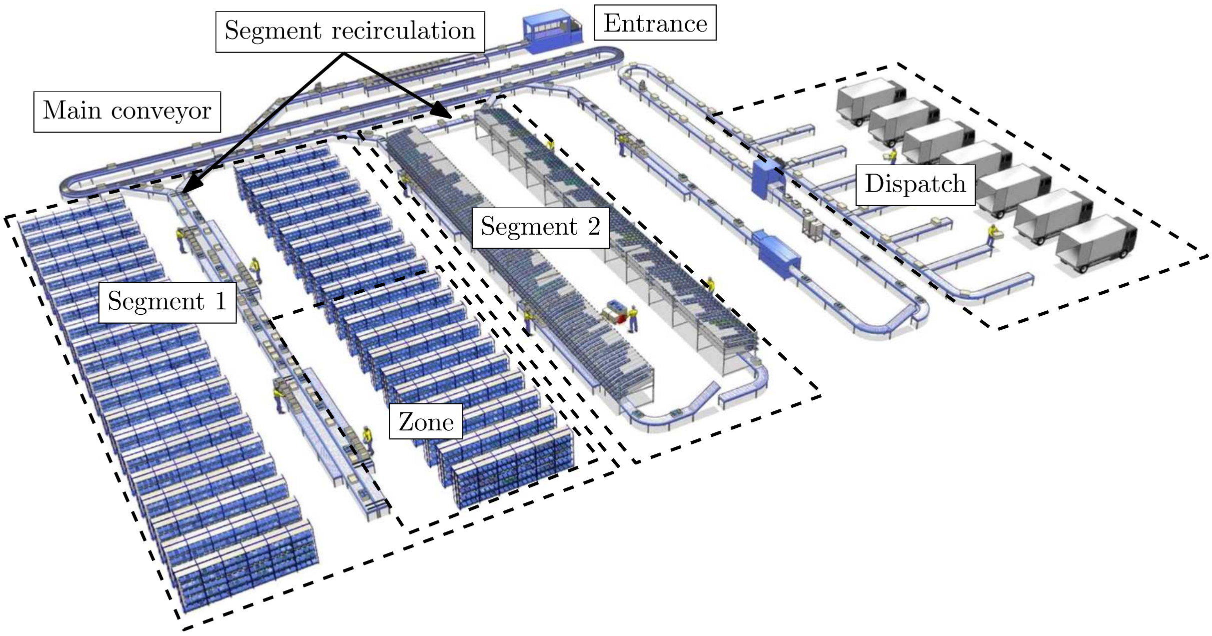


Figure 1 A sequential zone picking system with multi-segment routing. (Bakker 2007)

this paper is on the modeling and (approximate) analysis of sequential zone picking systems, with single or multi-segment routing.

Zone picking is a commonly used picker-to-parts order picking method, where the order picking area is zoned. In each zone, an order picker is responsible for picking from his or her dedicated part of the warehouse (Petersen 2002, Gu et al. 2010). In practice, the zones are often connected by conveyors. Zone picking systems are flexible in handling both small and large order volumes, different kinds of product sizes, with a different number of order pickers. These systems are often applied in warehouses handling customer orders with a large number of order lines and with a large number of different products kept in stock (Park 2012).

Zone picking systems can be divided in systems with parallel and sequential zone picking (De Koster et al. 2007). In a parallel zone picking system, a customer order, which can consist of several order lines, is picked simultaneously in multiple zones and a downstream sorting process consolidates the picked order lines into the customer orders after the picking process has finished. In sequential zone picking (or pick-and-pass), shown in Figure 1, an order is assigned to an order tote or order carton that travels on the conveyor and visits sequentially only those zones where products are stored that should be added to the order. At a zone, each picker picks for only one tote at a time. The advantage of sequential zone picking is that order integrity is maintained and no sorting and product consolidation is required (Petersen 2000).

There are two types of sequential zone picking systems that can be distinguished: *single-segment routing* and *multi-segment routing*. In single-segment routing, the conveyor forms one circular loop that connects all the zones, whereas in multi-segment routing, zones are grouped in segments and per segment the zones are connected to a conveyor with a recirculation loop, like in Figure 1. The different segments are then connected by a central (or main) conveyor that diverts totes to the required segments. Multi-segment routing can improve system throughput significantly due to shorter conveyor loops that avoid unnecessary long travel times. However, the investment costs and space requirements are higher compared to single-segment routing.

De Koster (1994), Yu and De Koster (2008, 2009) and Melacini et al. (2010) model a zone picking system as a network of queues. In order to estimate performance statistics, such as the utilization and throughput rate of a zone, and the mean and standard deviation of the throughput time of the totes, they use Whitt's queueing network analyzer (Whitt 1982). A crucial aspect, however, that is not taken into consideration is blocking. In most environments, the workloads of the zones exhibit highly variable behavior, due to differences in work profiles of the orders. In peak periods, zones can become congested, leading to blocking of order totes which may propagate through the entire network, such that zones become starved and order lead times significantly increase. Such blocking effects may have considerable impact on the performance of a zone picking system and cannot be ignored. Identification and quantification of the effects of blocking is very challenging and crucial in the design of a zone picking system.

Blocking and congestion occurs at zones as well as at segments. Zones can become congested due to finite buffer space, since only a limited number of totes can be stored before they are processed by the order picker. When the buffer is full, incoming totes from the conveyor cannot divert to this zone, thereby temporarily blocking all upstream totes. This is often resolved by providing a recirculation option for blocked totes, such that totes can visit other zones before returning to the blocked zone. Also segments can become congested if too many totes visit the segment at once. To resolve this, workload control can be applied by which the system will prevent incoming totes from entering the segment until sufficiently many totes have left.

Zone picking systems use a dynamic *block-and-recirculate* protocol: a blocked tote recirculates on the conveyor loop when the destination buffer is full or a segment is congested. The tote potentially visits other zones or segments before attempting to enter the place where it was blocked previously. Queueing networks that attempt to model systems with such blocking protocols are highly intractable: there are absolutely no exact results for the stationary distribution. In the literature, different blocking protocols have been investigated for various types of applications (see Schmidt and Jackman (2000), Hsieh and Bozer (2005), Osorio and Bierlaire (2009) for recent references in manufacturing systems with automated conveyors). For an extensive review on blocking in queueing networks, the reader is referred to the books of Perros (1994), Balsamo et al. (2001), Papadopoulos et al. (1993). A variation of the block-and-recirculate protocol for flexible manufacturing systems (FMS) was first studied by Yao and Buzacott (1987). According to their definition, a blocked part returns to the end of the queue where it came from such that it can try to enter for a second time. The authors derive product-form solutions for FMS networks with finite buffers and recirculation, including networks with a central server (e.g., the material handling system) and networks with zero-buffer stations. These networks, however, are not suitable to model zone picking systems, since totes can visit the required zones in a random order and they do not need to return immediately to the same blocked zone, but potentially visit other, less congested, zones before returning.

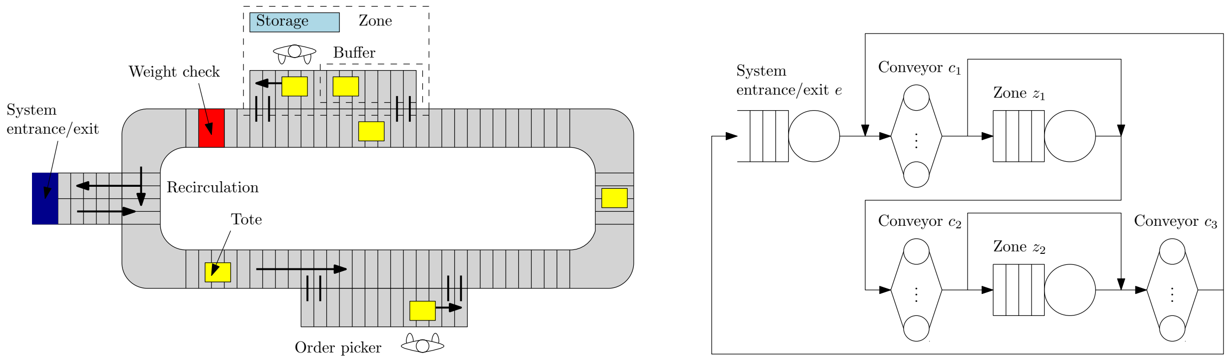
The objective of this paper is to develop an analytical model for sequential zone picking systems (hereafter zone picking) with either single-segment routing, or multi-segment routing. This model is used to study the effects of design choice, loading, and storage on blocking and congestion of this commonly used order picking method. It considerably extends the models of De Koster (1994), Yu and De Koster (2008, 2009) and Melacini et al. (2010) that only consider zone picking systems with single-segment routing and no blocking. We develop a queueing model that incorporates the dynamic block-and-recirculate protocol and use the model to estimate the key performance statistics. Because an exact analysis of the queueing model with blocking is not feasible, we iteratively estimate the blocking probabilities from a multi-class queueing network, with *jump-over* blocking (Van Dijk 1988, Economou and Fakinos 1998). This is a Markovian blocking protocol that will be shown to admit a product-form stationary queue-length distribution. Key to the approximation is to equip the jump-over queueing network with Markovian routing that correctly reflects the relation to the block-and-recirculate queueing network. Surprisingly, the jump-over queueing network provides very accurate estimates of the key performance statistics and allows us to study the sources of blocking and congestion in large systems containing many zones and many different classes. As such, it provides a powerful tool for system design, e.g., number of segments, number and size of the zones, buffer capacities, storage allocation of products to zones, in order to meet target performance levels.

The organization of this paper is as follows. Section 2 presents the model for single-segment routing zone picking systems and Section 3 discusses the corresponding approximation and analysis. The model is generalized to multi-segment routing zone picking systems in Section 4 and the approximation and analysis are given in Section 5. We extensively analyze the results of both models in Section 6 via computational experiments for a range of parameters and validate them for a real-life system. In the final section we conclude and suggest some extensions of the model and further research topics.

2. Modeling Zone Picking Systems with Single-segment Routing

In a zone picking system with single-segment routing, three different elements can be distinguished: the entrance/exit of the system, the conveyor, and the zones. These elements of a single-segment zone picking system with two zones are shown in Figure 2a. For now, the assumption is made that totes enter and leave the system at the same location.

The system works as follows. At the entrance/exit of the system, a customer order is assigned to an order tote. The tote is released into the system as soon as it is allowed by the workload



(a) A zone picking system with single-segment routing and two zones

(b) The corresponding queueing network with system entrance/exit station e, $\mathcal{C} = \{c_1, c_2, c_3\}$ and, $\mathcal{Z} = \{z_1, z_2\}$

Figure 2 A zone picking system with single-segment routing and its corresponding queueing network.

control mechanism (Park 2012). This mechanism keeps the number of totes in the system fixed over time and only releases a new tote when a tote with all required order lines leaves the system. This workload control mechanism prevents the conveyor to become the bottleneck of the system. After release, a tote moves to the buffer of a requested zone and diverts if the buffer of that zone is not full. A blocked tote will stay on the conveyor and visits potentially other zones before returning. When the picking process has finished, the picker pushes the tote back on the conveyor. The waiting time for a sufficiently large space on the conveyor is considered to be negligible. The conveyor then transports the tote to the next zone to be visited. A weight check at the end of the conveyor loop ensures that the tote contains all the required order lines; otherwise, the tote is sent back to the beginning of the loop. The blocked totes will return to the zones where they were blocked previously and attempt to enter a second time. When the tote has visited all the required zones, it leaves the system at the exit and a new order tote is immediately released into the system.

To model this system, a queueing network is proposed, the topology of which is shown in Figure 2b in case of two zones. The zone picking system is modeled as a closed queueing network with one entrance/exit, M zones and $M+1$ nodes that describe the conveyor between either two adjacent zones or between the entrance/exit and the first or last zone. The nodes are labeled in the following manner: denote the system entrance/exit as e, and let $\mathcal{Z} = \{z_1, \ldots, z_M\}$ denote the set of zones and $\mathcal{C} = \{c_1, \ldots, c_{M+1}\}$ the set of conveyors in the network. Finally, let $\mathcal{S} = \{e\} \cup \mathcal{C} \cup \mathcal{Z}$ be the union of all the nodes in the network. The following assumptions are adopted for the network:

- There is an infinite supply of totes at the entrance of the system. This means that a leaving tote can always be replaced immediately by a new tote. Each tote has a class $\boldsymbol{r} \subseteq \mathcal{Z}$, e.g., $\boldsymbol{r} = \{z_2, z_3\}$ means that the tote has to visit the second and third zone.
- The total number of totes in the system is constant N. As long as the total number of totes in the zones and conveyor nodes is less than N, new totes are released one-by-one at an exponential rate μ_e at the system entrance.
- The conveyor nodes are assumed to be delay nodes with a fixed delay of rate μ_i, $i \in \mathcal{C}$.
- Each zone has $d_i\,(\geq 1)$, $i \in \mathcal{Z}$ servers, which represent the order pickers in the zone. The order picking time is assumed to be exponentially distributed with rate μ_i, $i \in \mathcal{Z}$, that captures both variations in the pick time per tote and variations in the number of order lines to be picked; this assumption is relaxed in Subsection 3.4.
- When the order pickers are busy, incoming totes are stored in a finite buffer of size q_i, $i \in \mathcal{Z}$. Incoming totes are blocked when the total number of totes in the buffer equals q_i.

