Robotic Sorting Systems: Performance Estimation and Operating Policies Analysis

Bipan Zou,a René De Koster,b Yeming Gong,c Xianhao Xud Guwen Shen,e

aSchool of Business Administration, Zhongnan University of Economics and Law, Wuhan 430073, China; bRotterdam School of Management, Erasmus University, 3000 DR Rotterdam, Netherlands; cEMLYON Business School, 69134 Ecully Cedex, France; dSchool of Management, Huazhong University of Science and Technology, Wuhan 430074, China; eDeppon Express, Shanghai 201702, China

Contact: zbp0307@126.com, https://orcid.org/0000-0002-2908-9163 (BZ); rkoster@rsm.nl, https://orcid.org/0000-0001-9270-5507 (YG); xxhao@hust.edu.cn (XX); 474827034@qq.com (GS)

Abstract. Many distribution centers use expensive, conveyor-based sorting systems that require large buildings to house them. In areas with tight space, robotic sorting systems offer a new type of solution to sort parcels by destination. Such systems are highly flexible in throughput capacity and are now gradually being introduced, particularly in express companies. This paper studies robotic sorting system with two layouts. The first layout has two tiers: robots drive on the top tier and sort parcels by destination on spiral conveyors connected to roll containers at the lower tier. The second layout has a single tier with input and output points located at the perimeter, connected by robots. For each layout, we consider both the shortest path topology via dual-lane aisles and the detour path topology via single-lane aisles. We build closed queueing networks for performance estimation, design an iterative procedure to investigate robot congestion in the two-tier layout, and use a traffic flow function to estimate robot congestion in the single-tier layout. Random, closest, dedicated, and shortest-queue robot-to-loading-station assignment rules are examined. We validate analytical models by both simulation and a real case of Deppon Express and analyze the optimal system size and operating policies for throughput capacity and operating cost. The results show that the system throughput capacity is significantly affected by robot congestion in the single-tier layout with the detour path topology, but it is only slightly affected in the other systems. A square layout fits the shortest path and a rectangular layout fits the detour path. Both the random assignment rule and the shortest-queue assignment rule are superior for a large number of robots, whereas the dedicated assignment rule is superior for a small number of robots. We apply these insights at Deppon Express for different allocations in peak and off-peak hours. Our analysis shows that a robotic sorting system typically has lower overall annual cost than a traditional cross-belt sorting system when the required throughput capacity is not too large.

Funding: This research is partially supported by the National Natural Science Foundation of China [Grants 71801225, 72171233, 71620107002, and 71821001], the Hubei Provincial Natural Science Foundation of China [Grant 2018CFB166], and the Fundamental Research Funds for the Central Universities, Zhongnan University of Economics and Law [Grant 2722021B2026], Basic and Applied Basic Research Foundation of Guangdong Province [Grant 2019A1515111195].

Supplemental Material: The online appendix is available at https://doi.org/10.1287/trsc.2021.1053.

Keywords: automated warehouse • queueing network • performance estimation • simulation • robotic sorting

1. Introduction

Warehouses are increasingly using robotic technologies to help pick and deliver large volumes of customer orders, with short response times. These new technologies include vehicle-based solutions, such as autonomous vehicle storage and retrieval system systems (AVS/RS), robotic mobile fulfilment systems (RMFS), robotic compact storage and retrieval systems (RCSR) and puzzle-based storage systems. In all such systems, robotic autonomous vehicles transport the loads (see Azadeh, De Koster, and Roy 2019 for a detailed description). Sorting is a new application of robotic autonomous vehicles, in which robotic vehicles sort parcels and transport them from any origin to any destination point (see the solutions of Tompkins 2020). As examples, STO and Deppon Express, two of the biggest express delivery companies in Hangzhou and Shanghai, China, have implemented such systems on two vertical tiers (Wang 2016, You 2019). The upper tier is a mezzanine with evenly distributed holes (Figure 1(a)), the drop-off points, each of which corresponds to a delivery destination. Each hole is connected through a spiral conveyor with a roll container installed at the lower tier that collects the items per destination (Figure 1(b)). Loading stations located at
the system perimeter insert the items that need to be sorted. Autonomous robots are also used to sort items on the same tier, that is, a single tier system, including the Amazon Xanthus and Pegasus fulfillment robots (Ames 2019) and the baggage sorting robots used in Rotterdam The Hague Airport (Airport RTH 2018). Such robotic sorting systems (RSSs) have the following advantages compared with traditional conveyor-based high-speed sorting solutions:

1. The layout is flexible: the system can activate or deactivate input points based on sorting demands and change the assignment of destinations to outputs in control software.

2. The throughput capacity is scalable by adding or retracting robots and workers from the system.

3. They need much less floor space and built area.

There are also some disadvantages, such as the fact that expensive robots and complex control software are needed to control the many freely moving robots in a small area, without deadlocks or too much congestion. At the design stage of an RSS, various factors should be evaluated, which determine whether the system meets the required throughput capacity at a minimum investment. This paper aims to develop models to rapidly estimate the throughput capacity of an RSS for a given layout and path topology with given equipment. In turn, such models can be used to optimize the system layout and evaluate different robot operating policies, such as the robot-to-loading-station assignment, or to minimize the equipment needed, to achieve a certain throughput capacity. Because many robots operate in a compact area, it is vital to incorporate the impact of robot congestion on performance. We evaluate the following layouts and operating policies:

- System layout. Single-tier and two-tier layouts. In the single-tier layout, both input and output points are located at the perimeter, connected by autonomous robots. In the two-tier layout, input points are located at the perimeter and output points in the middle. Autonomous robots sort parcels on the top tier and release them into the roll containers at the lower tier.

- Robot-to-loading-station assignment rule. It determines to which loading station a robot should go when it finishes its order. We evaluate (1) a random assignment rule under which a robot will travel back to a random loading station, (2) the closest assignment rule under which a robot will travel back to the closest loading station, (3) the dedicated assignment rule under which the robot will travel back to its dedicated loading station, and (4) the shortest-robot-queue assignment rule under which a robot will travel back to the loading station with the minimum number of waiting robots.

- Robot path topology. We evaluate a shortest path topology via dual-lane aisles and a detour path topology via single-lane unidirectional aisles. In the system with the shortest path topology, each aisle and cross-aisle have two lanes with opposite driving directions, allowing shortest path travel. When each aisle and cross-aisle has a single unidirectional lane, a robot may need to take a detour to travel between loading stations and drops-off points.

We aim to answer the following research questions:

1. How to build accurate and efficient throughput capacity estimation models for an RSS while taking robot congestion into account?

2. Which robot-to-loading-station assignment rule gives the highest throughput capacity?

3. What layout performs the best, with which robot path topology, in terms of throughput capacity and operating cost? While the shortest path topology might yield shorter paths, it requires more space than the detour path topology, due to the multiple paths per aisle.

Figure 2 summarizes the different combinations of layouts, robot path topologies, and robot-to-loading-station

---

**Figure 1.** (Color online) Robotic Sorting System

Notes. (a) Main view. (b) Sorting robot.

---

Zou et al.: RSS: Performance Estimation and Operating Policies
Transportation Science, 2021, vol. 55, no. 6, pp. 1430–1455, © 2021 INFORMS
assignement rules studied in this paper. We build closed queueing networks (CQN) for these configurations to estimate the performance of an RSS. To investigate the effect of robot congestion in the two-tier layout, we capture the effective velocity of a robot by modeling each drop-off point as a single server and by approximating its expected utilization using an iterative algorithm. In the single-tier layout, we use a traffic flow function to estimate the effective velocity of a robot. We use the approximate mean value method to analyze the CQN and validate the analytical models by both simulation and a real case of Deppon Express. The results show that our analytical models can accurately estimate the maximum throughput capacity of the different RSSs. Then, we use the performance estimation model to analyze and optimize the system configuration and operating policies for throughput capacity and operating cost. Moreover, we also use a semiopen queueing network to obtain both the order cycle time and the expected waiting time of parcels for robots and validate it by simulation.

The results show that the effective velocity of robots decreases only slightly because of robot congestion in the two-tier RSS with shortest path and detour path topologies and in the single-tier RSS with the shortest path topology but decreases significantly in an RSS with the detour path topology, when the number of robots increases. A square layout fits the shortest path topology better and a rectangular layout fits the detour path topology better. Furthermore, when the number of robots exceeds a critical point, assigning a robot to either a random loading station or the loading station with the shortest robot queue provides a larger throughput capacity, otherwise, the closest and dedicated assignment rules are better. We apply these results to improve the performance in peak and off-peak hours at the sorting center of Deppon Express in Shanghai. Moreover, the cost analysis shows that the detour path topology needs lower investment than the shortest path topology when the required throughput capacity is below a critical point. We also show that, below a critical required throughput capacity, an RSS is cheaper than a traditional cross-belt sorter system, due to savings in space and sorter equipment.

Our contributions are twofold. We develop accurate closed queueing networks to estimate system performance for this new application, taking robot path topology and congestion into consideration. We then analyze system layouts, robot path topologies, and the robot-to-loading-station assignment rules both in terms of throughput capacity and operating cost, and indicate which RS configuration is the better choice, for example, compared with a traditional cross-belt sorter system.

2. Literature Review
Sorting items or parcels occurs in the order consolidation area or shipping area of warehouses or transshipment centers, typically by automated conveyor-based sorters (Boysen, Briskorn, and Fedtke 2019). Items to be sorted are supplied to the sorter machine through manual or automated induction stations, are uniquely identified, and then sorted onto different exit lanes (or chutes), from where items are collected for packing (item sorting) or shipping (parcel sorting). Depending on the items or parcels to be handled and the throughput capacity to be achieved, different sorting technologies are available, for example, cross-belt, bomb-bay,
till-tray, and sliding-shoe sorters. We here focus on studies of automated sorting systems that deal with the tactical level problems of throughput performance estimation and analysis of operating policies.

