World Academy of Science, Engineering and Technology
International Journal of Industrial and Manufacturing Engineering
Vol:11, No:5, 2017

Modelling a Hospital as a Queueing Network: Analysis for Improving Performance

Emad Alenany, M. Adel El-Baz

Abstract**—**In this paper, the flow of different classes of patients into a hospital is modelled and analyzed by using the queueing network analyzer (QNA) algorithm and discrete event simulation. Input data for QNA are the rate and variability parameters of the arrival and service times in addition to the number of servers in each facility. Patient flows mostly match real flow for a hospital in Egypt. Based on the analysis of the waiting times, two approaches are suggested for improving performance: Separating patients into service groups, and adopting different service policies for sequencing patients through hospital units. The separation of a specific group of patients, with higher performance target, to be served separately from the rest of patients requiring lower performance target, requires the same capacity while improves performance for the selected group of patients with higher target. Besides, it is shown that adopting the shortest processing time and shortest remaining processing time service policies among other tested policies would results in, respectively, 11.47% and 13.75% reduction in average waiting time relative to first come first served policy.

Keywords**—**Queueing network, discrete-event simulation, health applications, SPT.

I. INTRODUCTION

PERFORMANCE measurement of health care systems are vital for improving the service provided to patients. It is known that healthcare systems suffer from variability in demand and service times. This inherent variability in healthcare could adversely affect patient medical status and waiting times. Modelling the flow of different patients into a hospital under the conditions of uncertain demand and service environments can produce dramatic improvements in medical performance, patient satisfaction, and cost efficiency of the healthcare system [1].

There are three methodologies commonly used to analyze health care process: (1) Simulation modeling, (2) Flowcharting tools, and (3) Queueing models. Simulation models allow for detailed representation of complex, dynamic behavior; however, building a simulation model is very time-consuming and requires experienced modelers. Flowcharting tools represent the process structure, but are neither able to represent the dynamism of the process nor support quantitative analysis of process design alternatives. Queueing models can represent dynamic behavior, although not to the same level of detail provided by simulation models, but can be built in a fraction of the time that it takes to build a simulation model. Clearly, each modeling approach serves a purpose under different scenarios and stages of process analysis [2]. In order to improve performance in a complex environment such as a hospital system, its underlying dynamics at work need to be understood. To obtain such an understanding, queueing theory and simulation provide an ideal set of tools [1]. Networks of queues have proven to be useful models to analyze the performance of complex systems (see [3] and references therein). However, due to the unique features of healthcare systems, which makes queueing problems difficult to solve, the use of networks of queues in healthcare was relatively limited [1].

This paper is organized as follows. In Section II, the literature of using QNA algorithm in healthcare systems and simulation applications is reviewed. In Section III, the case study patient flow and data are presented and analyzed using the QNA algorithm. Discussion and comparison to discrete event simulation (DES) of the results is provided in Section IV, while Section V presents a set of intervention approaches and their results. Finally, conclusions and future research are suggested in Section VI.

II. LITERATURE REVIEW

A recent survey about queueing theory applications in healthcare can be found in [4]-[6]. A review and classifications of several references to healthcare examples that involve queueing networks are listed in [6]. Recent comprehensive treatment of the fundamental, methodological, and computational aspects of networks of queues is found in [7]. Few authors [8]-[10] used QNA to analyze and improve healthcare systems. In their research, [8] utilized QNA in outpatient clinics and demonstrated the advantage of using queueing networks in performing bottleneck analysis. While, in the paper, [10] constructed a queueing network of a general class of healthcare systems where hospital departments and their interdepartmental relationships are modeled as a network of GI/G/m queues. Through the analysis of this network, the impact of service interruptions and aggregating patient flows are explored and the optimal number of patients in a clinic session is determined. On the other hand, DESs, have been widely used to study complex systems [11]. One can see [12] for a comprehensive review of process simulation applications in healthcare areas, which summarizes projects applied in health care facilities like hospitals, emergency departments, intensive care units, surgical procedures, outpatient clinics, and facilities allocated in the health care supply chain. Within the extensive literature on using simulation in healthcare, studies that consider patients sequencing/prioritizing are the most relevant to this work. In systems where multiple