A new tote of class $\boldsymbol{r} \subseteq \mathcal{Z}$ is released at the system entrance with probability $\psi_{\boldsymbol{r}}$. These release probabilities correspond to a known order profile that can be obtained using e.g., historical data or forecasts. After release, a tote of class $\boldsymbol{r}$ moves from the system entrance to the first conveyor node c_1. At conveyor node c_i, the tote will either divert to zone i if $z_i \in \boldsymbol{r}$ or move to the next conveyor node c_{i+1}. In case the tote needs to enter and the buffer is full, the tote skips the zone and moves to the next conveyor c_{i+1}, while it keeps the same class. If the buffer is not full, the tote enters the zone and, after possibly waiting in the buffer, the order picker picks the required order lines. The class of the totes changes to $\boldsymbol{s} = \boldsymbol{r} \setminus \{z_i\}$ at departure from the zone, when the tote is routed to conveyor node c_{i+1}. After visiting the last conveyor node c_{M+1}, all the totes with $\boldsymbol{r} \neq \emptyset$ are routed to the first conveyor node c_1; the other totes move to the exit and are immediately replaced by a new tote which will wait for release at the entrance.

The stationary distribution of this queueing network is intractable due to the finite buffers and block-and-recirculate mechanism (Stidham Jr. 2002). This justifies the attempt to develop an approximate analysis of this queueing network.

3. Analysis of Zone Picking Systems with Single-segment Routing

In order to accurately estimate the performance statistics of the queueing network of Section 2, it is approximated in Subsection 3.1 by a network with the jump-over blocking protocol. This jump-over network exhibits a product-form steady state distribution, as will be shown in Subsection 3.2. Another property of the network, shown in Subsection 3.3, is that closed-form formulas of the visit ratios exist. The performance statistics of the jump-over network can be easily calculated exactly using e.g., mean value analysis (MVA), as in Subsection 3.4. Subsection 3.5 shows how the jump-over network is used to approximate the original network. Finally, the quality of the single-segment approximation is presented in Subsection 3.6.

3.1. Jump-over Network

Assume a tote that intends to visit zone z_i is "tagged" after z_i with either the label *visited* z_i or *skipped* z_i. In the real system, and hence in the queueing network of Section 2, a tote is tagged as *visited* z_i if the tote entered z_i and received service. On the other hand, a tote is tagged as *skipped* z_i if the tote skipped the zone because the buffer was full.

The idea of the jump-over network is to tag each tote that intends to visit z_i randomly, and independent of whether the tote actually visited z_i or not. The probability of tagging with either one of the two labels is taken as the fraction of totes receiving a specific tag in the original queueing network, such that the fraction of totes that are tagged with *skipped* z_i equals the blocking probability b_i, $i \in \mathcal{Z}$. Naturally, blocking probabilities are not known in advance, but will be estimated iteratively after an initial guess from the approximation, as shown in Subsection 3.5. Hereafter, b_i is assumed to be known beforehand in the jump-over network. Denote by $p_{i\boldsymbol{r},j\boldsymbol{s}}$ the routing probability in the jump-over network that a tote of class $\boldsymbol{r}$ is routed from node i to node j and enters as a class $\boldsymbol{s}$ tote. Then for each class $\boldsymbol{r}$ tote, independent of whether the tote visited or skipped

z_i (because of a full buffer), $1 - p_{z_i \boldsymbol{r}, c_{i+1}\boldsymbol{s}} = b_{z_i}$, $i = 1, \ldots, M$, where $\boldsymbol{s} = \boldsymbol{r} \setminus \{z_i\}$. This means that a tote of class $\boldsymbol{r}$ is tagged *skipped* z_i and routed to the next conveyor node c_{i+1} with the same class with probability b_{z_i}, and otherwise, with probability $1 - b_{z_i}$, the tote is tagged as *visited* z_i and the class of the tote changes to $\boldsymbol{s} = \boldsymbol{r} \setminus \{z_i\}$.

Since the tagging process is made independent of the state of the buffer, essentially the block-and-recirculate protocol is replaced by the *jump-over* blocking protocol (Van Dijk 1988). Under this protocol, each tote of class $\boldsymbol{r}$ leaving z_i, either after service or skipping, continues to follow the same Markovian routing. The advantage of the jump-over blocking protocol, also known as "overtake full stations, skipping, and blocking and rerouting", is that closed-form analytic results for single-class queueing networks are available in the literature (Pittel 1979, Schassberger 1984, Van Dijk 1988, Economou and Fakinos 1998).

3.2. Product-form of the Stationary Distribution of the Jump-over Network

A powerful tool to prove the existence of product-form solutions in the multi-class jump-over network is the concept of *quasi-reversibility* (Kelly 1979); when a queueing network with Markovian routing can be decomposed in terms of nodes that are quasi-reversible, the stationary distribution of the network can be written as the product of the stationary distributions of these individual nodes.

The state of the jump-over network is defined as $x = (x_i : i \in \mathcal{S})$, where $x_i = (\boldsymbol{r}_{i1}, \ldots, \boldsymbol{r}_{il}, \ldots, \boldsymbol{r}_{in_i})$ represents the state of node i with $\boldsymbol{r}_{i1}$ as the class of the first tote in the node, and $\boldsymbol{r}_{in_i}$ as the class of the last tote in the node. Let $\mathbb{S}(N)$ be the state space of the network, i.e., it is the set of states x for which the number of totes in the system is equal to $\sum_{i \in \mathcal{S}} n_i = N$ and where the number of totes in each zone satisfies $n_i \leq d_i + q_i$, $i \in \mathcal{Z}$.

Let $\lambda_{i\boldsymbol{r}}$ be the visit ratio of a class $\boldsymbol{r}$ tote to node i satisfying the traffic equations

$$\lambda_{i\boldsymbol{r}} = \sum_{j \in \mathcal{S}} \sum_{\boldsymbol{s} \subseteq \mathcal{Z}} \lambda_{j\boldsymbol{s}} p_{j\boldsymbol{s}, i\boldsymbol{r}}, \quad i \in \mathcal{S}, \boldsymbol{r} \subseteq \mathcal{Z}. \tag{1}$$

The equations (1) determine the visit ratios $\lambda_{i\boldsymbol{r}}$ up to a multiplicative constant.

THEOREM 1. *The jump-over network with state space $\mathbb{S}(N)$ has a product-form stationary distribution of the form*

$$\pi(x) = \frac{1}{G}\prod_{i\in\mathcal{S}}\pi_i(x_i), \tag{2}$$

where G is a normalization constant and

$$\pi_i(x_i) = \begin{cases} \prod_{l=1}^{n_i}\left(\frac{\lambda_{i\boldsymbol{r}_{il}}}{\mu_i}\right), & i = e, \\ \prod_{l=1}^{n_i}\left(\frac{\lambda_{i\boldsymbol{r}_{il}}}{\mu_i}\right)\cdot\frac{1}{n_i!}, & i\in\mathcal{C}, \\ \prod_{l=1}^{n_i}\left(\frac{\lambda_{i\boldsymbol{r}_{il}}}{\mu_i}\right)\cdot\frac{1}{\gamma(n_i)}, & i\in\mathcal{Z}, \end{cases} \tag{3}$$

with $\lambda_{i\boldsymbol{r}_{il}}$ is a solution of the traffic equations (1) *and*

$$\gamma(n_i) = \begin{cases} n_i!, & \text{if } n_i \le d_i, \\ d_i!\,(d_i)^{n_i-d_i}, & \text{if } n_i > d_i. \end{cases}$$

Proof: To verify that the jump-over network with transition rates $q(x,y)$, $x,y\in\mathbb{S}(N)$ has a product-form stationary distribution of form (2), it is sufficient to find non-negative numbers $\bar{q}(y,x)$ and a collection of positive numbers $\pi(x)$ summing to unity, such that the following two conditions are fulfilled (Kelly 1979)

$$\bar{q}(x) = q(x), \qquad x\in\mathbb{S}(N), \tag{4}$$

$$\pi(x)\,q(x,y) = \pi(y)\,\bar{q}(y,x), \quad x,y\in\mathbb{S}(N), \tag{5}$$

where $q(x) = \sum_{y\in\mathbb{S}(N)} q(x,y)$ and $\bar{q}(x) = \sum_{y\in\mathbb{S}(N)}\bar{q}(x,y)$. Then $\bar{q}(y,x)$ are the time-reversed transition rates and $\pi(x)$ is the product-form stationary distribution of the network.

Whenever state x does not contain a blocked zone, $\bar{q}(y,x)$ is defined the similar as a regular multi-class queueing network without jump-over blocking (Nelson 1995). Each node in the jump-over network has exponential service times and is either a multi-class single/multi-server node or a multi-class infinite server node which are well-known to be quasi-reversible. Then it can easily be verified that conditions (4) and (5) hold and $\pi(x)$ is given by the product form of Equation (2). In case state x contains a blocked zone and a transition that involves a tote skipping a zone occurs, the transition rates $q(x,y)$ and $\bar{q}(y,x)$ can be described as follows. When a tote of class $\boldsymbol{r}$ departs from the lth position of conveyor node c_i and moves to zone z_i with rate μ_{c_i}, it immediately jumps over the zone if it encounters a full buffer. The tote will move to the next conveyor node c_{i+1} with probability 1, where it joins the end of the node. When arriving in c_{i+1} the tote is either tagged as *visited* z_i with probability $1-b_i$ such that its class becomes $\boldsymbol{s}=\boldsymbol{r}\setminus\{z_i\}$ or as *skipped* z_i with probability b_i while the class of the tote remains the same.

The transition rates of a tote skipping zone i are given as follows, where state $x-\boldsymbol{r}_{c_il}+\boldsymbol{s}_{c_{i+1}n_{c_{i+1}}+1}$ denotes the removal of a class $\boldsymbol{r}$ tote at the lth position in c_i and an arrival of a class $\boldsymbol{s}$ tote in c_{i+1} at the last position,

$$q\left(x, x-\boldsymbol{r}_{c_il}+\boldsymbol{r}_{c_{i+1}n_{c_{i+1}}+1}\right)=\mu_{c_i}b_{z_i}, \qquad l=1,\ldots,n_{c_i}, \tag{6}$$

$$q\left(x, x-\boldsymbol{r}_{c_il}+\boldsymbol{s}_{c_{i+1}n_{c_{i+1}}+1}\right)=\mu_{c_i}\left(1-b_{z_i}\right), \quad l=1,\ldots,n_{c_i}. \tag{7}$$

In the time-reversed process a tote of class $\boldsymbol{r}$ or $\boldsymbol{s}$ departs from the last position in c_{i+1} with rate $\left(n_{c_{i+1}}+1\right)\mu_{c_{i+1}}$ and joins at position l in c_i as class $\boldsymbol{r}$ with probability $\lambda_{c_i\boldsymbol{r}_{c_il}}b_{z_i}/\lambda_{c_{i+1}\boldsymbol{r}_{c_{i+1}n_{c_{i+1}}+1}}$ if it was tagged as *skipped* z_i and $\lambda_{c_i\boldsymbol{r}_{c_il}}\left(1-b_{z_i}\right)/\lambda_{c_{i+1}\boldsymbol{s}_{c_{i+1}n_{c_{i+1}}+1}}$ otherwise,

$$\bar{q}\left(x-\boldsymbol{r}_{c_il}+\boldsymbol{r}_{c_{i+1}n_{c_{i+1}}+1}, x\right)=\mu_{c_{i+1}}\frac{n_{c_{i+1}}+1}{n_{c_i}}\frac{\lambda_{c_i\boldsymbol{r}_{c_il}}b_{z_i}}{\lambda_{c_{i+1}\boldsymbol{r}_{c_{i+1}n_{c_{i+1}}+1}}}, \quad l=1,\ldots,n_{c_i}, \tag{8}$$

$$\bar{q}\left(x-\boldsymbol{r}_{c_il}+\boldsymbol{s}_{c_{i+1}n_{c_{i+1}}+1}, x\right)=\mu_{c_{i+1}}\frac{n_{c_{i+1}}+1}{n_{c_i}}\frac{\lambda_{c_i\boldsymbol{r}_{c_il}}\left(1-b_{z_i}\right)}{\lambda_{c_{i+1}\boldsymbol{s}_{c_{i+1}n_{c_{i+1}}+1}}}, \quad l=1,\ldots,n_{c_i}. \tag{9}$$

Inserting (2)-(3) and (6)-(9) and canceling identical terms, it can be verified that (5) holds. By considering also transitions that do not skip a zone, it can be shown that $q(x)=\bar{q}(x)=\mu_e I_{(n_e>0)}+\sum_{i\in\mathcal{Z}}\min\{d_i,n_i\}\mu_i+\sum_{i\in\mathcal{C}}n_i\mu_i$, where $I_{(.)}$ is an indicator function. This means that the jump-over network has a product-form stationary distribution of the form of Equation (2). $\square$

Theorem 1 provides a detailed description of the state of the network by specifying the order of totes in the nodes. However, knowledge of the aggregate state, i.e., the total number of totes in a node, is sufficient to determine performance statistics as the throughput and waiting times in the zones. For the description of the aggregate state, it is convenient to transform the class dependent visit ratios $\lambda_{i\boldsymbol{r}}$ into chain visit ratios

$$V_i=\frac{\sum_{\boldsymbol{r}\subseteq\mathcal{Z}}\lambda_{i\boldsymbol{r}}}{\sum_{\boldsymbol{r}\subseteq\mathcal{Z}}\lambda_{e\boldsymbol{r}}}, \qquad i\in\mathcal{S}, \tag{10}$$

where V_i can be interpreted as the average number of times an arbitrary tote visits node i before moving to the exit node. Note that the jump-over network only has one chain of classes, with a population of N totes, due to the fact that every tote will be replaced by a tote of a possibly different class at the exit.

