NETWORKS OF QUEUES WITH MULTIPLE CUSTOMER TYPES: APPLICATION IN EMERGENCY DEPARTMENT

Jihène. JLASSI [a], Abederrahman. EL MHAMEDI [b] and Habib. CHABCHOUB[c]

[a], [c] ISGI de Sfax Route Mharza Km 1.5-B.P954- 3018 Sfax., Tunisia
[b] MGSI (Modélisation et Génie des Systèmes Industriels) IUT de Montreuil, 140 Rue de la Nouvelle France- 93100 MONTREUIL.

[a] jihene2000tn@yahoo.fr [b] a.elmhamedi@iut.univ-paris8.fr

ABSTRACT
Emergency department becomes a useful way to the access to hospital and it is a subject of study for many researchers. The research developed in this paper aims to improve the performance of the emergency department(ED) of Sfax Hospital by analytical method. So, a network of queues with multiple customer types is proposed. Different indicators of performance are used.

Keywords: Performance, Analytical method, networks of queues, multi class, priority patient class

1. INTRODUCTION

In general, evaluation techniques, that are methods by which performance evaluation indices are obtained, can be subdivided into two main categories: measurement (or empirical) techniques and modelling techniques. Empirical techniques require that the system or network to be evaluated exists and direct measurements of the evaluation target have to be taken. On the other hand, modelling techniques only require a model of the system. Modelling techniques are of two types: simulation and analytic.

The ED of hospital is a complex unit where the fight between life and death is always a hair's breath away, requiring a high degree of coordination and interrelations between human and material elements. (Jinn-Yi and Wen-Shan, 2007).

In the precedent paper (Jlassi et al, 2007), we proposed a simulation model which enables us to definite indicators to evaluate the performance of the ED of Sfax hospital. The study consisted of drawing a passage from a graphic model IDEF3x to a WITNESS model. Which reflect more the reality in a clearer way. Several numbers of indicators was defined. Different variation of the WITNESS model was proposed, showing the impact of variation in several parameters on the process performance.

This paper surveys the contributions and applications of queuing theory in the Emergency Department of Sfax hospital. It is organized as follows. We start with describing queuing theory and queuing network. In section3, we present literature review, in section 4, we describe the Emergency Department In section 5 we illustrate the proposed approach. Finally the last section presents concluding remarks and perspectives for future research.

2. QUEUING THEORY

Queuing theory has been a prominent analytical technique in operations research for more than half a century. A queuing model can save valuable time by providing analytical short cuts thereby improving the timeliness of management interventions and be helpful in the process of external scrutiny (e.g. via the Health Care Commission). In practice it may only be possible to do some of these things because of data and other limitations. Queuing theory provides an analytical approach to showing the variability in processes.

The theory permits the derivation and calculation of several performance measures including the average waiting time in the queue or the system, the expected number waiting or receiving service and the probability of encountering the system in certain states, such as empty, full, having an available server or having to wait a certain time to be served.

In many applications it is desirable to give certain classes of customers' preferential treatment; the queue is ordered and the higher priority customers are served first.

In a priority queueing system customers are divided into K ¸ 2 classes numbered 1; 2,…, K:

The lower the priority, the higher the class number. In other words, priority i customers are given preference over priority j customers if $i < j$.

The model with multiple stations is called a queueing network model. A variety of queueing network frameworks have been developed to represent various system mechanisms.

3. LITERATURE REVIEW

While there are many published healthcare research studies that apply queuing models, those that incorporate an Emergency Department concept are rare.

The study of Linda V. et al (2005) illustrates how data analysis and queuing models can be used to identify staffing changes that can decrease the delays in being seen by a provider and, thus, the fraction of patients who leave without being seen, without necessarily increasing capacity. It also highlights the need to establish patient delay standards, preferably by triage class, and to establish information systems to collect and track data on provider service times and patient delays in seeing a provider.

William et al (2005) explore conceptual development and practical application of the spreadsheet model. Particular features include: constructing lookup tables by hour of day containing estimates of minimum and maximum rates, using the rand between function to randomly select model inputs from a uniform distribution, developing frequency distributions to assist in output interpretation, illustrating conditional formatting, output graphing, etc. One can observe multiple samples of hourly patient fluctuations based on unit open beds and midnight census. Number of patients waiting can be shown at varying levels of system utilization. As utilization approaches approximately eighty percent, patient waiting time increases disproportionately. The spreadsheet model is a dynamic, visual illustration of how variation in individual process times can affect total process capability. Its use is primarily intended as a teaching tool for those new to simulation modeling.