Bozer and Sharp (1985) build a simulation model of a closed-loop sorting system to analyze the throughput impact of the number of output lanes and the degree of clustering (the probability of a group of incoming items being destined to a subset of the lanes). Bozer, Quiroz, and Sharp (1988) investigate the impact of a wave strategy, that is, batching of orders, on the system throughput of a closed-loop sorting system. Johnson (1997) study two classes of sorting strategies in a closed-loop sorting system: a fixed priority rule that sequences items to be sorted by a given priority and the next available rule that sequences items serially, in an accumulation sorting system. They find that the next available strategy shows throughput superiority over the fixed priority strategy in systems with little output lane blocking. Johnson and Meller (2002) focus on a split-case sorting system that is used to sort product cases, originating from batch picking, into orders with individual items. They develop a Bernoulli process to estimate the throughput of the induction process that is assumed to be the system bottleneck. Russell and Meller (2003) build a descriptive model to make the decision between manual and automated sortation and to design an automated sorting system with the lowest total annual cost that meets the required throughput. Gallien and Weber (2010) develop a queuing model to obtain control guidelines for a continuous (waveless) release policy and comprehensively compare the performance of a batch (wave) release policy and that of a waveless release policy. They find that the waveless policy has superior throughput capacity compared with the wave-based policy that increases with a diminishing number of output lanes.

Some studies on automated sorting systems deal with operational problems involving the optimization of the consolidation process, such as the sequence optimization of release bins analyzed by Boysen, Fedtke, and Weidinger (2018) and the scheduling of shipments on trays studied by Briskorn, Emde, and Boysen (2017). Several survey papers have been published, discussing the role of sorters inside facilities and their planning and control. We particularly refer to the papers by De Koster, Le-Duc, and Roodbergen (2007), Gu, Goetschalckx, and McGinnis (2010), and Boysen, Briskorn, and Fedtke (2019).

Although a robotic sorting system shares the same function and objectives with the sorting systems mentioned previously, its operations are different. It has more similarity with some vehicle-based robotic warehousing systems transporting items, such as the robotic mobile fulfillment systems (RMFS) or robotic compact storage and retrieval systems (RCSRS), autonomous vehicle-based storage and retrieval systems (AVS/RS), or automated container terminal vehicle systems (CT). We review studies on robotic and vehicle-based warehousing systems that focus on throughput performance estimation and operating policies, using queueing models and simulation. Table 1 gives an overview of key papers, with systems studied and model features.

An RMFS is a parts-to-pickers storage and order-picking system that uses robots to transport movable shelves with inventory. Because of its ability to handle large assortments of small products and to deal with strong demand fluctuations, the number of implementations in e-commerce fulfillment centers has grown rapidly. The system was introduced by Amazon in 2012 (Mountz 2012) and now many competing systems are on the market with implementations all around the globe. The pioneering paper by Lamballais, Roy, and De Koster (2017) builds a semiopen queueing network for performance estimation and investigates the system layouts and robot zoning strategies. They find that system performance is insensitive to the width-to-length ratio but sensitive to the location of loading stations. Yuan and Gong (2017) develop an open queueing network for performance estimation, considering both dedicated robot-to-pick station assignment and sharing protocols.

Several papers analyze operating policies in an RMFS, based on analytical performance estimation models. Lamballais, Roy, and De Koster (2020) optimize the inventory allocation by a cross-class matching multiclass semi-open queueing network (SOQN). They find that the system performance can be substantially improved by spreading inventory across multiple pods. Zou et al. (2017) investigate the robot-to-loading-station assignment rules, considering random, closest, and heuristic-based rules. They show that assigning robots to a proper loading station can significantly improve system performance. Because the vehicles are battery powered, it is important to take charging strategies into account in the design and operation. Zou et al. (2018) examine battery recovery by a SOQN, considering plug-in charging, inductive charging, and battery swapping strategies. They find that the inductive charging strategy performs the best, followed by battery swapping and plug-in charging, in terms of retrieval transaction throughput time. Besides queueing network modeling, simulation is also used to analyze operating policies in an RMFS. Merschmann et al. (2019) develop simulation models to study various decision problems in an RMFS, including the order-to-station assignment, pod selection, and pod storage assignment problems. Merschmann (2018) studies active pod repositioning in an RMFS by conducting a simulation-based experiment. The results show that active repositioning may boost RMFS throughput performance, whereas the cost of
repositioning is reasonable if the system faces regular idle periods. All these studies consider detour robot topology via unidirectional aisles and ignore robot congestion.

Different from an RMFS, an AVS/RS stores inventory items in a multitier storage rack and uses rack-rail-guided vehicles to transport the loads. In a tier-to-tier system, vehicles can transfer between tiers by using lifts. In a tier-captive configuration, a vehicle off-loads its load onto a lift and stays at the tier. Many papers have studied both configurations, using queuing networks or simulation. They all use shortest path topology via bidirectional aisles and cross-aisles and ignore congestion. Blocking prevention is sometimes taken into account as in Roy, Gupta, and De Koster (2011) and Roy and De Koster (2014). Azadeh, Roy, and De Koster (2019) explicitly consider vehicle congestion in an autonomous guided vehicle storage and retrieval system where the vehicles can autonomously move between tiers, without a lift. Malmborg (2002) develop an approximate cycle time model, considering the random storage and opportunistic pairing of storage and retrieval transactions. Heragu et al. (2011) build an open queuing network and use the manufacturing system performance analyzer to estimate the performance of a tier-captive AVS/RS. Marchet et al. (2012) build an open queuing network to estimate performance of a tier-captive AVS/RS for product totes, by additionally considering vehicle and lift acceleration and deceleration. These studies all assume sequential movements of vehicles and lift. Zou et al. (2016) consider parallel movements and build a fork-join queuing network to evaluate performance. Moving in parallel outperforms moving sequentially in terms of system throughput time when the retrieval transaction arrival rate is low.

Compared with a probabilistic travel model or an OQN model, a SOQN can more accurately capture the matching of vehicles and orders. Roy et al. (2012) develop a multiclass SOQN with class switching for a single-tier tier-to-tier AVS/RS and design a decomposition-based solution method. They investigate

<table>
<thead>
<tr>
<th>Reference</th>
<th>Functions</th>
<th>Model</th>
<th>Objective</th>
<th>Topology</th>
<th>Congestion</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMFS: Yuan and Gong (2017)</td>
<td>Storage, order-picking</td>
<td>OQN, simulation</td>
<td>Throughput</td>
<td>Detour path</td>
<td>(\times)</td>
</tr>
<tr>
<td>Lamballais et al. (2017);</td>
<td>Storage, order-picking</td>
<td>SOQN, simulation</td>
<td>Throughput</td>
<td>Detour path</td>
<td>(\times)</td>
</tr>
<tr>
<td>Lamballais et al. (2020); Zou et al. (2017,2018)</td>
<td>Storage, order-picking</td>
<td>Simulation</td>
<td>Throughput</td>
<td>Detour path</td>
<td>(\times)</td>
</tr>
<tr>
<td>Merschformann (2018); Merschformann et al. (2019)</td>
<td>Storage, order-picking</td>
<td>Simulation</td>
<td>Throughput</td>
<td>Detour path</td>
<td>(\times)</td>
</tr>
<tr>
<td>AVS/RS: Malmborg (2002);</td>
<td>Storage, order-picking</td>
<td>OQN, simulation</td>
<td>Throughput</td>
<td>Shortest path</td>
<td>(\times)</td>
</tr>
<tr>
<td>Fukunari and Malmborg (2008);</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heragu et al. (2011); Marchet et al. (2012); Zou et al. (2016)</td>
<td>Storage, order-picking</td>
<td>SOQN, simulation</td>
<td>Throughput</td>
<td>Shortest path</td>
<td>(\times)</td>
</tr>
<tr>
<td>Roy et al. (2012); Cai et al. (2014); Ekren et al. (2014); Tappia et al. (2017, 2019); Azadeh et al. (2019a)</td>
<td>Storage, order-picking</td>
<td>SOQN, simulation</td>
<td>Throughput</td>
<td>Shortest path</td>
<td>(\times)</td>
</tr>
<tr>
<td>Azadeh et al. (2019b)</td>
<td>Storage, order-picking</td>
<td>Simulation</td>
<td>Throughput</td>
<td>Shortest path</td>
<td>(\times)</td>
</tr>
<tr>
<td>Ekren and Heragu (2010); Ekren et al. (2010); Kumar et al. (2014); Ekren et al. (2015)</td>
<td>Storage, order-picking</td>
<td>Simulation</td>
<td>Throughput</td>
<td>Shortest path</td>
<td>(\times)</td>
</tr>
<tr>
<td>RCSRS: Beckschäfer et al. (2017)</td>
<td>Storage, order-picking</td>
<td>Simulation</td>
<td>Throughput</td>
<td>Shortest path</td>
<td>(\times)</td>
</tr>
<tr>
<td>Zou et al. (2018)</td>
<td>Storage, order-picking</td>
<td>SOQN, simulation</td>
<td>Throughput; cost</td>
<td>Shortest path</td>
<td>(\times)</td>
</tr>
<tr>
<td>CT: Roy et al. (2016); Roy and De Koster (2014)</td>
<td>Storage</td>
<td>OQN, simulation</td>
<td>Throughput; cost</td>
<td>Detour path</td>
<td>(\checkmark)</td>
</tr>
<tr>
<td>This paper: RSS</td>
<td>Sorting</td>
<td>CQN, simulation</td>
<td>Throughput; cost</td>
<td>Shortest path;</td>
<td>(\checkmark)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>detour path</td>
<td></td>
</tr>
</tbody>
</table>

Note. CQN, closed queueing network; OQN, open queueing network; SOQN, semiopen queueing network.
vehicle-to-load assignment rules and the allocation of vehicles among zones. The results reinforce the efficacy of multiple zoning and show that the random vehicle assignment rule and the preferred assignment rule have similar performance. Cai, Heragu, and Liu (2014) model a tier-to-tier AVS/RS with a SOQN and solve it using the approximate matrix-geometric method. Ekren et al. (2014) also analyze a SOQN for an AVS/RS, based on the method proposed by Jia and Heragu (2009). Tappia et al. (2017) study a multi-deep AVS/RS and develop a SOQN to evaluate the system performance of both specialized and generic shuttles and both continuous and discrete lifts. Tappia et al. (2019) build a SOQN to investigate an integrated upstream AVS/RS and a downstream picking system. The results show that an AVS/RS yields investment savings and provides a lower total throughput time compared with a crane-based AS/RS. Except analytical models, several papers also study the design of an AVS/RS by simulation, including Ekren and Heragu (2010), Ekren et al. (2010), Kumar, Roy, and Tiwari (2014), and Ekren, Sari, and Lerher (2015).