Corollary 1. *The jump-over network with aggregated state space $\bar{n} = (n_i : i \in \mathcal{S})$ where n_i is the number of totes in node i with $\sum_{i \in \mathcal{S}} n_i = N$ and $n_i \leq d_i + q_i$ for $i \in \mathcal{Z}$ has a product-form stationary distribution of the form*

$$\pi(\bar{n}) = \frac{1}{G} \prod_{i \in \mathcal{S}} \left(\frac{V_i}{\mu_i}\right)^{n_i} \prod_{i \in \mathcal{C}} \frac{1}{n_i!} \prod_{i \in \mathcal{Z}} \frac{1}{\gamma(n_i)}, \tag{11}$$

where G is a normalization constant.

Remark 1. The definition of quasi-reversibility by Kelly (1979) was further generalized by e.g. Chao and Miyazawa (2000), Henderson and Taylor (2001). They show that quasi-reversibility can also be applied to obtain product form results for queueing networks with signals, negative customers, transitions involving three or more nodes, and batch movements. This framework can also be used to describe the jump-over network.

3.3. Chain Visit Ratios of the Jump-over Network

To obtain the chain visit ratios V_i, $i \in \mathcal{S}$, first linear system (1) must be solved. This might, however, require a large computational effort if the queueing network consists of many nodes. Also, the number of different tote classes, 2^M, grows exponentially with the number of zones. Another way to obtain the chain visit ratios V_i is to calculate them directly per node type, i.e., entrance/exit, conveyor node or zone. Clearly, V_e is 1 by (10). Then the chain visit ratios of the conveyors nodes and the zones can be calculated as follows.

3.3.1. Conveyor Nodes. A tote visits all the conveyor nodes the same number of times during its stay in the system. As a result, the chain visit ratios of the conveyor nodes V_i, $i \in \mathcal{C}$ are equal and given by the average number of circulations of an arbitrary tote in the system before moving to the exit node.

To calculate the average number of circulations, an absorbing Markov chain $\{X_l, l \geq 0\}$ with a state space consisting of all subsets of $\mathcal{Z}$ and transition matrix Φ is defined. The chain starts in state $X_0 = \boldsymbol{r}$ with probability $\psi_{\boldsymbol{r}}$, which is the probability that a tote of class $\boldsymbol{r}$ is released in the system. State X_l defines the class of the tote at the last conveyor node after the lth circulation. Then the average number of circulations is equal to the expected number of transitions before entering the absorbing state $\emptyset$, i.e., the class for which all the order lines have been picked and that the tote should leave the system. During a circulation through the system, the transition probability from state $\boldsymbol{r}$ at the start of the circulation to $\boldsymbol{s}$ at the end of the circulation, $\Phi_{\boldsymbol{r},\boldsymbol{s}}$, is given by

$$\Phi_{\boldsymbol{r},\boldsymbol{s}} = \prod_{j \in \boldsymbol{s}} b_j \cdot \prod_{i \in \boldsymbol{r} \setminus \boldsymbol{s}} (1 - b_i), \quad \boldsymbol{s} \subseteq \boldsymbol{r} \subseteq \mathcal{Z},$$

and zero otherwise.

The Markov chain $\{X_l, l \geq 0\}$ has one absorbing state, and transitions are only possible to a state with fewer zones to visit, i.e., the states of transition matrix Φ can be ordered in such a way that Φ is an upper triangular matrix. This implies that transition matrix Φ can be rewritten in canonical form as

$$\Phi = \begin{bmatrix} \Theta & \Upsilon \\ \mathbf{0} & 1 \end{bmatrix}, \tag{12}$$

where Θ is an upper triangular sub-matrix of the transition probabilities between the transient states, and Υ is a column vector of the transition probabilities between the transient states and the absorbing state. The last row of Φ corresponds to the absorbing state of the Markov chain. The expected number of transitions until absorption in a Markov chain with one absorbing state is given by (Wolff 1989)

$$V_i = \psi (I - \Theta)^{-1} \mathbf{1}, \quad i \in \mathcal{C}, \tag{13}$$

where I is an identity matrix, $\mathbf{1}$ a column vector with ones, and $\psi = (\psi_{\boldsymbol{r}} : \boldsymbol{r} \subseteq \mathcal{Z} \setminus \emptyset)$ a row vector with the initial release probabilities ordered in the same way as Θ. Since $(I - \Theta)$ is an upper triangular matrix, its inverse can easily be determined by back-substitution. Denote $\omega = (I - \Theta)^{-1}\mathbf{1}$, then the jth element of ω can be found by the following recursion

$$\omega_j = (1 + \sum_{k=j+1}^{2^M-1} \Theta_{j,k}\omega_k) / (1 - \Theta_{j,j}), \quad j = 2^M - 1, 2^M - 2, \ldots, 1.$$

REMARK 2. When the location of the system entrance and exit do not coincide, the chain visit ratios of the conveyor nodes are not all equal. Now, a tote visits the conveyor nodes from the system entrance to the system exit a single time more often than the conveyor nodes going the opposite way. V_i, $i \in \mathcal{C}$ can still be found using the previous analysis, except that the Markov chain will now start in state $X_0 = \boldsymbol{r}'$ with corresponding probability $\tilde{\psi}_{\boldsymbol{r}'}$, where $\boldsymbol{r}' \subseteq \mathcal{Z}$ is the class of the tote when reaching the system exit for the first time. Then, Equation (13) will give the average number of recirculations in the system, which equals the chain visit ratios of the conveyor nodes from the system exit to the system entrance. These should be incremented with 1 for the other conveyor nodes.

3.3.2. Zones. The chain visit ratios of the zones V_i, $i \in \mathcal{Z}$ are equal to the mean number of times an arbitrary tote visits zone i before leaving the system. In the jump-over network, the number of times the tote visits zone i follows a geometric distribution with a probability of actually being tagged as *visited zone* i equal to $1 - b_i$. Hence, the chain visit ratios of zone i are

$$V_i = \sum_{\boldsymbol{r}:\, i \in \boldsymbol{r} \subseteq \mathcal{Z}} \psi_{\boldsymbol{r}} \cdot \frac{1}{1 - b_i}, \quad i \in \mathcal{Z}. \tag{14}$$

3.4. Mean Value Analysis of the Jump-over Network

A mean value analysis (MVA) algorithm (Reiser and Lavenberg 1980) can be formulated that efficiently computes exactly the key performance statistics of the jump-over network by iterating over the total number of totes $n = 1, \ldots, N$ in the system. MVA is based on the arrival theorem which can be shown to hold also for multi-class jump-over networks by exploiting their product-form distribution. The algorithm iteratively calculates the mean throughput time $E(T_i(n))$, which is the expected time a tote will spend in node i given that there are n totes in the system, the system throughput $X(n)$, the mean number of totes in a node $E(L_i(n))$, and the marginal queue length probabilities $\pi_i(j|n)$ of having j totes in zone i given that there are n totes in the network.

First, initialize $E(L_i(0))=0$, $i\in\mathcal{S}$ and $\pi_i(0|0)=1$, $\pi_i(j|0)=0$ for $j=1,\dots,d_i+q_i$ if $i\in\mathcal{Z}$. Then, the mean throughput time $E(T_i(n))$ of the entrance/exit and conveyor nodes can be calculated by

$$E(T_i(n))=\begin{cases}\frac{1}{\mu_i}\left(1+E(L_i(n-1))\right), & \text{if } i=e,\\ \frac{1}{\mu_i}, & \text{if } i\in\mathcal{C}.\end{cases} \tag{15}$$

This directly follows from the arrival theorem and the fact that the entrance/exit is a single-server node and the conveyor nodes are infinite servers nodes.

The mean throughput time of the zones can be calculated by

$$E(T_i(n))=\sum_{j=d_i}^{d_i+q_i-1}(j+1-d_i)\frac{1}{d_i\mu_i}\cdot\pi_i(j|n-1)+\frac{1}{\mu_i}\left(1-\pi_i(d_i+q_i|n-1)\right),\quad i\in\mathcal{Z}. \tag{16}$$

The first term of Equation (16) denotes the average waiting time conditioned on the number of totes j in the zone on arrival, and the second term is the tote's own average service time. When the buffer of the zone is full, the throughput time is 0, since the tote skips the zone.

The system throughput $X(n)$ can be calculated using $E(T_i(n))$, $i\in\mathcal{S}$, (Reiser and Lavenberg 1980)

$$X(n)=\frac{n}{\sum_{i\in\mathcal{S}}V_iE(T_i(n))}. \tag{17}$$

where the denominator denotes the average time a tote spends in the system or the system throughput time.

Applying Little's law gives the mean number of totes in a node

$$E(L_i(n))=V_iX(n)E(T_i(n)),\quad i\in\mathcal{S}. \tag{18}$$

Finally, the marginal queue length probabilities can be determined by balancing the number of transitions per time unit between state $j-1$ and j, where j is the number of totes in zone i. The rate from j to $j-1$ is given by $\min(j,d_i)\mu_i\pi_i(j|n)$ and, by the arrival theorem, the rate from $j-1$ to j is $V_iX(n)\pi_i(j-1|n-1)$. Hence,

$$\pi_i(j|n)=\frac{V_iX(n)}{\mu_i\cdot\min(j,d_i)}\pi_i(j-1|n-1),\quad j=1,\dots,d_i+q_i,\ i\in\mathcal{Z}, \tag{19}$$

and where $\pi_i(0|n)$ follows from normalization

$$\pi_i(0|n) = 1 - \sum_{j=1}^{d_i+q_i} \pi_i(j|n), \quad i \in \mathcal{Z}. \tag{20}$$

Equation (20) has often been reported as the cause of numerical instability in MVA (Chandy and Sauer 1980). An alternative approach is to use Equation 20 of Reiser (1981) which is known to be numerically stable.

Sequentially applying Equations (15)-(20) allows for an iterative procedure for obtaining the performance statistics. In the last step of the MVA, the system throughput time, the fraction of time zone i is blocked $b_i = \pi_i(d_i + q_i|N-1)$, and the utilization of the nodes ρ_i can be obtained. The utilization per node is calculated by

$$\rho_i = \begin{cases} X(N)/\mu_i, & \text{if } i = e, \\ V_i X(N)/\mu_i, & \text{if } i \in \mathcal{C}, \\ 1 - \sum_{j=0}^{d_i-1} \frac{d_i-j}{d_i} \pi_i(j|N), & \text{if } i \in \mathcal{Z}, \end{cases} \tag{21}$$

where ρ_e and ρ_i $i \in \mathcal{Z}$, are the fraction of time the entrance/exit or the zones are busy and where ρ_i, $i \in \mathcal{C}$ is the average number of totes at a conveyor node.

Remark 3. In a jump-over network with non-exponential picking times, MVA will no longer be exact. Still, closed queueing networks are known to be robust to the service distribution of a node (Bolch et al. 2006). Hence, the arrival theorem can be adopted as an approximation. When the picking times are non-exponentially distributed, the throughput time of a zone is equal to the tote's own service time if not all the order pickers are busy. If all the order pickers are busy, then a newly arriving tote has to wait for the first departure at the zone and then continues to wait for as many departures as there were totes waiting on arrival before it is served. The throughput time of a tote that encounters the zone full is still zero. Combining these results, Equation (16) is replaced in the approximative MVA with,

$$\begin{aligned} E(T_i(n)) = Q_i(n-1) \cdot \frac{E(R_i)}{d_i} + \sum_{j=d_i}^{d_i+q_i-1} (j+1-d_i) \frac{E(B_i)}{d_i} \cdot \pi_i(j|n-1) \\ + E(B_i)(1 - \pi_i(d_i+q_i|n-1)), \quad i \in \mathcal{Z}, \end{aligned} \tag{22}$$

where $E(B_i)$ is the expected service time of zone i, $E(R_i) = E(B_i^2)/2E(B_i)$ is the expected residual service time of zone i, and $Q_i(n-1) = \sum_{j=d_i}^{d_i+q_i} \pi_i(j|n-1)$ is the probability that all order pickers are busy in zone i upon an arrival instant. When the picking times are generally distributed, $\pi_i(j|n)$ can be approximated by the corresponding probabilities in a zone where each order picker has an exponential service rate $1/E(B_i)$.

3.5. Iterative Algorithm for Calculating Blocking Probabilities of the Jump-over Network

In the jump-over network, totes are tagged independently from the state of the buffer using blocking probabilities b_i, $i \in \mathcal{Z}$. However, these blocking probabilities are not known in advance. The probability b_i can be iteratively estimated by the probability that the buffer of zone i is full using the following algorithm.

First, initialize the blocking probabilities $b_i^{(1)}$, $i \in \mathcal{Z}$ to 0. Then, calculate the marginal queue length probabilities using Equations (19) and (20) and take the fraction of totes that find the zone containing $d_i + q_i$ totes as a new estimate for the blocking probability. Thus, by the arrival theorem,

$$b_i^{(m+1)} = \pi_i^{(m)}(d_i + q_i|N-1), \quad i \in \mathcal{Z}, \tag{23}$$

where the superscript indicates in which iteration the quantities have been calculated. Based on this new estimate for b_i, the routing probabilities and subsequently the chain visit ratios are updated for the zones and conveyor nodes. Equation (23) can be evaluated again after applying MVA in order to get a better estimate of the blocking probabilities and so on. By repeating this process until for all i the differences between $b_i^{(m+1)}$ and $b_i^{(m)}$ is less than a small ϵ, the algorithm terminates and the performance statistics are calculated. In our experience, convergence is reached fast, and does not depend on the initial starting values of b_i.

