Analytical model to estimate performances of Autonomous Vehicle Storage and Retrieval Systems for product totes

Abstract. In today's competitive scenario of increasingly faster deliveries and smaller order sizes, material handling providers are progressively developing new solutions. A recent solution consists in Autonomous Vehicle Storage and Retrieval System (AVS/RS). The present paper presents an analytical model to estimate the performances (i.e. transaction cycle time and waiting times) of AVS/RS for product tote movement. The model is based on an open queuing network approach. The model effectiveness in performance estimation is validated through simulation.

1. Introduction

Autonomous Vehicle Storage and Retrieval System (AVS/RS) represents a new automated material handling technology for unit load storage and retrieval. In traditional Automated Storage and Retrieval System (AS/RS), unit loads are handled using aisle-captive storage cranes that move simultaneously in the vertical and horizontal dimensions. Conversely, in AVS/RS unit loads are handled by vehicles moving horizontally along rails within storage racks with vertical movement provided by lifts mounted along the rack periphery (Malmborg 2003).

The main advantage of AVS/RS technology is the flexibility to allocate variable numbers of vehicle and lifts based on the level of transactions demand (Fukunary and Malmborg 2009). The main disadvantages lie in longer flow paths from sequential vertical and horizontal travel, and waiting times for lifts and vehicles. Based on the vehicle assignment to storage tiers, two main configurations may be defined (Heragu *et al.* 2009):

- AVS/RS with "tier to tier" vehicles;

- AVS/RS with "tier captive" vehicles.

In the "tier to tier" configuration vehicles may move from a tier of the storage racks to another by using lifts.

In the "tier captive" configuration each vehicle is dedicated to a single tier and therefore cannot move to another one. Lifts are used to move only the unit loads to the destination tiers. An AVS/RS with "tier captive" configuration is more expensive due to the higher number of vehicles being adopted, and offers better performances thanks to the fact that vehicle and lift movements are independent.

In industrial applications "tier to tier" configuration is usually adopted in case of pallet unit loads (Ekren *et al.* 2010). Conversely, "tier captive" configuration is more frequently used for product tote movement (as an example, please see the solutions offered by some of the main material handling providers, such as Multishuttle Captive by Dematic, OS-RS system by Knapp, SQS by Schafer, Quickstore HDS by Vanderlande).

As emerged from the literature review, previous research mainly focuses on AVS/RS with "tier to tier" configuration, whereas "tier captive" AVS/RS seems to be disregarded, notwithstanding its higher adoption in a number of industrial applications for tote handling.

The main objective of the present paper is to provide an analytical model to estimate the transaction cycle time in an AVS/RS for product tote, assuming a "tier captive" configuration. According to other studies on AVS/RS the main use of the model is in the early technology selection, or "conceptualisation" phase of system development (Fukunari and Malmborg 2008).

The remainder of the paper is organised as follows. We start by highlighting the main contributions in the literature regarding the estimate of AVS/RS performances (Section 2). Next, we describe the AVS/RS under consideration, as well

as the main data, assumptions and design variables (Section 3). In Section 4 we propose the model to estimate AVS/RS performances. In Section 5 we present the application of the model including a simulation-based validation of its accuracy. Conclusions and future developments are proposed in the final section (Section 6).

2. Literature review

The first study on the performances of an AVS/RS has been proposed by Malmborg (2002). With reference to a "tier to tier" configuration, a model has been proposed to estimate vehicle utilisation and cycle time as function of the number of storage columns, the number of tiers, vehicles and lifts. The following main assumptions have been taken into account: random storage policy, single-deep storage racks, and input/output point (I/O) located at the first tier. As such, the single command cycle time is computed as the sum of the vehicle average time to travel from I/O point to the storage position (i.e. including the expected waiting time for a lift), the vehicle average time to return to the I/O point, and a fixed time due to vehicle transfer between rails and lifts.

Kuo *et al.* (2007) present a cycle time model based on a storage rack configuration for random storage using the point-of-service-completion (POSC) dwell point policy. The formulation of the expected cycle time is determined by the vehicle movement elements and is based on the taxonomy of transaction types observed in the system. The vehicle waiting times for lift service is obtained using a G/G/L lift queuing system, where the lifts (L) are the servers and vehicles are the customers; the resolution of the model adopts Whitt's approximation.

Other studies compute the cycle time as a function of the resource movement elements, as well as previous research, but take different assumptions for queuing

modelling. In the model proposed by Heragu *et al.* (2009) the queuing theory is applied, by modelling an AVS/RS as an open queuing network. Zhang *et al.* (2009) propose a model representing storage and retrieval transactions as customers and the pairs vehicle/lift as parallel servers.

The performances of an AVS/RS are sometimes analysed by considering both single command and dual command, where either a storage and retrieval transaction are completed on the same cycle. In this latter case the expected cycle time is a weighted average of the single command and dual command cycle times. The key issue is the estimating of the proportion of dual command (often called α). In the literature three strategies have been proposed for the estimation of α: (i) state equation procedure (Malmborg 2003), (ii) the nested queuing procedure (Fukunari and Malmborg 2008, Kuo *et al.* 2007), and (iii) the network queuing approach (Fukunari and Malmborg 2009). According to Zhang *et al.* (2009) the benefit achievable in terms of efficiency depending on the increase of the proportion of dual command cycles is limited, due to several reasons: first, the difficulty of reaching high α values and second, as storage and retrieval transactions paired in the same cycle are usually associated with different storage tiers.

Some recent studies focus on the performance optimisation of AVS/RS by means of class-based storage policies adoption (Kuo *et al.* 2008), multiple load/unload points (Roy *et al.* 2009) or different dwell point policy (Ekren *et al.* 2010).

In summary, as highlighted in Table 1, since the early 2000s several studies have been performed on AVS/RS. Such studies focus on AVS/RS with "tier to tier" configuration, assuming that the vehicle moves vertically together with the lift and therefore performs sequentially horizontal and vertical movements. All the proposed

models allow an accurate estimate of the transaction cycle time. However, they do not provide an accurate estimate of the waiting times for lift service, whose average error ranges from 1% to 18% in Fukunary and Malmborg (2008), and is higher than 20% in Kuo *et al*. (2007).

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

Take in Table 1

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

3. AVS/RS description

This section provides the description of the examined AVS/RS for product totes, developed based on interviews with material handling providers. The transaction cycle time components are then presented, as well as the main assumptions for the AVS/RS performance estimation. The main notations used in the remainder of the paper are finally introduced.

