

Accepted Manuscript

Dynamic Allocation of Stochastically-Arriving Flexible Resources to Random Streams of Objects with Application to Kidney Cross-Transplantation

Yael Perlman , Amir Elalouf , Uri Yechiali

PII: S0377-2217(17)30711-7
DOI: [10.1016/j.ejor.2017.07.068](https://doi.org/10.1016/j.ejor.2017.07.068)
Reference: EOR 14618



To appear in: *European Journal of Operational Research*

Received date: 12 February 2017
Revised date: 25 July 2017
Accepted date: 29 July 2017

Please cite this article as: Yael Perlman , Amir Elalouf , Uri Yechiali , Dynamic Allocation of Stochastically-Arriving Flexible Resources to Random Streams of Objects with Application to Kidney Cross-Transplantation, *European Journal of Operational Research* (2017), doi: [10.1016/j.ejor.2017.07.068](https://doi.org/10.1016/j.ejor.2017.07.068)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Highlights

A dynamic allocation model of flexible resources to streams of objects is studied

Allocation probabilities depend on system's states

The model is applied to kidney cross-transplantation

Novel measure Expected Value of Transplantation based on human-leukocyte-antigen fit

Optimal probabilities of cross-transplantation are calculated

ACCEPTED MANUSCRIPT

Dynamic Allocation of Stochastically-Arriving Flexible Resources to Random Streams of Objects with Application to Kidney Cross-Transplantation

Yael Perlman¹ Amir Elalouf¹ Uri Yechiali⁴

³Department of Management, Bar Ilan University, Ramat Gan 52900, Israel

⁴Department of Statistics and Operations Research, School of Mathematical Sciences,
Tel Aviv University, Tel Aviv 69978, Israel

Abstract

Two distinct random streams of discrete objects flow into a system and queue in two separate lines. Concurrently, two distinct types of resources arrive stochastically over time. Upon arrival, each resource unit is matched with a waiting object. One resource type is 'flexible' and can be allocated to either one of the object types. However, units of the other, non-flexible, resource type can be allocated only to units of one specific object type. The allocation probabilities are not fixed and may depend on both queue sizes of the two objects. If a resource unit is not allocated immediately, it is lost. The goal is to find an optimal state-dependent probabilistic dynamic allocation policy. We formulate the system as a two-dimensional Markov process, analyze its probabilistic behavior, and derive its performance measures. We then apply the model to the problem of kidney cross-transplantation and propose a new measure of system effectiveness, called Expected Value of Transplantation (EVT), based on the histocompatibility between kidneys and candidates. We further show that it is possible to balance the objectives of achieving equity between the two groups and minimizing expected waiting times (EW) and maximizing EVT by equating the value of EW/EVT between the two groups.

1. Introduction

This paper studies a dynamic flexible-resource allocation problem that is encountered in numerous operational settings and is particularly salient in the context of live organ transplantation. We consider two random streams of discrete objects, denoted S_1 and S_2 which flow into a system and queue in two separate lines, denoted Q_1 and Q_2 respectively. In parallel, two distinct types of discrete resources, R_1 and R_2 arrive at the system stochastically over time. A unit of resource type R_1 can only be allocated to an S_1 -type object, whereas a unit of resource type R_2 can be allocated to either one of the two object types: It is allocated to an S_1 -type unit waiting in Q_1 (if the queue is not empty) with a probability that may depend on both queue sizes or, with the complementary probability, to a unit of an S_2 -object waiting (if any) in Q_2 . An arriving resource unit must be allocated to an object upon arrival, or it is lost. Thus, if a unit of resource R_2 arrives when there are no S_2 -type objects in Q_2 , it is discarded and disappears from the system, whereas an R_1 -type resource is discarded upon arrival only when both queues are empty (i.e., $Q_1 = Q_2 = 0$). This paper seeks to identify an optimal state-dependent probabilistic dynamic allocation policy in the setup described.

Flexible-resource allocation models characterizing manufacturing systems that can produce multiple products simultaneously have been addressed in the literature (see, e.g., Sethi and Sethi 1990; Buzacott and Shanthikumar 1993; and Perlman 2013). Additional applications include telecommunication networks (Ross 1995), in which incoming calls can be routed to multiple links. Similarly, in computer systems with multiple users and multiple servers, users can be dynamically routed to different servers, and computing capacity can be shared among different customers. Call centers are another important application of resource flexibility. As calls can vary by topic, urgency, duration, and level of difficulty, different call center operators may be trained to handle different subsets of call types (Koole and

Mandelbaum 2002; Shumsky 2004). Ahghari and Balcioglu (2009) performed extensive numerical studies based on realistic call center scenarios and showed that limited cross-training of operators (specifically, providing each agent with two additional skills) can considerably improve performance. Robbins and Harrison (2010) studied a call center queueing model with two customer types. They considered scenarios in which two separate teams were assigned to two different customer types, or in which a single cross-trained team could serve both types. They showed that cross-training a small number of agents results in substantial benefits as opposed to relying on agents who can only serve one type of customer.