3.6. An Example of the Single-segment Routing Model

In order to illustrate the performance and accuracy of the iterative algorithm of Subsection 3.5, consider the zone picking system with two zones shown in Figure 2. In total there are 2^2 different

Table 1 The performance statistics obtained for the example with varying number of totes N.

N	$X(N)$ (in hours^{-1})							$E(T_{\mathcal{Z}}(N))$ (in seconds)						
	Sim	Jump	CQN	YdK	Err (%)			Sim	Jump	CQN	YdK	Err (%)		
					Jump	CQN	YdK					Jump	CQN	YdK
10	104.4 (±0.16)	104.5	108.2	107.9	0.13	3.62	3.39	25.2 (±0.06)	25.3	27.1	28.6	0.12	7.27	13.15
20	182.8 (±0.25)	182.9	206.5	204.9	0.04	12.93	12.06	29.9 (±0.08)	29.8	41.8	46.4	0.28	40.00	55.29
30	234.3 (±0.52)	235.3	283.5	276.2	0.44	21.03	17.92	33.3 (±0.13)	33.1	72.8	86.0	0.58	119.07	158.47
40	268.8 (±0.58)	269.8	326.2	313.4	0.40	21.39	16.61	35.5 (±0.08)	35.5	132.4	154.5	0.02	273.31	335.67
50	291.5 (±0.39)	293.0	342.0	330.1	0.52	17.34	13.24	37.3 (±0.08)	37.3	216.8	240.4	0.06	481.65	544.82
100	336.4 (±0.65)	338.6	354.9	350.1	0.67	5.50	4.07	42.4 (±0.22)	42.3	704.6	723.4	0.17	1563.16	1607.49

N	$E(T_{\mathcal{C}}(N))$ (in seconds)							b_{z_1}			b_{z_2}		
	Sim	Jump	CQN	YdK	Err (%)			Sim	Jump	Err (%)	Sim	Jump	Err (%)
					Jump	CQN	YdK						
10	313.7 (±0.52)	313.4	300.0	300.0	0.08	4.37	4.37	0.01 (±0.00)	0.01	1.97	0.05 (±0.00)	0.05	0.62
20	357.4 (±0.54)	357.3	300.0	300.0	0.04	16.07	16.07	0.07 (±0.00)	0.07	1.92	0.18 (±0.00)	0.18	0.06
30	420.1 (±1.05)	418.7	300.0	300.0	0.34	28.59	28.59	0.16 (±0.00)	0.15	4.14	0.31 (±0.00)	0.31	0.27
40	491.9 (±0.96)	490.4	300.0	300.0	0.32	39.02	39.02	0.24 (±0.00)	0.23	3.40	0.41 (±0.00)	0.41	0.23
50	571.6 (±0.72)	568.8	300.0	300.0	0.49	47.51	47.51	0.32 (±0.00)	0.31	3.11	0.50 (±0.00)	0.50	0.41
100	1017.7 (±1.93)	1011.5	300.0	300.0	0.61	70.52	70.52	0.57 (±0.01)	0.55	4.41	0.72 (±0.00)	0.73	1.01

tote classes. The release probabilities are set to $\psi_{\{\emptyset\}} = 0$ and $\psi_{\{z_1\}} = \psi_{\{z_2\}} = \psi_{\{z_1,z_2\}} = 1/3$, and the mean service times are $\mu_e^{-1} = 5$ seconds for the entrance/exit, $\mu_{c_1}^{-1} = \mu_{c_2}^{-1} = \mu_{c_3}^{-1} = 100$ seconds for the conveyor nodes, and $\mu_{z_1}^{-1} = \mu_{z_2}^{-1} = 15$ seconds for the zones. The number of order pickers in both zones $d_{z_1} = d_{z_2}$ are equal to 1 and the buffer sizes of the zones are respectively $q_{z_1} = 2$ and $q_{z_2} = 1$.

Table 1 gives the average time in seconds a tote spends on the conveyor $E(T_{\mathcal{C}}(N))$ and at the zones $E(T_{\mathcal{Z}}(N))$, and the overall throughput rate per hour $X(N)$. These statistics are shown for the jump-over network (*Jump*), the same closed queueing network but with infinite buffers in the zones (*CQN*) and the approximation of Yu and De Koster (2008) (*YdK*). *YdK* uses an open queueing network in its analysis. Using bisection, the arrival rate of this approximation is set such that the average number of totes in the open network is equal to N. The results show that the jump-over network produces very accurate results compared to the simulation of the original queueing network (*Sim*), where the half width of the 95% confidence interval is given between brackets. In all cases, the algorithm stopped after 5 iterations with $\epsilon = 10^{-3}$. Both *CQN* and *YdK* assume infinite buffers, which means that they cannot estimate the blocking probabilities.

When the total number of totes in the system N is small, the errors of the jump-over network are negligible and relatively small for *CQN* and *YdK*. This is obvious since almost no blocking occurs in the system, e.g., only 5% of the totes that intends to visit the second zone are blocked. This means the jump-over network will have almost the same performance as *CQN* and the original queueing network. However, a higher N increases the probability of blocking and the number of times the totes have to recirculate.

When the number of totes in the system equals $N = 40$ or 50, blocking becomes more prominent. Since every zone is visited with the same frequency, the totes are more often blocked at zone 2 than at zone 1, due to the buffer sizes of the zones. Moreover, the system throughput time starts to increase rapidly, while the throughput rate stabilizes because all the zones become saturated. *CQN* and *YdK* produce large errors in the average time a tote spends at the zones and conveyor nodes, which is due to the assumption of infinite buffers in the zones. This does not happen in the jump-over network. Because of recirculation, the conveyor nodes act as buffers for totes that cannot enter a zone. When $N = 100$, blocking seriously impacts the performance of the system and totes spend twice as long in the system compared to $N = 50$.

4. Modeling Zone Picking Systems with Multi-segment Routing

In this section, the results of the single-segment routing model are extended to multi-segment routing. In a zone picking system with multi-segment routing, each segment consists of a number of zones connected by a conveyor with recirculation. The segments are connected to the main conveyor, which forms the center of the zone picking system. In order to analyze the system, again three different types of elements can be distinguished: the entrance/exit stations, the conveyors, and the zones. Furthermore, the entrance/exit stations are divided in the system entrance/exit and the segment entrance stations, and the conveyor nodes are split into main and segment conveyor nodes. An example of a zone picking system with multi-segment routing is shown in Figure 3, which has four segments that contain a different number of zones.

A zone picking system with multi-segment routing works very similarly to the system described in Section 2. Upon release at the system entrance, a tote is transported to the first segment where

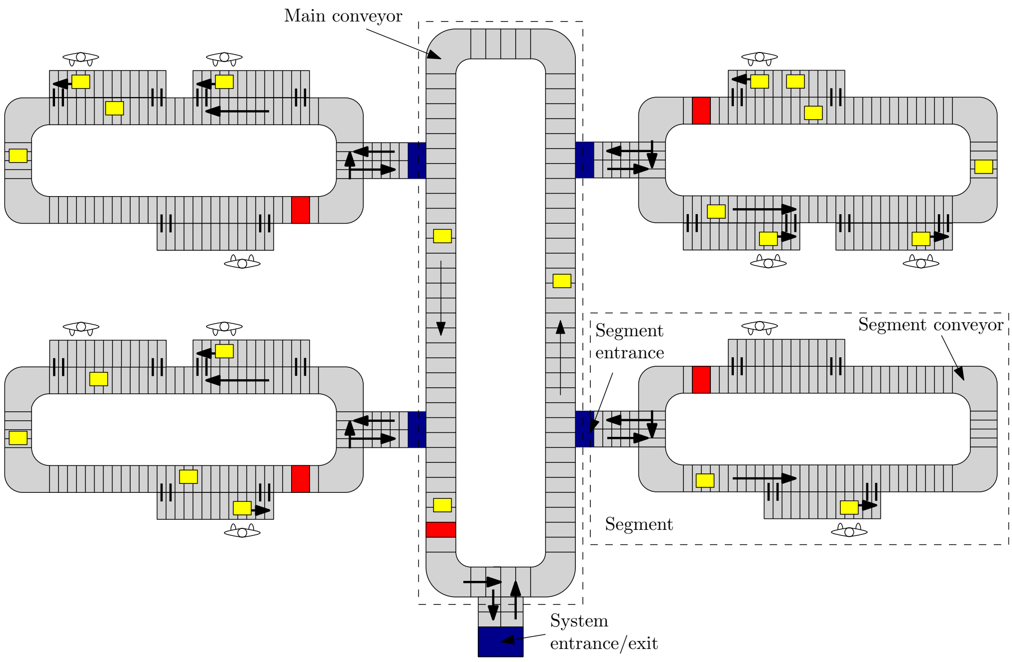


Figure 3 A zone picking system with multi-segment routing, four segments and eleven zones.

order lines have to be picked. The tote diverts to this segment via the segment entrance station and stays in the segment until it has visited all the required zones within the segment. When finished, the tote leaves the segment and is transported either to another segment or to the system exit in case the picking process has finished. When a segment is considered in isolation, it is equivalent to the single-segment model, except that the entrance station is now a delay station instead of a single-server node.

In a zone picking system with multi-segment routing, the workload control mechanism controls the maximum number of totes in the system and, in addition also, the maximum number of totes within each segment. If a tote tries to enter a segment that is fully saturated, the control mechanism withholds the tote from entering. The blocked tote will skip the segment and stay on the main conveyor, potentially visiting other segments before again attempting to enter this segment. This is very similar to the situation when a tote is blocked by a zone, but now blocking depends on the contents within an entire segment instead of a single zone.

The queueing network of Section 2 is extended to a zone picking system with multi-segment routing. Let the extended model consist of K segments. Denote the entrance/exit stations by

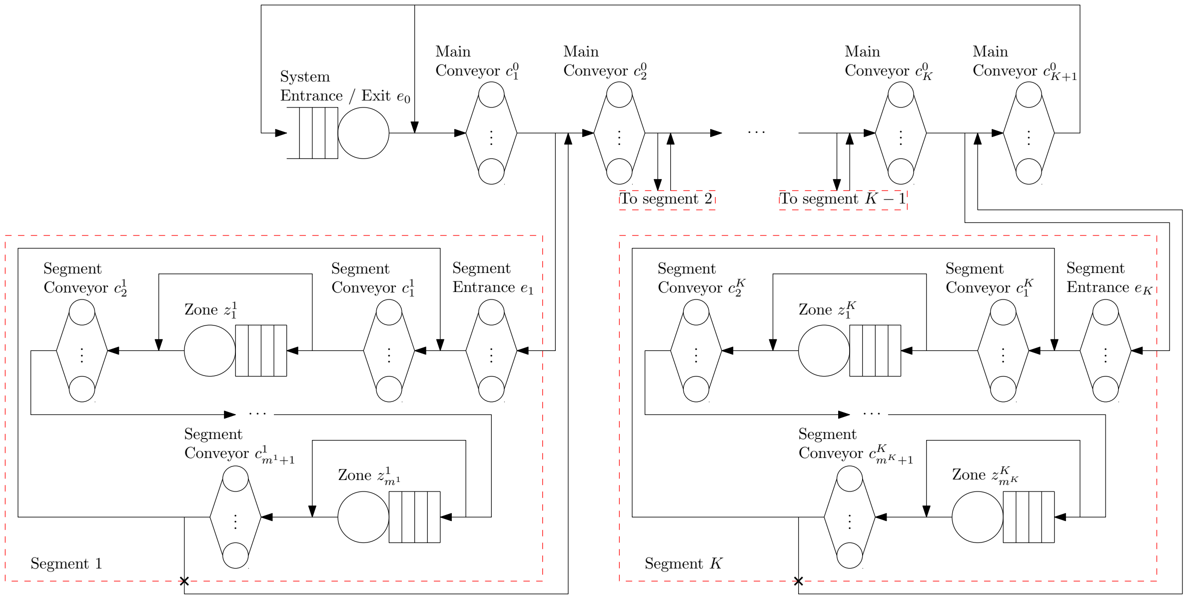


Figure 4 A multi-segment routing zone picking queueing network with K segments.

$\mathcal{E}=\{e_0,e_1,\dots e_k,\dots,e_K\}$ where e_0 is the system entrance/exit station and e_k the entrance station of segment k, that represents the conveyor connecting the segment with the main conveyor. Let $\mathcal{Z}=\cup_{k=1}^{K}\mathcal{Z}^k$ be the union of zones, where $\mathcal{Z}^k=\left\{z_1^k,\dots,z_{m^k}^k\right\}$ are the zones within segment k, such that $\sum_{k=1}^{K}m^k=M$. Denote $\mathcal{C}=\cup_{k=0}^{K}\mathcal{C}^k$ as the union of the conveyor nodes where $\mathcal{C}^0=\left\{c_1^0,\dots,c_{K+1}^0\right\}$ are the main conveyor nodes and $\mathcal{C}^k=\left\{c_1^k,\dots,c_{m^k+1}^k\right\}$ the segment conveyor nodes within segment k. Finally, let $\mathcal{S}=\mathcal{E}\cup\mathcal{C}\cup\mathcal{Z}$ be the union of all the nodes in the network. Figure 4 shows the topology of the multi-segment routing queueing network with K segments.