Emad M. Alenany and M. Adel El-Baz are with the Industrial Engineering Department at Zagazig University, Egypt (e-mail: emadenany@yahoo.com, elbaza@gmail.com).

customers are in the system at the same time (e.g., an emergency department with multiple patients in the exam rooms awaiting physician attention), the server (physician) faces a customer sequencing problem. For a related stream of works that consider customer sequencing that use queueing theory approach, one see [13]. For studies that consider using simulation to study the effect of sequencing heuristics, [14] used simulation to study effect of sequencing heuaristics for scheduling an Outpatient Procedure Center. Their analysis provides insight into the best scheduling heuristics and the tradeoff between patient and healthcare provider based criteria. References [15], [16] discussed using DES to create an efficient and equitable admission policy for patients arriving at a cardiothoracic intensive care unit.

III. QNA Model and Application to an Healthcare System

A. QNA Model

In this study, the flow of different patients into a hospital is modeled as a multiclass queueing network of $GI/G/m$ queues (general interarrival and service-time distribution with m identical servers in parallel), and analyzed using QNA algorithm [3]. In this model, the arrival process to each queue is assumed to be a generalized inter-arrival (GI) process. The service times may have any general distribution. The approximation made by this approach is that only the mean and the squared coefficient of variation ($SCV = variance / mean^2$) of the inter-arrival times and service times are required for our calculations - this is the reason why this kind of method is sometimes referred to as a two-moment method [17]. A summary of QNA algorithm is provided below [6].

λ_j: total arrival rate to dept. j for all patient classes, λ_{0j}: external arrival rate at department j, λ_{ij}: the flow rate from department i to j, c_{aj2}: squared coefficient of variation (SCV) of the arrival process at department j, m_j: number of doctors/beds at department j, τ_{oj}: expected service time at department j, n_k: number of nodes on route k, $c_{(s\_0\ j)2}$: SCV of the service time at department j, q_{ij}: the proportion of patients that go from i to j, ρ_j: resource utilization at each department j, EW_j: Expected waiting time at each department j, Here is a list of the input data for the k^{th} customer class of a network:

QNA Algorithm:

Step 1: Calculate the aggregate arrival rates at queue $j,\ \lambda_j$:

$$\lambda_j = \lambda_{0j} + \sum_{i=1}^{J} \lambda_j q_{ij} \tag{1}$$

Step 2: Calculate the load of a server at queue $j,\ \rho_j$:

$$\rho_j = \frac{\lambda_j \tau_j}{m_j}, \qquad 1 \le j \le n \tag{2}$$

Step 3: Calculate the flow from queue i to queue j, λ_{ij}

$$\lambda_{ij} = \lambda_i q_{ij}, \tag{3}$$

and the fraction of arrivals at queue j that come from queue i,

$$P_{ij} = \frac{\lambda_{ij}}{\lambda_j},\ i \ge 0 \tag{4}$$

Step 4: Calculate the SCV for the arrival process at queue j, c_{aj}^2:

$$c_{aj}^2 = a_j + \sum_{i=1}^{n} c_{ai}^2 b_{ij}, \qquad 1 \le j \le n \tag{5}$$

where a_j and b_{ij} are constants, depending on the input data:

$$a_j = 1 + w_j\{(P_{0j}c_{oj}^2 - 1) + \sum_{i=1}^{n} P_{ij}[(1-q_{ij}) + (1-v_{ij})q_{ij}\rho_i^2 x_i]\} \tag{6}$$

and

$$b_{ij} = w_j P_{ij} q_{ij}[v_{ij} + (1-v_{ij})(1-\rho_i^2)] \tag{7}$$

where c_{oj}^2 is the SCV of the external arrival process at queue j, and

$$x_i = 1 + m_i^{-0.5}(\max\{c_{si}^2, 0.2\} - 1) \tag{8}$$

$$w_j = \left[1 + 4(1-\rho_j)^2(v_j - 1)\right]^{-1} \tag{9}$$

and

$$v_j = \left[\sum_{i=0}^{n} P_{ij}^2\right]^{-1} \tag{10}$$

Step 5: The mean waiting time at queue

$$j, \mathrm{E}[W_j] = \mathrm{E}[W_{M/M/m}]\frac{c_{aj}^2 + c_{sj}^2}{2} = \left(\frac{c_{aj}^2 + c_{sj}^2}{2}\right)\left(\frac{\rho_j^{\sqrt{2(m_j+1)}-1}}{m_j(1-\rho_j)}\right)\tau_j \tag{11}$$