3.1. Layout

Figure 1 provides a representation of an AVS/RS for product totes (also indicated as 'unit loads' in the remainder of the paper), while Figure 2 synthesizes a single tier. As the figures illustrate, there are single-deep, double-sided storage racks. Each storage column *c* can hold one unit load. Lifts are mounted at fixed locations along the periphery of the storage. The input/output point (I/O) is located at the first tier beside each lift. Vehicles move along one dimension only, each of them within a specific tier *t* of a storage aisle A_i. The number of lifts installed in the system is equal to the number of aisles (*A*), whereas the number of vehicles is equal to product of the number of aisles times the number of tiers (*T*). The configuration is "tier captive", so the vehicles cannot move from a tier to another. To allow the mutual independence of

vehicle and lift, the first position at each side of the storage aisle in all tiers serves as

buffer and is implemented to transfer unit loads from vehicles to lifts (Figure 2). A

buffer (called *buffer out*) handles the unit loads which have been retrieved, the other

one (called *buffer in*), located in the other side of the storage aisle, handles the unit

loads to be stored. As such, unit loads (not vehicles) wait for the lift, so vehicle and

lift can work as parallels servers.

It should be noticed that the lift provide not only the vertical tote movement

for the transactions involving retrieval/storage positions located on different tiers, but

also the tote transfer from/to the buffer at the first tier.

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

Take in Fig. 1 and Fig. 2

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

3.2. Transaction cycle time components

As emerged from the literature review, performing dual command cycles in an

AVS/RS is rather difficult. Therefore, in the present model only single command

cycles have been considered.

With reference to a generic storage aisle and a retrieval transaction originating

in a tier t between 2 and T (number of tiers), the following movement elements may

be highlighted:

A. Transaction requests vehicle (FCFS dispatching)

B. Transaction waits for vehicle

C. Horizontal vehicle movement to retrieval address

D. Vehicle charges load at retrieval address

E. Horizontal vehicle movement from the retrieval address to buffer

F. Unit load discharges from and releases vehicle

G. Unit load requests lift (FCFS dispatching)

H. Unit load waits for the lift

I. Vertical lift movement to retrieval tier

J. Lift charges load at the buffer of the retrieval tier

K. Vertical lift movement from the retrieval tier to I/O point

L. Unit load discharges from and releases lift

In case the retrieval transaction is performed in the first tier ($t = 1$), lifts should only load (J) and discharge (L) the tote. Therefore, the vertical movements I and K should not be considered within the transaction.

Similarly, in case of storage in a generic aisle and in a tier t between 2 and T (number of tiers), the following movement elements may be highlighted:

M. Transaction requests lift (FCFS dispatching)

N. Transaction waits for lift

O. Vertical lift movement to I/O point

P. Lift charges load at the I/O point

Q. Vertical lift movement from tier 1 to destination tier

R. Unit load discharges from and releases lift

S. Transaction requests vehicle (FCFS dispatching)

T. Transaction waits for vehicle

U. Horizontal movement to buffer

V. Vehicle charges load at the buffer of its tier

W. Horizontal movement from buffer to storage address

X. Load discharges at storage address and release vehicle

In case the storage transaction is performed in the first tier ($t = 1$), lifts should only load (P) and discharge (R) the tote. Therefore, O and Q vertical movements should not be considered within the transaction.

Based on the movement elements and the above-mentioned observations, the retrieval transaction cycle time ($\tau_{SC,r}$) is:

$$\tau_{SC,r} = \tau_{v,r} + \tau_{l,r} + W_r \tag{1}$$

where:

$\tau_{v,r}$ is the component of cycle time related to the vehicle movement, which is equal to the sum of the movement elements C,D,E and F

$\tau_{l,r}$ is the component of cycle time related to the lift movement, which is equal to the sum of the movement elements I,J,K and L if $t > 1$, whereas, in case $t = 1$, the sum of the movement elements J and L only

W_r is the waiting time, corresponding to the movement element H

Similarly, the storage transaction cycle time ($\tau_{SC,s}$) may be computed as:

$$\tau_{SC,s} = \tau_{v,s} + \tau_{l,s} + W_s \tag{2}$$

where:

$\tau_{v,s}$ is the component of cycle time related to the vehicle movement, which is equal to the sum of the movement elements U,V, W and X

$\tau_{l,s}$ is the component of cycle time related to the lift movement, which is equal to the sum of the movement elements O, P, Q and R if $t > 1$, whereas, in case $t = 1$, the sum of the movement elements P and R only

W_s is the waiting time, corresponding to the movement element T

In both transaction types, the fact that one lift should simultaneously serve more than one vehicle may generate queues for transactions. From this viewpoint, the

estimates of the waiting times W_r and W_s are one of the most critical issues, and have been so far disregarded in previous literature for the type of AVS/RS being examined.

In the remainder of the paper only single retrieval transactions will be taken into account. The motivation of this choice is twofold. First, retrieval transactions represent the most critical activities from the organisational viewpoint as it is strictly related to the service level and – differently from storage transactions – cannot be postponed to a low-workload phase. Second, once a single retrieval transaction has been modelled, the process may be easily extended to the case of a single storage transaction.

3.3. Main assumptions

In most unit load SR systems, a space-conserving random storage policy is used because of capital cost considerations (Heragu 2008). Therefore, we assume a random storage policy. As a consequence, each storage aisle presents the same average transaction cycle time.

The other assumptions are as follows:

- number of unit loads handled per cycle by lift and vehicle: 1;
- vehicle dwell point policy: point-of-service-completion (POSC);
- lift dwell point policy: point-of-service-completion (POSC);
- the retrieval transactions demand to the system follows a Poisson distribution with parameter λ_r and therefore the time between two subsequent arrivals follows an exponential distribution with parameter $1/\lambda_r$, accordingly with previous studies in the literature, such as Fukunari and Malmborg (2008) and Ekren et al. (2010);
- the transaction demand to a storage aisle is evenly distributed among the storage aisles (A) and the different storage tiers (T). Indeed, under the

assumption of random storage, the probability that a retrieval transaction is required in a certain storage aisle is identical for each of them and is equal to $1/A$; similarly, the probability that a retrieval transaction is required in a certain tier is identical for each of them and is equal to $1/T$. As such, based on the properties of the exponential distribution, the time elapsed from two subsequent arrivals of retrieval transactions to a lift is still distributed as an exponential distribution with a mean arrival rate equal to λ_r/A,. Conversely, the time elapsed from two subsequent arrivals of retrieval transactions to a vehicle is described by an exponential distribution with parameter $\lambda_r / (A*T)$;

- lift and vehicle service times follow a uniform distribution. Indeed, under the assumption of random storage, the probability that the retrieval transaction is performed in one of the C storage columns is identical for each of them and is equal to $1/C$. Similarly, the probability that the retrieval transaction is performed in one of the T tier is identical for each of them and is equal to $1/T$.