An RCSRS (in particular the Autostore system) also uses vehicles to transport loads, but with much higher storage density than an RMFS or AVS/RS. It stores inventory items in bins that are organized as stacks in a grid. A fleet of robots drive on the roof of the system on a grid. They dig out a requested bin and transport it to a designated loading station, located at the perimeter of the system. Beckschäfer et al. (2017) develop a discrete event simulation model to examine bin storage policies. Zou, De Koster, and Xu (2018) build a semi-open queuing network to estimate performance of an RCSRS, considering both dedicated (one product per stack, reducing bin reshuffling time) and shared (multiple products per stack) storage policies. Their results show that the dedicated storage policy outperforms the shared storage policy in terms of system throughput time but that the shared storage policy results in lower system operating cost, due to smaller system size. For more research on robotic warehouse systems, we refer the readers to De Koster (2018).

Our literature review shows that robot congestion is often ignored in robotic warehousing systems. Specifically, studies on the RMFS assume detour path topology via unidirectional aisles to avoid robot blocking and additionally ignore congestion. Studies on AVS/RS and RCSRS assume shortest path topology via bidirectional aisles and ignore blocking and congestion (with the exception of Azadeh, Roy, and De Koster 2019). However, when many robots operate in a confined space, robot congestion cannot be ignored. In fact, in a recent study of an automated container terminal, Roy, Gupta, and De Koster (2016) find that the system throughput can decrease by up to 85% due to vehicle congestion, when vehicles operate in a loop path topology.

3. System Description and Operating Policies

Section 3.1 describes the system and states the assumptions and notations. Section 3.2 introduces the system layouts and the robot path topologies. Section 3.3 presents the robot-to-loading-station assignment rules.

3.1. System Description, Assumptions, and Notations

An RSS is a robotic sorting system that uses robots to transport items from loading stations to drop-off points to which specific destinations have been assigned. We consider two layouts for such a system: a two-tier RSS (Figure 3) and a single-tier RSS (Figure 4). In a two-tier system, the top tier is a mezzanine where items are placed on robots at loading stations located at the perimeter. The robots release the items into the drop-off points located in the middle of the mezzanine. The bottom tier stores roll containers that connect with the drop-off points to collect released items. In a single-tier system, input points and output points (equipped with outgoing conveyors) are located at the perimeter of the system, and the middle area is for robots traveling, parking, and charging. This study considers a general layout for an RSS with a size $W \times L$ ($W$ drop-off points in the width direction and $L$ drop-off points in the length direction) and $n_{ls}$ loading stations ($n_{ls}^w$ loading stations on the western and eastern sides and $n_{ls}^n$ loading stations on the northern and southern sides). The loading stations can be generally distributed at the perimeter by adjusting their coordinates. Drop-off points are evenly distributed over the space, separated by aisles and cross-aisles.

The following steps are needed to execute a sorting job:

1. The robot receives the item to be sorted at loading station $ls_{i}$, $i = 1, 2, \ldots, n_{ls}$. Then, it moves along aisles (indicated as $a_{i}$ in Figure 3) and cross-aisles (indicated as $ca_{j}$ in Figure 3) to the designated drop-off point (routing path will be specified in Section 3.2, depending on the system layout and robot path topology design). The robot may be blocked in its movement by other robots that are releasing items, which will cause congestion.

2. When the robot reaches the target drop-off point (two-tier RSS) or output point (single-tier RSS), it releases the item into the roll container or onto the conveyor.

3. The robot goes back to loading station $ls_{j}$, $j = 1, 2, \ldots, n_{ls}$ and waits there for a next order. The target loading station depends on the applied robot-to-loading-station assignment rule. Specifically, we consider the random, closest, dedicated and shortest-robot-queue assignment rules in this study (see Section 3.3).

We make the following assumptions in this study:

1. All lanes in aisles are unidirectional. A robot can only move in one direction in an aisle with only a
single-lane while it can move in either direction in an aisle with dual-lanes by choosing the correct one (Figures 3 and 4).

2. In a two-tier RSS with detour path topology, an eastern and western loading station face a rightward and leftward cross-aisle, respectively, a southern and northern loading station face an upward and downward aisle, respectively, allowing the shortest path travel (Figure 3). A robot may first take a detour and then travel by the path depicted in Figure 3(b), if this assumption is relaxed.

3. Sorting orders designated for drop-off point $d_{x,y}$ arrive at the system following a Poisson distribution with an arrival rate $\lambda_{x,y}$, which is drop-off point dependent.

4. A robotic sorting system generally handles items lighter than five kilograms (Tompkins 2020), a worker in a loading station can quickly put a parcel on a robot with the bar code upward for scanning, and a robot can release an item into a drop-off point by a short time $t_{lu}$ (assumed to be constant). The operating time in a loading station ($t_{ls}$) is assumed to follow a uniform

---

**Figure 3.** (Color online) Two-Tier RSS

**Notes.** (a) Shortest path topology. (b) Detour path topology.

**Figure 4.** (Color online) Single-Tier RSS

**Notes.** (a) Shortest path topology. (b) Single-ring path topology.
distribution $U[t_{l_{i}}, t_{r_{i}}]$. Other distributions can also be examined in the queuing network, by specifying the first two moments of the operating time.

5. We assume that the robot returns to a loading station based on some rules (such as random and dedicated). Such rules are also used in other robotic warehousing systems, such as the RMFS (Lamballais, Roy, and De Koster 2017, 2020) and the robotic compact storage and retrieval system (Zou, De Koster, and Xu 2018). We later relax this and also evaluate a dynamic shortest-queue allocation rule.

6. Because of the mechanism constraint, a robot can only release a load to its right side, which means the robot may need to rotate when the target drop-off point is on the left. This takes a time $t_{r}$.

7. To estimate the maximum throughput capacity of a system with a given number of robots and loading stations, we assume that the arrival rate of parcels is sufficiently large to keep loading stations busy. Closed queueing networks are built for performance analysis, in which the robots represent the circulating resource. Note that we do not intend to size the receiving area of the cross-dock. We assume that the receiving area is large enough, so that the parcels supplied by trucks can wait on pallets, roll cages or airplane containers (ULDs) in the receiving area, until they can be brought to the loading stations (e.g., by conveyor). The throughput capacity of a given system will then depend on the number of loading stations and robots, and the control rules deployed.

We present the main notations used in this paper in Table 2.

### Table 2. Main Notations

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W, L$</td>
<td>System width and length, by the number of drop-off points in the horizontal and vertical direction</td>
<td>$w_{aisle}, w_{dp}$</td>
<td>Width of an aisle/cross-aisle and a drop-off point ($m$)</td>
</tr>
<tr>
<td>$n_{a}, n_{c}$</td>
<td>Number of aisles and cross-aisles in the system</td>
<td>$n_{a_{e}}, n_{a_{w}}$</td>
<td>Number of loading stations on the east and north side and bottom of the system, total number of loading stations is $n_{a} = (n_{a_{e}} + n_{a_{w}})$</td>
</tr>
<tr>
<td>$R$</td>
<td>Number of robots</td>
<td>$d_{x,y}$</td>
<td>Time for a robot to drop off an item into a drop-off point ($s$)</td>
</tr>
<tr>
<td>$a$</td>
<td>Acceleration/deceleration rate of a robot ($m/s^2$)</td>
<td>$v_{m}$</td>
<td>Maximum velocity of a robot ($m/s$)</td>
</tr>
<tr>
<td>$\lambda_{x,y}$</td>
<td>Arrival rate of sorting orders (per hour) to drop-off point $d_{x,y}$, $x = 1, 2, …, W, y = 1, 2, …, L$</td>
<td>$p_{d_{x,y}}$</td>
<td>Probability that an order goes to drop-off point $d_{x,y}$, we have $p_{d_{x,y}} = \sum_{s=1}^{\sum_{y=1}^{y_{x}}} \sum_{x=1}^{x_{y}} x_{y}$</td>
</tr>
<tr>
<td>$t_{l}$</td>
<td>Time for a robot to drop off an item into a drop-off point ($s$)</td>
<td>$t_{r} \times t_{r}$</td>
<td>Time for a robot to take a 90° turn and to rotate 180°, we have $t_{r} = 2t_{r}$ ($s$)</td>
</tr>
<tr>
<td>$T_{l_{i}, A_{i}}$</td>
<td>Movement time of a robot from loading station $l_{i}$ to drop-off point $d_{x,y}$ ($s$)</td>
<td>$c_{a_{i}}$</td>
<td>Cross-aisle ($j = 1, 2, …, n_{c_{j}}$)</td>
</tr>
<tr>
<td>$T_{h_{i}, A_{i}}$</td>
<td>Horizontal movement time of a robot from loading station $l_{i}$ to drop-off point $d_{x,y}$ ($s$)</td>
<td>$T_{h_{i}, l_{i}}$</td>
<td>Movement time of a robot from drop-off point $d_{x,y}$ to loading station $l_{i}$ ($s$)</td>
</tr>
<tr>
<td>$T_{l_{i}, A_{i}}$</td>
<td>Remaining service time at drop-off point $d_{x,y}$ ($s$)</td>
<td>$T_{l_{i}, h_{i}}$</td>
<td>Vertical movement time of a robot from loading station $l_{i}$ to drop-off point $d_{x,y}$ ($s$)</td>
</tr>
<tr>
<td>$T_{r_{i}, A_{i}}$</td>
<td>Travel time from aisle $a_{i}$ to aisle $a_{i+1}$ through cross-aisle $a_{i}$ ($s$)</td>
<td>$T_{r_{i}, l_{i}}$</td>
<td>Travel time from cross-aisle $c_{a_{i}}$ to cross-aisle $c_{a_{i+1}}$ through aisle $a_{i}$ ($s$)</td>
</tr>
<tr>
<td>$T_{c_{i}, A_{i}}$</td>
<td>Time for a robot to take a 90° turn and to rotate 180°, we have $t_{r} = 2t_{r}$ ($s$)</td>
<td>$T_{c_{i}, c_{i}}$</td>
<td>Maximum throughput capacity of the RSS (per hour)</td>
</tr>
</tbody>
</table>

### 3.2. System Layouts and Robot Path Topologies

We consider both a two-tier (Figure 3) layout and a single-tier layout (Figure 4) for the RSS. The two-tier layout can provide large throughput capacity for many sorting destinations, whereas the single-tier layout may be more suitable for a small size system and better for maintaining and controlling of robots.

In an RSS, the robot path topology not only affects the travel distance of the robot between a loading station and a drop-off point, but also the congestion of robots in the system. We consider both a shortest path topology and a detour path topology for each layout.

