Department of Mathematics
Faculty of Mechanical Engineering
Slovak University of Technology in Bratislava

3rd International Conference
APLIMAT 2004

PLENARY LECTURE

QUEUEING NETWORKS WITH INFINITE SERVERS IN EACH NODE (AN APPLICATION IN LOGISTICS)

MANUEL ALBERTO M. FERREIRA (PT)

Abstract. After a report of results about queueing systems with infinite servers, namely considering its busy period, we intend to build a model, using networks of queues with infinite servers in each node, to study a two echelons repair system of a fleet of aircraft, of shipping or of trucking. The customers are the failures. And its service time is the time that goes from the instant at which they occur till the one at which they are completely repaired. The failures repairs occur in a base or in a remote station. The whole failures detected in the base are repaired there. Some of the failures detected in the station are repaired in the base, and the others in the station. The results referred above, about the infinite servers queues busy period, allow the determination of the two echelons repair system performance measures. In this application we work over models of Carrillo (1991) and Ferreira (1996) that we improve and complete. We will illustrate the theory with a very simple and short numerical example.

Key words. M|G|∞, network of queues, busy period, logistics.

1 Introduction

In the M|G|∞ queueing system

- The customers arrive according to a Poisson process at rate λ,
- Each of them receives a service whose length is a positive random variable with distribution function $G(\cdot)$ and mean value α. So

$$\alpha = \int_0^\infty [1 - G(t)]dt \qquad (1.1),$$

- There are infinite servers. So when a customer arrives it always finds an available server,
- The service of a customer is independent of the other customers services and of the arrival process.

An important parameter is the traffic intensity that we call ρ, being

$$\rho = \lambda\alpha \qquad (1.2).$$

It is obvious that in a M|G|∞ queueing system there are neither losses nor waiting. In fact there is not queueing in the formal sense of the word.

For these systems it is not so important to study the populational process as for other systems with losses or waiting.

Generally we are much more interested in the study of other processes as, for instance, the busy period.

The busy period of a queueing system begins when a customer arrives there, finding it empty, and ends when a customer leaves the system letting it empty. During the busy period there is always at least one customer in the system.

A network of queues is a collection of nodes arbitrarily connected by arcs, instantaneously traversed by customers, where

- An arrival process is associated to each node,
- There is a commutation process that commands the different customers paths.

We call J the network number of nodes. If $J < \infty$, the nodes are numbered $1,2,\ldots,J$ and we put $U = \{1,2,\ldots,J\}$.

The arrival processes may be the result of exogenous arrivals, from the outside of the collection, and of endogenous arrivals, from the other collection nodes.

A network is open if any customer can enter it or leave it. A network is closed if it has a fixed member of customers moving from node to node, with neither exterior arrivals nor departures. The networks opened for some customers and closed for others are called mixed.

We will consider in this paper open networks, with infinite servers in each node, and Poisson exogenous arrivals at rate λ. So

$$\Lambda = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_J \end{bmatrix} \qquad (1.3)$$

is the network exogenous arrival rate vector. λ_j is the exogenous arrival rate at node j, $j = 1,2,\ldots,J$ and $\sum_{j=1}^{J} \lambda_j = \lambda$.

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1J} \\ p_{21} & p_{22} & \cdots & p_{2J} \\ \vdots & \vdots & & \vdots \\ p_{J1} & p_{J2} & \cdots & p_{JJ} \end{bmatrix} \qquad (1.4)$$

is the commutation process matrix, being p_{jl} the probability of a customer, after ending its service at node j, going to node l, $j = 1,2,\ldots,J$, $l = 1,2,\ldots,J$. $q_j = 1 - \sum_{l=1}^{J} p_{jl}$ is the probability that a customer leaves the network from node j, $j = 1,2,\ldots,J$. We suppose that P does not change with t and is independent of everything that is happening in the network.