In Siddharthan et al article (1996), Investigates the increased waiting time costs imposed on society due to inappropriate use of the emergency department by patients seeking non-emergency or primary care. Proposes a simple economic model to illustrate the effect of this misuse at a public or not-for-profit hospital. Provides evidence that non-emergency patients contribute to lengthy delays in the ER for all classes of patients. Proposes a priority queuing model to reduce average waiting times.

Marianov et al (1995), presented, in their paper, a model which seeks to maximize the population covered by emergency vehicles with availability &. A probabilistic formulation is structured in which availability is computed utilizing queueing theory.

Au-Yeung et al (2000) report's developed a multiclass Markovian queuing network model of patient flow in the Accident and Emergency Department of a major London hospital.

4. CONTEXT OF THE STUDY

Habib Bourguiba hospital is presented as follows:

- Academic Hospital centre since 1985 ;
- Erected in public establishment of health since 1993.
- The missions of the hospital are:
- To lavish current pathology cares and essentially cares of reference;
- To assure the convenient formation of basis and retrain the medical and decorate-medical personnel
- To develop the activity of research in the medical domain and cares male nurses.

The hospital includes 18 Departments; the most important one is the Emergency Department that represents the entrance door of the hospital.

The geographical situation of Sfax City in the centre of the country, point of link between the south and the north of the country makes that the Emergency Department receives an important number of casualties of the public way. As second economic and industrial pole sheltering a lot of companies, the city knows a rather elevated number of work casualties.

The department includes:

- 2 rooms of care;
- 1 room of plaster;
- 2 post offices;
- 1 room of general surgery ;
- 1 room of orthopaedic;
- 1 room for medical visits;
- the overage number of patients per day is $\pm$ 300,
- the number of personnel is 13;

The process by wish a patient passes is: For the arrival of the patient, we distinguish 2 possibilities: either he will come by himself or, in dangerous cases, he will be transported by the ambulance. At the arrival, the patient will pass by a triage process that will define the Emergency degree and then the process that will be taken by the patient. In fact, there are two processes that can be taken by the patient, given the result of the triage process.

In both cases, the remaining process will be almost the same. In fact, the main difference concerns the administrative process. Thus, for a dangerous case, this process will be reported at the end of the patient process while for a non dangerous case, this process will be at the beginning of the whole process.

The remaining tasks of the process are the same as those of a normal consultation process. In fact, the patient will see a doctor who will determine the patient's state and if necessary ask for supplementary analysis. Given the analysis results, the doctor will either care the patient by him self or ask for a specialist. In both cases, the patient could be hospitalized or go home.

5. THE NETWORK MODEL

The queueing models considered here consist of J single-server stations and K customer types (Two patients types) , with each customer type having a fixed route through the network of stations. Customers of types k = 1, 2, . , K arrives to the network according to independent renewal processes. Each type k customer visits the sequence of stations

$$r_k(1), r_k(2), ..., r_k(s(k))$$

After which he leaves the system. This is an open network in the sense that all customers originate from external sources and all eventually leave the system. We require a feed forward routing pattern, meaning that the stations can be numbered j = 1, 2,. , J in such a way that for each k the route of a type k customer forms an increasing sequence. Intuitively this means that work flows in a directed fashion through the network. In particular, no customer may visit any work center more than once. The s (k) service times along the route of a given type k customer are allowed to have an arbitrary joint distribution, but are independent of the arrival processes and of the service times of any other customer. For each station j we define

$$C_j = \{K : r_k(S) = j \quad 1 \le S \le S(k)\}$$

We call the C, the "constituency" of station j; it is the set of customer types that visit j at some stage of their routes. The queue discipline at station j is defined by a partition of C, into subsets Gj and Hj, which represent equivalence classes with respect to priority level. Customers of type's $K \in H_j$ (high priority) have pre-emptive resume priority over customers of types $K \in G_j$, at station j. Within a priority class, customers are served in first-come-first-served (FCFS) order regardless of type. At a station j where there is only one priority level (i.e., the discipline is simply FCFS), we set $G_j = C_j$, and $H_j = \phi$.There is no bound on the number of customers that can be queued at any station.