In the two-tier RSS, we consider a design with shortest path topology via dual-lane aisles and cross-aisles (Figure 3(a)) and a design with detour path topology via single-lane aisles and cross-aisles (Figure 3(b)). In the system with the shortest path topology, each aisle and cross-aisle has two uni-directional lanes that run in opposite directions. A robot can take the shortest
path between a loading station and a drop-off point. The shortest path from a loading station to a drop-off point corresponds to the path with shortest Manhattan distance between these two points and at most one turn (red line in Figure 3(a)). In the shortest path from a drop-off point back to a loading station, the robot first moves to a point from where it can then travel by shortest Manhattan distance with at most one turn to the loading station (green line in Figure 3(a)). In the system with detour path topology, each aisle and cross-aisle has one uni-directional lane. A robot can take the shortest path from a loading station to a drop-off point, which has shortest Manhattan distance with at most one turn (red line in Figure 3(b)). Note that the robot may need to rotate \(180^\circ\) (rotation time \(t_r\)) to release the item, depending on the travel direction of the robot in the aisles or cross-aisles. To move from the drop-off point back to a loading station, the robot may need to take a detour, depending on the position of the loading station and drop-off point (see the green line for an example). Online Appendix C includes the detour path details. The shortest path topology design provides more efficient routing than the detour path topology, but it occupies more floor space.

In a single-tier RSS, we consider a design with the shortest path topology (Figure 4(a)) and a design with a single-ring topology (Figure 4(b)). In the first design, a robot can take a similar shortest path between an input point and an output point as that in the two-tier layout with shortest path topology (see the red and green lines as an example). In the second design, the robot travels between loading stations and output points on one unidirectional ring and the middle area is designed for parking and charging of robots.

### 3.3. Robot-to-Loading-Station Assignment Rules

The robot-to-loading-station assignment rule applied in the system determines to which loading station a robot returns after finishing a sorting order. We consider four robot-to-loading-station assignment rules in an RSS:

- **Random rule:** The robots are shared by all loading stations and they return to a loading station with probability \(p_{lsi} = \frac{1}{n_{lsi}}, i = 1, 2, \ldots, n_{lsi}\) after finishing a sorting order.

- **Closest rule:** The robots are shared by all loading stations and return to the closest loading station after finishing a sorting order. With this assignment rule, the distance between a drop-off point and any loading station is calculated to find the closest loading station for each drop-off point. We define the set of drop-off points whose closest loading station is \(ls_i\) as \(DPlsi\). We can then obtain the routing probability of a robot returning to loading station \(ls_i\) by

\[
λ_{lsi}/\sum_{w=1}^{W} \sum_{y=1}^{L} λ_{x,y}.
\]

- **Dedicated rule:** The robots are partitioned over the loading stations. Thereby, a robot will return to one of its dedicated loading stations randomly after finishing an order. In this paper, we assume that all loading stations have their dedicated robots, and the robots serve all drop-off points. The north-side loading stations \((ls_1^n, \ldots, ls_{nls}^n)\) have \(R_n\) robots, the south-side loading stations \((ls_1^s, \ldots, ls_{nls}^s)\) have \(R_s\) robots, the west-side loading stations \((ls_1^w, \ldots, ls_{nls}^w)\) have \(R_w\) robots, and the east-side loading stations \((ls_1^e, \ldots, ls_{nls}^e)\) have \(R_e\) robots, with \(R_n + R_s + R_w + R_e = R\).

- **Shortest queue rule:** The robots are shared by all loading stations and they return to a loading station with a minimum number of waiting robots.

The random assignment rule aims to balance the working load of loading stations, by assigning robots to all loading stations with the same probability. The closest assignment rule aims to reduce the expected travel distance from drop-off points to loading stations. The shortest queue rule works based on the work load of each loading station, ensuring the availability of robots at loading stations. Because of the intractability of the robot queue status in the closed queueing network, we investigate this rule by simulation.

To study the previous layouts and operating policies, the next section builds closed queueing networks to estimate the maximum system throughput capacity.

### 4. Analytical Models for an RSS

Based on our assumption that the loading stations will not overflow; that is, loads that have arrived by trucks on pallets, roll cages, or ULDs wait in the receiving area, until they can be brought to the load station, we focus on the most important issue in the design of the sorter and the examination of the effect of the operating policies: its maximum capacity at a given cost level. The CQN is the preferred model for this task because it releases a load to the system every time a load leaves the system. We build a single-class closed queueing network (Figure 5) for the two-tier RSS that uses the random and closest assignment rule and a multiclass closed queueing network (Figure 6) for the two-tier RSS that uses the dedicated assignment rule. We build a single-class closed queueing network for the single-tier RSS with the shortest path topology (Figure 7) and the single-tier RSS with the single-ring path (Figure 8), respectively. Our models can capture the robot operating policies by specific travel processes between loading stations and drop-off points (service nodes \(\mu_{lsid}\) and \(\mu_{lsid})\), and by specific visit ratios of robots to drop-off points \(p_{lsid}\) and loading stations \(p_{lsi}\). It is also possible to obtain the order cycle time and the parcel waiting time, but it requires the extension of our CQN model into a semi-open queueing network model (see Roy 2016 and Van...
der Gaast et al. 2020 on how to do this). We validate the performance of the SOQN in Online Appendix E based on the additional measures.

4.1. CQN for Two-Tier RSS with Random and Closest Robot-to-Loading-Station Assignment Rules

In the two-tier RSS, an order enters the system after it obtains access to a loading station, which corresponds to the waiting process at service node $\mu_{ls_i}, i = 1, 2, \ldots, n_{ls}$. Under the random and closest robot-to-loading-station assignment rules, the CQN contains one order class that may go to any loading station. The worker or an automated induct puts the item on a robot (service node $\mu_{ls_i}$) and then the robot moves to the designated drop-off point (service node $\mu_{ls_i,i}$). Because no waiting is incurred, we model the movement process as an infinite server (IS). The robot reaches the drop-off point $d_{x,y}$ with probability $p_{dx,x}$ and then releases the item at service node $\mu_{ls,dr}, x = 1, 2, \ldots, W; y = 1, 2, \ldots, L$. Because each drop-off point can be accessed by one robot at a time, we model $\mu_{ls,dr}$ as a single server. Finally, the robot returns to loading station $ls_i$ with probability $p_{ls_i,i}, i = 1, 2, \ldots, n_{ls}$, corresponding to the service node $\mu_{ls,dr}$. Under both random and closest robot-to-loading-station assignment rules, all robots are shared by all loading stations, thereby, a robot may serve any loading station.

The robot path topology in the two-tier RSS (shortest path topology and detour path topology) will affect the travel time of a robot between drop-off points and loading stations (service nodes $\mu_{ls_i,i}$ and $\mu_{ls,dr}$) and the release time of a robot at a drop-off point ($\mu_{ls,dr}$). We estimate the travel time of a robot between any two neighboring drop-off points in Online Appendix A, including acceleration and deceleration of the robot and the expected waiting at each drop-off point. To capture the effect of path topology on release time (whether there is rotation), we specify the cases that a robot rotates before releasing an item into a drop-off point (see the service time estimation for detour path topology). The robot-to-loading-station assignment rule affects the visit ratios of robots to loading stations ($p_{ls_i,i}, i = 1, 2, \ldots, n_{ls}$).

Next, we calculate the first two moments of the service time of the nodes in the CQN based on the robot path topology.

4.1.1. Service Time Estimation for the RSS with Shortest Path Topology. In the service node $\mu_{ls,t}$, the first two moments of the service time are given by

$$\mu_{ls,t}^{-1} = \frac{1}{c_{ls,t}} \sum_{j=1}^{\max (l_{ls,t}, y_{ls,t})} T_{ls,t,a_{ls,t}}^{a_{ls,t}} + g \cdot t_{ls,t}$$

Because all loading stations have the same service time $t_{ls,t}$ which is constant, we have $c_{ls,t}^2 = 0$.

After the robot receives the item to be sorted, it travels on the platform to the designated drop-off point, following the direction of aisles and cross-aisles. In the two-tier RSS with the shortest path topology, the robot can always find the shortest path to the destination by choosing the correct lane.

To travel from either a west side or east side loading station $ls_i$ to a drop-off point $d_{x,y}$, the robot first moves horizontally along cross-aisle $c_{a_{ls_i}}$ and then vertically along aisle $a_x$. The movement time from loading station $ls_i$ to the drop-off point $d_{x,y}$ can be calculated by Equation (2):

$$T_{ls_i,d_{x,y}} = T_{ls_i,d_{x,y}}^{a_{ls_i}} + T_{ls_i,d_{x,y}}^{c_{a_{ls_i}}} + \max (l_{ls,t}, y_{ls,t}) - 1 \sum_{j=1}^{\max (l_{ls,t}, y_{ls,t})} T_{ls,t,a_{ls,t}}^{a_{ls,t}} + g \cdot t_{ls,t}$$

where $\sum_{j=1}^{\max (l_{ls,t}, y_{ls,t})} T_{ls,t,a_{ls,t}}^{a_{ls,t}}$ is the movement time from aisle $a_1$ to aisle $a_x$ along the cross-aisle $c_{a_{ls_i}}$, and
\[ \sum_{j=\min(y, y_c)}^{\max(y, y_c)-1} T_{c y_c} \] is the movement time in cross-aisle \( c a_{y_c} \) to the drop-off point \( d_{x,y} \) (calculated in Online Appendix A). The binary variable \( g \) determines whether the robot needs to change its direction to drop off the item. If \( x \neq x_{ls} \) and \( y \neq y_{ls} \), a 90° turn is required and we have \( g = 1 \), otherwise, \( g = 0 \). The travel time between any two neighboring aisles (\( T_{c y_{ls}} \)) or cross-aisles (\( T_{c a_{ls}} \)) depends on the neighboring

**Figure 6.** (Color online) Closed Queueing Network for Performance Estimation of a Two-Tier RSS with Dedicated Robots

**Figure 7.** Closed Queueing Network for the Single-Tier RSS with Shortest Path Topology
Figure 8. Closed Queueing Network for Performance Estimation of a Single-Tier RSS with Single-Ring Topology Path

![Closed Queueing Network](image)

... drop-off points (Figure 1 in Online Appendix A). Specifically, the robot will travel freely (in constant velocity \(v_m\)) if both drop-off points are idle. It will first travel by \(v_m\) and then decelerate to zero if the first drop-off point is idle and the second one is busy. It will accelerate to \(v_m\) and then travel by \(v_m\) if the first drop-off point is busy and the second one is idle and will accelerate and then decelerate if both drop-off points are busy. We present the analysis of these cases and their corresponding travel time in Online Appendix A.