In addition, the system is partitioned into $K+1$ different subsystems: $\left\{\mathcal{H}^0,\mathcal{H}^1,\dots,\mathcal{H}^k,\dots,\mathcal{H}^K\right\}$, where $\mathcal{H}^0=\{e_0\}\cup\mathcal{C}^0$ consists of the system entrance/exit and the nodes on the main conveyor, and $\mathcal{H}^k=\{e_k\}\cup\mathcal{C}^k\cup\mathcal{Z}^k$ the set of nodes belonging to the kth segment. The following additional assumptions are adopted for the network:

- Each tote has a class $\boldsymbol{r}\subseteq\mathcal{Z}$ defining the zones the tote should visit. Let $\boldsymbol{r}^k\subseteq\mathcal{Z}^k$, $k=1,\dots,K$ describe the zones a class $\boldsymbol{r}$ tote has to visit within segment k. A tote will enter segment k if and only if $\boldsymbol{r}^k\neq\emptyset$.
- The entrance station e_k to segment k is assumed to be a delay station with a fixed delay of rate μ_{e_k}, $k=1,\dots,K$ that accounts for the time the tote needs for entering and leaving the segment.

- The maximum number of totes allowed in segment k, $k = 1, \ldots, K$ is N^k, which is controlled by the workload control mechanism.

At the system entrance, new totes of class $\boldsymbol{r} \subseteq \mathcal{Z}$ are released with probability $\psi_{\boldsymbol{r}}$. After release, a tote of class $\boldsymbol{r}$ moves from the system entrance to the first main conveyor node c_1^0. From c_k^0, the tote will either divert to segment entrance e_k if $\boldsymbol{r}^k \neq \emptyset$ or move to the next main conveyor node c_{k+1}^0. Whenever the number of totes in segment k equals N^k, the tote skips the segment and also moves to c_{k+1}^0, while its class remains the same. In case the tote actually enters the segment, it resides in the segment until it has visited all the required zones. Then the tote leaves the segment via $c_{m^k+1}^k$ and its class has changed from $\boldsymbol{r}$ to $\boldsymbol{s} = \boldsymbol{r} \backslash \boldsymbol{r}^k$. After visiting the last main conveyor node c_{K+1}^0, all totes with $\boldsymbol{r} \neq \emptyset$ are routed to the first main conveyor node c_1^0; the other totes are transported to the exit and are immediately replaced by a new tote which will wait for release at the entrance.

5. Analysis of Zone Picking Systems with Multi-segment Routing

Just as before, the first step of the analysis is to approximate the multi-segment queueing network of Section 4 by a network with jump-over blocking in Subsection 5.1 and 5.2. In Subsection 5.3 it is shown that the visit ratios of the multi-segment jump-over network again have closed form expressions. The performance statistics of this jump-over network are calculated in Subsection 5.4 using flow equivalent servers (Chandy et al. 1975) and MVA. The iterative algorithm for estimating the blocking probabilities is presented in Subsection 5.5.

5.1. Jump-over Network

In the multi-segment queueing network, totes can be blocked either by a zone or by a segment. A similar approach as the one described in Section 3 can be used to approximate segment blocking. In the real system totes of class $\boldsymbol{r}$ that have visit segment k are “tagged” after segment k with the labels *visited segment k* or *skipped segment k* depending on whether they received service or skipped the segment. This is now approximated by tagging a tote randomly and independently of whether the tote actually visited segment k or not. This approximation renders a jump-over network, where a tote skipping segment k will immediately move from the start to the end of the segment, which is indicated by a cross in Figure 4. Here, the skipping tote will act as a "regular" tote that actually visited the segment so $\boldsymbol{r}^k$ is set to $\emptyset$.

When a tote is tagged as *skipped segment k,* the class of the tote should revert to its class when it entered the segment. However, for all totes leaving the segment there is no knowledge about which zones the tote visited in the segment. Therefore, in the jump-over network, the classes are extended such that they also include the initial class of the tote when it entered the system. Denote the new classes by $\bar{\boldsymbol{r}} = \{\boldsymbol{h}, \boldsymbol{r}\}$, where $\boldsymbol{h} \subseteq \mathcal{Z}$ is the initial class of the tote and $\boldsymbol{r} \subseteq \mathcal{Z}$ the current set of zones the tote still needs to visit. The initial class $\boldsymbol{h}$ only changes when the tote is replaced by a new tote, whereas $\boldsymbol{r}$ changes to $\boldsymbol{s} = \boldsymbol{r} \setminus \left\{z_i^k\right\}$ if zone i in segment k is tagged as visited.

Let blocking probability B_k, $k = 1, \ldots, K$ be the fraction of totes that receive the *skipped segment k* tag in the real system. Then for each class $\bar{\boldsymbol{s}} = \{\boldsymbol{h}, \boldsymbol{s}\}$ tote leaving segment k, i.e., when $\boldsymbol{s}^k = \emptyset$, independent of whether the tote visited segment k or not (because of a fully saturated segment), $p_{c^k_{m^k+1}\bar{\boldsymbol{s}}, c^0_{k+1}\bar{\boldsymbol{r}}} = B_k$, $k = 1, \ldots K$, where $\bar{\boldsymbol{r}} = \left\{\boldsymbol{h}, \boldsymbol{s} \cup \boldsymbol{h}^k\right\}$ and $\boldsymbol{h}^k$ are the zones the tote was required to visit in segment k. This means that a tote is tagged as *skipped segment k* and is routed to next main conveyor node c^0_{k+1} leaving segment k with the same class $\bar{\boldsymbol{r}}$ as it entered the segment. Otherwise, the tote is tagged as *visited segment k* and the class of the tote does not change, i.e, $p_{c^k_{m^k+1}\bar{\boldsymbol{s}}, c^0_{k+1}\bar{\boldsymbol{s}}} = 1 - B_k$, $k = 1, \ldots K$. Just as in Subsection 3.1, the blocking probabilities B_k are not known in advance and need to be estimated, as is shown in Subsection 5.5.

5.2. Product-form of the Stationary Distribution of the Jump-over Network

In case of the multi-segment jump-over network, the of the network x is defined in the same way as in Subsection 3.2. Let $\bar{\mathbb{S}}(N)$ be the state space of the network where in each state x the number of totes in the system is N and where the number of totes in each zone and segment satisfy both $n_i \leq d_i + q_i$, $i \in \mathcal{Z}$ and $\sum_{i \in \mathcal{H}^k} n_i \leq N^k, k = 1, \ldots, K$ respectively. The existence of a product-form solution in the network can again be proven using conditions (4) and (5).

THEOREM 2. *The jump-over network with state space $\bar{\mathbb{S}}(N)$ has a product-form stationary distribution of the form*

$$\pi(x) = \frac{1}{G}\prod_{i\in\mathcal{S}} \pi_i(x_i), \tag{24}$$

where G is a normalization constant, $\pi_i(x_i)$ is the stationary distribution of node i, $i\in\mathcal{S}$ given by (3)*, where $\lambda_{i\boldsymbol{r}_{il}}$ is replaced by $\lambda_{i\bar{\boldsymbol{r}}_{il}}$.*

Proof: Whenever state x does not contain a blocked segment, it was shown in Theorem 1 that conditions (4) and (5) hold. In case state x contains a blocked segment and a transition that involves a tote skipping a segment occurs, the transition rate $q(x,y)$ is given as follows. A tote of class $\bar{\boldsymbol{r}}$ departs from the lth position of main conveyor node c_k^0 and it immediately jumps over the entire segment. The tote will move to the next main conveyor node c_{k+1}^0 with probability 1, where it joins the end of the node. When arriving in c_{k+1}^0 the tote is tagged with either as *visited segment k* or *skipped segment k*. Hence,

$$q\left(x, x - \bar{\boldsymbol{r}}_{c_k^0 l} + \bar{\boldsymbol{r}}_{c_{k+1}^0 n_{c_{k+1}^0}+1}\right) = \mu_{c_k^0} B_k, \qquad l = 1,\ldots,n_{c_k^0}, \tag{25}$$

$$q\left(x, x - \bar{\boldsymbol{r}}_{c_k^0 l} + \bar{\boldsymbol{s}}_{c_{k+1}^0 n_{c_{k+1}^0}+1}\right) = \mu_{c_k^0}(1-B_k), \quad l = 1,\ldots,n_{c_k^0}. \tag{26}$$

The time-reversed transition rates $\bar{q}(y,x)$ are analogous to (8) and (9). Then it can be verified, similarly as in Theorem 1, that conditions (4) and (5) also hold and the jump-over network has a product-form stationary distribution of the form of Equation (24). □

Theorem 2 again provides a detailed description of the state of the jump-over network. Since performance statistics as the throughput and mean waiting times in the zones are of interest, the aggregated state of the network will suffice and can be obtained similar as in Corollary 1.

5.3. Chain Visit Ratio of the Jump-over Network

The chain visit ratios of the jump-over network can be computed directly per node type similar as in Section 3.3. Let the chain visit ratio of the system entrance/exit be normalized to $V_{e_0} = 1$. Next, the chain visit ratios are derived in the following order for the main conveyor nodes, segment entrances, segment conveyor nodes, and the zones.

5.3.1. Main Conveyor Nodes and Segment Entrances. The chain visit ratios of the main conveyor nodes and the segment entrances can be computed as the conveyor nodes and the zones in the single-segment routing model. This follows from the fact that a tote needs to visit all the main conveyor nodes the same number of times during its stay in the network, and the number of visits to the segment entrances depends on the number of times a tote intends to visit a segment. The difference is now that the visit ratios depend on the blocking probabilities of the segments B_k, instead of those of the zones. The chain visit ratios of the main conveyor nodes V_i, $i \in \mathcal{C}^0$ are given by Subsection 3.3.1 and the segment entrances V_i, $i \in \mathcal{E} \setminus \{e_0\}$ by Subsection 3.3.2. For both, the tote classes $\boldsymbol{r} \subseteq \mathcal{Z}$ are replaced by the aggregated segment classes $\boldsymbol{k} \subseteq \{1, \dots, K\}$ that define the segments a tote should visit. The corresponding release probabilities of the segment classes $\hat{\psi}_{\boldsymbol{k}}$ can be obtained by summing over all the class specific release probabilities of totes that need to visit the segments contained in $\boldsymbol{k}$. Finally, the blocking probabilities of the zones b_j are replaced by the blocking probabilities of the segments B_k.

5.3.2. Segment Conveyor Nodes and Zones. Within a segment, the network behaves exactly the same as in the single-segment routing model. This means that the chain visit ratios of the nodes in $\mathcal{H}^k$, only depend on the blocking probabilities of the zones in $\mathcal{Z}^k$. The chain visit ratios of the segment conveyor nodes $V_i, i \in C^k$ are given by Subsection 3.3.1 where the tote classes are now replaced by $\boldsymbol{r}^k \subseteq \mathcal{Z}^k$. The corresponding release probabilities for segment k are $\psi^k = \left(\psi^k_{\boldsymbol{r}^k} : \boldsymbol{r}^k \subseteq \mathcal{Z}^k \backslash \emptyset\right)$, where $\psi^k_{\boldsymbol{r}^k} = \sum_{\boldsymbol{s} \subseteq \mathcal{Z}} \psi_{\boldsymbol{s}} \cdot I_{\left(\boldsymbol{s}^k = \boldsymbol{r}^k\right)} / \sum_{\boldsymbol{s} \subseteq \mathcal{Z}} \psi_{\boldsymbol{s}} \cdot I_{\left(\boldsymbol{s}^k \neq \emptyset\right)}$ is the normalized sum of all tote classes that need to visit the zones $\boldsymbol{r}^k \subseteq \mathcal{Z}^k$ and $I_{(.)}$ an indicator function. Then by calculating the expected number of transitions until entering the absorbing state, i.e., when the tote has to leave the segment, $V_i, i \in C^k$ is obtained by multiplying the number of times the tote intends to visit the segment V_{e_k} with the average number of circulations a tote makes within segment k

$$V_i = V_{e_k} \cdot \psi^k \left(I - \Theta^k\right)^{-1} \mathbf{1}, \quad i \in \mathcal{C}^k, k = 1, \dots, K, \tag{27}$$

where Θ^k is defined similar as in (12).

A similar argument holds for the chain visit ratios of the zones V_i, $i \in \mathcal{Z}$ which are given by Subsection 3.3.2. Hence,

$$V_i = V_{e_k} \cdot \sum_{\boldsymbol{r}^k:\, i \in \boldsymbol{r}^k \subseteq \mathcal{Z}^k} \psi^k_{\boldsymbol{r}^k} \cdot \frac{1}{1-b_i}, \quad i \in \mathcal{Z}^k,\, k = 1, \ldots, K. \tag{28}$$

5.4. Aggregation Technique

In order to analyze the jump-over network, an extended version of the MVA presented in Subsection 3.4 can be formulated. However, it is more efficient to aggregate the jump-over network by replacing all segments by flow equivalent server centers with load-dependent service rates (Chandy et al. 1975). Norton's theorem states that the stationary distribution of the rest of the network remains unchanged after this modification (Chandy et al. 1975, Walrand 1983, Boucherie 1998).

Based on the product-form of Subsection 5.2 and the fact that each tote enters or leaves a segment via a single input/output node, an equivalent queueing network can be analyzed along the same lines as the analysis of the single-segment routing model.

The first step of the aggregation technique is to replace all segments by flow equivalent servers with load-dependent service rates. These rates can be determined by calculating the throughput $X^k(n)$ of subsystem $\mathcal{H}^k$ in isolation when varying the number of totes n from 1 up to N^k. The isolated subsystem can be obtained by short-circuiting all nodes that are not in $\mathcal{H}^k$. As a result, a tote leaving subsystem $\mathcal{H}^k$ will instantaneously be routed back to the entrance of the subsystem. Since every subsystem $\mathcal{H}^k$, $k = 1, \ldots, K$ in isolation is equivalent to the single-segment routing model, it can be analyzed using the MVA presented in Section 3.4, where the entrance station is now a delay station.