We take as primitive a probability space (Ω, F, P) on which the following sequences of nonnegative random variables and vectors are defined. For k = 1, .. . , K, let $\{u_k^l \quad l \ge 1\}$ be mutually independent i.i.d. sequences with $E[u_k^l] = 1$. Independent = of these, we have K mutually independent i.i.d. sequences $\{\tilde{v}_k^l, \quad l \ge 1\}$ of random vectors

$$\tilde{v}_k^l = (v_{rk(1),k}^l, ..., v_{rk(s(k)),k}^l)$$

With $E[\tilde{v}_k^l] = (1,...,1)$ we define

$$\begin{aligned} &\alpha_k^2 = \mathrm{var}(u_k^l) \ge 0, \quad k = 1,...,K \\ &\beta_{jk}^2 = \mathrm{var}(v_{jk}^l) \ge 0, \quad j = 1,...,J \quad k \in C_j \\ &\beta_{ijk} = \mathrm{cov}(v_{ik}^l, v_{jk}^l) \quad j = 1,...,J \quad k \in C_i \cap C_j \end{aligned} \qquad (1)$$

The $\{u_k l\}$ represent normalized (i.e., unit mean) interarrival times, and the $\{v_{jk}^l\}$ represent normalized service times. The actual interarrival and service times for the queueing model are constructed from these sequences by specifying constants $\lambda_k \ge 0$ for k = 1 , .. . , K, and $m_{jk} \ge 0$ for j = 1,. . . , J and $k \in C_j$.We then let $(\lambda_k)^{-1} u_k^l$; be the interarrival time for the lth type k customer, and let $m_{jk} v_{jk}^l$ be this customer's service time at station j on his route. In this scheme, the λ_k , are the average arrival rates, and the m_{jk} are the mean service times. Observe that α_k^2 is the squared coefficient of variation (the variance divided by the square of the mean) of the interarrival time distribution for type k customers, and β_{jk}^2 is the squared coefficient of variation of the type k service times at station j. The parameters β_{ijk},, which for given (i, j, k) represent the covariance of the type k service times at station i and station j divided by the product of their means, are the analogous normalizations of the covariance in the service time vectors.

Let the renewal counting processes $N_k = \{N_k(t), \quad t \ge 0\}$ be defined for k = 1 , .. . ,K by

$$N_t(t) = \begin{cases} \max\{l \geq 0; u_k^l + \ldots + u_k^l \leq t\}, u_k^l \leq t \\ 0 \qquad\qquad u_k^l \geq t \end{cases}$$

(2)

The arrival counting process A, for type k customers is now given by $A_k(t) = N_k(\lambda_k t), \quad t \geq 0$ (we assume that the network is empty at time t = 0). Next, for j = 1 , ... ,J and $k \in C_j$, define the compound processes L_{jk} by

$$L_{jk}(t) = \sum_{l=1}^{N_k(t)} v_{jk}^l, \quad t \geq 0 \tag{3}$$

Then $m_{jk}\, l_{jk}(\lambda_k t)$ represents the amount of type k work for server j that arrives to
the network during [0, t]. We define the "netput" process Xj at station j by

$$X_j(t) = (\sum_{k \in C_j} m_{jk}\, l_{jk}(\lambda_k t)) - t \quad t \geq 0 \tag{4}$$

$$\hat{X}^n(t) = \varepsilon_n X^n(t/\varepsilon_n^2), \quad \hat{W}^n(t) = \varepsilon_n W^n(t/\varepsilon_n^2),$$
$$\hat{I}^n(t) = \varepsilon_n I^n(t/\varepsilon_n^2), \quad \hat{Z}^n(t) = \varepsilon_n Z^n(t/\varepsilon_n^2),$$
$$\hat{U}^n(t) = \varepsilon_n U^n(t/\varepsilon_n^2), \quad \hat{Q}^n(t) = \varepsilon_n Q^n(t/\varepsilon_n^2), \qquad 0 \leqslant t \leqslant 1.$$
$$\hat{S}^n(t) = \varepsilon_n S^n(t/\varepsilon_n^2),$$

