2018

An Integrated Network Modeling Framework for Analysis of Multi-line Order Pick Systems

Debjit Roy
Indian Institute of Management, Ahmedabad, debjit@iima.ac.in

Vishal Bansal
Indian Institute of Management, Ahmedabad, fpm16vishalb@iima.ac.in

An Integrated Network Modeling Framework for Analysis of Multi-line Order Pick Systems

Debjit Roy
Production and Quantitative Methods
Indian Institute of Management Ahmedabad
Ahmedabad, India
debjit@iima.ac.in

Vishal Bansal
Production and Quantitative Methods
Indian Institute of Management Ahmedabad
Ahmedabad, India
fpm16vishalb@iima.ac.in

***Abstract*—Due to demanding service levels in E-commerce order fulfillment, modeling and analysis of order picking processes in warehouses deserve special attention. With a particular focus on multi-line E-commerce orders, we develop modeling tools that are necessary to analyze the order consolidation delays in the downstream pick stations. We develop a queuing network modeling framework for integrated analysis of upstream (storage system) and downstream (pick system). We apply our modeling approach to an integrated order-pick system that includes a shuttle-based storage and retrieval system, and a single pick station. Using simulations, we test the effect of the storage system configuration on the order throughput time.**

***Index Terms*—E-commerce, order picking, multi-line orders, queuing**

I. Introduction and Background

E-commerce order fulfillment has always been challenging due to large SKU counts with a very long and slow-moving tail demand distribution, high demand variability (in particular extreme peak season volumes), and high penalty for poor delivery performance, potentially resulting in the brand damage. To address these challenges, innovative material handling solutions such as shuttle-based storage and retrieval systems and mobile fulfillment systems (developed by Kiva Systems, now part of Amazon Robotics) have been developed (see [11] for an overview of automated and robotized material handling solutions).

Warehouses must be able to provide fast unit-load operations and handle order fulfillment operations in an efficient, responsive, and flexible manner. Automated storage and order pick system technologies can aid to manage the picking process. In an automated E-commerce warehouse, pallets unloaded from incoming trucks are stored in reserve storage and pallets pick area. The storage system could be an Automated Storage and Retrieval System (AS/RS). These pallets are destacked and stored in the item tote storage area for order picking. When a customer orders a particular item, the corresponding item tote is retrieved and dispatched to the order pick station in a particular sequence. Finally, the items are picked from the totes at the pick stations (that are arranged in parallel or in series), packed, and shipped. There are several studies on analyzing the performance of automated pallet or tote storage systems, on developing system-specific designs, and on obtaining operational insights (e.g. [1]; [3]; [5]).

Several analytical models have been built for analyzing the performance of the upstream storage systems with shuttle-based storage and retrieval system (SBS/RS). For example, Malmborg [8] propose an analytical model to estimate the performance of SBS/R system as a function of the system characteristics such as storage space configuration, number of vehicles, and storage capacity. Fukunari and Malmborg [9] use a network queuing approach for estimating the performance measures of SBS/R system using opportunistic interleaving. Roy et al. [1] develop a queuing network model of a single-tier vehicle-based storage and retrieval system, and analyzed the effect of design choices such as tier depth-to-width ratio on throughput time performance. Roy et al. [13] also expand the model to incorporate blocking effects on the throughput time performance in a single tier. Several other studies have been performed on analyzing the performance of tier-captive systems with multiple-tiers (for example, see [12] and [13]). Ekren et al. [10] model the SBS/R system as a single-class, multiple server semi-open queuing network (SOQN) and solve it using matrix-geometric method (MGM) to estimate the performance measure. However, these studies only analyze the performance of isolated storage systems, without considering any interaction effects between the storage and the downstream pick station. The focus of this research is on the performance analysis of integrated tote storage and order-pick system (see Figure 1).

