L. Abolnikov, Prof.,
T. M. Zachariah, Ass. Prof.
Department of Mathematics, Loyola Marymount University
(Los Angeles, CA, USA, E-mail: Thomas.Zachariah@lmu.edu)

A Queueing Approach in Determining Optimal Number of Beds in a Hospital Serving Urgent and Non-Urgent Patients

(Recommended by Prof. E. Dshalalow)

In this paper, the authors consider a hospital as a multichannel queueing system that serves two different arriving Poisson flows of urgent and non-urgent customers (patients). The distribution of the service time is assumed to be exponential with different parameters for each kind of patients and priority for urgent patients. The results of the article are illustrated by several numerical examples obtained with the help of Mathematica program and can be easily applied to any hospital or its department.

Рассмотрена больница как многоканальная система массового обслуживания, которая обслуживает два различных входящих пуассоновских потока: срочных и несрочных пользователей (пациентов). Распределение времени обслуживания экспоненциальное с различными параметрами для каждого вида пациентов и приоритетом в обслуживании срочных пациентов. Приведены численные примеры, выполненные с помощью программы Mathematica. Полученные результаты могут быть применены к любой больнице или ее отделениям.

Key words: multichannel queueing system, optimal number of reserved beds in a hospital.

Introduction. A great deal of effort has been devoted to determine the allocation of the number of beds in a hospital department for the emergency patients so that the probability of refusal of such patients for immediate service is minimized. In allocating beds to different departments of a hospital, the overall objective is to balance the occupancy rates with the number of beds, subject to the constraint that emergency patients should be immediately admitted. G.Vassilacopoulos [1] developed a general priority queueing simulation model to determine the bed needs of hospital inpatient departments to meet daily community demand for inpatient care. By integrating queueing theory and compartmental models of flow F.Gorunescu [2] demonstrate how changing admission rates, length of stay and bed allocation influence bed occupancy, emptiness and rejection in departments of geriatric medicine. Their results show why 10–15 % bed emptiness is necessary to maintain service efficiency and provide a more responsive and cost effective service. Using discrete event simulation and hypothetical hospital data, A.Bagust [3] showed that regular bed crisis occur when medical bed occupancy is greater than 90 % and a risk of failure to admit occurs at occupancy rates above 85 %. S.Fomundam [4] surveys the contributions and applications of queueing theory in the field of healthcare. The authors provide sufficient information to analysts who are interested in using queueing theory to model a healthcare process and want to locate the details of relevant models.

In this paper, we consider a hospital (or each of its departments) as a large and complex queueing system with incoming flow of different kinds of arriving patients who need urgent or non-urgent help. Statistical analysis shows that both input flows of incoming patients and the service processes in the hospitals are usually rather complicated stochastic processes which makes researchers to use simulation as almost the only method of the analysis of the corresponding problems [1, 5].

However, analytical methods of queueing theory undoubtedly have an obvious advantage over simulation because they make it possible to find important characteristics of the system in an explicit form which is convenient for future analysis and optimization. These characteristics are, for example, the optimal number of beds for urgent and non-urgent patients [6], optimal distribution of other hospital facilities [7, 8], doctors and ambulance teams, virtual and real patient's waiting time and the length of queues of patients in different departments (which is especially important for the functioning of the emergency room) [9, 10], probabilities of refusals in hospitalizations of urgent and non-urgent patients, and others [6]. All this opens a way to solve many important optimization problems that arise in hospital practice that, in most cases, cannot be solved by simulation. Analytical queueing models 'are simpler, require less data and provide more generic results than simulation' [4].

In this paper, the authors describe a hospital in terms of queueing theory and use analytical methods of this theory to find the optimal number of hospital beds for urgent and non-urgent patients taking into account that urgent patients may need immediate help after their arrival to the hospital. Formulas obtained in the paper are illustrated by many numerical examples from hospital practice.

Arriving process. When considering a hospital as a queueing system, it is important to distinguish between non-urgent patients and patients who arrive to the hospital because of urgent needs (urgent patients) suffering from heart attack, stroke, accidents, etc. While non-urgent patients usually do not need immediate attention, the urgent patients are supposed to be served as soon as possible after their arrivals to the hospital. Another difference between these two kinds of patients is that, according to many authors investigating similar problems, arriving process of patients who need urgent help usually is statistically close to Poisson random process, while patients arriving to the hospital because of non-urgent needs may form a random process of any nature.

