International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 01, January-2021

Distribution of Occupied Resources on A Discrete Resources Sharing in A Queueing System

Toky Basilide Ravaliminoarimalalason
Ecole Doctorale en Sciences et Techniques de l'Ingénierie
et de l'Innovation,
University of Antananarivo,
Antananarivo, Madagascar

Falimanana Randimbindrainibe
Ecole Doctorale en Sciences et Techniques de l'Ingénierie
et de l'Innovation,
University of Antananarivo
Antananarivo, Madagascar

***Abstract*—In this paper, we study discrete resources sharing in a queueing system. We build analytical model of the distribution of occupied resources that can help for resources dimensioning. Both infinite and finite amount of discrete server resources are highlighted and validated with special cases of individual resource requirement following Poisson and Binomial distribution. It is found that there is a peak of usage near the average number of resources requested by customers, and other small peaks with low probability at multiples of this mean. The charging factor ρ of the queue impacts mostly on the resources occupation distribution.**

***Keywords*—Distribution; probability; queue; resource; sharing**

I. INTRODUCTION

An event flow called customers sequentially arrives in a system to claim a service [1]. The time between the arrival of these customers, called inter-arrival, and the length of service requested by customers are random variables. This queueing system is made up of a queueing area, of finite or infinite capacity, and servers handling the services requested by customers. With that comes a discipline dictating how or who in the queueing area customers are going to be served first.

Basic theories of queueing systems have focused on numbers (finite or not) of servers. Each customer will be served by a server, until all servers are busy. Once busy, customers in the queueing system will wait until one server is free.

However, in many real systems, some customers need more than one server, or more than one resource in a server. Some servers can take care of multiple customers at the same time, allocating some of their resources to customers, and some to other ones, depending on their needs. Researches have advanced on resource-based systems since Green's work on queues in which customers request multiple servers [2]. Others even wanted to generalize the notion of Erlang such as Romm [3] or Tikhonenko [4]. Studies and analyzes of queueing systems capacity have been advanced [6][7][8], on queues with multiple servers or multiple resources [9][10] but we will be interested in analytical models of a queueing system that can share its discrete resources with users requesting service from it. More precisely, we will explore the probability distributions followed by the occupied resources in such queues in order to be able to perform a resource dimensioning.

II. FROM BIRTH AND DEATH PROCESS

Given a queueing system a priori having k customers. It can handle between 0 and n customers. In this system, customers arrivals form a Poisson process of intensity λ_k, the service duration they request are random and are following an exponential distribution with average $1/\mu_k$.

We denote by N the state of this system, N describes the number of customers present on it. Let $(\lambda_0, \lambda_1, \dots, \lambda_{n-1})$ and $(\mu_1, \mu_2, \dots, \mu_n)$ be positive numbers.

Proposition 1

During an infinitely small time interval ε, knowing that there are k customers in the queue :

- The probability that a customer arrives in the queue is $\lambda_k\varepsilon + o(\varepsilon)$,
- The probability that a customer leaves the queue is $\mu_k\varepsilon + o(\varepsilon)$,
- All other eventualities have a probability $o(\varepsilon)$.

Proof:

The probability that a customer will arrive during an infinitely small time interval ε, which we denote by P_a, is equal to the probability that the inter-arrival T of the customers is less than ε. The arrival of customers form a Poisson process of intensity λ_k, so the inter-arrival T is random variable following an exponential distribution with parameter λ_k $P_a = P(T \leq \varepsilon) = 1 - e^{-\lambda_k\varepsilon}$. For ε infinitely small, we can write the expansion limited to order 1 of this probability in the neighborhood of ε by $P_a = 1 - (1 - \lambda_k\varepsilon) + o(\varepsilon) = \lambda_k\varepsilon + o(\varepsilon)$.