Here $X_j(t)$ represents the total workload input for server j over [0, t] minus the work which would be completed if the server were never idle. The vector netput process X = (XI,...,X,) is a fundamental building block for our theorem. For j = 1,. ..,J let $I_j(t)$ denote the cumulative time during [0, t] that server j is idle. Then the process defined by

$$W_j(t) = X_j(t) + I_j(t), \quad t \geq 0 \tag{5}$$

Represents the total amount of work for server j present anywhere in the network at time t. However, not all of the customers whose service times are accounted for in Wj.(t) will reach station j by time t; some will be in queue or in service at stations preceding j on their routes. Let $Z_j(t)$ denote the unfinished workload which is actually present at station j at time t. By our last comment, we have

$$Z_j(t) \leqslant W_j(t), \qquad t \geqslant 0.$$

The central result of this paper is a heavy traffic limit theorem for the vector process Z(t) = (Z,(t), .. .,Z,(t)). It turns out that the J-dimensional limit process for this vector of station workloads contains all the information necessary to describe the heavy traffic behavior of the system at the customer level. More precisely, let $Q_{jk}(t)$ denote the number of type k customers in queue or in service at station j at time t, and let $U_{jk}(t)$ denote unfinished work represented by these customers (of course, $\sum_k U_{jk}(t)$ is simply $Z_j(t)$). Also, let SK(t) denote the sojourn time in the network for the first type k customer arriving at time greater than or equal to t. Then heavy traffic limit results for the processes

$$Q(t) = (Q_{11}(t), \ldots, Q_{1K}(t), \ldots, Q_{J1}(t), \ldots, Q_{JK}(t)),$$
$$U(t) = (U_{11}(t), \ldots, U_{1K}(t), \ldots, U_{J1}(t), \ldots, U_{JK}(t)), \qquad t \geqslant 0,$$
$$S(t) = (S_1(t), \ldots, S_K(t)),$$

can be obtained jointly with the result for Z, with the limit processes for Q, U and S expressed as linear transformations of the limit process for Z.
The limit results are obtained under conditions of heavy traffic, which we now describe. Define

The parameters p_{jk} and p_j by:

$$p_{jk} = \lambda_k m_{jk} \quad j = 1, \ldots, J \quad k \in C_j$$
$$p_j = \sum_{k \in C_j} p_{jk} \quad j = 1, \ldots, J \tag{6}$$

The results gotten while applying the formulas of William Peterson (1991), present that the specialist physician and the generalist physician in the second passage of the patients are the most occupied step by which pass a patient (see Appendix A). And the patients in these two steps of process wait a lot to be served. What confirms the results gotten by the simulation model by Witness.

It is why we propose to add a physician (specialist or generalist who has experience) to be able to minimize the waiting time and therefore to increase the number of patients treaties

6. CONCLUSION

This paper studied the evaluation of the emergency department of Sfax hospital by using the multi class queuing networks technique.

First, we presented the queuing network. Then, we provided the primordial role of Emergency Department (ED) and we identified numerous studies which have been made in order to evaluate the out put of the ED. Next we presented the ED of Sfax Hospital. Finally, we presented the main theorem.

The results showed the necessity to add in case of need of a physician. To choose between what physicians to add, we proposed to apply a method of multiple criteria decision aid as a PROMETHEE II method.

APPENDIX

In the Emergency Department there are two patients Types

- type 1 : no urgent
- Classe 2 : urgent case (high priority)

In the networks there are 6 steps:
T1: administrative procedure;
T2: generalist physician (first)
T3 : Analyze
T4 : Radiology ;
T5 : generalist physician (second)
T6 : specialist physician.

$$p_{jk} = \lambda_k m_{jk} \quad j = 1,\dots,J \quad k \in C_j$$

$$p_j = \sum_{k \in c_j} p_{jk} \quad j = 1,\dots,J$$

$$p_{11} = \lambda_1 m_{11} = 55{,}59$$

$$p_{12} = \lambda_2 m_{12} = 0$$

$$p_{21} = \lambda_1 m_{21} = 107.008$$

$$p_{22} = 71.25$$

$$p_{31} = 146.3$$

$$p_{32} = 95$$

$$p_{41} = 96$$

$$p_{42} = 85.5$$

$$p_{51} = 133$$

$$p_{52} = 95$$

$$p_{61} = 188$$

$$p_{62} = 118$$

$$R = I \quad \Gamma + \Gamma^2 + \ldots + (\quad 1)^{J \quad 1} \Gamma^{J \quad 1}$$