B. Outlines of Model Input to Hospital

The first step for the modeling of a multiclass queueing network is to identify the different classes of patients entering the hospital and their routes through different departments. A population of patients coming to the hospital under study during a month is recorded. This is divided into 32 classes of patients that go through 20 different departments in the hospital. Table I lists hospital departments and the corresponding number of available servers in each of them. (note that servers may be doctors or beds or other resources depends on the department.) Table II provides a list of patients' classes along with their specific routes through hospital departments. Fig. 1 provides an overview of the aggregate flow of patients' classes between departments through the hospital.

III. QNA and DES Model Results

A. Model Results

A C# program was created by using Visual Studio 2010 to facilitate the computations of the QNA steps. Table III provides results of traffic intensity (utilization) and expected waiting time at $\mathrm{E}[W_j]$ at each department in the hospital, $j = 1,2,\ldots,20$.

It is shown in Table III that highest utilization departments are at department 9 (ICU), 10 (Pre/post-operative care), and 14 (Intermediate Burn Care Unit), their results are shown in bold in Table III. These units represent bottlenecks for the hospital system, and so where an intervention is needed. Improving flow at these areas promising to have a positive effect on later departments of the flow in the network.

B. Comparison of QNA Results to Simulation

This section compares QNA results for hospital performance measures with simulation. A simulation model for the hospital network is created by using C# programming language. Reference [18] provides a reference to build a DES model. The built DES model is created by using C# programming language. Simulation experiment terminates after 10^8 minutes with a transient interval for 10000 minutes. The overall simulation interval is chosen to be large enough so that the effects of the transients will be negligible and may be ignored [17]. The simulation runs until it reaches the termination time, and after that, no arrival is allowed to enter the system. Simulation continues to process the system until the system becomes empty.

Table III presents in the right two columns analytical results from QNA, expected waiting times from simulation model, and relative percent errors (RE) as a measure of error in approximate results of QNA. Percent relative errors are defined as

$$RE = 100\ (\text{Analytical.} - \text{Simul.})/\text{Simul.}$$

Based on the results in Table III, relative percent errors for large number of expected waiting times from QNA are within 0.15% percent error. Besides, relative percent errors are higher for QNA results compared to simulation results in a number of nodes. This could be due to the existence of almost immediate feedback. An example of this effect is a flow that returns relatively quickly after passing through one or more nodes; e.g., patients return to department 14 after passing through department 16, and also patients return to department 10 after passing through departments 9 and 10. Elimination of immediate (flows that return directly to the same node) or almost immediate feedback usually yields a better approximation (see Sections V and VII in [19]). Reference [19] suggests a refined procedure to eliminate those effects. Due to this difficulty arising from combining QNA with the network under study, and to test different service policies other than first-come first-served (FCFS), it is better to use simulation approach.