In a similar way, with a random variable of service duration following an exponential distribution with average $1/\mu_k$, we have the probability P_d of customer departure $P_a = 1 - e^{-\mu_k\varepsilon} = \mu_k\varepsilon + o(\varepsilon)$. ■

Proposition 2

The stationary distribution $\pi = (\pi_0, \dots, \pi_n)$ of this system is equal to :

$$\pi_k = \frac{\lambda_0\lambda_1 \dots \lambda_{k-1}}{\mu_1\mu_2 \dots \mu_k}\pi_0\,, \quad k = 1,2, \dots, n \tag{1}$$

Where

$$\frac{1}{\pi_0} = 1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_0\lambda_1}{\mu_1\mu_2} + \dots + \frac{\lambda_0\lambda_1 \dots \lambda_{n-1}}{\mu_1\mu_2 \dots \mu_n} \tag{2}$$

Proof:

As a discrete-time Markov chain, with time step ε, we can show the following transition probability p_{ij} :

- $p_{00} = 1 - \lambda_0\varepsilon + o(\varepsilon)$
- $p_{i,i-1} = \mu_i\varepsilon + o(\varepsilon)$ for $i = 1, \dots, n$
- $p_{ii} = 1 - (\lambda_i + \mu_i)\varepsilon + o(\varepsilon)$ for $i = 1, \dots, n-1$
- $p_{nn} = 1 - \mu_n\varepsilon + o(\varepsilon)$

- $p_{i,i+1} = \lambda_i \varepsilon + o(\varepsilon)$ for $i = 0, \dots, n-1$
- …

From these transition probabilities, the transition matrix corresponding to this Markov chain in discrete time is given by following equation:

$$P(\varepsilon) = \begin{pmatrix} 1-\lambda_0\varepsilon & \lambda_0\varepsilon & 0 & \dots \\ \mu_1\varepsilon & 1-(\lambda_1+\mu_1)\varepsilon & \lambda_1\varepsilon & \dots \\ 0 & \mu_2\varepsilon & 1-(\lambda_2+\mu_2)\varepsilon & \dots \\ \vdots & \vdots & \vdots & \dots \\ 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & \dots \end{pmatrix} + o(\varepsilon) \qquad (3)$$

This matrix can be written in the form $P(\varepsilon) = Id + A\varepsilon + o(\varepsilon)$ where A is the infinitesimal stochastic generator of the continuous-time Markov chain obtained as ε approaches 0.

$$A = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & \dots \\ \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & \dots \\ 0 & \mu_2 & -(\lambda_2+\mu_2) & \dots \\ \vdots & \vdots & \vdots & \dots \\ 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & \dots \end{pmatrix} \qquad (4)$$

The stationary distribution of the state of the system is obtained from the equation of state relation of a continuous-time Markov chain $\pi.A = 0$. So,

- $\lambda_0\pi_0 = \mu_1\pi_1$,
- $\lambda_0\pi_0 - (\lambda_1+\mu_1)\pi_1 + \mu_2\pi_2 = 0,\ \lambda_1\pi_1 = \mu_2\pi_2$
- …
- $\lambda_{i-2}\pi_{i-2} - (\lambda_{i-1}+\mu_{i-1})\pi_{i-1} + \mu_i\pi_i = 0,\ \lambda_{i-1}\pi_{i-1} = \mu_i\pi_i$, for $i = 2, \dots, n-1$
- …
- $\lambda_{n-1}\pi_{n-1} = \mu_n\pi_n$

From these equations, we can get $\pi_1 = \frac{\lambda_0}{\mu_1}\pi_0$, $\pi_2 = \frac{\lambda_1}{\mu_2}\pi_1 = \frac{\lambda_0}{\mu_1}\frac{\lambda_1}{\mu_2}\pi_0$, $\pi_3 = \frac{\lambda_2}{\mu_3}\pi_2 = \frac{\lambda_0}{\mu_1}\frac{\lambda_1}{\mu_2}\frac{\lambda_2}{\mu_3}\pi_0$, … , and then $\pi_k = \frac{\lambda_0\lambda_1\dots\lambda_{k-1}}{\mu_1\mu_2\dots\mu_k}\pi_0$ for $k = 1,2,\dots,n$. And knowing that $(\pi_0,\dots,\pi_n)$ is a probability distribution, $\pi_0 + \pi_1 + \dots + \pi_n = 1$, so π_0 is given by $\frac{1}{\pi_0} = 1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_0\lambda_1}{\mu_1\mu_2} + \dots + \frac{\lambda_0\lambda_1\dots\lambda_{n-1}}{\mu_1\mu_2\dots\mu_n}$. ∎

We note that $\pi_k = P(N = k)$, the probability to find k customers in the queueing system.