We shall see that a network of this kind is equivalent to a M|G|∞ system with Poisson process arrival at rate λ, where each customer service time is its sojourn time in the network.

But first we will present some results about the M|G|∞ systems busy period that are also useful for these networks of queues evidently.

Then we intend to build a model, using these networks with infinite servers in each node, to study a two echelons repair system of a fleet of aircraft, of shipping or of trucking. The customers are the failures. And its service time is the time that goes from the instant at which they occur till the one at which they are completely repaired. Here a busy period is a period in which there is at least one failure waiting for reparation or being repaired. The results referred above, about the M|G|∞ queue busy period, allow the determination of some system performance measures.

We will illustrate the theory with a very simple and short numerical example.

2 The M|G|∞ queue busy period

Let us call B to the M|G|∞ queueing system busy period length.

The mean value of B is, see Takács (1962), whatever is $G(\cdot)$,

$$E[B] = \frac{e^{\rho} - 1}{\lambda} \qquad (2.1).$$

Being $R(t)$ the mean number of busy periods that begin in $[0,t]$ (being $t = 0$ the beginning of a busy period) we have, Ferreira (1995),

$$e^{-\rho}(1 + \lambda t) \leq R(t) \leq 1 + \lambda t \qquad (2.2).$$

Let us call N_B the mean number of the customers served during a busy period in the M|G|∞ queueing systems. We have, Ferreira (2001),

- If $G(\cdot)$ is exponential

$$N_B^M = e^{\rho} \qquad (2.3)$$

- For any other service distribution function

$$N_B^M \cong \frac{e^{\rho(\gamma_s^2+1)}\left(\rho(\gamma_s^2 + 1) + 1\right) + \rho(\gamma_s^2 + 1) - 1}{2\rho(\gamma_s^2 + 1)} \qquad (2.4)$$

where γ_s is the variation coefficient of $G(\cdot)$.

3 The Sojourn Time of a Customer Laplace Transform in a Network of Queues with Infinite Servers in Each Node and Poisson Exogenous Arrivals

Note that for these queueing networks, that we described in **1.**,

- The sojourn times of a customer in each node are the service times, since there is not waiting,
- The sojourn times of a customer in the various nodes are independent.

The sojourn time of a customer in the network, if we know its path, is distributed as a convolution of the service times in the nodes for which it passes. So, the sojourn time distribution will be the mixture of the convolutions, related to each possible path, being the weight of each one given by the respective path probability. These probabilities are known and do not depend on time since we know and do not depend on time the exogenous arrival rates to the various nodes, the commutation probabilities and the probabilities to leave the network.

It will be, generally, difficult to get an efficient formula to the sojourn time distribution function, based on the whole possible paths direct enumeration because

- The number of paths may be infinite or enormously great, even in the situation of few nodes networks,
- The convolutions can lead to intractable analytic expressions.

Using matrixes whose shape is suggested by **(1.3)** and **(1.4)** we get a simple formula to the sojourn time Laplace Transform, as a function of the service times in each node Laplace Transforms (Ferreira and Ramalhoto (1990) and Ferreira (1996)).

Be *T* the network sojourn time of a customer, and S_j its service time at node j, $j = 1,2,\ldots,J$. Be $G(t)$ and $G_j(t)$, $j = 1,2,\ldots,J$ the T and S_j, $j = 1,2,\ldots,J$ distribution functions respectively, being $\overline{G}(s)$ and $\overline{G}_j(s)$, $j = 1,2,\ldots,J$ the T and S_j, $j = 1,2,\ldots,J$, Laplace Transforms, respectively. Puting

$$\Lambda(s) = \begin{bmatrix} \lambda_1 \overline{G}_1(s) \\ \lambda_2 \overline{G}_2(s) \\ \vdots \\ \lambda_J \overline{G}_J(s) \end{bmatrix} \tag{3.1}$$