Γ is strictly upper triangular guarantees that $(I + \Gamma) \equiv R$

$$\gamma_{ij} = \sum_{k \in G_i \cap C_j} p_{jk} \Big/ \sum_{k \in G_i} p_{ik} \quad i \le j$$

$$\Gamma = \begin{matrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1.9 & 1 & 0 & 0 & 0 & 0 \\ 1231 & 6.32 & 1 & 0 & 0 & 0 \\ 17.8 & .91 & 0.14 & 1 & 0 & 0 \\ 11 & 5.66 & 0.89 & 6.18 & 1 & 0 \\ 22.8 & 11.71 & 1.85 & 12.79 & 2.069 & 1 \end{matrix}$$

$$Z_j^*(t) = \begin{bmatrix} 94 \\ 199.3 \\ 975.7 \\ 920.15 \\ 1023.25 \\ 2092.15 \end{bmatrix}$$

<u>No urgent case</u>

- First step:

$$Q_{jK}^*(t) = \frac{\lambda_k}{P_{jG}} Z_j^*(t) = \frac{1}{m_{jk}} U_{jk}^*(t)$$

$$Q_{11}^*(t) = 7.068$$

- Others :

$$Q^*_{ik}(t) = \frac{\lambda_k}{P_{iG}} Z^*_i(t)\delta_{ik}$$

$$\delta_{ik} = \begin{cases} 1 & \text{si } k \in G_i \\ 0 & \text{others} \end{cases}$$

$$Q^*_{21}(t) = 7.78$$

$$Q^*_{3.1}(t) = 27.93$$

$$Q^*_{41}(t) = 40.06$$

$$Q^*_{51}(t) = 32.15$$

$$Q^*_{61}(t) = 46.51$$

High priority patient :

$$Q^n_{jk}(t) = A^n_{jk}(t) - A^n_{jk}(T^n_j(t))$$

$$A_k(t) = N_k(\lambda_k t), \quad t \geq 0$$

$$N_t(t) = \begin{cases} \max\{l \geq 0; u^l_k + ... + u^l_k \leq t\}, u^l_k \leq t \\ 0 \quad\quad u^l_k \geq t \end{cases}$$

$$T^n_j = \frac{1}{p_{jG}} Z^*_j$$

$$E(X) = \frac{1}{N}\sum X_i$$

$$V(X) = \left(X_i - \overline{X}\right)^2$$

Si $X_i \rightarrow N(\mu, \sigma)$

$$\frac{X - E(X)}{\sqrt{V(X)}} \rightarrow N(0,1)$$

Waiting time :

$$S^*_k(t) = \sum_{l=1}^{S(k)} S^*_{lk}(t) = \sum_{\{j: k \in G_j\}} \frac{Z^*_j}{P_{ij}}$$

$$S^*_k(t) = 1.69 + 1.86 + 9.58 + 7.69 + 11.12 = 31.94$$

REFERENCES

Carely. J. Jurishica. (2005), “Emergency department simulations: medicine for building effective models”, Proceedings of the 2005 Winter Simulation Conference.

Green, L. (2006) Queueing analysis in healthcare, in Patient Flow: Reducing Delay in Healthcare Delivery, Hall, R.W., ed., Springer, New York, 281-308.

Jinn-Yi Yeh, Wen-Shan Lin. (2007), “Using simulation technique and genetic algorithm to improve the quality care of a hospital emergency department”, Expert Systems with Application 32 (2007) 1073-1083

Linda V. Green, Joa˜o Soares, James F. Giglio, Robert Alan Green, Society for Academic Emergency Medicine November 2005

William A. Stout, Jr. Barbara Tawney Proceedings of the 2005 Systems and Information Engineering Design Symposium Ellen J. Bass, ed.

William, Peterson; A heavy traffic limit theorem for networks of queues with multiple customer types. Mathematics of operations research, vol, 16, No1. Feb, 1991, pp 90-118.