3.4. Main notations

The notation used in the AVS/RS model is as follows:

λ_r = total retrieval transaction demand rate to the system

A = number of storage aisles

T = number of storage tiers

t = tier involved in the transaction, with t between 1 and T

C = number of storage columns on each side of an storage aisle

c = storage column involved in the transaction, with c between 1 and C

n = total number of storage positions in each aisle ($2*T*C$)

μ_w = unit width clearance per storage position including allowances (m)

μ_h = unit height clearance per storage position including allowances (m)

v_h= maximum horizontal velocity of vehicle (m/s)

v_v = maximum vertical velocity of lift (m/s)

a_h = acceleration/deceleration rate of vehicle (m/s^2)

a_v = acceleration/deceleration rate of lift (m/s^2)

ε_1= time allowance for charging and discharging load from vehicle (s)

ε_2 = time allowance for charging and discharging load from lift (s)

$D_{v,c}$ = distance travelled by the vehicle to retrieve a unit load from the storage column c

$D_{l,t}$ = distance travelled by the lift to retrieve a unit load from tier t

$P_{v,c}$ = probability that the vehicle retrieves a unit load from the storage column c

$D_{l,t}$ = probability that the lift retrieves a unit load from tier t

4. Analytical model to estimate AVS/RS performances

In the subsequent paragraphs the model to estimate retrieval transaction cycle time is reported. As afore-mentioned in equation (1), transaction cycle time is computed as the sum of three components, respectively related to: (i) vehicle movement ($\tau_{v,r}$), (ii) lift movement ($\tau_{l,r}$), and (iii) waiting time to use a lift (W_r). A detailed description of the calculation method for each follows.

4.1. Estimation of the cycle time due to vehicle movement

The time required by the vehicle to retrieve one unit load ($\tau_{v,r}$) is equal to the weighted average of all the times to perform a complete retrieval transaction within one of the C storage columns:

$$\tau_{v,r} = \sum_{c=1}^{C} P_{v,c} * (\varepsilon_1 + TV_{v,c}) \tag{3}$$

where:

$$P_{v,c} = 1/C \qquad \forall c. \tag{4}$$

In turn, vehicle service rate ($\mu_{v,r}$) is the inverse of $\tau_{v,r}$.

As it may be noticed from (3), the time required to perform a retrieval transaction within the storage column c is composed of: a fixed part (ε_I), due to tote loading and unloading, and a variable part ($TV_{v,c}$), which represents the time required by the vehicle to cover the travel distance $D_{v,c}$ during each of the two horizontal movements (i.e. C and E). Under the POSC vehicle dwell point policy, at the end of each transaction the vehicle is in correspondence to the buffer. Therefore, independently from the storage column where unit load has to be retrieved, that position always represents the starting point of each retrieval transaction. The travel distance that the vehicle has to cover $D_{v,c}$ is given by:

$$D_{v,c} = (\mathrm{c} * \mu_w) \tag{5}$$

As for AS/RS, $TV_{v,c}$ may be estimated as the ratio between 2* $D_{v,c}$ to the average uniform velocity, or considering in the ratio an acceleration warm up (a_h), a period at maximum velocity v_h and a deceleration warm up (- a_h) (Hwang and Lee 1990, Lerher *et al.* 2006). Within the present study, the second approach has been adopted. Focussing on one of the two identical travel distances to be covered by the vehicle ($D_{v,c}$), two types of velocity profile can be distinguished depending on whether the peak velocity obtained is less than v_h. Being t_p the time necessary to reach the peak velocity and Y the arrival time at a destination, in case v_h is reached ($Y > 2v_h/a_h$) the vehicle velocity may be described as in equation (6) and Figure 3.

$$V(t) = \begin{cases} a_h t & t \leq t_p \\ v_h & t_p \leq t \leq Y - t_p \\ -a_h t & t \geq Y - t_p \end{cases} \tag{6}$$

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

Take in Figure 3

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

In case v_h is not reached ($Y < 2v_h/a_h$), the vehicle velocity is composed only of two components of acceleration (a_h) and deceleration ($-a_h$).

Therefore, $TV_{v,c}$ consists in the following expression:

$$TV_{v,c} = \begin{cases} 2 * \left[2\,v_h/a_h + \left(D_{v,c} - 2 * v_h^2/(2 * a_h)\right)/v_h\right] & Y > 2v_h/a_h \\ 2 * \left[2 * \sqrt{D_{v,c}/a_h}\right] & Y < 2v_h/a_h \end{cases} \tag{7}$$

4.2. Estimation of the cycle time due to lift movement

The average time required by a lift to retrieve one unit load ($\tau_{l,r}$) is a weighted average of the time components to perform a retrieval transaction in one of the T tiers:

$$\tau_{l,r} = \sum_{t=1}^{T} P_{l,t} * (\varepsilon_2 + TV_{l,t}) \tag{8}$$

where:

$$P_{l,t} = 1/T \qquad \forall t. \tag{9}$$

In turn, the lift service rate ($\mu_{l,r}$) is the inverse of $\tau_{l,r}$.

As observed in (8), the time required to perform a retrieval transaction in tier t is composed of: a fixed part (ε_2), due to tote loading and unloading, and a variable part ($TV_{l,t}$), which represents the time required by the lift to cover the travel distance $D_{l,t}$ during each of the two vertical movements (i.e. I and K).

Under the POSC lift dwell point policy, a lift always completes each retrieval cycle at the I/O point at tier 1. As such, independently from the tier containing the unit load to be retrieved, that position always represents the starting point of each retrieval transaction. The travel distance that the lift has to cover is given by:

$$D_{l,t} = (t - 1) * \mu_h \tag{10}$$

As for the vehicles, the lift velocity profile is composed of an acceleration warm up (a_v), a period at maximum velocity (v_v) and a deceleration warm up (- a_v) and the $TV_{l,t}$ computed similarly to the equation (7):

$$TV_{l,t} = \begin{cases} 2 * \left[2\,v_v/a_v + \left(D_{l,t} - 2 * v_v^2/(2 * a_v)\right)/v_v\right] & Y > 2v_v\,/a_v \\ 2 * \left[2 * \sqrt{D_{l,t}/a_v}\right] & Y < 2v_v\,/a_v \end{cases} \quad (11)$$

4.3. Estimation of the cycle time due to lift waiting time

Accordingly with the literature review, the waiting time component of the transaction cycle time has been modelled according to the queuing network approach. It should be noticed that the waiting time is related to the buffer located at the beginning of each storage tier. Such buffer is not present in a "tier to tier" AVS/RS, in which vehicle moves jointly with the lift. One possible solution consists in the use of a semi-open queuing network, whose application however requires a huge computational effort (Jia and Heragu 2009). As such, an open queuing network model (Heragu *et al.* 2009) has been selected with a two-fold aim: first, to achieve an easier modelling procedure to be implemented and, second, to fulfil the need of minimising the time required in the "conceptualisation" phase of an AVS/RS design (Malmborg 2002).

As synthesized in Figure 4, lifts and vehicles are assumed as servers, as they move unit loads vertically and horizontally respectively within the storage aisles. Retrieval transactions are assumed as customers. Each storage aisle is characterised by one lift and a number of vehicles equal to the number of tiers T. Therefore, the overall number of servers is $T + 1$.