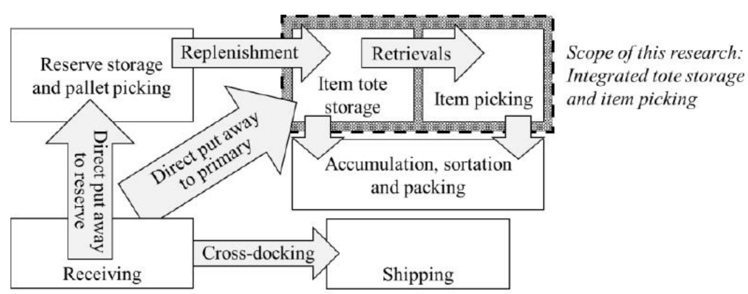

Fig. 1. Scope of this research (adapted from Tompkins et al. [2])

In goods-to-man warehouses, automated storage and retrieval systems (upstream) work together with semi-automated pick workstations (downstream). Upon order arrival, these automated storage systems fetch the totes and dispatch them to the pick modules via conveyors. There are several choices of upstream storage technologies such as crane-based or shuttle-based storage and retrieval systems. The impact of the system type on downstream order picking performance is largely unknown. Using a queuing network model, Tappia et al. [6] study the effect of storage system technology on order throughput times, and the effect of the picking station input buffer size on order picking performance. Their results indicate that using a shuttle-based technology instead of a crane-based storage systems yields investment cost savings (i.e., fewer aisles in the storage area and fewer picking stations), paired with a lower total throughput time at a given order arrival rate. However, their model assumes that all orders are single-line orders i.e., orders require only one item tote to be fulfilled. On the other hand, a multi-line order requires more than one item totes to complete an order.

Traditionally, E-commerce order pick systems were designed for handling single-line orders. However, E-commerce warehouses are witnessing a steady rise in the volume of multi-line orders. To provide free shipping, E-commerce providers are enforcing a minimum value on the orders. To avail free shipping, customers are now placing bundled orders. It is not clear if the system design parameter settings for a single line order are still optimal for a multi-line order. Further, it is not evident, if the optimal technology choice for storage system with single-line orders remains the same when the system operates with multi-line orders.

Using queuing network models, we analyze the design choices for integrated order pick systems. We develop an analytical framework to model integrated tote storage and order picking. The downstream order consolidation stations are modeled as synchronization queues. The remote Order Picking System (OPS) is modeled as a semi-open queuing network (SOQN). This queuing network captures the order waiting time at the external buffer and also allows to investigate the effect of the traffic intensity at the whole OPS system. While we consider shuttle-based storage and retrieval systems, our model can be adapted for analysis of other storage system technologies such as AS/RS.

The rest of the paper is organized as follows. In Section II, we discuss the system description and the modeling assumptions. We include the model notations and describe the model in Section III. We discuss the numerical experiments in Section IV and provide the concluding remarks in Section V.

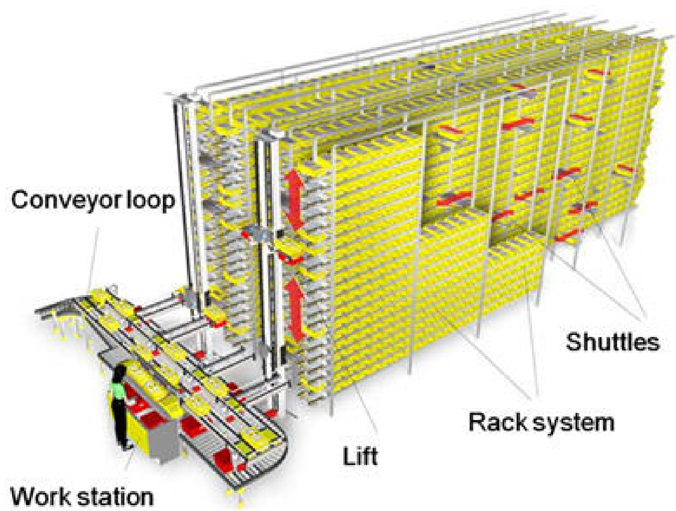

Fig. 2. Overall system view with storage and picking (Source: KNAPP Logistics Automation, Inc.)

Fig. 3. Closer view of an aisle with shuttles (Source: Vanderlande industries)

II. System Description and Modeling Assumptions

Figure 2 illustrates the full view of the remote order picking system. It consists of three major components: Shuttle-based storage and retrieval system (SBS/R), workstation for the human operator for order picking, and a closed loop conveyor connecting the storage and retrieval system to the order pick station. The storage and retrieval system composed of multiple aisles where each aisle consists of multiple tiers and each tier has multiple columns for storage. Figure 3 provides a closer look at a typical aisle which has one shuttle for every tier.