and

$$P(s) = \begin{bmatrix} p_{11}\overline{G}_1(s) & p_{12}\overline{G}_2(s) & \ldots & p_{1J}\overline{G}_J(s) \\ p_{21}\overline{G}_1(s) & p_{22}\overline{G}_2(s) & \ldots & p_{2J}\overline{G}_J(s) \\ \vdots & \vdots & & \vdots \\ p_{J1}\overline{G}_1(s) & p_{J2}\overline{G}_2(s) & \ldots & p_{JJ}\overline{G}_J(s) \end{bmatrix} \tag{3.2}$$

we have

$$\overline{G}(s) = \sum_{n=0}^{\infty} \lambda^{-1} \Lambda^T(s) P^n(s)(I - P)A \tag{3.3}$$

being *A* a column with *J* 1's.

Note that

- $\Lambda(0) = \Lambda$
- $P(0) = P$
- In **(3.3)** we consider the whole paths and the respective probabilities, associating them simultaneously its nodes service times Laplace Transforms product. Each path begins in node j with probability $\lambda_j \lambda^{-1}$ and will finish in node k with probability $1 - \sum_{j=1}^{J} p_{kj}$, $j = 1,2,\ldots,J$, $k = 1,2,\ldots,J$.

The formula that we want is got noting that **(3.3)** may be put in the form

$$\overline{G}(s) = \lambda^{-1} \Lambda^T(s)(I - P(s))^{-1}(I - P)A \tag{3.4}$$

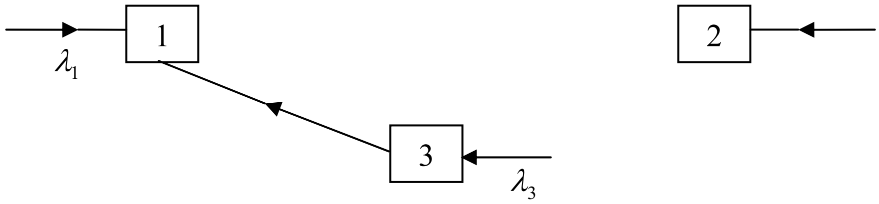
since $|I - P(s)| \neq 0$.

4 A two echelons repair system

The results presented may be applied in logistics. Suppose a fleet of aircraft, of shipping or of trucking whose failures repairs occur in a base or in a remote station.

The whole failures detected in the base are repaired there. Some of the failures detected in the station are repaired in the base with probability *p,* being necessary to transport them to the base, and the others in the station. Here the service time is the time that goes from the instant at which the failure occur till the one at which it is completely repaired. When it is necessary to transport an item with a failure from the remote station to the base we suppose that it is immediately possible, being the service time, now, the time that the transport lasts. We suppose still that the failures occur according to a Poisson process at rate λ, being some detected in the remote station with probability q and the others in the base.

So we will have a queueing network with three nodes



where

- 1 is the base,
- 2 is the remote station,
- 3 considers the necessary transports from the remote station to the base.

Representing the variables related with each node for the same letter, as in the former sections, but with an index corresponding to the node we have, obviously,

$$\Lambda = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_J \end{bmatrix} \begin{bmatrix} (1-q)\lambda \\ (1-p)q\lambda \\ pq\lambda \end{bmatrix} \quad (4.1),$$

$$P = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad (4.2).$$

Being the system, globally, a M|G|∞ queue and after **(3.4)** we get

$$\overline{G}(s) = \lambda^{-1}\left[(1-q)\lambda\overline{G}_1(s) \quad (1-p)q\lambda\overline{G}_2(s) \quad pq\lambda\overline{G}_3(s)\right]\cdot\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\overline{G}_1(S) & 0 & 1 \end{bmatrix}^{-1} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} =$$

$$= \left[(1-q)\overline{G}_1(s) \quad (1-p)q\overline{G}_2(s) \quad pq\lambda\overline{G}_3(s)\right]\cdot\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \overline{G}_1(S) & 0 & 1 \end{bmatrix}\cdot\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \left[(1-q)\overline{G}_1(s) + (1-p)q\overline{G}_2(s) + \right.$$

$$\left. + pq\overline{G}_1(s)\overline{G}_3(s)\right]\cdot$$

So the service time distribution function is

$$G(t) = (1-q)G_1(t) + (1-p)qG_2(t) + pqG_{13}(t) \quad (4.3)$$

being G_{13} the distribution function of the convolution of the service time distributions in nodes 1 and 3.

