Application the Queuing Theory in the Warehouse Optimization

Jaroslav Masek, Juraj Camaj, Eva Nedeliakova

Abstract—The aim of optimization of store management is not only designing the situation of store management itself including its equipment, technology and operation. In optimization of store management we need to consider also synchronizing of technological, transport, store and service operations throughout the whole process of logistic chain in such a way that a natural flow of material from provider to consumer will be achieved the shortest possible way, in the shortest possible time in requested quality and quantity and with minimum costs. The paper deals with the application of the queuing theory for optimization of warehouse processes. The first part refers to common information about the problematic of warehousing and using mathematical methods for logistics chains optimization. The second part refers to preparing a model of a warehouse within queuing theory. The conclusion of the paper includes two examples of using queuing theory in praxis.

Keywords—Queuing theory, logistics system, mathematical methods, warehouse optimization.

I. INTRODUCTION

MATHEMATICAL optimization methods are generally quite often used as a tool that allows you to shorten processing times and reduce the cost of implementation processes. Application of appropriate optimization methods can reduce the cost of implementing the activities of sub-logistics chain and subsequently the supply chain as a whole. Storage and transport processes together with the sub-logistics chain that are the most cost making and therefore specially checked [1], [3]. Choosing the appropriate method and its subsequent application to a particular process in the supply chain can reduce the time request of the implemented processes, thereby reducing costs.

In the context with the application of mathematical optimization methods, it should be noted that in addition to providing the benefits associated with them some disadvantages may occur. These include the basic condition using mathematical optimization method, which is the algorithm development [7]. This process necessarily leads to some simplification activities and tasks related to the solution. Another disadvantage can be a time-consuming and complicated mathematical use. Nevertheless, these methods often serve as a basis for proposed changes aimed at optimizing processes.

The logistics chain consists of subsystems and these in turn are made up of specific processes and activities. In the warehousing subsystem processes are mainly related to income and expenses by establishing the material, but also process such as packing, sorting, labelling and many others [2], [8]. To optimize these processes several mathematical methods can be applied to be used as a part of the decision making and management of logistics processes.

Among the most frequently used mathematical optimization methods, which are used as a tool to optimize warehouse management include:
- inventory management models;
- queuing theory;
- methods of graph theory, etc.

The models based on queuing theory have recently appeared as a suitable tool to optimize storage management [4]. This is mainly because they can deal with complex processes, respectively processes as part of the logistics system and not as a set of isolated subsystems.

II. THE CHARACTERISTICS OF QUEUE THEORY

Queue theory is the mathematical study of waiting lines, or queues. Queue theory deals with the phenomena and processes that are permanent having the character of large scale and with the creation servicing requirements and operating time are exposed to random influences. Queueing theory is the mathematical study of waiting lines, or queues. In queueing theory a model is constructed so that queue lengths and waiting time can be predicted [1]. Queueing theory is generally considered a branch of operations research because the results are often used when making business decisions about the resources needed to provide a service.

Queueing theory has its origins in research by Agner Krarup Erlang when he created models to describe the Copenhagen telephone exchange [9]. The ideas have since seen applications including telecommunication, traffic engineering, computing. Queue model consists of resource requirements and operating system itself. Queue theory provides information on the likely or average throughout the process queue.

The requirements entering the system queue theory are sequentially serviced by the channels (lines) of the services. If lines are busy demands waiting for served in the queue and will be serviced in the order which entered into the queue [9]. The intervals between the input requirements may be constant,

Jaroslav Masek, PhD. is with the Department of Railway Transport, University of Zilina, Univerzitna 8215/1, 010 26 Zilina, Slovakia (Corresponding author, phone: +421-41-531 3423; e-mail: juraj.camaj@fpedas.uniza.sk).

Juraj Camaj, PhD. and Eva Nedeliakova, PhD., Assoc. Prof. are is with The Department of Railway Transport, University of Zilina, Univerzitna 8215/1, 010 26 Zilina, Slovakia (E-mail: juraj.camaj@fpedas.uniza.sk, eva.nedeliakova@fpedas.uniza.sk).