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

Take in Figure 4

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

Retrieval transactions have been sub-divided based on the storage tier in which they have to be performed, and therefore on the specific vehicle to be used. As a consequence, given the number of tier T of the storage configuration (i.e. equal to the number of vehicles), the number of retrieval transaction types in the network is equal to T.

In the remainder of the paper, the storage tier t (being $t = 1,\ldots,T$) in which a retrieval transaction has to be performed also defines the customer type.

To solve the open queuing network model an approximate method has been adopted. Such approximate method was required to take into account non-exponential vehicle and lift service times.

Specifically, among the various algorithms proposed by previous literature, the *decomposition method* has been selected. This method is based on five sequential steps to be completed (Bolch *et al.* 2006).

With reference to the representation in Figure 4, the five steps of the methodology are explained as follows.

Step 1. Decompose the original open network in M/G/1 queues stochastically independent.

Step 2. Calculate the arrival rates, the utilisations and the mean service time of the individual servers.

The arrival rate for a generic server is obtained as the sum of the arrival rates related to all customer types visiting that server, and has to take into account the sequence followed by the customers when visiting the servers. In this case, the sequence is one only, since all the transaction types first require a vehicle, and then a lift. As such, the arrival rate of each vehicle λ_{vt} corresponds to the arrival rate of

customer type t only. Conversely, the arrival rate of the lift λ_l is equal to the sum of all customer type arrival rates.

$$\lambda_{v_1} = \lambda_{v_1,1} = \frac{\lambda_r}{A*T}$$

$$\vdots \tag{12}$$

$$\lambda_{v_T} = \lambda_{v_T,T} = \frac{\lambda_r}{A*T}$$

$$\lambda_l = \lambda_{l,1} + \lambda_{l,2} + \dots + \lambda_{l,T} = \frac{\lambda_r}{A}$$

The utilisation level of a server is obtained as the ratio of the arrival rate to the service rate. Therefore, the vehicle utilisation ρ_{vt} consists in the ratio of the arrival rate to the service rate related to customer type t only, whereas the lift utilisation is equal to the sum of the utilisation levels due to each customer type:

$$\rho_{v_1} = \frac{\lambda_{v_1,1}}{\mu_{v_1,1}}$$

$$\vdots \tag{13}$$

$$\rho_{v_T} = \frac{\lambda_{v_T,T}}{\mu_{v_T,T}}$$

$$\rho_l = \frac{\lambda_{l,1}}{\mu_{l,1}} + \frac{\lambda_{l,2}}{\mu_{l,2}} + \dots + \frac{\lambda_{l,T}}{\mu_{l,T}}$$

Finally, the mean service rate of each server may be expressed as the weighted average of the service rates related to each customer type:

$$\mu_{v_1} = \frac{1}{\left(\frac{\lambda_{v_1,1}}{\lambda_{v_1}} * \frac{1}{\mu_{v_1,1}}\right)}$$

$$\vdots \tag{14}$$

$$\mu_{v_T} = \frac{1}{\left(\frac{\lambda_{v_T,T}}{\lambda_{v_T}} * \frac{1}{\mu_{v_T,T}}\right)}$$

$$\mu_l = \frac{1}{\left(\frac{\lambda_{l,1}}{\lambda_l} * \frac{1}{\mu_{l,1}} + \dots + \frac{\lambda_{l,T}}{\lambda_l} * \frac{1}{\mu_{l,T}}\right)}$$

Step 3. For each server, compute the square of the variation coefficient of the service times c_B^2, by aggregating the variation coefficients of the service time associated to each customer type visiting the server. The variation coefficients of the lift service time associated to each customer type is null, as the time required by the lift is fixed, given the tier t in which the retrieval transaction has to be performed (corresponding to the customer type t).

$$c_{B,v_1}^2 = -1 + \frac{\lambda_{v_1,1}}{\lambda_{v_1}} * \left(\frac{\mu_{v_1}}{\mu_{v_1,1}}\right)^2 * \left(c_{B_{v_1,1}}^2 + 1\right)$$

$$\vdots \tag{15}$$

$$c_{B,v_T}^2 = -1 + \frac{\lambda_{v_T,T}}{\lambda_{v_T}} * \left(\frac{\mu_{v_T}}{\mu_{v_T,T}}\right)^2 * \left(c_{B_{v_T,T}}^2 + 1\right)$$

$$c_{B,l}^2 = -1 + \left[\frac{\lambda_{l,1}}{\lambda_l} * \left(\frac{\mu_l}{\mu_{l,1}}\right)^2 + \ldots + \frac{\lambda_{l,T}}{\lambda_l} * \left(\frac{\mu_l}{\mu_{l,T}}\right)^2\right]$$

Step 4. Compute the square of the variation coefficient of the inter-arrival times of each server. This step is done iteratively by completing the subsequent three phases:

- **Phase 1_*Merging***: in this phase, the different arrival processes to each server are merged into a single arrival process; this latter process is characterised by the arrival rate and square of the variation coefficient of the inter-arrival time c_{Ai}^2. The decomposition method of Pujolle (Pujolle and Ai 1986) has been selected, which consists in:

$$c_{Ai,t}^2 = \frac{1}{\lambda_{i,t}} * \left(\sum_j c_{ji,t}^2 * \lambda_{j,t} * p_{ji,t} + c_{0i,t}^2 * \lambda_{0,t} * p_{0i,t}\right) \tag{16}$$

$$c_{Ai}^2 = \frac{1}{\lambda_i} * \sum_{t=1}^T c_{Ai,t}^2 * \lambda_{i,t} \tag{17}$$

Being i and j the indexes of the server (i and $j = v_1$, v_2, ..., v_T, l), and $c_{0i,t}^2$ the square of the variation coefficient of the inter-arrival time of customer type t at server i (equal to 1, as the distribution is exponential):

$$c_{0v_1,1}^2 = c_{0v_2,2}^2 = \cdots = c_{0v_T,T}^2 = 1 \tag{18}$$

- **Phase 2_*Flow***: in this phase, the square of the variation coefficient of the inter-departure times c_{Di}^2 is computed. It depends on the square of the variation coefficient of the inter-arrival time c_{Ai}^2, as well as the square of the coefficient variation of the service times $c_{B,i}^2$. The decomposition method of Whitt (Whitt 1983a 1983b) has been selected, which consists in:

$$c_{Di}^2 = 1 + \frac{\rho_i^2 * (c_{B,i}^2 - 1)}{\sqrt{m_i}} + (1 - \rho_i^2) * (c_{Ai}^2 - 1) \tag{19}$$

- **Phase 3_*Splitting***: in this latter phase, a splitting of the departure process is performed. The computation is reported as follows:

$$c_{ij,1}^2 = 1 + c_{Di}^2 - 1$$
$$\vdots \tag{20}$$
$$c_{ij,P_T}^2 = 1 + c_{Di}^2 - 1$$

Coherently with the sequence followed by the customers when visiting the servers, the variables $c_{ij,t}^2$ that may be defined are:

$$c_{v_1-l,1}, c_{v_2-l,2}, \dots, c_{v_T-l,T} \neq \tag{21}$$

The initialisation of this three-phase iterative method presumes that the initial values of all $c_{ij,t}^2$ are assumed equal to 1. The iteration k after which the algorithm stops has to fulfill the following condition:

$$\left(\frac{c_{ij,t}^2(k)-c_{ij,t}^2(k-1)}{c_{ij,t}^2(k)}\right) < \varepsilon \tag{22}$$

for all customer types t and with ε sufficiently small.