III. STATE DISTRIBUTION IN RESOURCE SHARING

Now, we consider that the queueing system has discrete resources in its servers. In total, C resources are available on it, $C \in \mathbb{N} \cup \{+\infty\}$. It means that the quantity of system resources can be finite or infinite in our study, relative to the case.

Each customer needs an amount r of resources, a discrete random variable. If the amount r of resources requested by the customer is available, the server can allocate them while serving it. If they are not available, then the requesting customer remains in the queueing area until the requested resources are released for use.

A. *Infinite capacity queueing system with infinite server resources*

The system queueing is M/M/∞.

1) General case

Proposition 3

The stationary distribution of this queueing system is:

$$\pi_k = \frac{1}{k!}\rho^k e^{-\rho} \text{ where } \rho = \frac{\lambda}{\mu} \qquad (5)$$

ρ is called charging factor.

Proof:

In this case, $n \to \infty$, $\lambda_k = \lambda$ for all $k = 0, 1, 2, \dots$ and $\mu_k = k\mu$ for all $k = 1,2,\dots$

$$\pi_k = \frac{\frac{\lambda_0\lambda_1\dots\lambda_{k-1}}{\mu_1\mu_2\dots\mu_k}}{1+\frac{\lambda_0}{\mu_1}+\frac{\lambda_0\lambda_1}{\mu_1\mu_2}+\dots+\frac{\lambda_0\lambda_1\dots\lambda_{n-1}}{\mu_1\mu_2\dots\mu_n}+\dots} = \frac{\frac{\lambda\lambda\dots\lambda}{\mu.2\mu\dots k\mu}}{1+\frac{\lambda}{\mu}+\frac{\lambda\lambda}{\mu.2\mu}+\dots+\frac{\lambda\lambda\dots\lambda}{\mu.2\mu\dots k\mu}+\dots}$$

$$= \frac{\frac{\lambda^k}{\mu^k.k!}}{1+\frac{\lambda}{\mu}+\frac{\lambda^2}{\mu^2.2!}+\dots+\frac{\lambda^n}{\mu^n.n!}+\dots} = \frac{\left(\frac{\lambda}{\mu}\right)^k}{k!e^{\frac{\lambda}{\mu}}} = \frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k e^{-\frac{\lambda}{\mu}} \quad ∎$$

Since the server has infinite capacity, then all customers in the system are served simultaneously. Customers arriving to the queueing area are immediately served after entering the system. If the system has k customers served, they are using at the same time r_k amount of resources. Let denote $P(R = r)$ the probability distribution of R, the random variable describing the total amount of occupied resources, and $P_k(r)$ the probability that the k customers use r ressources.

$$\begin{aligned} P(R = r) &= \textstyle\sum_{k=1}^{+\infty}(P(N = k).P_k(r)) \\ &= \textstyle\sum_{k=1}^{+\infty}\left(\frac{1}{k!}\rho^k e^{-\rho}.P_k(r)\right) \end{aligned} \qquad (6)$$

2) Individual resource usage following Poisson distribution

We consider that resource usage by individual customer follows Poisson distribution with parameter $\alpha : \mathcal{P}(\alpha)$.

Proposition 4

Resource usage by k customers is a random variable also following the Poisson distribution with parameter $k\alpha$: $P_k \rightsquigarrow \mathcal{P}(k\alpha)$.