There are two major blocks of this integrated system - *Upstream System* and *Downstream System*. An upstream system performs the fetching of item totes from storage locations and then transfers them to the downstream order picking system (OPS), a pick station. We illustrate the model for a single choice of the upstream system - a tier-captive SBS/R system. Once the item tote reaches the OPS, the following pick activities are performed: item picking from item totes, consolidating the multi-line items for an order, and then dispatching the completed orders to the packaging station. Note that the proposed modeling framework can be deployed for other types of storage system as well.

To model the integrated system, we include the following assumptions:

- The item totes are stored and fetched in the high-bay storage locations in (from) a random location.
- Each item tote holds a sufficient number of units for each item.
- The processing of jobs at shuttles, lifts, and pick station resources is carried out on a First Come First Serve (FCFS) basis.
- The order arrival process is Poisson.
- All incoming orders are of multi-line category and they require exactly two item totes to be fulfilled. Note that we assume two items per order only for the ease of illustration.
- An order is characterized by its *Order Profile* which contains the information about the item totes required by the order and how many units of an item should be picked from the corresponding item tote.
- There is only one order picking station (OPS) in the downstream network.
- The conveyor takes a deterministic time for transferring an item tote from entrance node to the picking station.
- The processing time at an entrance node to conveyor system and at the picking station follows an exponential distribution.

III. Model Description

We adopt the notations used by Tappia et al. [6]. K, K_u, and K_d denote the maximum number of customer orders in the integrated system, maximum number of customer orders in the upstream system, and maximum number of customer orders in the downstream system, respectively. The term λ, denotes the order arrival rate to the system. The terms N_a, N_t, and N_c denote the number of aisles, tiers, and columns in the upstream system, respectively, while the term N_p denotes the number of picking stations in the downstream system. In the upstream system, μ_S^{-1}, σ_S^2 and μ_L^{-1}, σ_L^2 correspond to the mean and variance of the service times for the shuttles and lift, respectively. FT_L and FT_S denote the fixed time required for loading/unloading the item tote to/from the lift and shuttle, respectively. Also, the notations v_L, a_L, v_S, a_S denote the velocity and acceleration/deceleration rates of lift and shuttle, respectively. On the other hand, loc_w and loc_h denote the unit width clearance and unit height clearance per storage location. In the downstream system, μ_{ent}^{-1}, t_{conv}, and μ_{pick}^{-1} denote the mean service time at the entrance node for conveyor, the mean service time of the conveyor, and the mean service time at the pick station, respectively.

We permit a maximum of K order tokens in the system. If there are already K orders flowing in the system, then the customer orders (on arrival) queue at Buffer B_1. If a token is available, the order on arrival is matched with an order token available in B_2 and they visit first the upstream system. The item totes corresponding to an order is fetched from the upstream system. After fetching the item tote, the item tote is transferred along with the token to the downstream pick station. Once the item picks corresponding to all line items for an order is complete, the token is released to Buffer B_2 and is available to process another transaction. The consolidated order is released for packaging. Note that we consider a dual-command cycle in the upstream station i.e., the returning item tote (after picking) is first stored at an open location present in the destination tier and aisle before fetching the item tote for another order. This network is a special class of queuing network known as a semi-open queuing network (see Figure 4). The network can be viewed as an open network with respect to the external arrival of the customer orders and the network can also be viewed as a closed network because the population of customers in the system is constant, K order tokens.

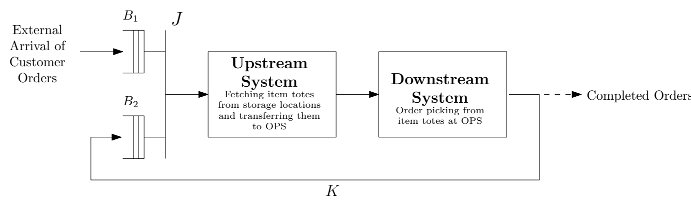

Fig. 4. Integrated queuing network model (adapted from Tappia et al. [6])

In the integrated system, modeling the upstream network can be quite complex depending on the type and number of resources and routing of tokens in the network. Likewise, modeling the downstream network can be quite complex depending on the configuration of the pick stations (series, parallel, or a combination). To analyze the integrated system, we adopt an aggregation approach. We first develop two detailed closed queuing network models of the upstream and the downstream system, respectively. We then obtain the throughput of the closed queuing networks for varying population using approximate mean value analysis algorithm. Using network reduction techniques, we substitute the upstream and the downstream system with two load-dependent queues operating in tandem. We now discuss the queuing network models corresponding to the upstream and the downstream system in the following subsections.