World Academy of Science, Engineering and Technology
International Journal of Industrial and Manufacturing Engineering
Vol:9, No:11, 2015

but most are random variables. A schematic illustration of the process in queue system is shown in Fig. 1.

Typical examples might be:

- banks/supermarkets - waiting for service;
- computers - waiting for a response;
- failure situations - waiting for a failure to occur e.g. in a piece of machinery;

- public transport - waiting for a train or a bus;
- logistics problems – warehousing.

To analyze every sub-system we need information relating to [10]:

- arrival process;
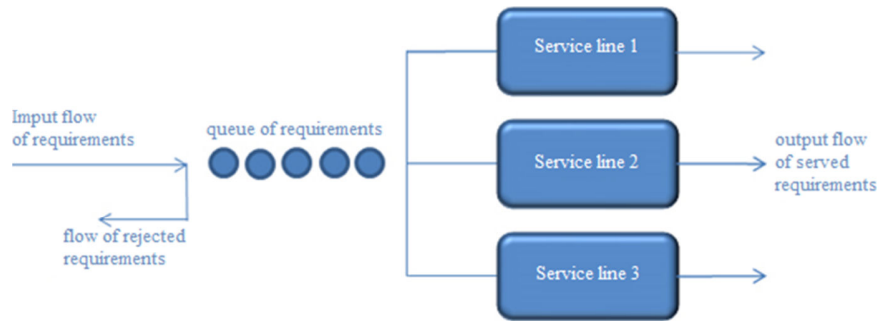- service mechanism;
- queue characteristics.

Fig. 1 Queue system with one front and three service lines

The simplest arrival process is one where we have completely regular arrivals (i.e. the same constant time interval between successive arrivals). A Poisson stream of arrivals corresponds to arrivals at random [5]. In a Poisson stream successive customers arrive after intervals which independently are exponentially distributed. The Poisson stream is important as it is a convenient mathematical model of many real life queuing systems and is described by a single parameter - the average arrival rate. Other important arrival processes are scheduled arrivals; batch arrivals; and time dependent arrival rates (i.e. the arrival rate varies according to the time of day) [10]. Arrival process includes:

- how customers arrive e.g. singly or in groups (batch or bulk arrivals);
- how the arrivals are distributed in time (e.g. what is the probability distribution of time between successive arrivals (the interarrival time distribution));
- whether there is a finite population of customers or (effectively) an infinite number.

Assuming that the service times for customers are independent and do not depend upon the arrival process is common. Another common assumption about service times is that they are exponentially distributed. Service mechanism is [10]:

- a description of the resources needed for service to begin;
- how long the service will take (the service time distribution);
- the number of servers available;
- whether the servers are in series (each server has a separate queue) or in parallel (one queue for all servers);
- whether pre-emption is allowed (a server can stop processing a customer to deal with another "emergency" customer).

Queue characteristics considered:

- how, from the set of customers waiting for service, do we choose the one to be served next (e.g. FIFO (first-in first-out) - also known as FCFS (first-come first served); LIFO

(last-in first-out); randomly) (this is often called the queue discipline) [10];
- do we have:
- balking (customers deciding not to join the queue if it is too long);
- reneging (customers leave the queue if they have waited too long for service);
- jockeying (customers switch between queues if they think they will get served faster by so doing);
- a queue of finite capacity or (effectively) of infinite capacity.

Changing the queue discipline (the rule by which we select the next customer to be served) can often reduce congestion. Often the queue discipline "choose the customer with the lowest service time" results in the smallest value for the time (on average) a customer spends queuing.

Critical factors (characteristics) affect the queuing system are:

- the average number of requests entering the per unit time;
- the probability distribution of inputs requirements to the system (constant or random);
- the average number of services executed by one line of the service per unit of time.