Next, an equivalent queueing network can be constructed by replacing each subsystem $\mathcal{H}^k$, $k = 1, \ldots, K$ in the jump-over network by a flow equivalent server center. The server rates of the kth flow equivalent server are equal to the throughputs $X^k(n)$ of the isolated subsystems, so

$$\mu_{FES_k}(n) = X^k(n), \quad n = 1, \ldots, N^k,\, k = 1, \ldots, K. \tag{29}$$

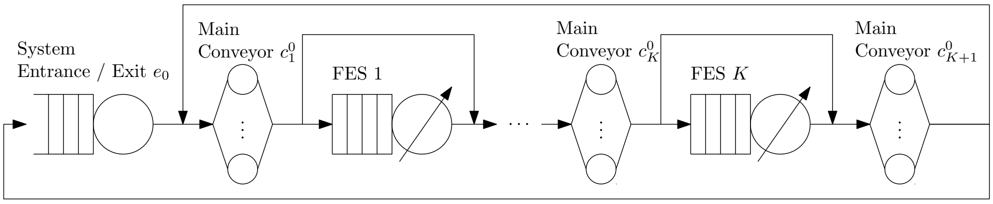


Figure 5 The equivalent network of the jump-over network. Segments are replaced by flow equivalent service centers with load-dependent service rates.

In Figure 5 the equivalent network of Figure 4 is shown, which is identical to Figure 2b except that the zones are replaced by flow equivalent service centers. The MVA of Section 3.4 can be applied to analyze the system, where (16) should be replaced by the mean throughput time of a tote in subsystem $\mathcal{H}^k$ (Reiser 1981),

$$E\left(T_{FES_k}(n)\right) = \sum_{j=1}^{N^k-1} \frac{j}{\mu_{FES_k}(j)} \cdot \Pi_{FES_k}(j-1|n-1), \quad k=1,\dots,K, \tag{30}$$

where the marginal queue length probabilities $\Pi_{FES_k}(j|n)$ of having j totes in the kth flow equivalent server in a network with n circulating totes. These can be obtained similar as in (19) by

$$\Pi_{FES_k}(j|n) = \frac{V_{FES_k} X(n)}{\mu_{FES_k}(j)} \Pi_{FES_k}(j-1|n-1), \quad j=1,\dots,N^k, k=1,\dots,K, \tag{31}$$

where V_{FES_k} is the visit ratio of the kth flow equivalent server, which is equal to the visit ratio of segment entrance e_k. Equation (30) is obtained by application of Little's law to the kth flow equivalent server and substitution of (31).

The performance statistics obtained from the equivalent (aggregate) network correspond with the aggregated performance statistics of the segments in the original jump-over network, e.g., the mean throughput time of subsystem $\mathcal{H}^k$, $E\left(T_{FES_k}(N)\right)$, and the marginal queue length probabilities $\Pi_{FES_k}(j|N)$ of having j totes in subsystem $\mathcal{H}^k$ when there are N totes in the system.

The detailed performance statistics of the nodes within the segments can be obtained by a disaggregation step. Let the marginal queue length probabilities of subsystem $\mathcal{H}^k$ analyzed in isolation be $\pi_i^k(j \mid n)$, where j is the number of totes in node i in segment k, given the number of totes in segment k is n. Then, the detailed marginal queue length probabilities $\pi_i(j|N)$ are given by (Baynat and Dallery 1993)

$$\pi_i(j|N) = \begin{cases} \Pi_i(j|N), & \text{if } i \in \mathcal{H}^0, \\ \sum_{l=1}^{N^k} \pi_i^k(j \mid l)\, \Pi_{FES_k}(l|N), & \text{if } i \in \mathcal{H}^k. \end{cases} \tag{32}$$

The performance statistics of all the nodes can now be calculated. The utilization of the system entrance (with $d_i = 1$) and zones can be calculated as

$$\rho_i = 1 - \sum_{j=0}^{d_i-1} \frac{d_i - j}{d_i} \pi_i(j|N), \quad i \in e_0 \cup \mathcal{Z}. \tag{33}$$

The mean number of totes in a node is given by

$$E(L_i(N)) = \sum_{j=1}^{\sigma_i} j \cdot \pi_i(j|N), \quad i \in \mathcal{S}, \tag{34}$$

where σ_i equals the number of totes in the system N if $i \in \mathcal{H}^0$, the segment capacity N^k if $i \in \mathcal{H}^k \backslash \mathcal{Z}^k$ and $d_i + q_i$ if $i \in \mathcal{Z}$.

Applying Little's law gives the mean throughput time in a node

$$E(T_i(N)) = E(L_i(N))/V_i X(N), \quad i \in \mathcal{S}, \tag{35}$$

where $X(N)$ is the overall throughput rate from the equivalent aggregate network.

5.5. Iterative Algorithm for Calculating the Blocking Probabilities of the Jump-over Network

As in Subsection 3.4, totes are tagged independently from the state of the network using the blocking probabilities; b_i, $i \in \mathcal{Z}$ and B_k, $k = 1, \ldots, K$. These blocking probabilities are not known in advance, but they can be iteratively estimated by the probabilities that the buffer of the zone is full or a segment is saturated.

First, blocking probabilities $b_i^{(1)}$, $i \in \mathcal{Z}$ and $B_k^{(1)}$, $k = 1, \ldots, K$ are initialized to 0. Then, calculate the marginal queue length probabilities of the equivalent network and use the fraction of arrivals that see a segment being saturated as a new estimate for the blocking probabilities of the segments, so by the arrival theorem,

$$B_k^{(m+1)} = \Pi_{FES_k}^{(m)}\left(N^k | N - 1\right), \quad k = 1, \ldots, K, \tag{36}$$

where the superscript denotes the iteration number. Using the detailed marginal queue length probabilities of Equation (32) and by calculating the fraction of "real" arrivals in segment k that encounter a full buffer in zone i, the new estimates for the blocking probabilities of the zones are given by

$$\begin{aligned} b_i^{(m+1)} &= \pi_i^{(m)}\left(d_i+q_i|N^k-1, N-1\right), \\ &= \sum_{l=1}^{N^k-1} \pi_i^k\left(d_i+q_i \mid l\right)\Pi_{FES_k}\left(l|N-1\right), \quad i \in \mathcal{Z}. \end{aligned} \tag{37}$$

which is the probability of encountering a full buffer in zone i in a network containing $N-1$ totes, where in segment k a maximum of N^k-1 totes is allowed. Note the remarkable feature that a tote arriving at a zone sees the network in equilibrium without itself *and* in which the capacity of the segment is reduced by 1.

With (36) and (37), the routing probabilities and subsequently the visit ratios are updated for all nodes in the network. Applying the MVA equations, (36) and (37) are updated to obtain better estimates. By repeating this process until for all i, the differences $B_i^{(m+1)}-B_i^{(m)}$ and $b_i^{(m+1)}-b_i^{(m)}$are less than a small ϵ, the algorithm terminates and the performance statistics are calculated.

6. Numerical Results

In order to investigate the performance of the (extended) jump-over network, the approximation is evaluated for a large test set and compared with the results of a discrete-event simulation of the real queueing network. This section is split into three parts. Subsections 6.1 and 6.2 discuss the accuracy of the approximation for a large test set for single-segment and for multi-segment routing systems. In Subsection 6.3, the quality of the approximation is investigated for the performance of a real zone picking system of a large wholesale company supplying non-food.

Both the jump-over network and the discrete-event simulation were implemented in Java. For each case, the simulation model was run 10 times for 1,000,000 seconds, preceded by 10,000 seconds of initialization for the system to become stable, which guaranteed that the 95% confidence interval width of the system throughput time is less than 1% of the mean value for all the cases.

Table 2 Parameters of the single-segment routing model test set.

(a) Balanced test set (9,600 cases)

Parameter	Value
Nr. of zones, M	1,2,3,4,5,6,7,8
Nr. of totes, N	10,20,30,40,50,60,70,80
Mean conveyor times, $\mu_i^{-1}, i \in \mathcal{C}$	20,30,40,50,60
Mean zone times, $\mu_i^{-1}, i \in \mathcal{Z}$	10,15,20,25,30
Nr. of order pickers, $d_i, i \in \mathcal{Z}$	1,2,3
Buffer size of a zone, $q_i, i \in \mathcal{Z}$	0,1

(b) Imbalanced test set (224 cases)

Parameter	Value
Nr. of zones, M	2,3,4,5,6,7,8
Nr. of totes, N	10,20,30,40,50,60,70,80
Mean zone times, $\mu_i^{-1}, i \in \mathcal{Z}$	1) 10,10,10,...
	2) 10,12,14,...
	3) 10,15,20,...
	4) 10,20,30,...

6.1. Numerical Results Single-segment Routing Models

In this section the performance of the approximation is investigated for the single-segment routing model. A test set was generated for which the parameters are listed in Table 2. The number of zones in the system M varied between 1 and 8 and the number of totes N between 10 and 80. Furthermore, it is first assumed that every zone and every conveyor node are identical to the other nodes of the same type and that all possible tote classes are released into the system with the same probability. This ensures that the work-load of all zones in the system is balanced. In the test set, the mean conveyor times, $\mu_i^{-1}, i \in \mathcal{C}$ are varied between 20 and 60 seconds and the mean zone times, $\mu_i^{-1}, i \in \mathcal{Z}$ between 10 and 30 seconds. The number of order pickers $d_i, i \in \mathcal{Z}$ and buffer places $q_i, i \in \mathcal{Z}$ varied between 1 and 3, and 0 and 1, respectively. In total, this leads to 9,600 different cases.

In addition, the effect of work-load imbalance among order pickers and zones is tested. Imbalance can be introduced by either changing the release probabilities or the parameters of a zone, e.g., the mean zone times. The latter approach was chosen for the imbalanced test set. In this test set, the mean conveyor times $\mu_i^{-1}, i \in \mathcal{C}$ are equal to 30 seconds and both the number of order pickers $d_i, i \in \mathcal{Z}$ and buffer places $q_i, i \in \mathcal{Z}$ are equal to 1. Four different scenarios were created for the mean zone times. In the first scenario, the mean zone times are equal, whereas in the other three scenarios they increase by either 2, 5, or 10 seconds per subsequent zone. This leads to an additional 224 cases.

Table 3 Results of the balanced test set with a varying number of zones M.

M	Error (%) in system throughput				Error (%) in nr. of circulations				Error (%) in throughput times zones			
	Avg	$0-1$	$1-5$	>5	Avg	$0-1$	$1-5$	>5	Avg	$0-1$	$1-5$	>5
1	0.08	100.0	0.0	0.0	0.08	100.0	0.0	0.0	0.09	100.0	0.0	0.0
2	0.67	70.0	29.8	0.2	0.78	69.0	29.8	1.3	0.44	83.9	16.1	0.0
3	0.78	68.2	31.7	0.2	0.94	67.2	30.3	2.5	0.44	86.2	13.8	0.0
4	0.73	71.9	27.8	0.3	0.90	71.3	25.9	2.8	0.38	90.3	9.8	0.0
5	0.64	76.6	23.3	0.2	0.80	75.0	22.4	2.6	0.32	93.2	6.8	0.0
6	0.54	80.4	19.5	0.1	0.68	78.6	18.9	2.5	0.28	94.9	5.1	0.0
7	0.45	83.8	16.2	0.0	0.57	82.4	15.8	1.8	0.25	96.9	3.1	0.0
8	0.38	86.7	13.3	0.0	0.48	85.2	13.5	1.3	0.23	97.7	2.3	0.0

Table 4 Results of the balanced test set with a varying number of totes in the system N.

N	Error (%) in system throughput				Error (%) in nr. of circulations				Error (%) in throughput times zones			
	Avg	$0-1$	$1-5$	>5	Avg	$0-1$	$1-5$	>5	Avg	$0-1$	$1-5$	>5
10	0.24	95.0	4.8	0.2	0.29	93.1	5.9	1.0	0.21	99.8	0.3	0.0
20	0.40	86.8	12.9	0.3	0.53	85.5	12.3	2.2	0.21	97.9	2.1	0.0
30	0.52	81.7	18.1	0.3	0.67	80.0	17.5	2.5	0.24	95.6	4.4	0.0
40	0.59	77.6	22.3	0.2	0.74	76.7	20.8	2.6	0.28	93.0	7.0	0.0
50	0.62	75.3	24.8	0.0	0.77	74.2	23.7	2.2	0.32	91.8	8.3	0.0
60	0.64	74.1	25.9	0.0	0.76	73.0	25.1	1.9	0.36	89.6	10.4	0.0
70	0.64	73.6	26.4	0.0	0.75	72.9	25.7	1.4	0.38	88.8	11.3	0.0
80	0.64	73.6	26.4	0.0	0.72	73.3	25.7	1.0	0.41	86.7	13.3	0.0

The results of the balanced test set are summarized in Tables 3, 4, 5, and 6. Each table lists the average of the relative error between the approximation and the simulation for the system throughput in hour^{-1}, the average number of circulations a tote makes in the system before moving to the exit, and the mean of the sum of throughput times of the zones. Each table also gives the percentage of cases that fall in three different error-ranges. From the tables it can be concluded that the approximation produces very accurate results for the three performance statistics. The overall average error in the system throughput is 0.54%, it is 0.65% for the mean number of circulations, and 0.30% for the average mean throughput times of the zones. Almost all of these errors fall between $0-1$%, with only a few larger than 5%.