Service time. In queueing theory, 'service time' means random time necessary to serve a single customer arriving to the system. If we consider a patient as a customer and a bed as a server then the service time in such a queueing system will be the duration of time of a patient being in a hospital. This random variable may have different distributions for each department.

Determining these distributions is an important part of the analysis of any queueing system. If as a result of this statistical investigation the distribution of the service time turns out to be in a rather general form, the analytical solution of the corresponding queueing problem may be difficult, or even impossible. That is why in practice sometimes the real distribution of service time is approximated by exponential for which an analytical solution can be found. In this paper, we suppose that the service time has an exponential distribution (but different for urgent and non-urgent patients).

Service discipline. Since it is extremely important to have an available server at the moment of an urgent patient's arrival, the service discipline is arranged according to the following rule (we will call it 'one-threshold level admittance policy rule'). The total number of beds is divided in two parts, n basic and m reserved, such that if at the moment of a patient's arrival, the number of occupied beds is r, $r < n$, then the arriving patient is immediately accepted for service. However, if at the moment of a patient's arrival, the number r of occupied beds is such that $n \le r < n + m$, then the patient is accepted for service only if he is an urgent patient; otherwise, he is refused for service and leaves the system. If at the moment of a patient's arrival all beds are occupied, then he is refused for service and leaves the system regardless if he is an urgent or non-urgent patient.

The probabilities of the refusals in service of non-urgent and urgent patients are denoted by π_1 and π_2, respectively. They are the most important parameters of the system performance (in queueing theory such systems are called systems with refusals). We suppose that both urgent and non-urgent patients arrive to the hospital according to two different and independent Poisson processes with parameters λ_1 and λ_2, respectively. We also suppose that the duration of a patient staying in the hospital (service time) has exponential distributions with parameters μ_1 and μ_2 for urgent and non-urgent patients, respectively. Here the parameters λ_1 and λ_2 are the average number of patients daily arrived to the department and $1/\mu_1$, and $1/\mu_2$ are the average number of patients being served per day in the department.

The goal of the paper. The goal of this paper is to solve the following two optimization problems:

1. Given the total number of beds in a hospital (or in one of the departments), find the number of reserved beds such that the probability of the refusal in immediate service for urgent patients is minimized provided that the probability of the refusal of non-urgent patients is reasonably small.

2. Given the values of π_1 and π_2, find the minimum number of beds in the department so that optimal division of the number of beds for urgent and non-urgent patients is possible.

Method of solution and criteria of optimality. In this paper, a hospital (as well as each of its departments) is considered as a multichannel queueing system that consists of a certain number of identical servers (beds). Each server can serve only one customer (patient) at a time. The patients arrive to the system according to two independent Poisson processes with parameters λ_1 and λ_2 and are served according to the service discipline described above.

The distribution of the service time is exponential with parameters μ_1 and μ_2 for non-urgent and urgent patients, respectively. It can be seen that under these assumptions the main queueing process (the number of beds in the system at the instance *t*) is a discrete Markov process (a Markov chain). It enables us to use some of the results obtained in [6] where a similar queueing system is considered. Here the authors derive a system of Chapman-Kolmogorov differential equations describing the main queueing process and solve it using the method of generating functions. The ergodic probabilities of the states of the system are found and it is shown that the probabilities π_1 and π_2 of refusals in service of non-urgent and urgent patients are:

$$\pi_1 = (\rho_1 + \rho_2)^n \sum_{k=0}^{m} \frac{\rho_1^k}{(n+k)!} \rho_0 \pi_1 \tag{1}$$

and

$$\pi_2 = \frac{\rho_1^m (\rho_1 + \rho_2)^n}{(n+m)!} \rho_0, \tag{2}$$

respectively. Here

$$\rho_0 = \frac{1}{\sum_{k=0}^{n} \frac{(\rho_1 + \rho_2)^k}{k!} + \left(\frac{\rho_1 + \rho_2}{\rho_1} \right)^n \sum_{j=n+1}^{n+m} \frac{\rho_1^j}{j!}}, \quad \rho_1 = \frac{\lambda_1}{\mu_1}, \quad \rho_2 = \frac{\lambda_2}{\mu_2}.$$