To characterize a queueing system we have to identify the probabilistic properties of the incoming flow of requests, service times and service disciplines [9]. The arrival process can be characterized by the distribution of the interarrival times of the customers, denoted by A (t), that is:

$$A\ (t) = P\ (\text{interarri val time} < t) \qquad (1)$$

In queueing theory these interarrival times are usually assumed to be independent and identically distributed random variables. The other random variable is the service time, sometimes it is called service request, work. Its distribution function is denoted by B (x), that is:

World Academy of Science, Engineering and Technology
International Journal of Industrial and Manufacturing Engineering
Vol:9, No:11, 2015

$$B\ (x) = P\ (service\ \ time\ < x) \qquad (2)$$

The service times, and interarrival times are commonly supposed to be independent random variables. The structure of service and service discipline tell us the number of servers, the capacity of the system that is the maximum number of customers staying in the system including the ones being under service [9]. The service discipline determines the rule according to the next customer is selected. The most commonly used laws are [11], [12]:

- first in first out - this principle states that customers are served one at a time and that the customer that has been waiting the longest is served first;
- last in first out - this principle also serves customers one at a time, but the customer with the shortest waiting time will be served first, also known as a stack [11];
- processor sharing - service capacity is shared equally between customers;
- priority - customers with high priority are served first. Priority queues can be of two types, non-preemptive (where a job in service cannot be interrupted) and preemptive (where a job in service can be interrupted by a higher priority job). No work is lost in either model;
- shortest job first - the next job to be served is the one with the smallest size [12];
- preemptive shortest job first - The next job to be served is the one with the original smallest size;
- shortest remaining processing time - The next job to serve is the one with the smallest remaining processing requirement;
- service facility:
- customer's behavior of waiting.

The aim of all investigations in queueing theory is to get the main performance measures of the system, which are the probabilistic properties (distribution function, density function, mean, variance) of the following random variables: number of customers in the system, number of waiting customers, utilization of the server/s, response time of a customer, waiting time of a customer, idle time of the server, busy time of a server [9]. Of course, the answers heavily depend on the assumptions concerning the distribution of interarrival times, service times, number of servers, capacity and service discipline. It is quite rare, except for elementary or Markovian systems, that the distributions can be computed. Usually their mean or transforms can be calculated.

The resulting parameters, such as the number of requests in the system, using the service time of line, waiting time of the requirements in the queue, and so are the basis for the decision, and then to select the appropriate option [6].

Economic balance between the cost of acquisition and operation of lines and losses of service of waiting requirements in the queue determines the optimal number of lines. The solution can be obtained analytically (using calculation relations) or by simulation. By analytically way is possible to finalize the results of more simple processes (determining the optimal number of parallel lines of service), the simulation is designed to deal with more complex processes (some combination of serial and parallel lines). In addition, the simulation can also be used for training the staff, and also for cases involving crisis management.

III. APPLICATION OF QUEUING MODELS

Queue models can be used to determine the density of the terminal networks, the size and capacity of the warehouses, determine the types of handling equipment and others. By application of queue theory the process can be addressed and realized in the warehouse management, such as activities in the central warehouse, which deals with assembling of shipments from multiple suppliers and then distributing them according to a combination of specific customer requirements.

In developing the model should first define the warehouse elements, which will enter the queuing system. In the model the queuing process consists of resource requirements which are the stock items in the form of goods packed in cartons. Front of awaiting requirements is made by items that are collected for the roller track (stack). Stock items are handled manually by workers from pallets to the roller conveyor.

Service lines in this model consist of e-points (computers) that are used to enter the input information about the weight, number, or other characteristics of items and income records. Each e-point is operated by one employee – storekeeper, operator. Over serviced is every item transfer by the roller conveyor and then dispatched in a direction orthogonal to conveyor belt (Fig. 2). Next is the loading place, where are the items serviced again by the operator and sorted by identifiers. Typical are for identification used bar codes or RFID tags. In this model of storage solutions the queue system can be used twice (income and outcome goods), but as an example to illustrate the application queue theory it will only be used to optimize income items.

Queue system provides information about probable or average course of the procedure and is therefore suitable for the detection of information about:

- the number of items in the stack (front);
- average waiting time over to serve;
- changes caused by increasing or decreasing the number of service lines (e-points);
- changes caused by using one stack for supply two lines of service.