To travel from a bottom loading station \(l_{si}\) to a drop-off point \(d_{xy}\), the robot first moves in aisle \(a_{xy}\) to cross-aisle \(c_{xy}\) and then moves in cross-aisle \(c_{xy}\) to the drop-off point. The movement time can be calculated by Equation (3):

\[
T_{b_{i,d_{xy}}} = T_{b_{i,a_{xy}}} + T_{b_{i,c_{xy}}} = \sum_{j=1}^{y-1} T_{c_{i,y},c_{i,j}} + \sum_{j=\min(s,x)}^{\max(s,x,y)-1} T_{a_{i,y},a_{i,j}} + g \cdot t_r, \tag{3}
\]

Movement time from a top loading station \(l_{si}\) to a drop-off point \(d_{xy}\) can be calculated in the same way as in Equation (3). The travel time between two neighboring drop-off points depends on the probability \(p_{xy, \text{busy}}\) that drop-off point \(d_{xy}\) is occupied by another vehicle. We design an iterative procedure (Algorithm 1 in Online Appendix A) to approximate \(p_{xy, \text{busy}}\). Intuitively, more robots in the system will lead to larger \(p_{xy, \text{busy}}\) and result in lower effective robot velocity.

The probability that a robot returns to loading station \(l_{si}\) depends on the robot-to-loading-station assignment rule applied (see Section 3.3). The first two moments of the robot travel time from loading station \(l_{si}\) to a drop-off point (mean value \(\mu_{b_{i,d_{xy}}}^{-1}\) and squared coefficient of variation \(\mu_{b_{i,d_{xy}}}^{-1} \cdot \mu_{b_{i,d_{xy}}}^{-2}\)) can now be calculated by Equation (4):

\[
\mu_{b_{i,d_{xy}}}^{-1} = \sum_{i=1}^{n_i} \sum_{s=1}^{W} \sum_{y=1}^{L} P_{b_{i}} \cdot p_{d_{xy}} \cdot T_{b_{i,d_{xy}}} \cdot \mu_{b_{i,d_{xy}}}^{-2} = \sum_{i=1}^{n_i} \sum_{s=1}^{W} \sum_{y=1}^{L} P_{b_{i}} \cdot p_{d_{xy}} \cdot T_{b_{i,b_{i,d_{xy}}}} - 1. \tag{4}
\]

Calculating the robot movement time from a drop-off point to loading station \(l_{si}\) is similar to that from loading station \(l_{si}\) to a drop-off point. Therefore, we can obtain the first two moments of the service time of the node \(d_{xy, i}\) by Equation (5):

\[
\mu_{d_{xy, i}}^{-1} = \sum_{s=1}^{W} \sum_{y=1}^{L} P_{d_{xy}} \cdot T_{d_{xy, d_{xy}}} \cdot \mu_{d_{xy, i}}^{-2} = \sum_{s=1}^{W} \sum_{y=1}^{L} P_{d_{xy}} \cdot T_{d_{d_{xy, d_{xy}}}} - 1. \tag{5}
\]

At each drop-off service node \(d_{xy, i}\), we have \(\mu_{d_{xy, i}}^{-1} = \frac{1}{\mu_{d_{xy, i}}} \cdot \mu_{d_{xy, i}}^{-2} = 0\).

4.1.2. Service Time Estimation for the RSS with Detour Path Topology. In the two-tier RSS with the detour path topology, a robot can travel from a loading station to a drop-off point by the shortest path, whereas it may first take a detour first to travel from a drop-off point back to a loading station.

To travel from the loading station \(l_{si}\) to the drop-off point \(d_{xy}\), the robot first moves into the aisle or the cross-aisle facing the loading station, and then turns into the aisle or the cross-aisle that neighbors the drop-off point (see the blue lines in Figure 3(b)). Thereby the movement time \(T_{b_{i,d_{xy}}}\) can be calculated in the way as in Equations (2) and (3) (see Online Appendix A for the travel time between any two neighboring drop-off points).

To release an item into the drop-off point, the robot may first rotate 180° with probability of 0.5 that is included in each drop-off service node \(d_{xy, i}\). Depending on the position of the drop-off point and the loading station position, we have the following cases for the service time at the service node \(d_{xy, i}\):

\[
T_{d_{xy, d_{xy}}} = \begin{cases} t_{i_r} & x \text{ is odd and } y \text{ is odd,} \\ t_{i_r} + t_r & x \text{ is even and } y \text{ is even,} \\ t_{i_r} & \text{otherwise}, \end{cases} \tag{6}
\]

where rotation is not required for the first case but always required for the second case. For the third case, the drop-off point has two sides at which the item can be directly released and two sides which require rotation. Thereby, the expected rotation time is \(\frac{t_r}{2} = t_i\).

Then, the first two moments of the service time at service node \(d_{xy, i}\) can be calculated by Equation (7):

\[
\mu_{d_{xy, i}}^{-1} = \sum_{i=1}^{n_i} P_{d_{xy}} \cdot T_{d_{xy, d_{xy}}} \cdot \mu_{d_{xy, i}}^{-2} = \sum_{i=1}^{n_i} P_{d_{xy}} \cdot \left(\frac{T_{d_{xy, d_{xy}}}^2}{\mu_{d_{xy, i}}} - 1\right). \tag{7}
\]
To travel from the drop-off point $d_{k,g}$ to loading station $ls_i$, the robot may first take a detour and then take the shortest path to the loading station, depending on the location of the loading station, the drop-off point location, and the robot position. Online Appendix C gives the details of detour path calculation.

With the first two moments of the service time at all service nodes and the visiting probabilities in the CQN, we can estimate the system performance using the AMVA method (Buitenheek, van Houtum, and Zijm 2000).

### 4.2. CQN for the Two-Tier RSS with Dedicated Robot-to-Loading-Station Assignment

For the two-tier RSS that uses the dedicated robot-to-loading-station assignment rule, we assume that loading stations at the same side of the system share robots, that is, loading stations located at the north side of the system share $R_n$ robots, those located at the south side of the system share $R_s$ robots, those located at the west side of the system share $R_w$ robots, and those located at the east side of the system share $R_e$ robots. This leads to four classes of customers in the CQN. First, the worker or an automated induct puts the item on a robot (service nodes $\mu_{ls}, \mu_{ls}, \mu_{ls}, \mu_{ls}$ for north side, south side, west side, and east side loading stations). Then, the robot transports the item from the loading station to the designated drop-off point (service nodes $\mu_{ls,dr}, \mu_{ls,dr}, \mu_{ls,dr}, \mu_{ls,dr}$ correspond to the robot movement from north side, south side, west side, and east side loading stations to drop-off points, respectively). Because no waiting is required, we model this process as an IS.

The service time of service nodes $\mu_{ls}, \mu_{ls}, \mu_{ls}, \mu_{ls}$ and $\mu_{ls,dr}$ can be obtained in the same way as in Section 4.1. Because we assume random assignment among loading stations located at the same side of the system, the conditional visiting probability of a north-side loading station is $p_{ls}^n = \frac{1}{n^l}, i = 1, 2, \ldots, n^l$, the conditional visiting probability of a south-side loading station is $p_{ls}^s = \frac{1}{n^l}, i = 1, 2, \ldots, n^l$, the conditional visiting probability of a west-side loading station is $p_{ls}^w = \frac{1}{n^l}, i = 1, 2, \ldots, n^l$, the conditional visiting probability of an east-side loading station is $p_{ls}^e = \frac{1}{n^l}, i = 1, 2, \ldots, n^l$. The service nodes $\mu_{ls,dr}$ share the same visiting probabilities with service nodes $\mu_{ls,dr}$. Then, we can use the multiclass AMVA method (Buitenheek, van Houtum, and Zijm 2000) to estimate the system performance under the dedicated robot-to-loading-station assignment rule.

### 4.3. CQN for the Single-Tier RSS with Shortest Path Topology

For the single-tier RSS, build a CQN for the design with shortest path topology (Figure 7). In the system with shortest path topology, the robot travels between an input point and an output point by the shortest path (service nodes $\mu_{I,O}$ and $\mu_{O,I}$). Because no waiting is required for the movement, we model service nodes $\mu_{I,O}$ and $\mu_{O,I}$ as an infinite servers. Each input point and output point is modelled as a single server (service nodes $\mu_I$ and $\mu_O$). The number of input points is $n_I = n_I$ and the number of output points is $n_O = 2(W + L)$. The service rate of an input point is $\mu_I = \mu_{ls}$ and the service rate of an output point is $\mu_O = \mu_{ls}$. No rotation is required to release an item.

Next, we calculate the service time expression of the service nodes in the CQN and give the solution approach. In the single-tier RSS, we first specify the coordinates of all input and output points. Then, we calculate the first two moments of movement time between input and output points. Similar to the loading stations in the two-tier RSS, we index the input/output points from the top left in the clockwise direction. As we mentioned in Section 3, the number of drop-off points between any two neighboring loading stations is the same, that is, both $\frac{W}{n^I}$ and $\frac{L}{n^O}$ are integer. Then, we can calculate the coordinates of input points by the Equations (8) and (9).

$$\begin{align*}
x_I = & \begin{cases} 
\left( i - \frac{\lfloor W/n^I + 1 \rfloor}{n^I} \right) \cdot \left( W/n^I + 1 \right), & i = 1, 2, \ldots, n^I \\
W + n^I, & i = n^I + 1, \ldots, n^I + n^w \\
(2n^w + n^w - i + 1) \cdot \left( W/n^I + 1 \right), & i = n^I + n^w + 1, \ldots, 2n^I + n^w \\
0, & i = 2n^I + n^w + 1, \ldots, n_I 
\end{cases} \\
y_I = & \begin{cases} 
\left( \frac{L + n^w}{2} - i + 1 \right) \cdot \left( \frac{L + n^w}{2} + 1 \right), & i = 1, 2, \ldots, n^w \\
(2n^w - i + 1) \cdot \left( \frac{L + n^w}{2} + 1 \right), & i = n^w + 1, \ldots, n^w + n^e \\
0, & i = 2n^w + n^w + 1, \ldots, n_O 
\end{cases}
\end{align*}$$