Since an urgent patient's health condition may be life threatening, we want the probability π_2 of refusal in immediate service for urgent patients to be less than a certain predetermined small number (say 10^{-5}). This number is supposed to be chosen by the administrator of the hospital (department). At the same time we want to keep the probability π_1 of refusal in immediate service for non-urgent patients to be reasonably small (for example, we can require $\pi_1 \leq .15$. This means that more than 85 % of non-urgent patients are hospitalized on the day of their arrival). Therefore, in this situation there are two optimization problems related to the hospital practice.

Optimization criteria 1. Given $m + n = c$, c = constant, find m such that $\pi_1 \leq .15$ and $\pi_2 \leq 10^{-5}$, where c is the total number of beds in the department, and π_1 and π_2 are given in (1) and (2) (see example 1 below).

Optimization criteria 2. Given the optimization restrictions $\pi_1 \leq .15$ and $\pi_2 \leq$ $\leq 10^{-5}$, find the minimum total number of beds "c" for which the optimal division of the number of beds is possible (see example 4).

Finding the optimal number of reserved beds. We developed a Mathematica program that will output the values of π_1 and π_2 for any values of arrival and service rates for different capacities of the facility. The obtained results are illustrated below.

Mathematica program. The Mathematica program used for the calculations and one sample output are given below. The input values for the Mathematica program are: k $(= n + m)$, λ_1, λ_2, μ_1, μ_2 and the number of iterations in the "Do" loop.

The Mathematica Program: $k = 30$, $\lambda_1 = 6$, $\lambda_2 = 3$, $\mu_1 = \mu_2 = 1/2$

$\lambda_1 = 6;\ \lambda_2 = 3;\ \mu_1 = 1/2;\ \mu_2 = 1/2; k = 35;$

$\rho_1 = \frac{\lambda_1}{\mu_1};\ \rho_2 = \frac{\lambda_2}{\mu_2};$ temp = {"m", "π_1", "π_2"};

Do $\Big[\Big\{ n = k = m;$

$$p_0 = \frac{1}{\left(\sum_{k=0}^{n} \frac{(\rho_1 + \rho_2)}{k!} + \left(\frac{\rho_1 + \rho_2}{\rho_1}\right)^n \sum_{j=n+1}^{n+m} \frac{\rho_2^j}{j!}\right)};$$

$$\pi_2 = \frac{\rho_2^m \cdot (\rho_1 + \rho_2)^n}{(n+m)!} \cdot p_0;$$

$$\pi_1 = (\rho_1 + \rho_2)^n \cdot p_0 \cdot \sum_{k=0}^{m} \frac{\rho_2^k}{(n+k)!};\quad \text{AppendTo}[\text{temp}, \{m, \pi_1, \pi_2\}]\Big\}, \{m, 0, 15\}\Big].$$

TableForm[Partition[Platten[N[temp]], {3}], (3, 4)]

Output. Total number of beds $c = m + n = 30$, the probability of refusal in service of non-urgent patient is π_1 and the probabilities of refusal in service of urgent patient is π_2.

Sample Mathematica output:

m	π_1	π_2
0.	0.000126782	0.000126782
1.	0.000331055	0.0000845248
2.	0.000686448	0.0000563561
3.	0.00131175	0.0000375793
4.	0.0023938	0.0000250638
5.	0.00421514	0.0000167226
6.	0.00718351	0.0000111641
7.	0.0118571	7.46057×10^{-6}
8.	0.0189566	4.99345×10^{-6}
9.	0.0293539	3.35014×10^{-6}
10.	0.0440297	2.25549×10^{-6}
11.	0.0639965	1.52606×10^{-6}
12.	0.0901967	1.03955×10^{-6}
13.	0.123388	7.14515×10^{-7}
14.	0.164042	4.96775×10^{-7}
15.	0.212266	3.5034×10^{-7}

Example 1. (An optimal solution can be found). A 35-bed pediatric department of a hospital serves arriving urgent and non-urgent patients. It is necessary to determine whether it is possible to divide the total number of beds into two parts m and n, $m + n = 35$, according to the service discipline mentioned above in section 4 such that

(a) the probability of the refusal of urgent patients π_2 is less than 10^{-5} and

(b) more than 85 % of the non-urgent patients are served (i.e. $\pi_1 < .15$).