A. Upstream System with SBS/R Technology

We consider SBS/R technology choice for the storage system in the upstream network. It is modeled as a closed queuing model with a single customer class (Figure 5). There are N_a aisles where each aisle has N_t tiers. Each tier is equipped with a dedicated shuttle while each aisle has a dedicated lift for storage and retrieval of item totes. All shuttles and lift resources are modeled as single server stations with generally distributed service time. In case of a retrieval transaction, the item tote is fetched from the tier of a particular aisle using the corresponding shuttle. Then, through the lift corresponding to that aisle, it is brought to the entrance of the conveyor, which can dispatch only one item tote at a time to the conveyor. The entrance node of the conveyor is modeled as a single server station with exponentially distributed service time. As the distance to be traveled from the entrance point of the conveyor to the pick station is known and fixed, it will take a deterministic time for the item tote to travel to the pick station on the conveyor. Hence, the travel time on the conveyor is modeled using an Infinite Server (IS) node with deterministic service times. Therefore, the total number of nodes in an upstream stream with SBS/R technology is equal to $N_a N_t + N_a + 2$. There are a maximum of K_u orders allowed in the upstream system. As we already assume that there will be exactly two items required by all orders, so a maximum of $2K_u$ item tokens are allowed in the upstream system corresponding to the K_u orders.

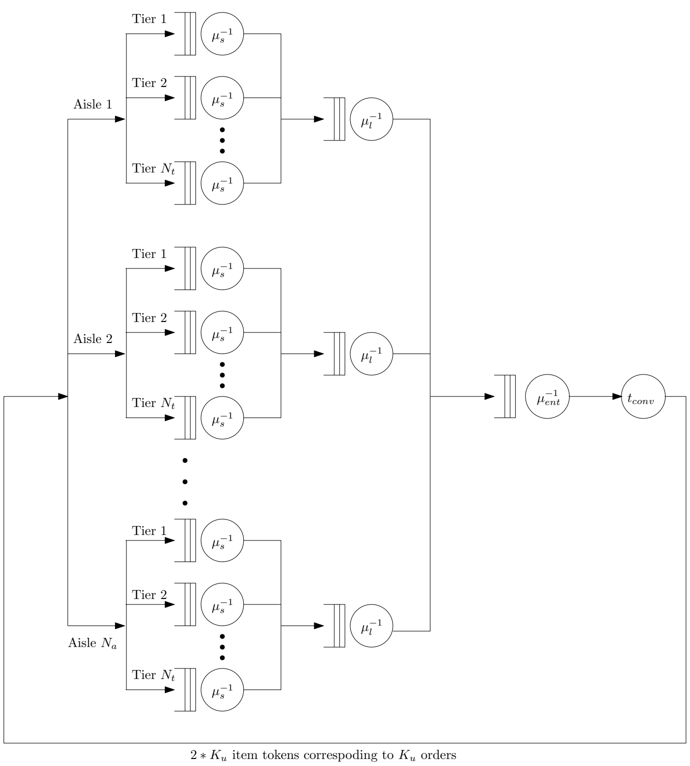

Fig. 5. Upstream network with SBS/R for dense tote storage and retrieval (adapted from Tappia et al. [6])

We assume that the lift dwells at the LU point of first tier after completion of every storage/retrieval transaction. Likewise, the shuttle dwells at the LU point of its tier. Now, the service time parameters (mean and standard deviation) for the lift and the shuttle can be calculated using the following equations provided by Marchet et. al. [15]. Here, $D_L(t)$ and $D_S(c)$ represent the distance traveled by the lift and the shuttle for performing a storage/retrieval transaction involving tier t and column c, respectively. For further details on the service time expressions, see [15].