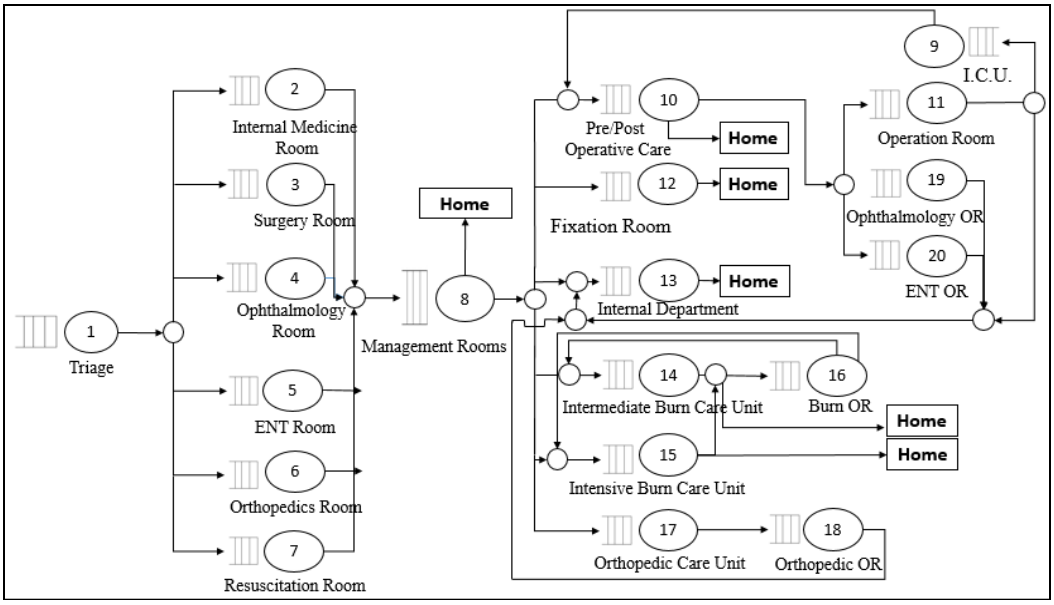

Fig. 1 Diagram of Hospital operations

TABLE I
DIFFERENT DEPARTMENTS IN THE HOSPITAL UNDER STUDY

N	Station	m	N	Station	m
1	Triage	2	11	Operation Room	4
2	Internal Medicine Room	1	12	Fixation Room	2
3	Surgery Room	1	13	Internal Department	120
4	Ophthalmology Room	1	14	Intermediate Burn Care Unit	12
5	Ear/Nose/Throat (ENT) Room	1	15	Intensive Burn Care Unit	2
6	Orthopedics Room	1	16	Burn OR	1
7	Resuscitation Room	2	17	Orthopedic Care Unit	4
8	Management Rooms	9	18	Orthopedic OR	2
9	I.C.U.	12	19	Ophthalmology OR	2
10	Pre/Post-Operative Care	18	20	ENT OR	1