Suppose that statistical analysis shows that the stochastic processes of arriving patients and their service times can be described by Poisson distributions with parameter $\lambda_1 = 6, \lambda_2 = 3$ and exponential distributions with $\mu_1 = \mu_2 = 1/2$, respectively. Using the Mathematica program and formulas (1) and (2) we obtain that $m = 7$ meets our criteria of optimality (Table 1).

Therefore, the optimal division of the total 35 beds is: the number of beds for urgent patients is 7 and the number of beds for non-urgent patients is 28. In this case the probability of the refusal of non-urgent patients will be $\pi_1 = 0.0118571$ and the probability of the refusal of urgent patients will be $\pi_2 = 7.46057\times10^{-6}$.

Example 2. (An optimal division does not exist). In some cases an optimal division of the total number of beds that satisfies the given conditions does not exist. For example, consider a 40-bed cardiology department of a hospital. Suppose $\lambda_1 = 12$, $\lambda_2 = 8$, $\mu_1 = 1/2$ and $\mu_2 = 1/3$. The computed values of the probabilities of the refusals of the non-urgent and urgent patients for this case are given in the Table 2 below. It is clear from the Table 2 that in this case, it is not possible to find an optimal division of beds satisfying the given criteria.

Table 1. **Optimality is achieved when $m = 7, m + n = 35, \lambda_1 = 6, \lambda_2 = 3, 1/\mu_1 = 2, 1/\mu_2 = 2$**

Number of reserved beds m	Probability of refusal of non-urgent patients π_1	Probability of refusal of urgent patients π_2
0	0.000126782	0.000126782
1	0.000331055	0.0000845248
2	0.000686448	0.0000563561
3	0.00131175	0.0000375793
4	0.0023938	0.0000167226
5	0.00421514	0.0000167226
6	0.00718351	0.0000111641
7	0.0118571	7.46057×10^{-6}
8	0.0189566	4.99345×10^{-6}

Table 2. **Optimality is not achieved for any m; $m + n = 40, \lambda_1 = 12, \lambda_2 = 8, 1/\mu_1 = 2, 1/\mu_2 = 3$**

Number of reserved beds m	Probability of refusal of non-urgent patients π_1	Probability of refusal of urgent patients π_2
0	0.224392	0.224392
1	0.336999	0.126375
2	0.408457	0.0759919
3	0.4613925	0.0477473
4	0.505033	0.0310328
5	0.54367	0.0207582
6	0.579445	0.0142538
7	0.613467	0.0100339
8	0.646304	0.00723672
9	0.678226	0.00534589
10	0.709335	0.00404448
11	0.739623	0.00313364
12	0.769015	0.00248628
13	0.797383	0.00201974
14	0.824564	0.00167946
15	0.850364	0.0014288

Therefore, we conclude that in this case a 40-bed facility for cardiology department is not large enough to adequately serve the cardiology needs of the local community.

Example 3. (Finding the minimum number of beds for which the optimal solution exists.) In example 2, we saw that the number of beds in the facility is not large enough to get the optimum division of the total number of beds. In this situation, it is important to find how many beds the facility should have so that the optimum division of the number of beds is possible. In the following example, we consider this problem. Solving optimization problem for different values of c, $c = m + n$, we conclude that the minimum number of beds in this department for which the optimal division is possible is $c = 61$ (Table 3).

In this case, $m = 11$ and $n = 50$ and the probability of refusal of non-urgent patients will be $\pi_1 = 0.144832$ and the probability of refusal of urgent patients will be $\pi_2 = 7.08885 \times 10^{-6}$.