Tables 3 and 4 show that the largest errors occur when the system has three or four zones and when the number of totes in the system is high. An explanation is that the blocking probabilities increase when the number of zones M decreases or if the number of totes in the system N increases.

Table 5 Results of the balanced test set with varying mean conveyor times $\mu_i^{-1}, i \in \mathcal{C}$.

$\mu_i^{-1}, i \in \mathcal{C}$	Error (%) in system throughput				Error (%) in nr. of circulations				Error (%) in throughput times zones			
	Avg	$0-1$	$1-5$	>5	Avg	$0-1$	$1-5$	>5	Avg	$0-1$	$1-5$	>5
20	0.91	66.5	33.0	0.5	1.21	65.3	28.1	6.7	0.47	85.4	14.6	0.0
30	0.64	74.1	25.9	0.0	0.79	72.7	25.2	2.2	0.33	91.0	9.0	0.0
40	0.47	81.4	18.6	0.0	0.55	80.3	19.4	0.4	0.27	94.4	5.6	0.0
50	0.36	86.3	13.7	0.0	0.41	85.5	14.5	0.0	0.23	96.3	3.8	0.0
60	0.29	90.2	9.8	0.0	0.31	89.2	10.8	0.0	0.21	97.3	2.7	0.0

Table 6 Results of the balanced test set with varying mean zone times $\mu_i^{-1}, i \in \mathcal{Z}$.

$\mu_i^{-1}, i \in \mathcal{Z}$	Error (%) in system throughput				Error (%) in nr. of circulations				Error (%) in throughput times zones			
	Avg	$0-1$	$1-5$	>5	Avg	$0-1$	$1-5$	>5	Avg	$0-1$	$1-5$	>5
10	0.24	93.0	7.0	0.0	0.23	92.7	7.3	0.0	0.17	98.2	1.8	0.0
15	0.40	84.9	15.1	0.0	0.46	83.5	16.1	0.4	0.23	95.5	4.5	0.0
20	0.55	78.4	21.6	0.0	0.67	76.8	21.7	1.5	0.30	92.7	7.3	0.0
25	0.68	72.9	27.0	0.2	0.86	71.8	25.3	2.9	0.37	90.1	9.9	0.0
30	0.80	69.3	30.3	0.4	1.05	68.1	27.3	4.5	0.43	87.8	12.2	0.0

Moreover, if blocking is prevalent, a higher M means the approximation needs to estimate more blocking probabilities, which creates more room for error. Eventually, M is high enough that blocking is almost fully absent for any N. The approximation becomes exact since the network will behave precisely as the original queueing network where totes are never blocked.

Tables 5 and 6 show that the largest errors occur with low mean conveyor times and high mean zone times. Here the product-form assumption that each node can be analyzed in isolation does not describe the real behavior sufficient. For example, if a tote is blocked by a zone, it can circulate through the whole system and eventually encounter the zone still working on the same tote. This will create dependencies between succesive visits to the nodes which are not captured by the approximation. However, this situation is very unlikely in practice. The total recirculation time is usually much higher than the time a tote will spend in a zone.

Table 7 presents the results of the imbalanced test set. The errors of the three performance statistics are slightly larger than those of the balanced test set. In particular, the errors increase when there is more imbalance between the zones. Totes that need to visit the slowest zone now spend more time in the system since the probability of being blocked is higher, which increases

Table 7 Results of the imbalanced test set with varying mean zone times $\mu_i^{-1}, i \in \mathcal{Z}$.

$\mu_i^{-1}, i \in \mathcal{Z}$	Error (%) in system throughput				Error (%) in nr. of circulations				Error (%) in throughput times zones			
	Avg	$0-1$	$1-5$	>5	Avg	$0-1$	$1-5$	>5	Avg	$0-1$	$1-5$	>5
10,10,10,...	0.20	100.0	0.0	0.0	0.20	100.0	0.0	0.0	0.15	100.0	0.0	0.0
10,12,14,...	0.23	100.0	0.0	0.0	0.24	100.0	0.0	0.0	0.16	100.0	0.0	0.0
10,15,20,...	0.35	98.2	1.8	0.0	0.36	94.6	5.4	0.0	0.21	100.0	0.0	0.0
10,20,30,...	0.40	89.3	10.7	0.0	0.45	85.7	14.3	0.0	0.32	100.0	0.0	0.0

Table 8 Parameters of the multi-segment routing model test set (1,320 cases).

Parameter	Value	Parameter	Value
Nr. of segments, K	2,3,4,5,6	Nr. of zones per segment, m_k	1) 9,9
Nr. of totes, N	10,20,30,40,50,60,70,80		2) 6,6,6
Mean main conveyor times, $\mu_i^{-1}, i \in \mathcal{C}^0$	10,20,30,40,50,60		3) 3,6,3,6
Segment capacity, N^k, $k = 1, \ldots, K$	10,15,20,25,30,35,40		4) 3,3,6,3,3
			5) 3,3,3,3,3,3

errors, as seen in the previous tables. Still, on average the errors for the three statistics are well below 1%.

6.2. Numerical Results Multi-segment Routing Models

For the multi-segment routing model a new test set is used, the parameters of which are listed in Table 8. In all test cases, the number of zones M equals 18, but the number of zones per segment m_k can vary between 3, 6, and 9. Furthermore, it is assumed that within every segment the zones and conveyor nodes are identical, i.e., $\mu_i^{-1} = 30, i \in \mathcal{C} \backslash \mathcal{C}^0$, $\mu_i^{-1} = 15, i \in \mathcal{Z}$, and $q_i = d_i = 1, i \in \mathcal{Z}$. The release probabilities $\psi_{\boldsymbol{r}}$ are assumed to be the same for all $\boldsymbol{r}$, and the service means of all entrances are equal to $\mu_i^{-1} = 5, i \in \mathcal{E}$. The number of totes in the system is varied between 10 and 80 and the capacities of the segments N^k between 10 and 40 totes as long as $N \geq N^k$. Finally, the mean main conveyor times, $\mu_i^{-1}, i \in \mathcal{C}^0$ varied between 10 and 60. This leads to 1,320 different test cases.

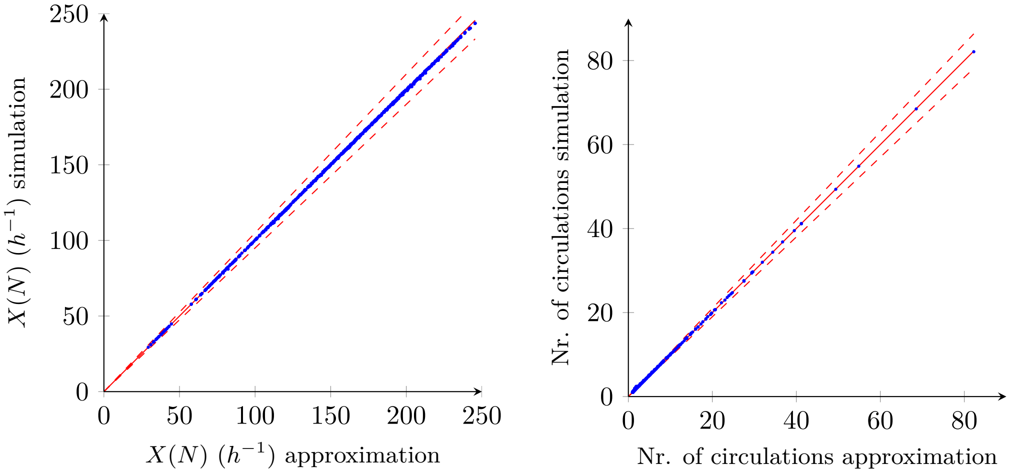
The results of the multi-segment test set are summarized in Tables 9 and 10. The overall average error in the system throughput is 0.21%, for the mean number of circulations on the main conveyor 0.93%, and for the average throughput times of the zones 0.24%. The tables show that the errors are the largest with a low number of segments and a fast main conveyor. In these cases, totes are

Table 9 Results multi-segment routing test set with a varying number of zones per segment m_k.

m_k	Error (%) in system throughput				Error (%) in nr. of circulations				Error (%) in throughput times zones			
	Avg	$0-1$	$1-5$	>5	Avg	$0-1$	$1-5$	>5	Avg	$0-1$	$1-5$	>5
9,9	0.21	99.6	0.4	0.0	2.27	66.3	20.1	13.6	0.23	100.0	0.0	0.0
6,6,6	0.25	100.0	0.0	0.0	1.23	72.0	21.6	6.4	0.23	100.0	0.0	0.0
6,3,6,3	0.19	100.0	0.0	0.0	0.43	84.8	15.2	0.0	0.24	100.0	0.0	0.0
3,3,6,3,3	0.16	100.0	0.0	0.0	0.28	89.8	10.2	0.0	0.25	100.0	0.0	0.0
3,3,3,3,3,3	0.23	99.2	0.8	0.0	0.43	86.7	12.5	0.8	0.23	100.0	0.0	0.0

Table 10 Results multi-segment routing test set with varying mean conveyor times $\mu_i^{-1}, i \in \mathcal{C}^0$.

$\mu_i^{-1}, i \in \mathcal{C}^0$	Error (%) in system throughput				Error (%) in nr. of circulations				Error (%) in throughput times zones			
	Avg	$0-1$	$1-5$	>5	Avg	$0-1$	$1-5$	>5	Avg	$0-1$	$1-5$	>5
10	0.28	99.1	0.9	0.0	2.23	66.8	22.3	10.9	0.23	100.0	0.0	0.0
20	0.24	100.0	0.0	0.0	1.15	75.5	18.6	5.9	0.23	100.0	0.0	0.0
30	0.21	100.0	0.0	0.0	0.77	77.3	18.2	4.5	0.24	100.0	0.0	0.0
40	0.19	99.5	0.5	0.0	0.58	83.6	14.5	1.8	0.24	100.0	0.0	0.0
50	0.18	100.0	0.0	0.0	0.45	86.8	11.8	1.4	0.24	100.0	0.0	0.0
60	0.16	100.0	0.0	0.0	0.37	89.5	10.0	0.5	0.24	100.0	0.0	0.0



(a) System throughput (b) Nr. of circulations on the main conveyor

Figure 6 Scatter plots of the approximation and simulated performance statistics of the multi-segment routing test set; a point between the dashed lines indicates that the relative error is less than 5%.

more likely to be blocked by a segment such that they need to recirculate on the main conveyor multiple times. As seen in the previous results, the errors increase when there is more blocking in the system. When varying the segment capacities N^k similar results are obtained.

Figures 6a-6b show scatter plots of the system throughput $X(N)$ and mean number of circulations on the main conveyor for the approximation and the simulation of the multi-segment routing model test set. On the x-axis the results of the approximation are given and on the y-axis those of the simulation. The approximation often overestimates the system throughput, since it typically gives conservative estimates of the blocking probabilities. This can be seen in Figure 6b where most of the points lay above the 45° line. This implies that totes spend less time in the network which increases the system throughput. This result is not related to the initial starting values of the blocking probabilities in the algorithm; different starting values produce the same results.

6.3. Validation Approximation for a Real Zone Picking System

In order to further validate the approximation, its results are compared to the performance of a real zone picking system at a large wholesaler supplying non-food items to supermarkets. The part of the warehouse dedicated to zone picking is divided into four interconnected segments. Table 11 lists a selection of the data used in this example. The first segment consists of three pallet picking zones, while the other three segments contain eight zones each and use pick-by-light. On a normal working day, the system deals with 220 totes simultaneously. On busy days, this number can increase to 280 before the conveyor becomes congested. Up to 95 totes can be present in each segment at the same time. By analyzing the log files from the Warehouse Management System (WMS) for several representative picking days, data about release probabilities and service times of the zones are obtained.

In each zone, the number of order pickers d_i equals 1, while the buffer size q_i depends on the location of the zone. For the zones that use pick-by-light, $q_i = 11$, except for the first and last zone in each segment for which q_i is either 8 or 9. The buffer sizes of the zones in the first segment are 12, 17, and, 19 respectively. In addition, the mean and coefficient of variation of the empirical picking time distribution obtained from the log files is given in Table 11b for each zone. Notice that all coefficients of variation are close to 1. Finally, the mean conveyor times vary between 25 up to 180 seconds depending on the location of the conveyor, and the mean time spent in one of the entrance stations equals 5 seconds.

Table 11 Parameters used in the test of the real zone picking system.