TABLE II
PATIENTS' CLASSES ARRIVING AT THE HOSPITAL AND THEIR ROUTES

Pt. No.	Pt. Code	Patient Class	Patient Route
1	A1	Chest Pain Case I	$\mathbf{1 \to 2 \to 8}$
2	A2	Chest Pain Case II	$\mathbf{1 \to 2 \to 8 \to 13}$
3	B1	Abdominal Pain Case I	$\mathbf{1 \to 2 \to 8}$
4	B2	Abdominal Pain Case II	$\mathbf{1 \to 2 \to 8 \to 13}$
5	C	Hypertensive Case	$\mathbf{1 \to 2 \to 8}$
6	D1	Coma Case I	$\mathbf{1 \to 2 \to 8}$
7	D2	Coma Case II	$\mathbf{1 \to 2 \to 8 \to 13}$
8	E	Convulsion Case	$\mathbf{1 \to 2 \to 8}$
9	F1	Trauma Case I	$\mathbf{1 \to 7 \to 8}$
10	F2	Trauma Case II	$\mathbf{1 \to 7 \to 8 \to 10 \to 11 \to 9 \to 10}$
11	G1	First Degree Burn	$\mathbf{1 \to 7 \to 8}$
12	G2	Second Degree Burn Case I	$\mathbf{1 \to 7 \to 8 \to 14}$
13	G3	Second Degree Burn Case II	$\mathbf{1 \to 7 \to 8 \to 14 \to 16 \to 14}$
14	G4	Third Degree Burn Case I	$\mathbf{1 \to 7 \to 8 \to 15}$
15	G5	Third Degree Burn Case II	$\mathbf{1 \to 7 \to 8 \to 15 \to 16 \to 15}$
16	H	Drowning Case	$\mathbf{1 \to 7 \to 8}$
17	I	Simple Wound Case	$\mathbf{1 \to 3 \to 8}$
18	J	Complicated Wound Case	$\mathbf{1 \to 3 \to 8 \to 10 \to 11 \to 9 \to 10}$
19	K1	Urine Retention Case I	$\mathbf{1 \to 3 \to 8}$
20	K2	Urine Retention Case II	$\mathbf{1 \to 3 \to 8 \to 10 \to 11 \to 13}$
21	L	Acute Appendicitis Case	$\mathbf{1 \to 3 \to 8 \to 10 \to 11 \to 9 \to 10}$
22	M	Acute Cholecystitis Case	$\mathbf{1 \to 3 \to 8 \to 10 \to 11 \to 13}$
23	N	Acute Intestinal Obstruction I.O. Case	$\mathbf{1 \to 3 \to 8 \to 10 \to 11 \to 13}$
24	O1	Renal Colics Case I	$\mathbf{1 \to 3 \to 8}$
25	O2	Renal Colics Case II	$\mathbf{1 \to 3 \to 8 \to 10 \to 11 \to 13}$
26	P1	Simple Fracture Case I	$\mathbf{1 \to 6 \to 8}$
27	P2	Simple Fracture Case II	$\mathbf{1 \to 6 \to 8 \to 12}$
28	Q	Complicated Fracture Case	$\mathbf{1 \to 6 \to 8 \to 17 \to 18 \to 13}$
29	R	F.B. in Eye Case	$\mathbf{1 \to 4 \to 8}$
30	S	Eye Trauma Case	$\mathbf{1 \to 4 \to 8 \to 10 \to 19 \to 13}$
31	T	E.N.T. Trauma Case	$\mathbf{1 \to 5 \to 8 \to 10 \to 20 \to 13}$
32	U	E.N.T. F.B. Case	$\mathbf{1 \to 5 \to 8}$

TABLE III
SUMMARY TABLE OF THE MODEL RESULTS*

i	Arrival rate $(Pt./day)$	Avg. Service time (min)	Utilization		Expected Waiting Time (min)		RE
			QNA	DES	QNA	DES	
1	211.27	10.00	0.734	0.732	11.98	11.59	+0.03
2	60.25	16.31	0.682	0.681	37.38	37.11	+0.01
3	54.33	16.38	0.618	0.618	33.26	33.35	+0.00
4	30.00	40.00	0.833	0.832	200.00	198.79	+0.01
5	3.00	90.00	0.188	0.187	20.77	20.64	+0.01
6	30.00	30.00	0.625	0.623	50.00	49.86	+0.00
7	33.69	45.11	0.528	0.526	19.18	17.54	+0.09
8	211.27	60.00	0.978	0.977	291.62	281.22	+0.04
9	4.00	2.75 (d)	0.917	0.917	**2.87 (d)**	**4.14 (d)**	**-0.45**
10	23.33	0.75 (d)	0.971	0.971	**6.74 (d)**	**5.09 (d)**	**+0.24**
11	12.33	360.00	0.771	0.770	0.23 (d)	0.69 (d)	-1.96
12	19.00	60.00	0.396	0.395	12.96	11.10	+0.14
13	26.33	3.68	0.808	0.807	12.11	16.65	-0.37
14	2.61	4.26 (d)	0.873	0.934	**5.73 (d)**	**4.05 (d)**	**+0.29**
15	0.21	3.33 (d)	0.350	0.362	0.72 (d)	0.51 (d)	+0.29
16	1.40	248.57	0.242	0.241	84.48	108.40	-0.28
17	5.00	480.00	0.417	0.415	30.99	20.49	+0.34
18	5.00	300.00	0.521	0.519	121.61	110.42	+0.09
19	6.00	360.00	0.750	0.750	0.41 (d)	0.84 (d)	-1.03
20	1.00	480.00	0.333	0.333	249.97	330.08	-0.32

* times in min, except where followed by (d) in days.