The coordinates of output points can be calculated by the Equations (10) and (11):

$$\begin{align*}
x_O = & \begin{cases} 
i + 1 \cdot \left[ \frac{i + n^I}{W} \right], & i = 1, 2, \ldots, W \\
W + n^I, & i = W + 1, \ldots, W + L \\
(2W + L - i + 1) \cdot \left( W/n^I \right), & i = W + L + 1, \ldots, 2W + L \\
0, & i = 2W + L + 1, \ldots, 2(W + L)
\end{cases}
\end{align*}$$
In the single-tier RSS with the shortest path topology, the travel time between an input point \( (I_i) \) and an output point \( (O_j) \) consists of two parts: width direction distance \( D_{I_i,O_j}^w = |x_i - x_0| \cdot (w_{dp} + w_{side}) \) and length direction distance \( D_{I_i,O_j}^l = |y_i - y_0| \cdot (w_{dp} + w_{side}) \). In calculating the travel time, we take the acceleration and deceleration of robots into account. Online Appendix B explains how they affect the system throughput. Then, we validate our analytical models by both simulation and a real case of Deppon Express. We build simulation models for the robotic sorting system using Flexsim (Education version 2017). For the two-tier RSS, we consider two system scenarios: \( W \times L = 40 \times 25 \) with \( n_{lw}^w = 3, n_{lw}^l = 9 \) and \( W \times L = 30 \times 30 \) with \( n_{lw}^w = 6, n_{lw}^l = 6 \). For the single-tier RSS, we also consider two system scenarios: \( W \times L = 80 \times 48 \) with \( n_{lw}^w = 3, n_{lw}^l = 9 \) and \( W \times L = 56 \times 56 \) with \( n_{lw}^w = 6, n_{lw}^l = 6 \). For each system scenario, the number of robots takes three values: \( R = 80, 100, 120 \) and both random robot-to-loading-station assignment rule and closest robot-to-loading-station assignment rule are examined. Under the dedicated assignment rule, we apply the closest robot-to-loading-station assignment rule in each dedicated zone. Thereby, we only consider random and closest assignment rules in simulation. In the single-tier RSS with a single-ring path topology, the random and closest assignment rules work similarly. Therefore, we only consider the closest assignment rule for the path topology. This leads to 42 system scenarios in total. The
The parameters are obtained from the robotic sorting system of Depon Express, an express company in Shanghai, China.

### Table 3. System Parameters Used in Simulation

<table>
<thead>
<tr>
<th>$w_{wp}$</th>
<th>$w_{wide}$</th>
<th>$w_{we}$</th>
<th>$v_{we}$</th>
<th>$a$</th>
<th>$t_{lu}$</th>
<th>$t_{l}$</th>
<th>$t_{u}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(m)</td>
<td>(m)</td>
<td>(m/s)</td>
<td>(m/s$^2$)</td>
<td>(s)</td>
<td>(s)</td>
<td>(s)</td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>[3,8]</td>
</tr>
</tbody>
</table>

*Note.* The parameters are obtained from the robotic sorting system of Depon Express, an express company in Shanghai, China.

arrival of sorting orders for each drop-off point is generated following a uniform distribution, and the mean arrival rates are sufficiently large to make all loading stations and robots busy. The simulation models are described in detail in Online Appendix D and the other system parameters are presented in Table 3.

### 5.1. Congestion and Blocking Investigation

In this section, we first check the allocation frequency of a traveling point (a control point in the simulation) to zero, one, and two robots to investigate the probability of robot blocking in the two-tier RSS with the shortest and detour path topologies and the single-tier RSS with the shortest path topology. Then, we measure the robot congestion by the robot’s effective velocity in the simulation models. We use this to estimate the key parameter $k_0$ in the traffic flow function in the single-tier RSS with a single-ring path topology.

In a two-tier RSS, each drop-off point has four control points at four sides. The robot loads/unloads items at a control point. To collect the blocking frequency and avoid circular waiting of robots that may cause blocking, we set the allocation capacity of all control points to two in the simulation model, that is, two robots can claim a drop-off point at the same time. Although this setting prevents the blocking of robots to a large degree, it helps to capture the blocking frequency. Once a control point is allocated by two robots at the same time, robots may be blocked.

For the single-tier RSS with a single-ring path topology, we approximate robot congestion by a traffic flow function. The random robot-to-loading-station assignment rule and the random distribution of demands among drop-off points are used. We consider the two-tier RSSs (shortest path topology and detour path topology) with $R = 120$ robots and the single-tier RSS with shortest path topology and $R = 120$ robots. For each system scenario, we first collect the allocation frequency of eight selected control points (six in the middle and two at perimeters), and then collect the allocation frequency of eight most congested control points. The simulation model runs 100 hours with a warmup period of 12 hours. We collect the number of times these control points are allocated to zero, one or two robots and present the frequency in Figures 9 and 10.

The results show that the frequency at which these selected drop-off points are allocated by two robots is less than 6%, and the mean value of this frequency is 2.83%. The frequency at which the eight most congested control points are allocated to two robots is less than 10%, and the mean value of this frequency is 8.37%. Therefore, although the robot blocking in the most congested area is significant, robot blocking in the entire system occurs infrequently. In addition, blocking duration is very short, as dropping off an item take only $t_{lu}$ seconds. Therefore, simulating the system without blocking is acceptable.

Next, we examine the robot’s effective velocity (denoted by $EV_R(m/s)$) in the system. For both single-tier and two-tier RSSs, we vary the number of robots $R$ from 80 to 200 with a stepsize of 20. We collect the time percentage of empty traveling and loaded traveling of robots and the total travel distance of a robot. We obtain the effective velocity of robots by dividing the total travel distance by the travel time. The results are presented in Figure 11. The results suggest that $EV_R$ decreases slightly with the number of robots $R$ in the two-tier RSS with either the shortest or detour path topology and the single-tier RSS with the shortest path topology, while it decreases significantly with the number of robots $R$ in the single-tier RSS with the single-ring path topology. This again suggests that we can neglect the effect of robot blocking, except for the single-tier layout with a single-ring path topology.

We use the method of capturing the expected occupancy of drop-off points to approximate robot congestion in the two-tier RSS (see Online Appendices B and C), and adopt the traffic flow function to estimate robot congestion in the single-tier RSS (see Equation (14)). We ignore the level-$k$, $k \geq 2$ blocking in our analytical model, that is, a robot that is already blocked also causes blocking, since we rarely find these situations in simulation.

### 5.2. Analytical Model Validation by Simulation

We also use the 42 system scenarios specified in the last section to validate the accuracy and effectiveness of our analytical models. The simulation model produces the topology path of robots, following the logic of the AGVNetwork Navigator (see Online Appendix D). It directs robots dynamically, based on the AGV path properties and the current status of all robots. To avoid robots congestion and deadlock, a robot may take more turns than the static topology path proposed in our analytical models. This will not only increase the turning time but also decrease the effective velocity of the robot (because of acceleration/deceleration).

For each scenario, we ran 30 replications with a warm-up period of 12 hours and a run time of 100 hours per replication, leading to 95% confidence intervals where the half-width is within 2% of the average. We collect four performance measures, including the total throughput capacity $TC$ (per hour), the travel
distance of robots \((m)\), the time percentage of empty robot traveling, and the time percentage of loaded robot traveling. The accuracy of the analytical models is measured by the absolute relative errors of analytical results to simulation results, denoted by \(\delta_{TC}\). 

\[
\delta_{TC} = \frac{|TCA - TCS|}{TCS} \times 100\%,
\]

where \(TCA\) and \(TCS\) are the analytical throughput capacity and the average simulation throughput capacity, respectively.

Figure 12 presents the distributions of relative errors between analytical and simulation results. Tables 4 and 5 present the detailed results of \(\delta_{TC}\). The results show that the relative errors between analytical results and simulation results are acceptable, the maximum relative error is about 13.2% and the average relative error is 4.6%. We can also extend our CQN into an SOQN to validate other performance measures, using the method described by Roy (2016) or Van der Gaast et al. (2020). Specifically, we validate the SOQN model for order cycle time and expected waiting time of orders for robots in Online Appendix D.

5.3. Analytical and Simulation Model Validation by a Real Case

In this section, we investigate the real case of Deppon Express, in Shanghai, China, to validate both our analytical model and a simulation model. The layout of the real system is depicted in Figure 6(a) in Online Appendix F. The system uses six loading stations and 170 robots to serve the orders of 108 destinations (allocated to \(W \times L = 18 \times 6\) drop-off points). Each aisle and cross-aisle has at least two lanes that ensures the robot can travel by the shortest path between loading stations and drop-off points. The system allocates the largest demands to the drop-off points that are far...
from loading stations to prevent robot congestion. It also assigns a robot that finishes an order back to a fast loading station. We approximate this robot-to-loading-station assignment rule by setting \( p_{ls} = \mu_{ls} / \sum_{i=1}^{nls} \mu_{ls} \) in both the analytical and simulation models. We use the demands of drop-off points during peak hours in both the analytical and simulation models. The simulation model is similar to that of the previous section and also implemented in Flexsim. The simulation runs for 100 hours with a warm up period of 12 hours. Table 6 presents the validation results, where \( \delta_{A} \) and \( \delta_{S} \) represent the relative error between the throughput capacity of the Deppon system, the analytical model and the simulation model, respectively. \( \delta_{D} = 100\% \). \( \frac{TC_{A} - TC_{S}}{TC_{A}} \) is the relative error between the throughput capacity of the real system and that of the analytical model, \( \delta_{D} = 100\% \). \( \frac{TC_{D} - TC_{S}}{TC_{D}} \) is the relative error between the throughput capacity of the real system and that of the simulation model. Results show that the simulation model and the analytical model can provide an accurate performance estimation of an RSS.

6. Numerical Experiments

In this section, we conduct numerical experiments to optimize system size, investigate operating policies, and minimize system cost.