Proof:

The sum of independent Poisson random variable $\mathcal{P}(\alpha)$ and $\mathcal{P}(\beta)$ is a random variable following Poisson distribution with parameter $\alpha + \beta : \mathcal{P}(\alpha + \beta)$.In other hand, the total amount of occupied resources is additive (sum of resources amount of used by each customers). ∎

Proposition 5

The number of occupied resources in the server has the following probability distribution:

$$P(R = r) = \frac{\alpha^r}{r!}e^{-\rho}\textstyle\sum_{k=1}^{+\infty}\frac{k^r}{k!}(\rho e^{-\alpha})^k \qquad (7)$$

Proof:

From (6), and knowing that $P_k(r) = \frac{1}{r!}(k\alpha)^r e^{-k\alpha}$, we get

$$P(R = r) = \textstyle\sum_{k=1}^{+\infty}\frac{1}{k!}\rho^k e^{-\rho}\frac{1}{r!}(k\alpha)^r e^{-k\alpha} = \frac{\alpha^r}{r!}e^{-\rho}\sum_{k=1}^{+\infty}\frac{k^r}{k!}(\rho e^{-\alpha})^k \quad ∎$$

B. *Infinite capacity queueing system with finite server resources*

1) General case

We still have infinite capacity of queueing system, $n \to \infty$. But the server has finite resources, so the number of customers that can be served simultaneously is limited. Let s be the maximum number of customers served by the server. We note that s is variable according the amount of resources required by the customers. The stationary distribution of the queueing system having capacity s in terms of number of customers is denoted by $\pi_{k,s}$.

Proposition 8

The stationary distribution $\pi_{k,s}$ is given by :

$$\pi_{k,s} = \begin{cases} \dfrac{\frac{1}{k!}\rho^k}{\sum_{i=0}^{s}\frac{1}{i!}\rho^i+\sum_{i=1}^{+\infty}\frac{1}{s!s^i}\rho^{s+i}} & for\ k \leq s \\ \dfrac{\frac{1}{s!s^{k-s}}\rho^k}{\sum_{i=0}^{s}\frac{1}{i!}\rho^i+\sum_{i=1}^{+\infty}\frac{1}{s!s^i}\rho^{s+i}} & for\ k > s \end{cases} \tag{8}$$

Proof:

We get this stationary distribution from the following parameters : $\lambda_k = \lambda\ (0 \leq k)$, $\mu_k = k\mu\ (0 < k < s)$ and $\mu_k = s\mu\ (s \leq k)$, because the number of customers that can exit the system is limited by s.

So, for $k \leq s$:

$$\pi_{k,s} = \frac{\frac{\lambda}{\mu}.\frac{\lambda}{2\mu}\cdots\frac{\lambda}{k\mu}}{1+\frac{\lambda}{\mu}+\frac{\lambda}{\mu}.\frac{\lambda}{2\mu}+\cdots+\frac{\lambda}{\mu}.\frac{\lambda}{2\mu}\cdots\frac{\lambda}{s\mu}+\frac{\lambda}{\mu}.\frac{\lambda}{2\mu}\cdots\frac{\lambda}{s\mu}\frac{\lambda}{s\mu}\cdots}$$

$$= \frac{\frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k}{1+\frac{\lambda}{\mu}+\cdots+\frac{1}{s!}\left(\frac{\lambda}{\mu}\right)^s+\sum_{i=1}^{+\infty}\frac{1}{s!s^i}\left(\frac{\lambda}{\mu}\right)^{s+i}} = \frac{\frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k}{\sum_{i=0}^{s}\frac{1}{i!}\left(\frac{\lambda}{\mu}\right)^i+\sum_{i=1}^{+\infty}\frac{1}{s!s^i}\left(\frac{\lambda}{\mu}\right)^{s+i}}$$

And for $k > s$:

$$\pi_{k,s} = \frac{\frac{\lambda}{\mu}.\frac{\lambda}{2\mu}\cdots\frac{\lambda}{s\mu}\frac{\lambda}{s\mu}\frac{\lambda}{s\mu}}{1+\frac{\lambda}{\mu}+\frac{\lambda}{\mu}.\frac{\lambda}{2\mu}+\cdots+\frac{\lambda}{\mu}.\frac{\lambda}{2\mu}\cdots\frac{\lambda}{s\mu}+\frac{\lambda}{\mu}.\frac{\lambda}{2\mu}\cdots\frac{\lambda}{s\mu}\frac{\lambda}{s\mu}\cdots}$$