V. Improvement Interventions

A. Partitioning Patients into Service Groups

TABLE IV
PERFORMANCE TARGETS

Type	Arrival rate (patient/day)	% of patients	Performance target-excess (tail) probability
1	31.73	15 %	80 % < 1 hour
2	179.54	85 %	80 % < 4 hours

The general perception appears to exist that it would be better to merge two (or multiple) queues into a single line, in order to use capacities more efficiently. Reference [20] lists two reasons (situations) where it may be more beneficial to separate patients to be served in two or more queues than to pool them in a single line system: (1) Different service characteristics; and (2) Different service targets (workload ratio).

For the first situation, pooling servers introduces variability due to the mix ratio of different means. This may essentially cause a negative effect according to the Pollaczek-Khintchine's formula. An extensive analytic and numerical treatment of this counter-intuitive phenomenon can be found in [20] and more recently in [21]. Reference [20] focused exclusively on the second situation involving different service targets. They provided a theoretical and practical illustration to answer the question of whether and when it is beneficial to combine different patients with different performance targets.

In the current case study, performance target, e.g. average waiting time, is not the same for all patients. For example, patients going through department 8 (management rooms), which provides services for scanning and laboratory measurements, could be classified into two types. This classification of patients is based on the number of departments that a specific patient will go through when leaves department 8. From this definition, all patients who go through department 8 are categorized into the following two types:

Type 1. A small percentage of patients which after passing through department 8, they still have to visit another two or more departments in the hospital in their routes. This group of patients is expected to need higher performance target, in terms of waiting times.

Type 2. The remaining number of patients (represents majority of patients) who will have to go through one department at most after leaving department 8. This group of patients is assumed to be able to accept a lower performance target than the first group because they assumed to have less severity in their health status.

In practice, particularly in health care, excess or tail probabilities are often used as a performance measure instead of average waiting times. Table IV provides data for the two types of patients, and target performance measure. It is questionable to test which of the two situations, pooling or separation, requires less capacity to meet the required performance targets. Type 1 patients are given a tight performance target, while type 2 customers are required to meet a substantially lower target. The two situations are illustrated in Fig. 2.

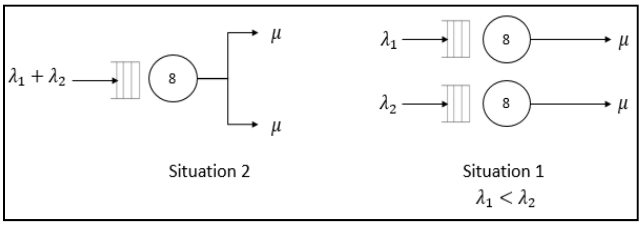

Fig. 2 Representation of situations

The results of simulation of the two situations are presented in Table V. In the pooled situation, 11 servers are necessary to meet the high-performance target for type 1 patients (80 % of the patients need to have their tests within three hours). The situation of two separate queues also required three plus eight (11 servers) in order to meet the performance targets from Table IV. These results show that the pooled situation results coincide with the separation situation in this case. Furthermore, in the pooled situation on average 86 % of the patients will wait less than one hour. In the situation with separate queues, 87% of type 1 patients will wait less than 1 hour and 88.9% of the second type patients will wait less than four hours. As capacities are unchanged in the two situations, separate queues will lead to higher percentages which meet their critical value of the performance targets. Therefore, keeping the capacities separate can be considered superior in this case.

TABLE V
PERFORMANCE MEASURES FOR POOLED AND SEPARATION CASE FOR DEPARTMENT 8

Situation	Number of servers	Utilization (ρ)	Tail probability
Pooled	11	$\rho = 79\ \%$	$P(W_p < 1\ hour) = 0.859$
Separation	3 (type 1)+ 8 (type 2)	$\rho_1 = 43.6\ \%$ $\rho_2 = 91.4\ \%$	$P(W_1 < 1\ hour) = 0.87$ $P(W_2 < 4\ hours) = 0.889$

B. Adopting Other Service Policies

Considering different service policies other than FCFS rule for patients to be seen by physician or scheduled for an operation for example, can offer benefits regarding efficient use of these care resources in hospital. Patient sequencing is an important activity in any service system [22].