Based on these data (model characteristics) it is possible to choose the best variant - the number of service lines, what is resulting into the determination of the necessary number of the operators (workers).

Queue system is characterized by two main parameters:

- number of items entering the system (x),
- number of items served per unit of time (y).

World Academy of Science, Engineering and Technology
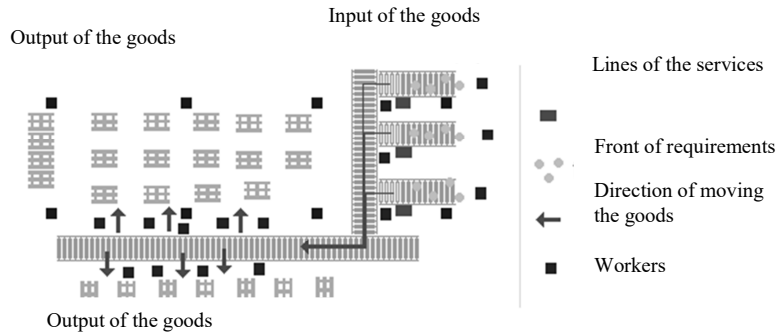International Journal of Industrial and Manufacturing Engineering
Vol:9, No:11, 2015

Fig. 2 Processes solution in the warehouse by queue theory application

The basic condition for the solution is to reach such a state, when entering the number of items is less than the number of items served. Other important characteristics necessary for the calculation are:

- type of distribution of a random variable, which manage the input flows and the operation,
- number of service stations;
- limitation of waiting inputs;
- the order of the items in the front.

Methods queue theory contains more parameters and calculations appropriate for determining the optimal variant. Using of queue theory may be to demonstrate on the models mentioned warehouse, but only with one e-point.

On average, to the system enters 80 items within an 8-hour workday, and these inputs are exponentially distributed. Served one item on entrance e-point takes 5 minutes and this time has an exponential distribution too.

Summary model depends on this information:
- input (x = 10 items per hour);
- services capacity (y = 12 items per hour).

Operating costs of one e-point is 110 Euro per day. Wage of the worker who serve e-point is 4.00 Euro per hour. Wage of the worker who fills the stack (front) is 7.00 Euro per hour. One year has 250 working days.

Traffic density is calculated from the relationship:

$$\rho = x / y \tag{3}$$

The average number of items in the stack is determined from the relationship:

$$\Phi L = 2\rho / (1 - \rho) \tag{4}$$

The average waiting time in the stack is:

$$\Phi W = \rho / [y . (1 - \rho)] \tag{5}$$

A. Example 1: The One-Line Service and One Front

The calculation (1) of the density of traffic:

$$\rho = 10/12 = 0.833$$

Calculate (2) the average number of items in the stack:

$$\Phi L = 0.833^2 / (1 - .833) = 4.166$$

Calculate (3) the average waiting time in the tank:

$$\Phi W = 0.833 / [12 (1 - .833)] = 0.413 \text{ h} = 25 \text{ min}$$

Economic evaluation of example 1: Based on data on traffic density, it is necessary to employ 5 workers serving supply stack (5 x 7.00 x 8 x 250 = 70 000 Euro/year). The annual cost of e-point is the (110.00 x 250) 27 500 Euro per worker, which it serves 8 000 Euro/year (1 x 4.00 x 8 x 250). The total annual cost is therefore 105 500 Euro.

B. Example 1: Two Lines and Two Service Fronts

The calculation (1) of the density of traffic:

$$\rho = 5/12 = 0.416$$

Calculate (2) the average number of items in the stack:

$$\Phi L = 0.416^2 / (1 - .416) = 0.297$$

Calculate (3) the average waiting time in the tank:

$$\Phi L = 0.416 / [12 (1 - .416)] = 0.059 = 3.6 \text{ min}$$

Economic evaluation of example 2: Thanks the significant time savings in storage waiting items, we can think only with of two workers serving supply stack with the annual salary (2 x 7.00 x 8 x 250) 28 000 Euro. Annual operating costs of two e-Point is 55 000 Euro and wages of workers who serve there are 16 000 Euro/year (2 x 4.00 x 8 x 250). Total annual costs in this case amount to 99 000 Euro.