Step 5. Calculate the mean queue length and the mean waiting time for each customer type t at each server. The approximation by Allen (1990) has been adopted, which consists in the following expressions:

$$\bar{Q}_{v_1,1} = \bar{Q}_{v_1,1,M/M/1} * \frac{\left(c_{Av_1,1}^2 + c_{B,v_1,1}^2\right)}{2}$$

$$\vdots \tag{23}$$

$$\bar{Q}_{v_T,T} = \bar{Q}_{v_T,T,M/M/1} * \frac{\left(c_{Av_T,T}^2 + c_{B,v_T,T}^2\right)}{2}$$

$$\bar{Q}_{l,t} = \bar{Q}_{l,t,M/M/1} * \frac{\left(c_{Al,t}^2 + c_{Bl,t}^2\right)}{2}$$

Since i is the index referred to the server ($i = v_1, v_2, v_T, l$), for each transaction type t the following expressions have to be used to obtain $\bar{Q}_{i,T}$:

$$\bar{Q}_{i,tM/M/1} = \frac{\rho_{i,t}}{1-\rho_i} * P_{1,i} \tag{24}$$

$$P_{1,i} = \frac{\rho_i}{(1-\rho_i)} * \pi_{0i} \tag{25}$$

$$P_{0,i} = \left(1 + \frac{\rho_i}{1-\rho_i}\right)^{-1} \tag{26}$$

where:

$P_{1,i}$ probability that an arriving customer at the server i has to wait in the queue;

$P_{0,i}$ probability of no customers at the server i.

Finally, the mean waiting time at the server i ($i = v_1, v_2, v_T, l$) of the customer type t $\bar{W}_{i,t}$ is computed by means of Little's theorem (Little 1961):

$$\bar{W}_{i,t} = \frac{\bar{Q}_{i,t}}{\lambda_{i,t}} \tag{27}$$

Once the different $\overline{W}_{i,t}$ are known, the mean waiting time for the lift may be computed using:

$$W_r = \frac{\sum_{t=1}^{T} \overline{W}_{i,t}}{T} \tag{28}$$

5. Model application

We now present an illustrative example with a twofold aim: first, to highlight the value of the proposed analytical model, and, second, to study AVS/RS for product tote performances varying the examined context. To validate the model, an Arena (version 13.0) simulation of an AVS/RS has been created. The application of the analytical model has been performed on a spreadsheet using Visual Basic macros.

Similarly to previous studies (Kuo et al. 2007, Dallari et al. 2009), the analysis has been performed by considering a number of combinations of storage capacity and transaction demand level.

The number of tiers has been taken as constant equal to 12. Such number allows to reach the maximum height capacity for an industrial building (usually between 10 and 11 m). The maximum length of the storage aisles has been considered equal to 70 m, which corresponds to the maximum value for AVS/RSs currently in place. Under these assumptions, for each storage aisle the service rate of the whole vehicle fleet is always higher than the lift service rate. As a consequence, the single vehicle utilisation is generally not high.

Two different storage capacity values have been examined, namely: 1,000 unit loads and 10,000 unit loads. In both cases, three possible retrieval transaction demand levels have been considered, thus to analyse the performances with different lift utilisation level (i.e. low, medium, high). Therefore, six scenarios have been analysed,

as reported in Table 2. The number of storage aisles – and therefore their length – has been computed to reach the different levels of lift utilisation.

Utilisation values reported in Table 2 have been estimated by using the proposed analytical model. According to equations (13), the expected lift utilisation ρ_l may be defined as the ratio of retrieval transaction arrival rate within one storage aisle to the weighted average of the lift service rate of each retrieval transaction type t. As the number of tiers remains constant varying the examined scenario, the lift utilisation is considered exclusively as a function of the retrieval transaction arrival rate at one storage aisle.

According to equations (13), the expected vehicles utilisation ρ_{vt} may be defined as the ratio of retrieval transaction arrival rate within one tier of a storage aisle to the vehicle service rate. The vehicle utilisation level depends not only from the retrieval arrival rate to a storage aisle, but also from the service rate. Being equal the storage capacity, as the number of aisles (A) increases, the average vehicle travel distance decreases, and therefore service rate rises. For all six scenarios, data reported in Table 3 have been considered.

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

Take in Tables 2 and Table 3

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

5.1. System simulation modelling

Figure 5 illustrates the simulation flow chart of retrieval transactions. A retrieval transaction requires the vehicle in the storage tier where the unit load to be retrieved is located. The vehicle first retrieves the unit load and then discharges it at *buffer out*. The lift moves the unit load to the output point. Within the above-described cycle, it may happen that, given a retrieval transaction to be performed, the vehicle is already

busy or the vehicle has to wait that the *buffer out* is empty before discharging the unit

load (i.e. limited buffer capacity), or else that the tote has to wait for the lift.

The assumptions of the simulation model are the same as those of the

analytical model reported in Paragraph 3.2. The simulation model is assumed to be a

non-terminating system, allowing us to conduct a steady state analysis (Ekren and

Heragu 2010) and, similarly to Fukunary and Malmborg (2008), the length of each

simulation run is 48 hours. The model is run for 20 independent replications. The

warm up is computed following the procedure by Welch (Law and Kelton 2000), and

observations belonging to the warm up, (i.e. 3 hours length on average) have been

omitted from the analysis.

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

Take in Figure 5

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

5.2. Experimental results

The validation of results has been performed with respect to the transaction cycle

time, as well as lift and vehicle utilisation (Table 4). The model yields an accuracy

level in estimating the transaction cycle time and the resource utilisation level.

With respect to the transaction cycle time, the percentage deviation is always

lower than 1%. The transaction cycle time values for the examined six scenarios are

influenced by the number of storage aisle. In those scenarios with storage capacity

equal to 1,000 (scenario A, B, C) only one storage aisle is present. Therefore, the

storage aisle length is constant, as well as the vehicle service rate. As such, as the

retrieval transaction arrival rate increases, queues also tend to increase, as well as the

overall cycle time. Those scenarios with storage capacity equal to 10,000 are

characterised by a different number of storage aisles – 3 in scenario A, 6 in scenario B, and 9 in scenario C –, and therefore a different storage aisle length. In this context, the effect of queuing (as the arrival rate increases) on cycle time is balanced by a reduction in the vehicles service time, due to the presence of a higher number of storage aisles.