$$= \frac{\frac{1}{s!s^{k-s}}\left(\frac{\lambda}{\mu}\right)^k}{1+\frac{\lambda}{\mu}+\cdots+\frac{1}{s!}\left(\frac{\lambda}{\mu}\right)^s+\sum_{i=1}^{+\infty}\frac{1}{s!s^i}\left(\frac{\lambda}{\mu}\right)^{s+i}} = \frac{\frac{1}{s!s^{k-s}}\left(\frac{\lambda}{\mu}\right)^k}{\sum_{i=0}^{s}\frac{1}{i!}\left(\frac{\lambda}{\mu}\right)^i+\sum_{i=1}^{+\infty}\frac{1}{s!s^i}\left(\frac{\lambda}{\mu}\right)^{s+i}} \quad \blacksquare$$

We denote C $(C \in \mathbb{N}^*)$, the number of resources that the server can allocate. The capcity s customers varies from 1 (if the customer require all C resources at the same time) to C (each customers consume a single resource). Let P_s be the probability that this capacity in terms of customers is s during an observation of the system.

Proposition 6

The probability distribution P_s is given by :

$$P_s = \sum_{r=1}^{C} P(R_s = r).P(R_1 > C - r) \tag{9}$$

Proof:

The capacity s of the system is achieved in case of r resources are used by these s customers , and the $s + 1$-th following customer requests an amount of $R > C - r$ resources. $P(R_k = r) = P_k(r)$ denotes the probability that k customers are using r resources of the server. ■

Proposition 7

The probability that any resource is used is:

$$P(R = 0) = \frac{\sum_{r=1}^{C} P(R_s=r).P(R_1>C-r)}{\sum_{i=0}^{s}\frac{1}{i!}\rho^i+\sum_{i=1}^{+\infty}\frac{1}{s!s^i}\rho^{s+i}} \tag{10}$$

Proof:

This event is obtained when no customer is found in the queueing system, that is, with a probability $\pi_{0,s}$, for all possible values of s.

$$P(R = 0) = \sum_{s=1}^{C} \pi_{0,s} . P_s = \frac{\sum_{r=1}^{C} P(R_s=r).P(R_1>C-r)}{\sum_{i=0}^{s}\frac{1}{i!}\rho^i+\sum_{i=1}^{+\infty}\frac{1}{s!s^i}\rho^{s+i}}. \ \blacksquare$$

Proposition 8

The probability distribution of the number of occupied resources is equal to:

$$P(R = r) = \sum_{s=1}^{C}\left(\sum_{k=1}^{s} \pi_{k,s}.P(R_k = r) + \sum_{k=s+1}^{+\infty} \pi_{k,s}.P(R_s = r)\right).P_s \ \text{ for } 0 < r \leq C \tag{11}$$

Proof:

If the capacity in terms of customer is $s = 1$, then the usage of resources amount r is limited to this only one customer. We have the probability : $\pi_{1,s}.P(R_1 = r) + \pi_{2,s}.P(R_1 = r) + \pi_{3,s}.P(R_1 = r) + \cdots = \pi_{1,s}.P(R_1 = r) + \sum_{k=2}^{+\infty} \pi_{k,s}.P(R_1 = r)$.

If the capacity in terms of customer is $s = 2$, then the usage of resources amount r is limited to these 2 customers. We have the probability : $\pi_{1,s}.P(R_1 = r) + \pi_{2,s}.P(R_2 = r) + \pi_{3,s}.P(R_2 = r) + \cdots = \sum_{k=1}^{2} \pi_{k,s}.P(R_k = r) + \sum_{k=3}^{+\infty} \pi_{k,s}.P(R_2 = r)$.

For any s, the utilization of resources amount r is limited to s customers : $\pi_{1,s}.P(R_1 = r) + \pi_{2,s}.P(R_2 = r) + \cdots + \pi_{s,s}.P(R_s = r) + \pi_{s+1,s}.P(R_s = r) + \pi_{s+2,s}.P(R_s = r) + \cdots = \sum_{k=1}^{s} \pi_{k,s}.P(R_k = r) + \sum_{k=s+1}^{+\infty} \pi_{k,s}.P(R_s = r)$.