IV. CONCLUSION

Based on the economic evaluation of the two examples it has been shown that while increasing the number of service lines, the variant with two e-points seems better in terms of operating costs.

It is necessary to mention that the values of the parameters may be changed due to change in the input data. However, it is important that the application of queuing theory on the created model determines the basic characteristics and the expected course of these processes and the behaviour of the warehouse.

World Academy of Science, Engineering and Technology
International Journal of Industrial and Manufacturing Engineering
Vol:9, No:11, 2015

quality of education and development of the human resources as pillars of the knowledge society at the Faculty PEDAS", ITMS project code 26110230083, **University of Zilina**.

Modern education for the knowledge society / Project is co-financed by funds from the EC

Európska únia
Európsky sociálny fond

REFERENCES

[1] Camaj, J., Dolinayova, A., Lalinska, J., Bariak, M.: The Technological Problem of Simulation of the Logistics Centre. IN: ICLT 2015 : 17th International Conference on Logistics and Transportation on April, 8-9, 2015 at Dubai, UAE. World Academy of Science, Engineering and Technology, International Science Index, Economics and Management Engineering, eISSN: 1307-6892, p. 2613-2617.

[2] Camaj, J., Lalinska, J., Masek, J.: Simulations of continental logistics centre from the perspective of technologist. IN: ICIMSA 2014 International Conference on Industrial Engineering, Management Science and Applications 2014. Proceeding book of conference. Beijing: IEEE, 2014. p. 305-308.. ISBN 978-1-4799-6541-0.

[3] Camaj J, Dolinayova, A,: The organization of the wagon flows with respect to the economic criteria of railways companies. In: LOGI 2011: 12th international scientific conference, November 24th, 2011 in Pardubice. Czech Republic - ISBN 978-80-263-0094-6. p. 140-148.

[4] Gašparik, J., Meško, P., Zitrický, V.: Role of rail cargo companies in logistic chains. In: Railway logistics and rail cargo: proceedings of the 5th SoNorA University Think Tank Conference: 17th of June 2010, (Erfurt). Erfurt: Institut Verkehr und Raum der Fachhochschule Erfurt, 2010. S. 19-26. ISSN 1868-8411.

[5] Sztrik, J.: Basic Queueing Theory University of Debrecen, Faculty of Informatics,

[6] Kendra M., Lalinska J., Camaj, J.: Optimization of transport and logistics processes by simulation. In: ISTEC: 3rd international science, technology and engineering conference, December 13-14-15, 2012, Dubai, United Arab Emirates (UAE). - ISSN 2116-7383. - p. 886-892.

[7] Klapita, V., Mašek, J.: Processes solution in the warehouse by the queueing theory application. In: Horizons of railway transport: scientific papers. Vol. 3, no. 1 (2012), s. 64-69. ISSN 1338-287X.

[8] Mašek, J., Zitrický, V.: Systematic design of stores as part of the logistics, In: LOGI 2010: 11th international scientific conference, November 19th, 2010, Pardubice, Czech Republic, Brno: Tribun EU, 2010. ISBN 978-80-7399-205-7.

[9] Sundarapandian, V. "7. Queueing Theory". Probability, Statistics and Queueing Theory. PHI Learning (2009). ISBN 8120338448.

[10] Beasley, J. E.: Operations research - notes. Brunel University, West London. http://people.brunel.ac.uk/~mastjjb/jeb/or/contents.html

[11] Penttinen A.: Chapter 8 – Queueing Systems, Lecture Notes: S-38.145 - Introduction to Teletraffic Theory.

[12] Harchol-Balter, M.: Scheduling: Non-Preemptive, Size-Based Policies. Performance Modeling and Design of Computer Systems. (2012) p. 499.