$$D_L(t) = (t-1) * loc_h$$
$$\mu_L^{-1}(t) = \begin{cases} 2 * [2.v_L/a_L + (D_L(t) - 2.v_L^2/(2.a_L))/v_L] + 2.FT_L, \\ \text{for } D_L(t) > v_L^2/a_L \text{ , and} \\ 2 * [2 * \sqrt{D_L(t)/a_L}] + 2.FT_L, \\ \text{for } D_L(t) < v_L^2/a_L \end{cases}$$

$$\mu_L^{-1} = \frac{1}{N_t} * \sum_{t=1}^{N_t} \mu_L^{-1}(t) \tag{1}$$

$$\sigma_L^2 = \frac{1}{N_t - 1} * \sum_{t=1}^{N_t} [\mu_L^{-1}(t) - \mu_L^{-1}]^2 \tag{2}$$

$$D_S(c) = c * loc_w$$
$$\mu_S^{-1}(c) = \begin{cases} 2 * [2.v_S/a_S + (D_S(c) - 2.v_S^2/(2.a_S))/v_S] + 2.FT_S, \\ \text{for } D_S(c) > v_S^2/a_S \text{ , and} \\ 2 * [2 * \sqrt{D_S(c)/a_S}] + 2.FT_S, \\ \text{for } D_S(c) < v_S^2/a_S \end{cases}$$

$$\mu_S^{-1} = \frac{1}{N_c} * \sum_{c=1}^{N_c} \mu_S^{-1}(c) \tag{3}$$

$$\sigma_S^2 = \frac{1}{N_c - 1} * \sum_{c=1}^{N_c} [\mu_S^{-1}(c) - \mu_S^{-1}]^2 \tag{4}$$

B. Downstream System

The downstream system includes the pick station. There are two operations included in the downstream system: 1) Picking the item from the item tote and placing it into the corresponding order tote, and 2) Waiting for the completion of an order. There are a maximum of K_d orders allowed in the downstream system. As we assume that each order requires exactly two item totes to be fulfilled, there can be a maximum of $2K_d$ item tokens in the system corresponding to those K_d orders. To incorporate the human picker operation, there is a node with exponentially distributed service time in Figure 6. The synchronization station M models the waiting time to consolidate all items and complete an order.

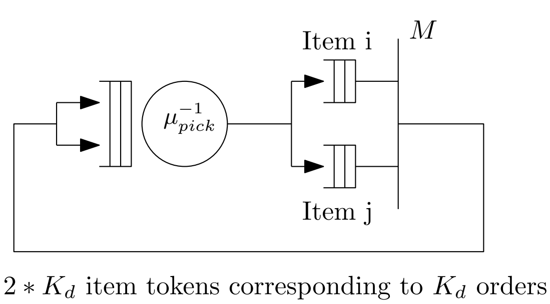

Fig. 6. Downstream network for order picking

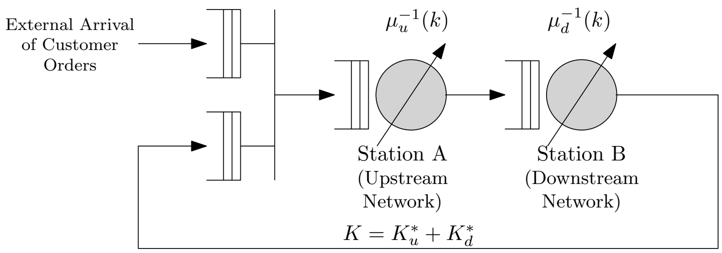

Fig. 7. Reduced semi-open queuing network for the integrated system

Figure 7 depicts the solution approach for the integrated model. It is a single class semi-open queuing network with maximum K number of orders allowed in the system. Station A is an equivalent load-dependent server corresponding to the upstream system, while station B is an equivalent load-dependent server for the downstream system. Here, K_u^* and K_d^* are the maximum number of orders allowed in the upstream and downstream system, respectively such that the corresponding throughput estimates in the respective systems are at a maximum level.