Thus, the probability of usage of r amount of resources is obtained for all possible values of s : $P(R = r) = \sum_{s=1}^{C}\left(\sum_{k=1}^{s} \pi_{k,s}.P(R_k = r) + \sum_{k=s+1}^{+\infty} \pi_{k,s}.P(R_s = r)\right).P_s$ ■

Blocking without loss applies in the following cases:

- If the remaining amount of server resources is less than the next customer requirement, it cannot enter the server area and must wait there until the resources become available,
- If the total number of server resources (finite) is less than the customer's requirement, the system will hang forever.

Proposition 9

The blocking probability is given by:

$$B = \sum_{r=0}^{C} \sum_{s=1}^{C} P_s.\pi_{s,s}.P(R_s = r).P(R_1 \geq C - r) \tag{12}$$

Proof:

The system is blocked for any capacity s in terms of customers whose server already occupies s customers consuming r amount of resources, and another customer requests an amount of resources greater than $C - r$. ■

2) Individual resource usage following Binomial distribution

Let's take the example of binomial case to not fall in an eternal blockage of the system (we can limit the maximal number or required resources). A customer can request now a random quantity of resources amount at most M according to his needs, and following a binomial distribution with mean m.

Proposition 10

The number of required resources by one customer follows a binomial distribution with parameters $n = M$ and $p = m/M$: $\mathcal{B}(n, p)$.

Proof:

The maximal number of resources that a customer can request is M, so the n parameter of the binomial distribution is $n = M$. His mean is $np = m$, so $p = m/M$. It is a binomial distribution $\mathcal{B}(n, p) = \mathcal{B}(n = M, p = m/M)$. ■

Proposition 11

The amount of resources requested by k customers follows also a binomial distribution with parameters kM and m/M.

Proof:

Due to the additivity of the amount of resources, and that the sum of k independents binomial distributions $\mathcal{B}(n,p)$ is still a binomial distribution $\mathcal{B}(kn,p)$. ■

Using (9), (10), (11) and (12), we have:
The probability of usage of r resources by k customers:

$$P(R_k = r) = \binom{kM}{r} p^r q^{kM-r} \text{ where } q = 1 - p \qquad (13)$$

The probability that the system has a capacity of s customers :

$$P_s = \sum_{r=1}^{C} P(R_s = r)\,.\,P(R_1 > C - r)$$
$$= \sum_{r=1}^{C}\left(\binom{sM}{r} p^r q^{sM-r}.\sum_{i=C-r+1}^{C}\binom{M}{i} p^i q^{M-i}\right) \qquad (14)$$

The probability that r amount of resources are occupied in the system :

$$P(R = r) = \sum_{s=1}^{C}\left(\sum_{k=1}^{s} \pi_{k,s}.P(R_k = r) + \sum_{k=s+1}^{+\infty} \pi_{k,s}.P(R_s = r)\right).P_s$$
$$= \sum_{s=1}^{C}\left(\frac{1}{\sum_{i=0}^{s}\frac{1}{i!}\rho^i + \sum_{i=1}^{+\infty}\frac{1}{s!s^i}\rho^{s+i}}\left(\sum_{k=1}^{s}\frac{1}{k!}\rho^k.\binom{kM}{r}p^r q^{kM-r} + \sum_{k=s+1}^{+\infty}\frac{1}{s!s^{k-s}}\rho^k.\binom{sM}{r}p^r q^{sM-r}\right)\right).P_s \qquad (15)$$

IV. Results and Discussions

A. Validation of the analytical model

To validate our theoretical results, simulations were carried out under Matlab-Simulink model of resource sharing.

The first step was to share discrete resources from a queue to server of infinite resources. Customers arrive with an arrival rate $\lambda = 1/1.2$ s^{-1}, request service with a random duration following exponential distribution of average $1/\mu = 0.8$ s, and require a random number of resources amount following a Poisson distribution with average $\alpha = 5$ resources.