Example 4. (The optimal solution can be achieved with fewer numbers of beds.) In some cases, the optimal solution can be achieved with fewer numbers of beds than those available in a department. Our approach enables us to reduce the number of beds without loosing the quality of service (see optimization crite-

Table 3. **Finding minimum number of beds for which optimal solution exists,** $\boldsymbol{m + n = 61, \lambda_1 = 12, \lambda_2 = 8, 1/\mu_1 = 2, 1/\mu_2 = 3}$

Number of reserved beds m	Probability of refusal of non-urgent patients π_1	Probability of refusal of urgent patients π_2
0	0.0104001	0.0104001
1	0.0185131	0.00522724
2	0.0261056	0.00263804
3	0.0341019	0.00133646
4	0.04301	0.000679823
5	0.0531225	0.000347382
6	0.0646132	0.00017843
7	0.0775838	0.0000921937
8	0.0920877	0.0000479572
9	0.108142	0.0000251359
10	0.125735	0.0000132864
11	0.144832	7.08885×10^{-6}
12	0.165378	3.82113×10^{-6}
13	0.187305	2.08279×10^{-6}
14	0.210531	1.149×10^{-6}
15	0.234967	6.42091×10^{-6}

ria 2). Consider, for example, an Ob-gyn department of a hospital with the total number of beds $m+n=85$, and parameters $\lambda_1=10, \lambda_2=7, 1/\mu_1=3.5$, $1/\mu_2=$ $=2.5$. In this case the calculations show that the chosen criteria of optimality ($\pi_1 \leq .15$ and $\pi_2 \leq 10^{-5}$) is satisfied even without any reserved beds (i.e. $m=0$). It follows that to meet the conditions of optimality we do not need any reserved beds which means that the number of beds in the department is not optimal (in fact, it is too large; see Table 4) and can be reduced without reducing the quality of service.

Calculations using Mathematica program show that the total number of beds in this situation can be reduced to 71 without reducing the quality of service. When c = 71, m = 7 and n = 64 and the probability of refusal of non-urgent patients will be $\pi_1=$ 0.011715 and the probability of refusal of urgent patients will be $\pi_2=6.60387\times10^{-6}$ probability (Table 5).

Remark 1. It follows from Examples 2, 3 and 4 that the approach developed in this paper makes it possible to solve another kind of optimization problems important in hospital practice (see Optimization Problem 2, Service discipline

Table 4. **The facility already has excess number of beds, $m+n=85$, $\lambda_1=10$, $\lambda_2=7$, $1/\mu_1=3.5$, $1/\mu_2=2.5$**

Number of reserved beds m	Probability of refusal of non-urgent patients π_1	Probability of refusal of urgent patients π_2
0	1.24183×10^{-6}	1.24183×10^{-6}
1	2.75963×10^{-6}	6.27188×10^{-7}
2	5.01244×10^{-6}	3.16762×10^{-7}
3	8.59926×10^{-6}	1.59981×10^{-7}
4	1.43947×10^{-5}	8.0799×10^{-8}
5	2.37183×10^{-5}	4.08079×10^{-8}
6	3.85624×10^{-5}	2.06103×10^{-8}
7	6.19030×10^{-5}	1.04094×10^{-8}
8	9.81213×10^{-5}	5.25745×10^{-8}
9	0.000154	2.65541×10^{-9}
10	0.000237	1.34122×10^{-9}
11	0.000362	6.77461×10^{-10}
12	0.000545	3.42213×10^{-10}
13	0.000809	1.72882×10^{-10}
14	0.001185	8.73489×10^{-11}
15	0.001714	4.41415×10^{-11}

and Example 4). Namely, given the optimality conditions, we can find the minimum number of total beds in the department for which the desired level of service can be achieved. For example, in the situation of Example 2, we cannot achieve the desired level of service because the number of beds in the department is not large enough to provide needed parameters of service. However, the situation may be improved if the number of beds is at least 61. Therefore, $m + n =$ $= 61$ is the minimum number of beds for which the required level of service ($\pi_1 \leq$ $\leq .15$ and $\pi_2 \leq 10^{-5}$) can be achieved.