We can consider also three M|G|∞ queues

- One related with the repairs in the base, whose failures were detected there:

$$\begin{aligned} \lambda_b &= (1-q)\lambda \\ G_b(t) &= G_1(t) \\ \rho_b &= (1-q)\lambda\alpha_1 \end{aligned} \quad (4.4)$$

- Other related with the repairs in the remote station:

$$\begin{aligned} \lambda_{st} &= (1-p)q\lambda \\ G_{st}(t) &= G_2(t) \\ \rho_{st} &= (1-p)q\lambda\alpha_2 \end{aligned} \quad (4.5)$$

- And still other related with the repairs in the base after transport from the remote station

$$\begin{aligned} \lambda_{tr} &= pq\lambda \\ G_{tr}(t) &= G_{13}(t) \\ \rho_{tr} &= pq\lambda(\alpha_1 + \alpha_3) \end{aligned} \quad (4.6).$$

Concerning the application of **(2.4)**, being σ_1^2, σ_2^2 and σ_3^2 the variances corresponding to $G_1(\cdot)$, $G_2(\cdot)$ and $G_3(\cdot)$, respectively we have

$$\begin{aligned} \gamma_{sb} &= \gamma_{s1} \\ \gamma_{sst} &= \gamma_{s2} \\ \gamma_{str} &= \frac{\sqrt{\sigma_1^2 + \sigma_3^2}}{\alpha_1 + \alpha_3} \end{aligned} \quad (4.7)$$

and for the coefficient of variation corresponding to the distribution function given for **(4.3)**

$$\gamma_s = \sqrt{\frac{(1-q)(\sigma_1^2 + \alpha_1^2) + (1-p)q(\sigma_2^2 + \alpha_2^2) + pq(\sigma_1^2 + \sigma_3^2 + (\alpha_1 + \alpha_3)^2)}{((1-q)\alpha_1 + (1-p)q\alpha_2 + pq(\alpha_1 + \alpha_3))^2} - 1} \quad (4.8).$$

5 Example

Suppose that $G_1(\cdot)$ and $G_2(\cdot)$ are both exponential with mean 1 week and $G_3(\cdot)$ is constant with value 1 week. $q = 0.3$ and $\lambda = \frac{1}{4\ week}$ (1 per month). Now p=0,9. Considering 1 year (52 weeks) of operation we would like to conclude if decreasing *p* (that is: increasing the station capacity of repairs) there would be any advantage.

Making *p*=0,9; 0,8; ...0,1 we have

- For the global system

p	$\frac{e^{\rho}-1}{\lambda}$	$e^{-\rho}(1+\lambda 52)$	$(1+\lambda 52)$	N_B
0,9	1,51	10	14	2,04
0,8	1,45	10	14	2,03
0,7	1,40	10	14	2,02
0,6	1,40	10	14	2,03
0,5	1,35	10	14	2,02
0,4	1,29	11	14	2,01
0,3	1,24	11	14	1,99
0,2	1,24	11	14	2,00
0,1	1,19	11	14	1,99

- For the remote station

P	$\frac{e^{\rho_{st}}-1}{\lambda_{st}}$	$e^{-\rho_{st}}(1+\lambda_{st}52)$	$(1+\lambda_{st}52)$	N_B
0,9	1,00	1,38	1,39	1,01
0,8	1,01	1,75	1,78	1,02
0,7	1,01	2,12	2,17	1,02
0,6	1,02	2,48	2,56	1,03
0,5	1,02	2,84	2,95	1,04
0,4	1,02	3,19	3,34	1,05
0,3	1,03	3,54	3,73	1,05
0,2	1,03	3,88	4,12	1,06
0,1	1,03	4,22	4,51	1,07