6.1. Optimal System Size of an RSS

In an RSS with \( N_{dp} \) drop-off points and \( n_{ls} \) loading stations, the system size \( W, L \) may affect throughput capacity. We conduct numerical experiments for both the single-tier and two-tier RSS to optimize system size, in terms of system throughput capacity. We consider a series of scenarios by varying the number of drop-off points \( N_{dp} \), the number of loading stations \( n_{ls} \), the number of robots \( R \), and the velocity of robot \( v_{m,a} \). We present the details in Table 7. We allocate
the loading stations along the perimeter of the system, based on
\[ \frac{n_{ls}}{W L} = \frac{\lfloor n_{ls}/2 \rfloor}{W L}. \]
In total, we have 240 system scenarios to investigate the optimal system size. For each system scenario, we consider both single-tier (shortest path topology and single-ring path topology) and two-tier (shortest path topology and detour path topology) RSSs, coupled with the random, closest and dedicated robot-to-loading-station assignment rules. To explore the optimal system size, we vary the system width \( W \) from 0 to \( 0.45 \cdot N_{dp} \) with a stepsize of one for each system scenario.

We present the statistical results of optimal system size in Table 8. We have the following observations based on the results:

1. For the single-tier RSS and the two-tier RSS with the shortest path topology, a square layout \( (r^* \approx 1) \) may maximize system throughput capacity. For the two-tier RSS with the detour path topology, the optimal length-to-width ratio is \( L^*/W^* \approx 1.25 \). The results are stable under the confidence level \( \alpha = 0.05 \).

2. Although we can find an optimal width-to-length ratio \( r^* \), the maximum system throughput capacity \( TC \) decreases only slightly if we take a width-to-length ratio that is different but still close to \( r^* \). This means that the designer of an RSS can take a width-to-length ratio of around one, and it will not go wrong too much.

3. The optimal system layout is insensitive to the number of robots \( R \) and the number of loading stations \( n_{ls} \), but can be significantly affected by the allocation of loading stations \( n_{ls}^{\text{ref}} \) and \( n_{ls}^{\text{adj}} \) and the robot-to-loading-station assignment rule. We compare the performance of robot-to-loading-station assignment rules in more detail in next section.

### 6.2. Operating Policies in an RSS

One of the important decisions that needed to be taken at Deppon Express was the assignment of robots to loading stations. In peak hours, about 150 robots are deployed (maximum 170 are available), whereas at off-peak hours, only 80 robots are used. Therefore, in
this section, we investigate the operating policies examined in this paper by numerical experiments, including the robot-to-loading-station assignment rule and the path topologies.

We first compare the performance of the random (Ra), dedicated (De), closest (Cl), and shortest-robot-queue (Sq) assignment rules in both two-tier and single-tier RSSs. Under the dedicated assignment rule, we assign a robot to the closest loading station in each dedicated zone (see Section 4.2 for the determination of dedicated zones). The shortest-robot-queue assignment rule cannot be evaluated by a CQN model; instead, we use Figure 12.

![Figure 12](Color online) Distribution of Relative Errors Between Analytical and Simulation Results

![Diagram](a) Two-tier RSS with shortest path topology

![Diagram](b) Two-tier RSS with detour path topology

![Diagram](c) Single-tier RSS with shortest path topology

![Diagram](d) Single-tier RSS with single-ring path topology

Notes: (a) Two-tier RSS with shortest path topology. (b) Two-tier RSS with detour path topology. (c) Single-tier RSS with shortest path topology. (d) Single-tier RSS with single-ring path topology.

### Table 4. Simulation Validation Results of Two-Tier RSS

<table>
<thead>
<tr>
<th>Topology path</th>
<th>40 × 25</th>
<th>30 × 30</th>
<th>40 × 25</th>
<th>30 × 30</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Shortest path topology</td>
<td>Detour path topology</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TC_A (h)</td>
<td>TC_S (h)</td>
<td>δ_{TC} (%)</td>
</tr>
<tr>
<td>Assignment rule</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random</td>
<td>R = 80</td>
<td>1,903.9</td>
<td>2,041.7</td>
<td>6.8</td>
</tr>
<tr>
<td></td>
<td>R = 100</td>
<td>2,385.9</td>
<td>2,445.0</td>
<td>2.4</td>
</tr>
<tr>
<td></td>
<td>R = 120</td>
<td>2,867.8</td>
<td>2,779.8</td>
<td>3.2</td>
</tr>
<tr>
<td>Closest</td>
<td>R = 80</td>
<td>2,665.5</td>
<td>2,728.0</td>
<td>6.8</td>
</tr>
<tr>
<td></td>
<td>R = 100</td>
<td>3,340.3</td>
<td>3,636.7</td>
<td>8.2</td>
</tr>
<tr>
<td></td>
<td>R = 120</td>
<td>4,015</td>
<td>4,076.7</td>
<td>1.5</td>
</tr>
</tbody>
</table>
Random Closest comparison results:

curve (ber of robots. The closest assignment rule is inferior for a large num-
signment rules are superior, but differences are small.

number of robots, the random and shortest-queue as-

shortest path topology, and the system scenario is
10. For the single-tier RSS, we only consider the

W × L /equals /equals /equals

Layout

Two-tier 18 6 6 170 7,163 6,907 7,054 3.4% 1.5%

Table 6. Real Case Validation Result

Table 5. Simulation Validation Results of Single-Tier RSS

<table>
<thead>
<tr>
<th>Topology path</th>
<th>Shortest path topology</th>
<th>Single-ring path topology</th>
</tr>
</thead>
<tbody>
<tr>
<td>W × L</td>
<td>80 × 48</td>
<td>56 × 56</td>
</tr>
<tr>
<td>Assignment rule</td>
<td>TC_A (h)</td>
<td>TC_S (h)</td>
</tr>
<tr>
<td>Random</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R = 80</td>
<td>1,411.9</td>
<td>1,522.9</td>
</tr>
<tr>
<td>R = 100</td>
<td>1,871.6</td>
<td>1,871.1</td>
</tr>
<tr>
<td>R = 120</td>
<td>2,191.4</td>
<td>2,158.6</td>
</tr>
<tr>
<td>Closest</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R = 80</td>
<td>2,397.9</td>
<td>2,345.6</td>
</tr>
<tr>
<td>R = 100</td>
<td>2,976.6</td>
<td>2,911.5</td>
</tr>
<tr>
<td>R = 120</td>
<td>3,394.5</td>
<td>3,593.9</td>
</tr>
</tbody>
</table>

Note. The parameter k_0 = 0.56 in the single-tier RSS.

Flexsim simulation to evaluate it. For the two-tier RSS, we consider the system with W = 20, L = 20 and n^r_2 = 3, n^s_2 = 3. Both the detour path topology and the shortest path topology are considered. We vary the number of robots R from 0 to 250 with a stepsize of 10. For the single-tier RSS, we only consider the shortest path topology, and the system scenario is W = 50, L = 50, n^r = 3, n^s = 3. We vary the number of robots R from 0 to 150 with a stepsize of 10. We present the relationship between the throughput capacity TC and the number of robots in Figure 13.

We have the following observations based on the comparison results:

1. For a small number of robots, the dedicated assignment rule is superior for all configurations, for a large number of robots, the random and shortest-queue assignment rules are superior, but differences are small. The closest assignment rule is inferior for a large number of robots.

2. There exists an intersection between the random assignment rule curve and the closest assignment rule curve (R^CS). When R ≤ R^CS, the closest assignment rule performs better. This can be explained as the closest assignment rule can reduce robot travel time, but increases the expected waiting time of robots at loading stations. When the number of robots is small, the reduction in travel time dominates, otherwise, the situation reverses.

3. We can find a similar tradeoff between the shortest-robot-queue assignment rule and the closest assignment rule or dedicated assignment rule. The shortest-robot-queue assignment rule provides larger throughput capacity than either the closest or dedicated assignment rules when the number of robots used is larger than the intersection point (R^CS or R^CS).

Based on these insights, Deppon tested both random and closest assignment rules in its sorting system and found that allocating robots randomly to loading stations in peak hours (typically 150 robots are used) outperforms the closest assignment by 12.4%, whereas the closed assignment outperforms the random assignment in off peak hours (80 robots) by 8.9%. Deppon has therefore adopted this dynamic allocation for a sustained high performance.

Next, we analyze robot path topologies in both the single-tier and two-tier RSSs. For the two-tier RSS, we consider two system scenarios: N^dp = 600, n^s = 16 and N^dp = 800, n^s = 20. The number of robots varies from 100 to 250 with a stepsize of 50. For the single-tier RSS, we consider two system scenarios: N^dp = 200, n^s = 8 and N^dp = 400, n^s = 16. The number of robots R varies from 50 to 200 with a stepsize of 50. For each system scenario, we first find the best robot-to-loading-station assignment rule and the optimal system size in the same way as in Section 6.1, and then compare the maximum system throughput capacity under different assignment rules. The results are presented in Table 9. It shows that the detour path topology outperforms the shortest path topology in terms of system throughput capacity in the two-tier RSS, whereas the shortest path topology dominates the detour path topology in the single-tier RSS.

Table 7. System Scenarios for Optimal Size Investigation

<table>
<thead>
<tr>
<th>N^dp</th>
<th>n^s</th>
<th>R</th>
<th>v_m (m/s), a_m (m/s^2)</th>
<th>W</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>8,12,16,20</td>
<td>50,70,90,110</td>
<td>(3,1.5),(2.1),(1,0.5)</td>
<td>5–50</td>
</tr>
<tr>
<td>400</td>
<td>12,16,20,24</td>
<td>80,100,120,140</td>
<td>(3,1.5),(2.1),(1,0.5)</td>
<td>10–100</td>
</tr>
<tr>
<td>600</td>
<td>16,20,24,28</td>
<td>110,130,150,170</td>
<td>(3,1.5),(2.1),(1,0.5)</td>
<td>15–150</td>
</tr>
<tr>
<td>800</td>
<td>20,24,28,32</td>
<td>140,160,180,200</td>
<td>(3,1.5),(2.1),(1,0.5)</td>
<td>20–200</td>
</tr>
<tr>
<td>1,000</td>
<td>24,28,32,36</td>
<td>170,190,210,230</td>
<td>(3,1.5),(2.1),(1,0.5)</td>
<td>25–250</td>
</tr>
</tbody>
</table>
6.3. Sensitivity Analysis on the Arrival Rate of Sorting Orders