We provide and analyze a general model of the flexible-resource queueing system described above, and then focus on its application in the context of live kidney transplantation. Stanford et al. (2014) discussed the problem of allocating stochastically-arriving kidneys to random streams of transplantation candidates, who form separate queues according to their blood types. Their model considers two blood types δ type O and type B δ where type O kidneys can be given to any candidate, and type B kidneys can only be given to candidates with blood type B. The authors propose that it is possible to allocate a fixed fraction of blood type O kidneys to blood type B candidates such that the expected waiting times (EW) for transplantation for the two different types of candidates are the same. However, the fixed probability assumption overlooks the situation in which a type O kidney arrives while there are only type O candidates in the system, but no type B candidates. Our model addresses this issue by assuming that, in such a situation, an arriving type O kidney is allocated exclusively to a type O candidate. More generally, a key contribution of our paper is in letting the probability of cross transplantation depend dynamically on the actual number of B candidates present in the system, rather than assuming that the probability is fixed.

Human tissue cells contain antigens that are immunologically relevant to specific candidate and donor. The system of these antigens is known as the Human Leukocyte Antigen (HLA) system. HLA matching is one of the most important factors when deciding on kidney allocation. A review of the determinants of successful kidney transplantation is given in Bendersky and David (2016). Proper HLA

matching decreases the risk of graft lost by about 40% (Takemoto et al. 2004). In applying our model to the kidney transplantation context, we propose a new means of measuring the effectiveness of the allocation system, beyond traditional metrics such as mean waiting times or mean queue sizes. This measure, which we refer to as the Expected Value of Transplantation (EVT), takes into account the extent to which the candidates and the kidneys they receive are compatible, i.e., matched in terms of their Human Leukocyte Antigen (HLA) groups. This measure constitutes another important contribution of our study. It is straightforward to generalize this measure to more traditional manufacturing systems, by attributing a value to each object-resource pair, representing the utility obtained from that specific pairing.

In a numerical analysis, we observe that long queues and long waiting times are associated with higher EVT values, as they increase the likelihood that an incoming kidney will find a well-matched candidate. On the other hand, long waiting times are expected to lead to deterioration of health, potentially culminating in death. Thus, we propose an additional measure of system effectiveness: the ratio of EW to EVT. The measure EW/EVT quantifies the rate of change in expected waiting time attributable to a change in EVT. Thus, this measure balances the two goals of achieving equitable waiting times and maximizing the overall quality of transplants. We show that only a small fraction of type O blood kidneys should be cross-transplanted to blood type B candidates in order to optimize the effectiveness of the system.

While the importance of having flexible resources or flexible servers in manufacturing and call center operations has long been recognized, in this paper we extend the scope of analysis by assuming that the number of available servers is not fixed but changes dynamically and randomly. Specifically, resources (e.g., kidneys, or servers) do not stay and wait for objects to arrive. Rather, individual resources arrive randomly over time, and each one must serve a waiting object (e.g., a transplantation candidate) as soon as it arrives; otherwise (i.e., if no appropriate objects are available), it is lost (i.e., disappears from the system).

The remainder of the paper is structured as follows. Section 2 presents the model formulation. In Section 3 we define and construct probability-generating functions (PGFs).

two-dimensional boundary probabilities, as well as its marginal state probabilities. In Section 4 we employ matrix geometric methods to further analyze the system and derive the system stability condition and various performance measures. In Section 5 we formulate the EVT as a measure of the effectiveness of a system for the dynamic allocation of kidneys for (cross-) transplantation. In Section 6 we perform numerical analysis and conclude that only a small fraction of the flexible resource should be allocated to cross-transplantation. Section 7 concludes the paper.

2. Model formulation

2.1. Problem description

Two Poisson streams of discrete objects, A and B , flow into a system at rates of λ_A and λ_B respectively, and queue in two separate lines, S_A and S_B . Concurrently, two types of discrete resources, R_1 and R_2 , arrive stochastically with Poisson rates μ_1 and μ_2 respectively. All four processes are mutually independent. When a unit of resource R_1 arrives, it is allocated to an object waiting in S_A . If S_A is empty, the unit is lost. However, an R_2 -type resource can be allocated to an object of either type. The probability that a unit of R_2 is allocated to an object is assumed to depend both on the number of S_B objects in S_B and on the number of S_A objects in S_A . If both queues are empty, the R_2 is lost.

Let n_A and n_B denote, respectively, the number of A objects and the number of B objects present in the system. It is assumed that the number of A objects is bounded such that it cannot exceed a given value N_A . Let π_{n_A, n_B} be the steady-state probabilities, and let α be the probability that an arriving R_2 resource is allocated to an object when the system is in state (n_A, n_B) . We assume that $\alpha = 1$ for all (n_A, n_B) , since when an R_2 resource is allocated only to an A object. We further set $\alpha = 0$ for $(n_A, n_B) = (0, 0)$ since when no object is present, an arriving R_2 resource is allocated with probability 1 to an object (if present).

Thekinen, J., J. H. Panchal. 2016. Journal of Manufacturing Systems. "Resource allocation in cloud-based design and manufacturing: A mechanism design approach.

Zenios, S.A. 1999. Modeling the transplant waiting list: A queueing model with renegeing. Queueing Syst. 31 2396251.

Zenios, S.A., G. M. Chertow, L. M. Wein. 2000. Dynamic allocation of kidneys to candidates on the transplant waiting list. Oper. Res. 48 5496569.

ACCEPTED MANUSCRIPT