With respect to lift and vehicle utilisations, the percentage deviation is always lower than 1%.

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

Take in Table 4

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

Table 5 illustrates the results with reference to the queue creation. Two queues are considered: (i) the queue at *buffer out*, which coincides with the component W_r of the transaction cycle time and with the movement element H, and (ii) the queue related to the vehicle, which corresponds to the movement element B and it is connected to the system response time.

The percentage deviation between model results and simulation results is less than 5%, thus presenting better results with respect to others in previous literature on AVS/RS with "tier to tier" configuration. As expected, as the lift utilisation increases, the queue at *buffer out* also tends to increase (e.g. from 0.31 s to 5.76 s in case of storage capacity equal to 1,000, and from 0.31 s to 6.11 s in case of storage capacity equal to 10,000). With reference to waiting time for the vehicles (i.e. waiting time while the vehicle is completing the previous transaction), estimation errors are similar. Looking at results, the waiting time for a vehicle is higher in scenarios D, E and F with respect to scenarios A, B and C, because of a lower vehicle service rate due to longer storage aisles. Finally, comparing lift and vehicle waiting time, it should

be noticed that in scenarios A, B, D and E vehicle waiting time is higher; conversely, in the other scenarios lift waiting time is higher. Therefore, it is possible to have higher criticalities due to vehicles or lifts as a function of the rack configuration and transaction demand rate.

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

Take in Table 5

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

6. Conclusion and further research

The present study focuses on the analysis of AVS/RS performances for product tote, with a "tier captive" configuration. We propose an analytical model by expanding previous contributions in literature about AVS/RS with "tier to tier" configuration. The model is based on analytical models to estimate the travel distances and queuing network theory to analyse the average waiting times. By means of the proposed model, an overall estimate of the AVS/RS performances (i.e. transaction cycle time, waiting time) can be provided in a short time. The model is mainly useful within the "conceptualisation" stage in warehousing design.

The model has been validated through simulation. Results confirm the effectiveness of the model in estimating the transaction cycle time, and specifically the waiting time, which is indeed its more critical component to be evaluated. Additionally, results highlight the existing complexity for a warehouse designer to evaluate the AVS/RS performances because of the combined effect of the kinematic behaviour of vehicles and lift, as well as the creation of queues deriving from the interaction of the two above-mentioned resources. In turn, such complexity may be viewed as a function of the rack configuration (i.e. number and length of storage

aisles and number of tiers), and make it difficult to evaluate the suitability of these systems.

As such, further development of the present research may go towards the development of a general framework for AVS/RS rack configuration, and study the cost effectiveness of AVS/RS for product totes compared to other automated systems (e.g. miniload). Finally, the present analytical model may be extended by considering dual command cycles.

References

Allen A., 1990. *Probability, Statistics and Queueing Theory with Computer Science Applications.* 2nd ed. New York: Academic Press.

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

Dallari F., Marchet G. and Melacini M., 2009. Design of Order Picking System. *International Journal of Advance Manufacturing Technology*, 42 (1), 1–12.

Ekren B.Y. and Heragu S.S., 2010. Simulation Based Regression Analysis for Rack Configuration of Autonomous Vehicle Storage and Retrieval Systems. *International Journal of Production Research*, 48 (21), 6257–6274.

Ekren B.Y., Heragu S.S., Krishnamurthy A. and Malmborg C.J., 2010. Simulation Based Experimental Design to Identify Factors Affecting Performance of AVS/RS. *Computers & Industrial Engineering*, 58 (1), 175–185.

Fukunari M. and Malmborg C.J., 2008. An Efficient Cycle Time Model for Autonomous Vehicle Storage and Retrieval Systems. *International Journal of Production Research*, 46 (12), 3167–3184.

Fukunari M. and Malmborg C.J., 2009. A network approach for evaluation of performance measures in Autonomous Vehicle Storage and Retrieval Systems. *European Journal of Operational Research* 193 (1), 152–167.

Heragu, S.S., 2008. *Facilities design*, Clermont, FL: CRC Press.

Heragu S.S., Cai X., Krishnamurthy A. and Malmborg C.J., 2009. *Analysis of Autonomous Vehicle Storage and Retrieval System by Open Queueing Network*. 5th ed. Annual IEEE Conference on Automation Science and Engineering, Bangalore (India).

Hwang H. and Lee S.B., 1990. Travel time model considering the operating characteristics of the storage and retrieval machine. *International Journal of Production Research*, 28 (10), 1779–1789.

Jia J. and Heragu S., 2009. Solving semi-Open Queuing Networks. *Operation Research*, 57 (2), 391–401.

Kuo P.H., Krishnamurthy A. and Malmborg C.J., 2007. Design Models for Unit Load Storage and Retrieval Systems Using Autonomous Vehicle Technology and Resource Conserving Storage and Dwell Point Policies. *Applied Mathematical Modelling*, 31 (10), 2332–2346.

Kuo, P. H., Krishnamurthy, A. and Malmborg, C.J. 2008. Performance Modelling of Autonomous Vehicle Storage and Retrieval Systems Using Class-Based Storage

Policies. *International Journal of Computer Applications in Technology*, 31 (3/4), 238–248.

Law A.M. and Kelton W.D., 2000. *Simulation Modeling and Analysis*. 3rd ed. New York: McGraw-Hill.

Lerher T., Sraml M., Kramberger J., Potrc I., Borovinsek M. and Zmazek B., 2006. Analytical travel time models for multi aisle automated storage and retrieval system. *International Journal of Advance Manufacturing Technology*, 30 (3/4), 340–356.

Little J., 1961. A Proof for the Queueing Formula: L =λW. *Operations Research* 9 (3), 383–387.

Malmborg C.J., 2002. Conceptualizing Tools for Autonomous Vehicle Storage and Retrieval Systems. *International Journal of Production Research*, 40 (8), 1807–1822.

Malmborg C.J., 2003. Interleaving Dynamics in Autonomous Vehicle Storage and Retrieval Systems. *International Journal of Production Research*, 41 (5), 1057–1069.

Pujolle G. and Ai W., 1986. A Solution for Multiserver and Multiclass Open Queueing Networks. *INFOR*, 24 (3), 221–230.

Roy D., Krishnamurthy A., Heragu S.S. and Malmborg C.J., 2009. *Impact of Zones on Throughput and Cycle Times in Warehouses with Autonomous Vehicles*. 5th ed. Annual IEEE Conference on Automation Science and Engineering, Bangalore (India).

Whitt. W., 1983a. Performance of the Queueing Network Analyzer. *Bell System Technical Journal*, 62 (9), 2817–2843.

Whitt W., 1983b. The Queueing Network Analyzer. *Bell System Technical Journal*, 62 (9), 2779–2815.