In this paper, five different sequencing rules are compared and they are: FCFS, shortest processing time (SPT), Largest processing time (LPT), Shortest Remaining processing time (SRPT), Largest Remaining Processing Time (LRPT). In the following, each of the previous policies is defined.

- FCFS policy: in which a physician selects his/her next patient in order of their arrival to the specific department.
- SPT policy: to give priority to waiting patients with shortest service time.
- SRPT policy: to give priority to waiting patients based on their expected remaining service times including current

and following departments.

- LPT policy: opposite of SPT policy.
- LRPT policy: the opposite of SRPT.

Simulation experiment of the DES model for the hospital will be run with five times. In each run, one of service disciplines defined above will be considered as the service discipline in every department in the hospital.

In each of the simulation runs performed, average waiting times are computed. Average waiting time is used as a proxy for efficiency of use of care resources. An obvious advantage of using simulation models over queueing models is that many performance measures could be obtained with minimum effort.

Comparison of Service Policies Performance

Fig. 3 presents average waiting and maximum time for the five different sequencing policies for the three bottleneck departments 9, 10, 14. It can be seen that, for the average waiting measure, SPT and SRPT perform better than FCFS policy with 11.47% and 13.75%, respectively.

It is shown that for the average waiting measure, SPT and SRPT perform better than FCFS policy with 11.47% and 13.75%, respectively.

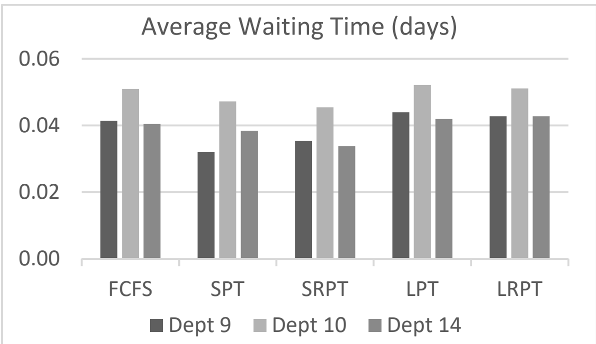

Fig. 3 Average waiting time for the five different service policies (days)

VI. Conclusion and Future Research

This paper models a hospital system for queueing network using QNA algorithm and DES. Use of QNA algorithm helps to rapidly calculate approximate congestion measures for the network of queues in a hospital system and to identify bottlenecks. A number of improvement techniques are then investigated to improve performance at bottlenecks and the healthcare system in general. Suggested improvement techniques used in this paper represent general solutions and could be applied to partial areas of the hospital. These solutions inherited from operations management show how medical and hospital systems could make use of operations management and operations research tools used in manufacturing systems to improve system performance in healthcare systems. A significant insight of this paper is that applying simple heuristics for sequencing patients into different departments can improve performance measures for a medical system.

A future research effort could be devoted to make modifications to QNA algorithms to make it able to apply to a wider range of environments with less assumptions. Modifications include tuning QNA to allow multiple classes with random routes rather than deterministic routes. Also, there is a need to extend QNA to make it workable for networks with finite capacity and different service disciplines other than FCFS, e.g. SPT or other assigned priorities. Also, an important point is to consider other sequencing rules that could incorporate the delay cost per unit time or severity of patient case. It is also beneficial to study effect of eliminating almost immediate feedback with QNA and to develop an automatic procedure for eliminating immediate feedback in order to incorporate in QNA.

Acknowledgment

Thanks are presented to Ivan Kuckir, Faculty of Mathematics and Physics, Charles University in Prague. Matrix class in C# written by him was used in solving the system of linear equations in QNA.