(a) Parameters of the segments

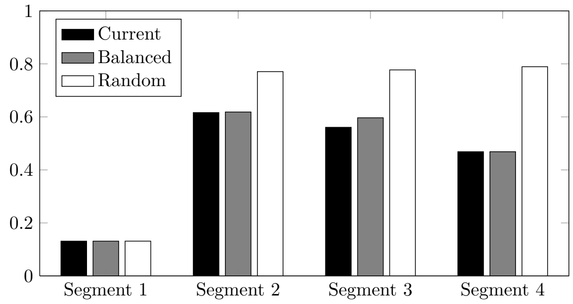
Parameter	Value
Number of totes, N	220,240,260,280
Number of zones, M	27
Number of segments, K	4
Number of zones per segment, m^k, $k=1,\dots,K$	3,8,8,8
Segment capacities, N^k, $k=1,\dots,K$	95

(b) Parameters of the individual zones

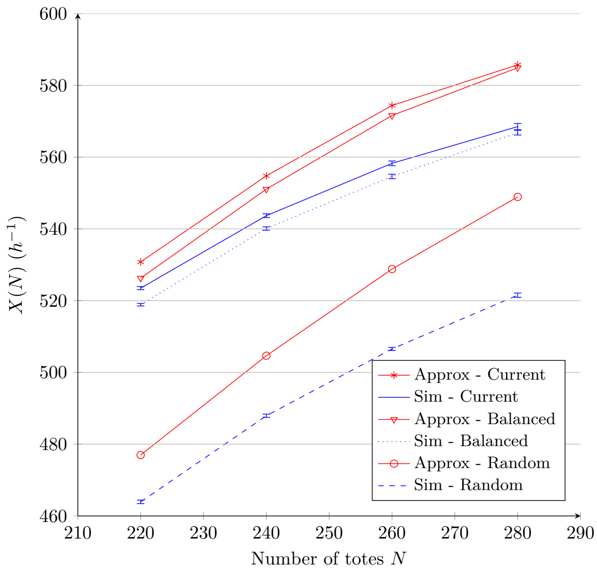
Zone	d_i	q_i	Empirical dist		Zone	d_i	q_i	Empirical dist		Zone	d_i	q_i	Empirical dist	
			Mean	CV				Mean	CV				Mean	CV
z_1^1	1	12	26.5	1.06	z_7^2	1	11	21.5	0.98	z_8^3	1	8	24.5	0.91
z_2^1	1	17	28.4	1.11	z_8^2	1	8	23.1	0.96	z_1^4	1	8	25.2	0.89
z_3^1	1	19	32.9	1.16	z_1^3	1	9	24.6	0.87	z_2^4	1	11	26.7	0.86
z_1^2	1	9	18.3	0.79	z_2^3	1	11	25.4	0.87	z_3^4	1	11	26.0	0.92
z_2^2	1	11	19.8	0.89	z_3^3	1	11	26.2	0.92	z_4^4	1	11	25.3	0.85
z_3^2	1	11	21.6	0.95	z_4^3	1	11	24.2	0.83	z_5^4	1	11	22.2	0.78
z_4^2	1	11	21.0	0.93	z_5^3	1	11	24.6	0.87	z_6^4	1	11	21.9	0.88
z_5^2	1	11	22.5	0.82	z_6^3	1	11	24.0	0.90	z_7^4	1	11	23.4	0.83
z_6^2	1	11	21.2	0.91	z_7^3	1	11	26.4	0.91	z_8^4	1	8	25.4	0.84

In the current storage policy of the company, the products in the pick-by-light zones are stored in such a way that the mean number of segments a tote has to visit is minimized. The advantage of this policy is that it reduces the total travel distance, since large parts of the system can be cut short. However, it may cause congestion in a segment, since certain zones are visited more often than others. If these zones are in the same segment, the probability of being blocked by a segment increases. Another policy is to store products in the zones such that the picking work-load between the segments is balanced. In this case, the total expected number of totes that visit a segment will be about the same for the different segments which reduces the probability of being blocked by a segment. A downside of this policy is that a tote can spend more time in the system, since the probability of visiting more than one segment increases. The performance of both storage policies is compared with a "random" storage policy. For this storage policy, a single allocation was generated where zones are randomly interchanged between segments.

The simulation uses the empirical picking time distributions of Table 11b. The quality of the



(a) The aggregated probabilities $\tilde{\psi}_k$ of a tote visiting segment k for the three different storage policies



(b) The system throughput of the three storage policies for the approximation and the simulation when varying the number of totes in the system N

Figure 7 The aggregated visiting probabilities and the results of the approximation and the simulation for the real zone picking system.

three storage policies is compared by varying the number of totes in the system N from 220 up to 280. The three policies are generated by interchanging zones between (pick-by-light) segments. The *current* and the *balanced* policies are obtained by solving an allocation problem which takes as input the orders in the log files, whereas in the *random* policy zones are randomly interchanged between the segments. Figure 7a shows for the three policies the aggregated visiting probability $\tilde{\psi}_k = \sum_{\boldsymbol{r} \subseteq \mathcal{Z}} \psi_{\boldsymbol{r}} I_{\left(\boldsymbol{r}^k \neq 0\right)}$ of segment k. The probability of visiting the first pick-by-pallet segment is unaltered, since pallet-pick zones cannot be interchanged with pick-by-light zones in the other segments. On average, totes will visit 1.73 segments in case of the *current* policy, 1.81 for the *balanced* policy, and 2.47 for *random* policy. The aggregated visiting probabilities $\tilde{\psi}_k$ of the *current* policy of the second and third segment (0.61 and 0.56) are slightly lower than the probabilities of the *balanced* policy (0.62 and 0.60). The aggregated visiting probabilities of the *random* policy are considerably higher for the three pick-by-light segments.

Figure 7b shows for the three different policies the impact on the system throughput $X(N)$

when increasing the number of totes in the system. The 95% confidence interval is shown for the simulation results of each policy. The mean average error in the system throughput is 2.28% for the *current* policy, 2.37% for the *balanced* policy, and 3.81% for the *random* policy. The errors are slightly higher than in Subsection 6.1-6.2, possibly due to the use of empirically instead of exponentially distributed picking times in the simulation (cf. Remark 2). For the three policies, the approximated system throughput is always higher than the one obtained from simulation, since the approximation tends to underestimate the blocking probabilities as seen in Subsection 6.1.

When $N = 220$ zones and segments become rarely congested. This implies that minimizing the mean number of segments a tote has to visit, which also minimizes the mean total travel distance of a tote, yields the highest throughput. Clearly, this is the case for the *current* policy, whereas the throughput of the *balanced* and the *random* policy are lower due to totes traveling larger distances. Increasing N, increases the probability that a tote is blocked by a zone or a segment. When $N = 280$, the system throughput of the *current* and the *balanced* policy are close to each other. Under the *current* policy, totes are more often blocked by a segment than for the *balanced* policy. Since a blocked tote has to recirculate on the main conveyor, the mean total travel distance of a tote increases and the system throughput decreases. By balancing the work-load between segments like under the *balanced* policy, congestion starts to occur later compared to the *current* policy. Therefore, when blocking is very prevalent and total recirculation distances are high, use of the *balanced* policy will yield a higher system throughput compared to the *current* policy. Again, the results show that both the *current* and the *balanced* policies can significantly improve the system throughput compared to the *random* policy.

7. Conclusion and further research

In this paper, we developed an analytical model for sequential zone picking systems with either single-segment, or multi-segment routing. The model provides a valuable tool for the design of complex zone picking systems in order to meet specific performance levels and it can be used to study and reduce the sources of blocking and congestion. We developed a queueing model to estimate the performance of the system. Because an exact analysis of this queueing model is not feasible, we applied product-form results, MVA, and an aggregation technique to obtain very accurate estimates of the key performance statistics of a zone picking system. Comparison of the approximation results to simulation for a wide range of parameters showed that the mean relative error for statistics as the system throughput and the mean number of circulations in the system is less than 1%.

The model lends itself to several modifications and extensions left for future research. The approximation can be used to evaluate and compare operational policies such as order batching and order splitting on system performance, like in Yu and De Koster (2008). In addition, a general optimization framework can be formulated for allocating products to zones in order to maximize, for example, the system throughput. Also, when the zone picking system is too heavily loaded, congestion also occurs when two conveyor streams merge, e.g., totes flowing out of a zone with totes on the conveyor. Totes should wait for a sufficiently large space on the conveyor before merging, which decreases the performance of the system due to long waiting times. By incorporating this kind of blocking-after-service helps to set achievable targets, e.g., the system throughput time, and predict when the system is overloaded. Another relevant extension is the situation where order pickers can help each other when the workload in one zone is high or leave when there is little work such that one order picker becomes responsible for picking products at multiple zones. Furthermore, the model may provide a starting point in order to approximate higher moments or the distribution of performance statistics such as the zone, segment, and system throughput time.

References

Bakker, R.F. 2007. Performance analysis of zone picking systems. Master's thesis, Eindhoven University of Technology.

Balsamo, S., V.D.N. Personé, R.O. Onvural. 2001. *Analysis of Queueing Networks with Blocking*. Springer Netherlands.

Baynat, B., Y. Dallery. 1993. A unified view of product-form approximation techniques for general closed queueing networks. *Performance Evaluation* **18**(3) 205–224.

Bolch, G., S. Greiner, H. de Meer, K.S. Trivedi. 2006. *Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications*. 2nd ed. Wiley-Interscience.

Boucherie, R.J. 1998. Norton's equivalent for queueing networks comprised of quasireversible components linked by state-dependent routing. *Performance Evaluation* **32**(2) 83–99.

Chandy, K.M., U. Herzog, L. Woo. 1975. Parametric analysis of queuing networks. *IBM Journal of Research and Development* **19**(1) 36–42.

Chandy, K.M., C.H. Sauer. 1980. Computational algorithms for product form queueing networks. *Communications of the ACM* **23**(10) 573–583.

Chao, X., M. Miyazawa. 2000. Queueing networks with instantaneous movements: a unified approach by quasi-reversibility. *Advances in Applied Probability* **32**(1) 284–313.

De Koster, R. 1994. Performance approximation of pick-to-belt orderpicking systems. *European Journal of Operational Research* **72**(3) 558–573.

De Koster, R., T. Le-Duc, K.J. Roodbergen. 2007. Design and control of warehouse order picking: A literature review. *European Journal of Operational Research* **182**(2) 481–501.

Drury, J. 1988. Towards more efficient order picking. Tech. Rep. 1, The Institute of Materials Management, Cranfield.

Economou, A., D. Fakinos. 1998. Product form stationary distributions for queueing networks with blocking and rerouting. *Queueing Systems* **30**(3) 251–260.

Gu, J., M. Goetschalckx, L.F. McGinnis. 2010. Research on warehouse design and performance evaluation: A comprehensive review. *European Journal of Operational Research* **203**(3) 539–549.

Henderson, W., P.G. Taylor. 2001. State-dependent coupling of quasireversible nodes. *Queueing Systems* **37**(1) 163–197.

Hsieh, Y-J., Y. Bozer. 2005. Analytical modeling of closed-loop conveyors with load recirculation. *Computational Science and Its Applications - ICCSA 2005*, *Lecture Notes in Computer Science*, vol. 3483. Springer Berlin, 838–838.

Kelly, F.P. 1979. *Reversibility and Stochastic Networks*. 1st ed. Wiley, London.

Le-Duc, T., R.M.B.M. de Koster. 2007. Travel time estimation and order batching in a 2-block warehouse. *European Journal of Operational Research* **176**(1) 374–388.

Melacini, M., S. Perotti, A. Tumino. 2010. Development of a framework for pick-and-pass order picking system design. *The International Journal of Advanced Manufacturing Technology* **53**(9) 841–854.

Nelson, R. 1995. *Probability, Stochastic Processes, and Queueing theory: The Mathematics of Computer Performance Modelling*. Springer.

Osorio, C., M. Bierlaire. 2009. An analytic finite capacity queueing network model capturing the propagation of congestion and blocking. *European Journal of Operational Research* **196**(3) 996–1007.

Papadopoulos, HT, C. Heavey, J. Browne. 1993. *Queueing Theory in Manufacturing Systems Analysis and Design*. Springer.

Park, B.C. 2012. *Warehousing in the Global Supply Chain: Advanced Models, Tools and Applications for Storage Systems*. 1st ed. Springer-Verlag, New York.

Perros, H.G. 1994. *Queueing Networks with Blocking*. Oxford University Press, Inc.

Petersen, C.G. 2000. An evaluation of order picking policies for mail order companies. *Production and Operations Management* **9**(4) 319–335.

Petersen, C.G. 2002. Considerations in order picking zone configuration. *International Journal of Operations & Production Management* **22**(7) 793–805.

Pittel, B. 1979. Closed exponential networks of queues with saturation: The Jackson-type stationary distribution and its asymptotic analysis. *Mathematics of Operations Research* **4**(4) 357–378.

Reiser, M. 1981. Mean-value analysis and convolution method for queue-dependent servers in closed queueing networks. *Performance Evaluation* **1**(1) 7–18.

Reiser, M., S.S. Lavenberg. 1980. Mean-value analysis of closed multichain queuing networks. *Journal of the ACM* **27**(2) 313–322.

Schassberger, R. 1984. Decomposable stochastic networks: Some observations. *Modelling and Performance Evaluation Methodology* **60** 135–150.

Schmidt, L.C., J. Jackman. 2000. Modeling recirculating conveyors with blocking. *European Journal of Operational Research* **124**(2) 422–436.

Stidham Jr., S. 2002. Analysis, design, and control of queueing systems. *Operations Research* **50**(1) 197–216.

Van Dijk, N.M. 1988. On Jackson's product form with "jump-over" blocking. *Operations Research Letters* **7**(5) 233–235.

Walrand, J. 1983. A note on Norton's theorem for queuing networks. *Journal of Applied Probability* **20**(2) 442–444.

Whitt, W. 1982. Approximating a point process by a renewal process, I: Two basic methods. *Operations Research* **30**(1) 125–147.

Wolff, R.W. 1989. *Stochastic Modeling and the Theory of Queues*, vol. 14. 1st ed. Prentice Hall, London.

Yao, D.D., JA Buzacott. 1987. Modeling a class of flexible manufacturing systems with reversible routing. *Operations Research* **35**(1) 87–93.

Yu, M., R. De Koster. 2008. Performance approximation and design of pick-and-pass order picking systems. *IIE Transactions* **40**(11) 1054–1069.

Yu, M., R. De Koster. 2009. The impact of order batching and picking area zoning on order picking system performance. *European Journal of Operational Research* **198**(2) 480–490.