Zhang L., Krishnamurthy A., Malmborg C.J. and Heragu S.S., 2009. Variance-based Approximations of Transaction Waiting Times in Autonomous Vehicle Storage and Retrieval Systems. *European Journal Of Industrial Engineering*, 3 (2), 146–169.
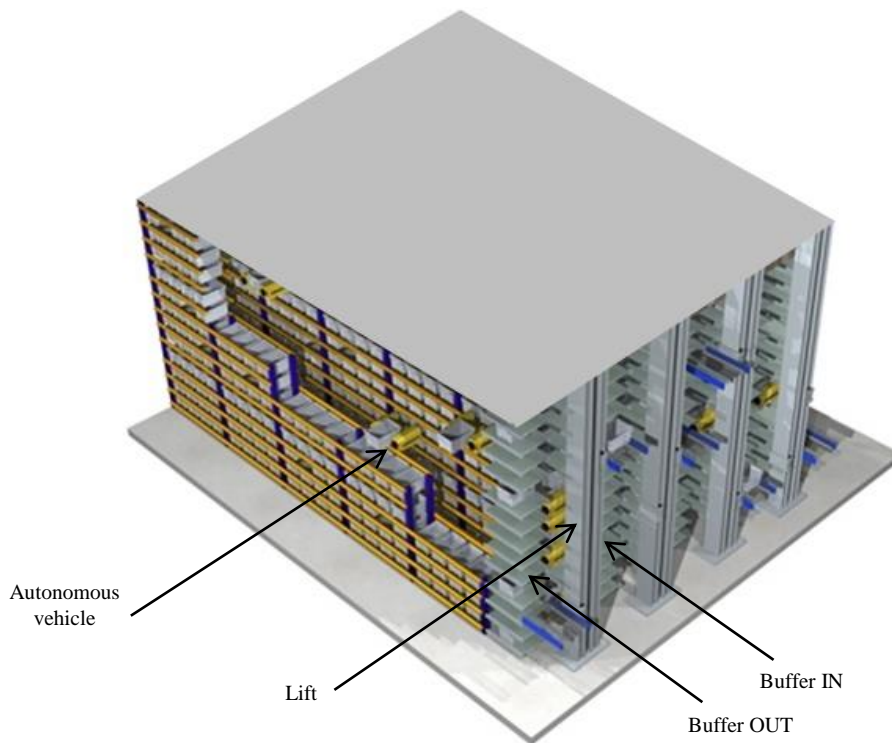
Figure 1. AVS/RS for product tote with "tier captive" configuration.
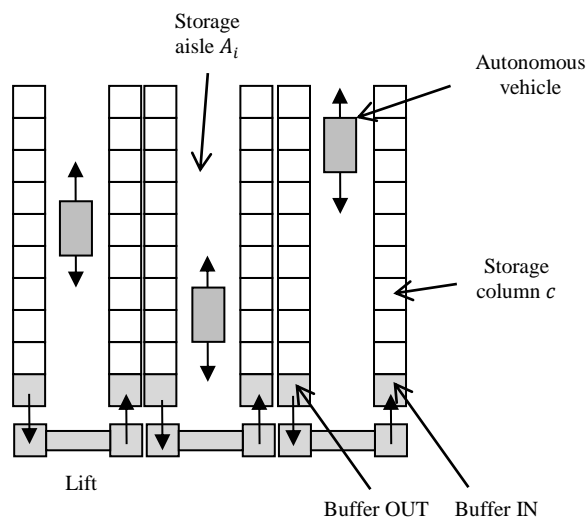

Figure 2. Image illustrating a tier of AVS/RS for product tote with "tier captive" configuration.
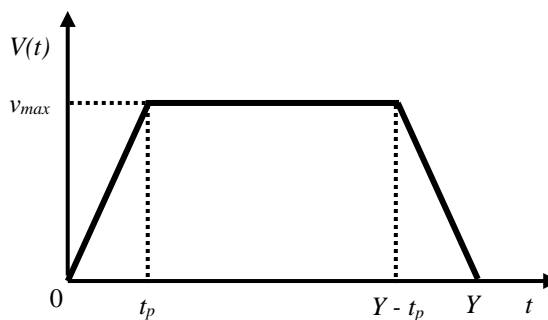