References

[1] Creemers S., Lambrecht M.R., Modeling a healthcare system as a queueing network: the case of a Belgian hospital. In: FBE publications: Research Reports and Discussion papers, 2007.
[2] Jiang L., Giachetti R.E., A queueing network model to analyze the impact of parallelization of care on patient cycle time. Health Care Management Science 11(3):248- 261, 2008.
[3] Whitt W, The queuing network analyzer. Bell Syst Tech J 62:2779-2815. 1983.
[4] Bhattacharjee P., Ray, P.K. Patient flow modelling and performance analysis of healthcare delivery processes in hospitals: A review and reflections. Computers & Industrial Engineering, 2014.
[5] Lakshmi C., Sivakumar Appa Iyer, Application of queueing theory in health care: A literature review, Oper. Res. for Health Care, 2013
[6] Zondeland M. E. "Curing the Queue." Ph.D., University of Twente, Enschede, the Netherlands, 2012.
[7] Boucherie R.J., van Dijk NM (eds) Queueing networks: a fundamental approach. Springer, New York, NY, USA, 2011.
[8] Albin L. S., Jerrey Barrett, David Ito, John E. Mueller, A queueing network analysis of a health center. Queueing Systems, 1990, Volume 7, Issue 1, pp 51-61.
[9] Zondeland M. E., Boer F., Boucherie R.J., De Roode A., Kleef J. W. van. Redesign of a University Hospital Preanesthesia Evaluation Clinic Using a Queuing Theory Approach. Anesthesia & Analgesia: November 2009 - Volume 109 - Issue 5- pp 1612-1621.
[10] Creemers S, Lambrecht M, Modeling a hospital queueing network. In: Boucherie RJ, van Dijk NM (eds) Queueing networks: a fundamental approach. Springer, New York, NY, USA, 2011.
[11] Law AM, Kelton WD: Simulation Modelling and Analysis. Third Edition. New York, NY,The McGrow-Hill, 1999.
[12] Thorwarth, Michael and Arisha, Amr, "Application of Discrete-Event Simulation in Health Care: a Review", Reports. Paper 3. http://arrow.dit.ie/buschmanrep/3, 2009. Accessed at: 13/3/2015.
[13] Saghafian S., Hopp W. J, Desmond S. Jeffry, Van Oyen, Kronick, L. Steven, Patient Streaming as a Mechanism for Improving Responsiveness in Emergency Departments, Operations Research, 1080-1097, 2012.
[14] Gul S., Denton B., Fowler J., Huschka T., BI-Criteria Scheduling of Surgical Services for An Outpatients Procedure Center, Production and Operations Management, 2011.
[15] Yang, M., Fry, M. J., & Surlock, C. The ICU will see you now:. IIE Transactions, 2015.
[16] Yang, M., Fry, M. J., Raikhlekar J., Chin C., Anyanwu A., Brand J., Scurlock C., Efficient-equitable admission control policies for a surgical ICU with batch arrivals, ccmjournal, volume 41, number 2, Februrary 2013.

[17] Bose K. S., An introduction to queueing systems, Springer Science+Business media New York, 2002.
[18] Lawerance Leemis, Steve Park, Discrete-Event Simulation: A First Course; ISBN: 0-13-142917-5 .
[19] Whitt W., Performance of the queuing network analyzer. Bell Syst Tech J 62:2817-2843. 1983.
[20] Whitt W., Partitioning customers into service groups. Management Science, 45, 579–1592.
[21] P. Joustra, E. van der Sluis, N.M. van Dijk, To pool or not to pool in hospitals: a theoretical and practical comparison for a radiotherapy outpatient department, Ann of Oper Res, 2010.
[22] Hopp W. J., Lovejoy W. S., Hospital Operations: Principles of High Efficiency Health Care, FT Press Operations Management, November 9, 2012.

Emad M. Alenenay is a research assistant in the Department of Industrial Engineering at Zagazig University. His research interests are in applications of industrial engineering to healthcare systems.

M. Adel Elbaz has more than 25-year experience in Industrial and Manufacturing systems engineering and the author of many scientific publications in Genetic Algorithms, Neural Networks, Fuzzy logic applied in the field of Scheduling, Facilities layout, Quality control, and Supply chain management. Following his Ph.D. in industrial engineering and production, he took the academic positions trend and he is now the chairman of the Industrial Engineering Department, Faculty of Engineering, Zagazig University, Egypt. He also acted as a consultant in engineering management at many of the companies in industrial sector in Egypt. Dr. M. Adel Elbaz worked industrial planning consultant at the ministry of economy and planning, Kingdom of Saudi Arabia during the year 2012-2013.