This section conducts a sensitivity analysis on the arrival rate of sorting orders, during both off-peak hours and peak hours. We aim to investigate the effect of the arrival rate of sorting orders on both the minimum number of robots required (denoted as $R$) and the preferred assignment of robot-to-loading-station. The system scenarios are the same as those in last section, that is, $W = L = 20, n_{ls}^{in} = n_{ls}^{out} = 3$ in the two-tier layout with 150 robots and $W = L = 50, n_{ls}^{in} = n_{ls}^{out} = 3$ in the single-tier layout with 100 robots. In each layout, we use three uniform distributions to generate the mean arrival rate $\lambda_{x,y}$ (per hour) of each drop-off point, corresponding to a low-demand, a middle-demand and a high-demand (Table 10). We obtain $R$ and the preferred assignment rule by the following method: For a given required throughput capacity $TC_{min}$, we first look for the minimum number of robots required for the random, closest and dedicated assignment rule (denoted as $R_{r}, R_{c}, R_{d}$) by increasing the number of robots $R$ from a very small value until the system throughput capacity $TC \geq TC_{min}$. Then, we choose the assignment rule that provides the minimum number

**Table 8.** System Scenarios for Optimal Size Investigation

<table>
<thead>
<tr>
<th>System layout</th>
<th>Topology path</th>
<th>Assignment rule</th>
<th>$r^* = \frac{W}{L}$ ($\sigma = 0.05$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two-tier</td>
<td>Shortest path</td>
<td>Random</td>
<td>1.05 (0.91,1.18)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Closest</td>
<td>1.09 (0.91,1.26)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dedicated</td>
<td>1.07 (0.95,1.20)</td>
</tr>
<tr>
<td></td>
<td>Detour path</td>
<td>Random</td>
<td>0.85 (0.77,0.94)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Closest</td>
<td>0.84 (0.79,0.88)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dedicated</td>
<td>0.82 (0.79,0.88)</td>
</tr>
<tr>
<td>Single-tier</td>
<td>Shortest path</td>
<td>Random</td>
<td>1.07 (0.95,1.20)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Closest</td>
<td>1.09 (0.95,1.04)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dedicated</td>
<td>0.94 (0.73,1.27)</td>
</tr>
<tr>
<td>Single-ring</td>
<td>Random</td>
<td>1.13 (0.85,1.30)</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 13.** (Color online) Performance Comparison of Assignment Rules in an RSS

Notes. (a) Two-tier RSS with shortest path topology. (b) Two-tier RSS with detour path topology. (c) Single-tier RSS with shortest path topology. Ra, random assignment rule; Sq, shortest-robot-queue assignment rule; Cl, closest assignment rule; De, dedicated assignment rule.
of robots $R = \min\{R_r, R_c, R_d\}$ as the preferred assignment rule. The CQN particularly fits such an exploration, because it runs fast to estimate the maximum throughput capacity of a system (less than one second) with a given number of resources (loading stations, robots), for a given system size, layout, and control rules. It can therefore also be used in a reverse form: to evaluate for a desired throughput capacity, the expected number of required robots, the required number of sorting orders, both for peak and nonpeak loads, or the effect of different control rules on throughput capacity.

The results show that the demand level significantly affects the number of robots needed to meet the required number of sorting orders. Moreover, the preferred assignment of robot-to-loading-station also depends on the required system throughput capacity.

### 6.4. Cost Minimization of an RSS

In this section, we investigate the selection between the shortest path topology and the detour path topology in terms of system cost, considering the cost minimization problem of an RSS with a requirement on throughput capacity $TC \geq TC_{\text{min}}$.

The number of destinations ($N_{dp}$) and the corresponding order arrival rate $\lambda_{x,y}$ are both given. The system throughput capacity should satisfy the requirement $TC \geq TC_{\text{min}} = \sum_{x=1}^{W} \sum_{y=1}^{L} \lambda_{x,y}$, and the minimum number of loading stations $n_{ls}$ is $\lceil TC_{\text{min}} / \mu_{ls} \rceil$. The objective is to design the system to minimize the system total annual cost $TC$, comprising the cost of robots, loading stations, and floor space. For each combination of layout with path topology, we investigate the total annual cost minimization model (Equations (16) and (17)):

$$
\min C(W, L, R, AR, n_{ls}) = C_R \cdot R + C_{ls} \cdot n_{ls} + C_{FS} \cdot W \cdot L \cdot (w_{dp} + w_{aisle})^2
$$

(16)

$$
\begin{align*}
N_{dp} \leq W \cdot L, \\
N_{dp}^{ST} \leq 2 \cdot (W + L), \\
TC \geq TC_{\text{min}}, \\
n_{ls} \geq \left\lceil \frac{TC_{\text{min}}}{\mu_{ls}} \right\rceil, \\
W \leq \bar{W}, L \leq \bar{L}, \\
N_{dp}, w_{aisle}, w_{dp} & \text{ are given}
\end{align*}
$$

(17a-17f)

Table 10. Sensitivity Analysis on the Arrival Rate of Sorting Orders

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$R$</td>
<td>48</td>
<td>72</td>
<td>145</td>
<td>51</td>
<td>92</td>
<td>142</td>
<td>34</td>
<td>48</td>
<td>96</td>
</tr>
<tr>
<td>AR</td>
<td>Dedicated</td>
<td>Dedicated</td>
<td>Random</td>
<td>Dedicated</td>
<td>Dedicated</td>
<td>Dedicated</td>
<td>Dedicated</td>
<td>Dedicated</td>
<td>Random</td>
</tr>
</tbody>
</table>

*Note. AR, preferred assignment rule.*
annualized cost per square meter floor space. Constraints (17a) and (17b) ensure that the total number of drop-off points in both the two-tier layout (NTTdp) and the single-tier layout (NSTdp) is adequate for all destinations. Constraint (17c) ensures that the system throughput capacity meets the requirement $TC \geq T_{Cmin}$. Constraint (17d) ensures that the capacity of all loading stations satisfies the throughput capacity requirement. Constraints (17e) limit the size of the platform. The decision variables are the system size $(W, L)$, the number of robots $R$, the number of loading stations $nls$, and the robot-to-loading-station assignment rule.

As an example, we consider the cost minimization of a two-tier RSS with $N_{dp} = 400$ and a single-tier RSS with $N_{dp} = 200$. Moreover, we compare the performance of an RSS with a conventional cross-belt sorter (CBS; Online Appendix F). Considering the variation on the price of warehouse floor space per square meter (denoted as $P_{FS}$), we vary $P_{FS}$ from €350 to €800 with a stepsize of 150, annualized in 30 years. The price of a robot and a loading station is €20,000 and €10,000, respectively, annualized in 10 years. We consider an interest rate $IR = 0.5\%$. Then, the annual costs of a robot, a loading station, and a square meter warehouse floor space are

$$C_R = \sum_{i=1}^{10} \frac{20000(1 + IR)^{-i}}{10}, \quad C_{ls} = \sum_{i=1}^{10} \frac{10000(1 + IR)^{-i}}{10},$$

$$C_{FS} = \sum_{i=1}^{30} \frac{P_{FS}(1 + IR)^{-i}}{30}.$$

Figure 14 presents the results of the two-tier RSS, and Table 4 in Online Appendix F includes the detailed costs. We have the following observations:

- A critical required throughput capacity exists for the shortest path topology and the detour path topology curves (denoted by $T_{CSP}$). When $T_{Cmin} \leq T_{CSP}$, the detour path topology outperforms the shortest path topology in terms of system cost, otherwise, the shortest path topology performs better. This can be explained because the saving on floor space cost of the detour path topology dominates the incremental investment in robots when $T_{Cmin} \leq T_{CSP}$. Conversely, the saving in robot investment dominates the incremental investment in floor space, when $T_{Cmin} \geq T_{CSP}$. 

Figure 14. (Color online) Cost Minimization of Two-Tier RSS and CBS
With the increase of $P_{FS}$, $T_{SP}^{DP}$ becomes larger. This implies that the advantage of the detour path topology will become less obvious with an increase in floor space cost.

We can also find a critical throughput capacity point for the RSS and the CBS curves (denoted by $T_{CBS}^{SP}$ for the RSS with shortest path topology and $T_{CBS}^{DP}$ for the RSS with detour path topology). When $T_{min} \leq T_{CBS}^{DP}$ or $T_{min} \leq T_{CBS}^{SP}$, the RSS with shortest path topology outperforms the CBS in terms of system cost. Otherwise, the CBS is cheaper. This can also be explained by the difference between the investment in transporting capacity of the two systems.

### 7. Conclusions and Further Work

This study considers a new RSS. In such a system, robots run on a platform to transport items from loading stations located at the perimeter of the system to the drop-off points located in the middle or along the perimeter, depending on the system layout.

We focus on performance estimation and operating policies of the RSS, considering robot congestion. We examine both single-tier layout and two-tier layout. For each layout, we investigate the shortest path topology and the detour path topology and examine random, closest, dedicated and shortest-queue-based robot-to-loading-station assignment rules. Closed queueing
networks are built to estimate the maximum throughput capacity of an RSS under different combinations of layout with robot path topology. We design an iterative algorithm to approximate robot congestion in a two-tier layout and use a traffic flow function to estimate robot congestion in a single-tier layout. Maximum traffic density is obtained through simulation. We use simulation and a real case to investigate robot congestion and to validate the analytical models. The results show that robot congestion plays a significant role in the single-tier RSS with the single-ring path topology. The relative errors show that our analytical models can accurately estimate the system performance.

We use the analytical model to optimize the system size and analyze the operating policies in terms of throughput capacity and operating cost. The results show that a square layout is preferred for the shortest path topology and a rectangular layout is beneficial for the detour path topology. The comparison among different robot-to-loading-station assignment rules shows that assigning a robot to either a random loading station or that with the shortest robot queue is superior in throughput capacity when the number of robots is smaller than a critical point, otherwise, the closest and dedicated assignment rules perform better. These insights are also implemented in the sorting system of Deppon Express to improve the throughput capacity in peak and off-peak periods. The cost minimization results show that while the shortest path topology outperforms the detour path topology in terms of throughput capacity, the detour path topology is cheaper when the required system throughput capacity is smaller than a critical point. Moreover, we compare the RSS with a cross-belt sorter system in terms of operating cost. Our results show that the RSS is cheaper than the cross-belt sorter system when the required throughput capacity is smaller than this control point.

For future studies, it is interesting to analyze the distribution of demand among destinations, the tradeoff between expected travel time saving by assigning large demands to drop-off points that are close to loading stations, and robot congestion relief caused by assigning large demands to drop-off points that are far loading stations. Moreover, it may be interesting to study the assignment of parcels to loading stations, as this impacts the load balance and the travel distance of robots, and different robot battery recovery policies as these impact throughput capacity and parcel throughput time.

Acknowledgment
The authors thank Deppon Express for providing the AGV sorting system data.

References