Figure 3. Velocity – time relationship of the vehicles traveling (with $Y > 2 \, v_h / a_h$).
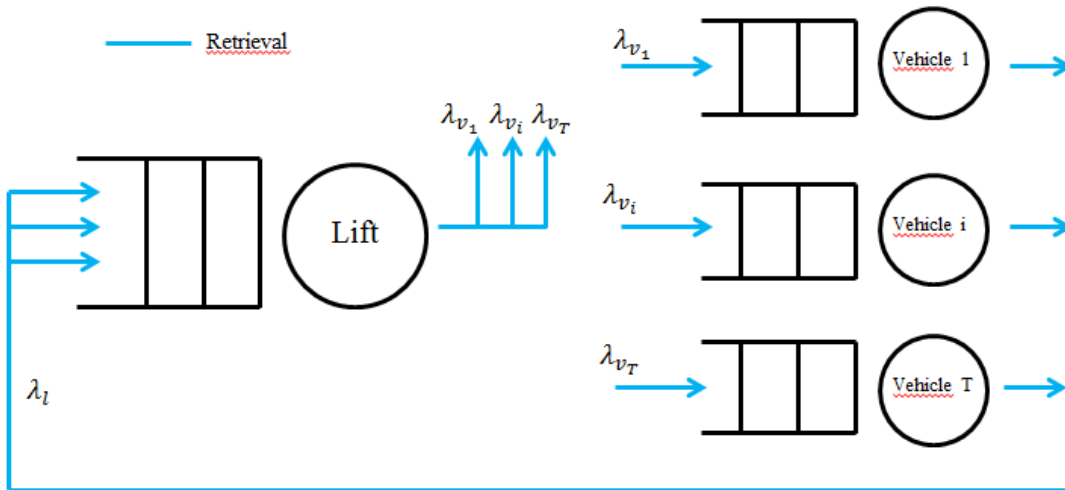
Figure 4. Modelling a single aisle of AVS/RS as an open queuing network.
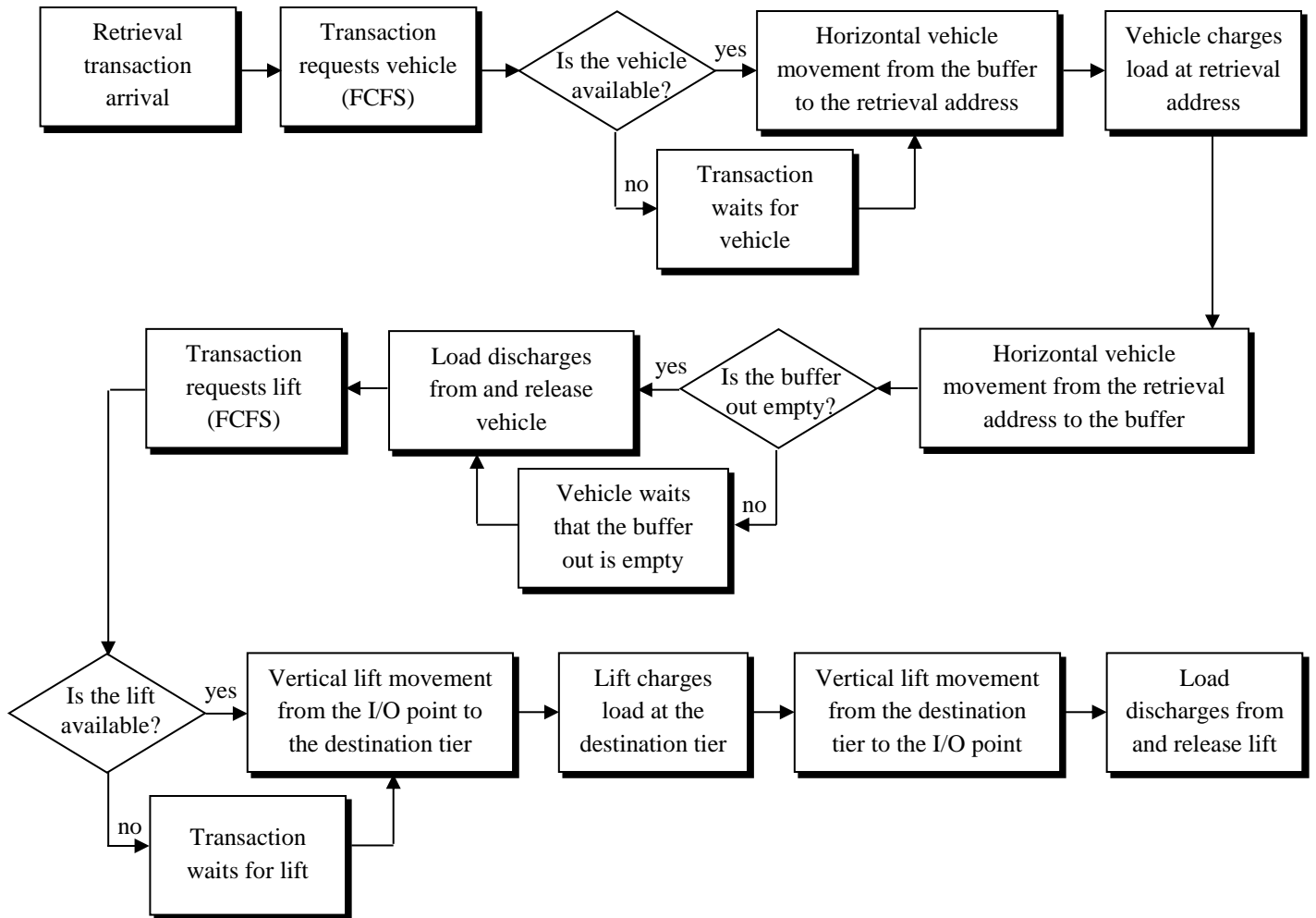

Figure 5. Simulation flow chart for retrieval transaction.

Paper	Configuration	I/O number and location	Storage policy	Type of cycles
Malmborg (2002)	"tier to tier"	1, corner	random	SC/DC
Malmborg (2003)	"tier to tier"	1, corner	random	SC/DC
Kuo et al. (2007)	"tier to tier"	1, corner	random	SC
Fukunari and Malmborg (2008)	"tier to tier"	1, corner	random	SC/DC
Kuo and Malmborg (2008)	"tier to tier"	1, corner	zoning	SC
Fukunari and Malmborg (2009)	"tier to tier"	1, corner	random	SC/DC
Roy et al. (2009)	"tier to tier"	More than 1, in the middle	random	SC
Zhang et al. (2009)	"tier to tier"	1, corner	random	SC
Ekren and Heragu (2010)	"tier to tier"	More than 1	random	SC
Ekren et al. (2010)	"tier to tier"	More than 1	random	SC/DC

Legend: SC: Single Command cycle; DC: Dual Command cycle

Table 1. Overview of the main studies on AVS/RS.

Scenario	Number of storage locations (n)	Number of tiers (T)	Number of storage aisles (A)	Number of storage columns (C)	λ [retrievals/hour]	ρ_l (lift utilisation)	ρ_{vt} (vehicle utilisation)
A	1,000	12	1	42	100	0.122	0.040
B	1,000	12	1	42	400	0.489	0.170
C	1,000	12	1	42	600	0.734	0.260
D	10,000	12	3	139	300	0.120	0.118
E	10,000	12	6	70	2,500	0.510	0.270
F	10,000	12	9	47	5,500	0.748	0.289

Table 2. Main data for the examined scenarios.

Variable	Unit of measure	Data
μ_w	m	0.5
μ_h	m	0.8
v_h	m/s	1.5
v_v	m/s	5
a_h	m/s^2	1
a_v	m/s^2	7
ε_1	s	3
ε_2	s	2

Table 3. Main data for the scenario analysis.

Scenario	Transaction cycle time [s/transaction]			Lift utilisation [%]			Vehicle utilisation [%]		
	Model	Simulation	Percentage deviation [%]	Model	Simulation	Percentage deviation [%]	Model	Simulation	Percentage deviation [%]
A	23.545	23.500	0.197%	0.122	0.122	0.122%	0.044	0.043	0.332%
B	25.300	25.102	0.784%	0.490	0.489	0.144%	0.174	0.173	0.913%
C	29.016	28.786	0.793%	0.735	0.734	0.165%	0.262	0.260	0.752%
D	55.869	55.771	0.175%	0.123	0.122	0.531%	0.118	0.117	0.753%
E	34.726	34.532	0.559%	0.510	0.509	0.218%	0.272	0.270	0.621%
F	31.003	30.810	0.622%	0.748	0.747	0.216%	0.290	0.288	0.582%

Table 4. Comparison between analytical model and simulation results with respect to lifts utilisation, vehicles utilisation and transaction cycle time.

Scenario	Lift waiting time [s]			Vehicle waiting time [s]		
	Model	Simulation	Percentage deviation [%]	Model	Simulation	Percentage deviation [%]
A	0.308	0.296	3.796%	0.508	0.509	-0.122%
B	2.064	1.972	4.433%	2.354	2.295	2.528%
C	5.760	5.681	1.375%	3.960	3.866	2.380%
D	0.308	0.299	2.673%	4.375	4.283	2.100%
E	2.166	2.056	5.079%	6.447	6.333	1.780%
F	6.109	6.011	1.607%	4.997	4.922	1.509%

Table 5. Comparison between analytical model and simulation results with respect to vehicle and lift waiting times.