- For the repairs in the base after transport

p	$\frac{e^{\rho_{tr}}-1}{\lambda_{tr}}$	$e^{-\rho_{tr}}(1+\lambda_{tr}52)$	$(1+\lambda_{tr}52)$	N_B
0,9	2,14	3,94	4,51	3,49
0,8	2,12	3,65	4,12	3,13
0,7	2,11	3,36	3,73	2,81
0,6	2,09	3,05	3,34	2,54
0,5	2,08	2,74	2,95	2,30
0,4	2,06	2,41	2,56	2,10
0,3	2,05	2,07	2,17	1,91
0,2	2,03	1,73	1,78	1,76
0,1	2,02	1,37	1,39	1,62

- For the base

$\frac{e^{\rho_b}-1}{\lambda_b}=1{,}09,\ e^{-\rho_b}(1+\lambda_b 52)=8{,}48,\ 1+\lambda_b 52=10{,}1$ and $N_B^M \cong 1{,}19$, whatever is p.

Of course, in the operation of a fleet, we are interested in big idle periods and in little busy periods. And if these occur it is good that they are as rare as possible, with a short number of failures.

So we can say that the global system improves its performance as p decreases but very lightly. The remote station grows worse has it was expected and the repairs in the base after transport improve its performance.

Note, for instance, that if we want to guarantee less than two busy periods for the repairs in the base after transport we have to get $p = 0{,}2$. But, for this, it is necessary to spend money in staff and material in the remote station. It is necessary to balance these expenses with the savings in transports.

6 Conclusions

To apply this model is necessary to verify if the failures occur according to a Poisson process. This hypothesis has to be tested. After this it is very simple to apply the other results.

Carrillo (1991) studied a model looking like the one presented here. But he did not consider either the possibility of transport from the station to the base or the busy period.

So these conceptions allow, in a very simple way, to evaluate the performance of a fleet as the example presented has shown.

References

[1] Carrillo, M. J., (1991), "Extensions of Palm's Theorem: A Review", Management Science, vol. 37, nº 6, p. 739-744.

[2] Ferreira, M. A. M., (1995), "Comportamento transeunte e período de ocupação de sistemas de filas de espera sem espera", PhD Thesis discussed in I.S.C.T.E., Supervisor: Prof. Augusto A. Albuquerque.

[3] Ferreira, M. A. M., (1996), *"*Sistemas de filas de espera com infinitos servidores-Uma aplicação em logística*", Revista de Estatística, vol. 1, I.N.E., 1996.*

[4] Ferreira, M. A. M., (2001)*, "*Mean number of the customers served during a busy period in the M|G|∞ queueing system*", Statistical Review, vol. 3, I.N.E., 2001.*

[5] Ferreira, M. A. M. e Ramalhoto, M. E., (1990), *"*Algoritmo para o cálculo da Transformada de Laplace-Stieltjes do tempo de permanência de um cliente em redes de filas de espera abertas, com distribuições de equilíbrio da forma produto, com tempos de permanência independentes em cada nó*", Actas da 1ª Conferência em Estatística e Optimização. C.E.A.. (I.N.I.C.) e D.E.I.O.C.. Universidade de Lisboa.*

[6] Takács, L., (1962), "An Introduction to Queueing Theory", Oxford University Press, New York, 1962.

Contact Address

Manuel Alberto M. Ferreira, Associate Professor, ISCTE, Departamento de Métodos Quantitativos
Avª das Forças Armadas, 1649-026 LISBOA, Telefone: +351 21 790 30 46,
Fax: +351 21 790 39 64
e-mail: vice.presidente@iscte.pt