Remark 2. The service policy described in this paper "one-threshold level admittance policy rule" can be generalized to the case where the total number of patients arriving to the hospital is divided into three or more categories. Consider, for example, two-threshold level admittance policy rule. In this case, the arriving customers are divided into three categories: non-urgent, urgent and extremely urgent. Suppose also that there are two numbers m and s such that if at a moment of a patient's arrival the number of occupied beds is r, and $r < n$, then the patient is immediately accepted for service no matter to which category he belongs. However, if $n \leq r < n+m$ then the patient is accepted for service only if he is an urgent patient of the first or second category. Finally, if $n+m \leq r < n + m + s$

Table 5. **Reducing the number of beds so that the optimal solution is still possible,** $\boldsymbol{m + n = 71,\ \lambda_1 = 10,\ \lambda_2 = 7,\ 1/\mu_1 = 3.5,\ 1/\mu_2 = 2.5}$

Number of reserved beds m	Probability of refusal of non-urgent patients π_1	Probability of refusal of urgent patients π_2
0	0.000776988	0.000776988
1	0.00150747	0.000392569
2	0.00233971	0.000198415
3	0.00338495	0.000100326
4	0.00474236	0.0000507547
5	0.00651151	0.000025694
6	0.008798	0.0000130187
7	0.011715	6.60387×10^{-6}
8	0.0153823	3.35474×10^{-6}
9	0.0199238	1.70731×10^{-6}
10	0.0254634	8.70867×10^{-7}
11	0.0321205	4.45446×10^{-7}
12	0.0400052	2.28607×10^{-7}
13	0.0492138	1.1779×10^{-7}
14	0.059825	6.09754×10^{-8}
15	0.0718969	3.17353×10^{-8}

the arriving customer is accepted for service only if he is an extremely urgent patient. If $r = n + m + s$, then the arriving customer is refused for service and leaves the system. In this case, the problem of finding the optimal values of threshold levels can be solved by using the approach similar to one presented in this paper.

Розглянуто лікарню як багатоканальну систему масового обслуговування, яка обслуговує два різних вхідних пуасонівських потоки: термінових і нетермінових користувачів (пацієнтів). Розподіл часу обслуговування є експоненціальним з різними параметрами для кожного виду пацієнтів і пріоритетом в обслуговуванні термінових пацієнтів. Наведено числові приклади, виконані за допомогою програми Mathematika. Отримані результати можуть бути застосовані у будь-якій лікарні або її відділеннях.

1. *Vassilacopoulos G.* A simulation model for bed allocation to hospital inpatient departments. //Simulations. — 1985. — **45.** — P. 233—241.
2. *Gorunescu F., McClean S., Millard P. H.* Using a queueing model to help plan bed allocation in a Department of Geriatric Medicine. //Health Care Management Science. — 2002. — **5.** — P. 307—312.
3. *Bagust A., Place M., Posnett J. W.* Dynamics of bed use in accommodating emergency admissions: Stochastic simulation model. // British Medical Journal. — 1999. — **319**. — P. 155— 158.
4. *Fomundam S., Herrmann J. W.* A Survey of Queuing Theory Applications in Healthcare. Technical report submitted for publication in IIE Transactions. — 2007. — Available at http://hdl.handle.net/1903/7222.
5. *Eldabi T., Paul R.J., Young T.* Simulation modeling in healthcare: Reviewing legacies and investigating futures.//J. Opl Res. Soc. — 2007. — **58**. — P. 262—270.
6. *Abolnikov L. M., Yasnogorodskiy R. M.* A class of queueing problems with priorities when there are urgent orders. // Engineering Cybernetics. — 1974. — **12**, №4. — P. 62—72.
7. *Harper P. R., Shahani A. K.* Modelling for the planning and management of bed capacities in hospitals.// J. Oper. Res. Soc. — 2002. — **53** (1). — P. 11—18.
8. *Green L. V., Nguyen V.* Strategies for Cutting Hospital Beds: The Impact on Patient Service // Health Serv. Res. — 2001. — **36** (2). — P. 421 — 442.
9. *Preater, J.* (2002). 'Queues in health.' Health Care Manage Sci 5 (4): 283.
10. *Au-Yeung S. W., Harder U., McCoy E. J., Knottenbelt W. J.* Predicting patient arrivals to an accident and emergency department // Emergency Medical J. — 2009. — **26.**— P. 241— 244.

Submitted 03.06.09