IV. Simulation Model and Results

We consider a 10-tier high and a 100 columns deep, shuttle-based storage and retrieval system. We also vary the number of aisles in the system from 1 to 3. We analyze the system with different levels of transaction order arrival rates. We develop a discrete-event simulation model of the integrated queuing network model discussed in Section III using Arena Simulation software. We use 15 replications to obtain a 95% CI. The parameter settings for the design of experiments is included in Tables I and II. We vary the order arrival rate and observe the change in the order throughput time. For this set of experiments, we assumed an infinite input buffer space at the order picking station (OPS) in the downstream system. As expected, we observe that the average order throughput time increases with increase in the arrival rate (Figure 8). Also, we observe that the throughput time decreases with an increase in the number of aisles (Figure 9).

Parameter	Value
Number of replication	15
Warm-up period	25000 sec
Replication length	1250000 sec
loc_w	0.5 m
loc_h	0.6 m
FT_L	3.7 sec
FT_S	13 sec
v_L,a_L	3 m/s, 4 m/s^2
v_S,a_S	3 m/s, 1 m/s^2
N_t	10
N_c	100
N_p	1
K	28
μ_{pick}^{-1}	16.7 sec
t_{conv}	30 sec
μ_{ent}^{-1}	5 sec

TABLE I
PARAMETER SETTING

Scenario	N_a	λ (per hour)
1	1	50
2	1	60
3	1	70
4	1	80
5	1	90
6	1	100
7	1	105
8	2	50
9	2	60
10	2	70
11	2	80
12	2	90
13	2	100
14	2	105
15	3	50
16	3	60
17	3	70
18	3	80
19	3	90
20	3	100
21	3	105

TABLE II
DESIGN OF EXPERIMENTS - I

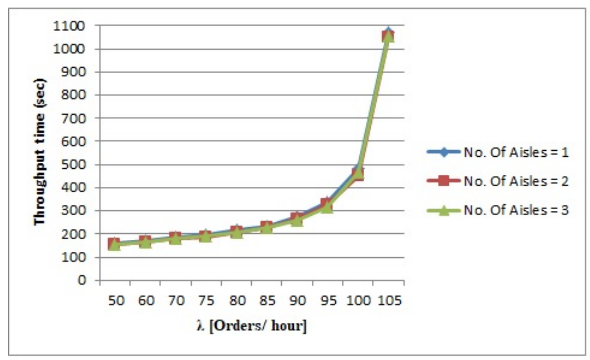

Fig. 8. Effect of the order arrival rate on throughput time of the system with infinite input buffer at the OPS

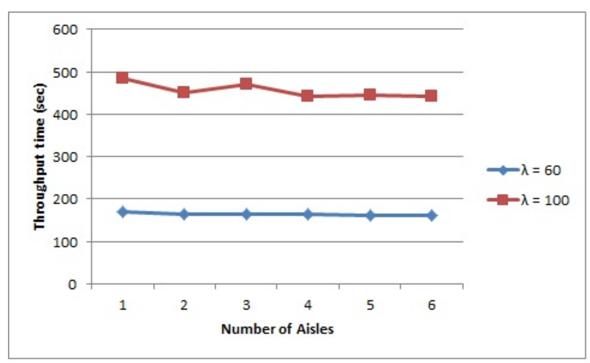

Fig. 9. Effect of the number of aisles on throughput time of the system with infinite input buffer at the OPS

We also explored the effect of input buffer size at OPS, on the throughput time. The parameter setting is retained from Table I; while the design of experiments is included in Table III. We observe that the average throughput time decreases with increase in the input buffer size at OPS (Figure 10).

Scenario	N_a	Input buffer size at OPS
1	1	1
2	1	2
3	1	3
4	1	4
5	1	5
6	1	6
7	1	7
8	1	8
9	2	1
10	2	2
11	2	3
12	2	4
13	2	5
14	2	6
15	2	7
16	2	8
17	3	1
18	3	2
19	3	3
20	3	4
21	3	5
22	3	6
23	3	7
24	3	8

TABLE III
DESIGN OF EXPERIMENTS - II

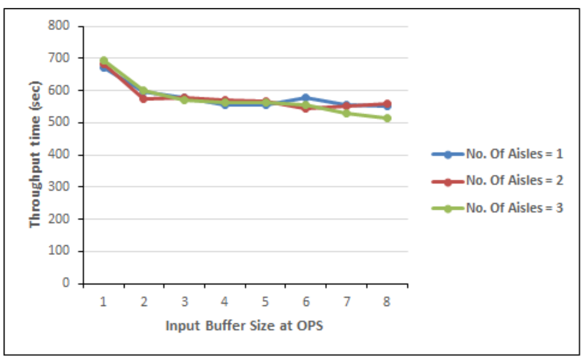

Fig. 10. Effect of the input buffer size at OPS on throughput time of the system with $\lambda = 100$ orders/hour

V. CONCLUSIONS AND FUTURE WORK

In this research, we develop a queuing network model for performance analysis of integrated order pick stations handling multi-line order fulfillment. We apply the model for a storage system operated using shuttle-based storage and retrieval system and a pick station. We evaluate the model using discrete-event simulations and show the effect of order throughput and a number of storage aisles on average order throughput time. We plan to evaluate the model using a combination of analytical and simulation methods. We also intend to apply the framework for a wider selection of storage system technology choices such as AS/RS.

REFERENCES

[1] D. Roy, A. Krishnamurthy, S. S. Heragu, and C. J. Malmborg, “Performance analysis and design trade-offs in warehouses with autonomous vehicle technology,” IIE Transactions, vol. 44, pp. 1045–1060, 2012.
[2] J. A. Tompkins, J. A. White, Y. A. Bozer, and J. M. A. Tanchoco, Facilities planning, John Wiley & Sons, 2010.
[3] N. Zaerpour, Y. Yu, and R. De Koster, “Storing fresh produce for fast retrieval in an automated compact cross-dock system,” Production and Operations Management, vol. 24, pp. 1266–1284, 2015.
[4] E. Tappia, D. Roy, R. De Koster, and M. Melacini, “Modeling, analysis, and design insights for shuttle-based compact storage systems,” Transportation Science, vol. 51, pp. 269–295, 2016.
[5] Y. Wang, S. Mou, and Y. Wu, “Task scheduling for multi-tier shuttle warehousing systems,” International Journal of Production Research, vol. 53, pp. 5884–5895, 2015.
[6] E. Tappia, D. Roy, M. Melacini, and R. De Koster, “Integrated storage-order picking systems: technology, performance models, and design insights,” unpublished.
[7] J. A. Pazour, D. Roy and, V. Dhingra, “Batching decisions for e-commerce order fulfillment: technology, models, and data insights,” unpublished.
[8] C. J. Malmborg, “Conceptualizing tools for autonomous vehicle storage and retrieval systems,” International Journal of Production Research, vol. 40, pp. 1807–1822, 2002.
[9] M. Fukunari and C. J. Malmborg, “A network queuing approach for evaluation of performance measures in autonomous vehicle storage and retrieval systems,” European Journal of Operational Research, vol. 193, pp. 152–167, 2009.
[10] B. Y. Ekren, S. S. Heragu, A. Krishnamurthy, and C. J. Malmborg, “Matrix-geometric solution for semi-open queuing network model of autonomous vehicle storage and retrieval system,” vol. 68, pp. 78–86, 2014.
[11] K. Azadeh, R. De Koster, and D. Roy, “Robotized and Automated Warehouse Systems: Review and Recent Developments,” unpublished.
[12] S. S. Heragu, X. Cai, A. Krishnamurthy, and C. J. Malmborg, “Analytical models for analysis of automated warehouse material handling systems,” International Journal of Production Research, vol. 49, pp. 6833–6861, 2011.
[13] D. Roy, A. Krishnamurthy, S. S. Heragu, and C. J. Malmborg, “A multi-tier linking approach to analyze performance of autonomous vehicle-based storage and retrieval systems,” Computers & Operations Research, vol. 83, pp. 173–188, 2017.
[14] D. Roy, A. Krishnamurthy, S. S. Heragu, and C. J. Malmborg, “Blocking effects in warehouse systems with autonomous vehicles,” IEEE Trans. Automation Science and Engineering vol. 11, pp. 439–451, 2014.
[15] G. Marchet, M. Melacini, S. Perotti, and E. Tappia, “Analytical model to estimate performances of autonomous vehicle storage and retrieval systems for product totes, “ International Journal of Production Research, vol. 50(24), pp. 7134–7148, 2012.