Fig. 1. shows comparison of the simulation result (histogram in blue) and the analytical expression of (7) (red curve). The histogram in blue indicates the normalized representation of observed number of resources occupied by customers. The observation is at each 0.1s for a duration of 5,000s. We note that the result of simulation follows the analytical expression of the probability distribution. Other simulations have been carried out for different values of λ, μ and ρ and we found the same facts. This validates our expression in (7).

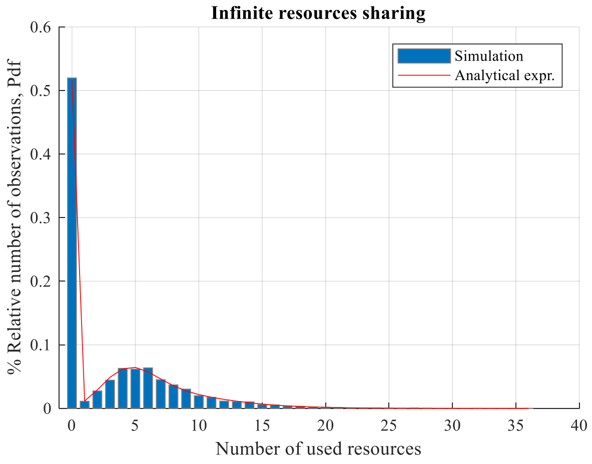

Fig. 1. Occupied resources in an infinite resources sharing

For the sharing of finite resources, Fig. 2. represents the result of simulation for customers arriving with an arrival rate $\lambda = 1/1.2$ s^{-1}, requesting service with a random duration following exponential distribution of average $1/\mu = 0.8$ s. The resources requested by these customers are random following binomial distribution of maximum value $M = 6$ resources, and of mean $m = 3$ resources. The capacity of the server is $C = 10$ resources.

Fig 2. shows the comparison of the results of the simulation and the analytical expression in (15).

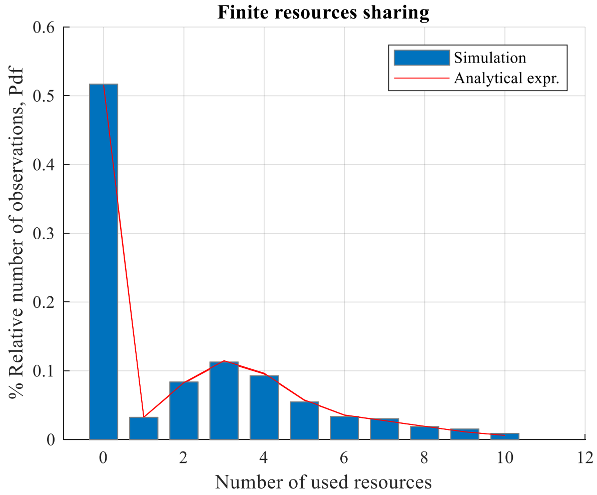

Fig. 2. Occupied resources in a finite resources sharing

We reach the same conclusion from the validation of our analytical model of finite resource sharing after multiple simulations with different values of λ, μ, M and m.

B. Abacus of occupied resources

From the analytical expressions that we have just validated, we can draw up the following charts based on the queueing system parameters values, whether finite resources or not.

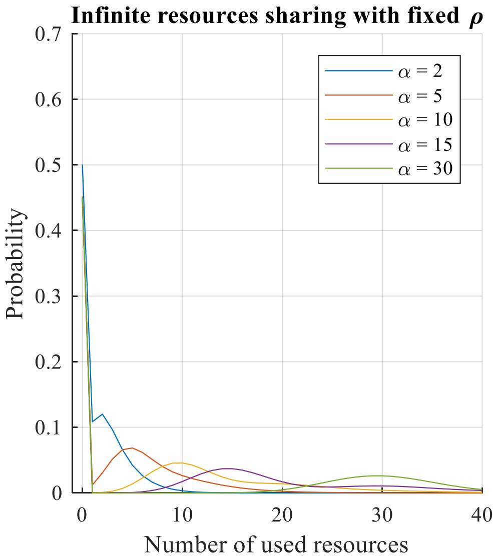

Fig. 3. Distribution of occupied resources for fixed ρ in an infinite resources sharing

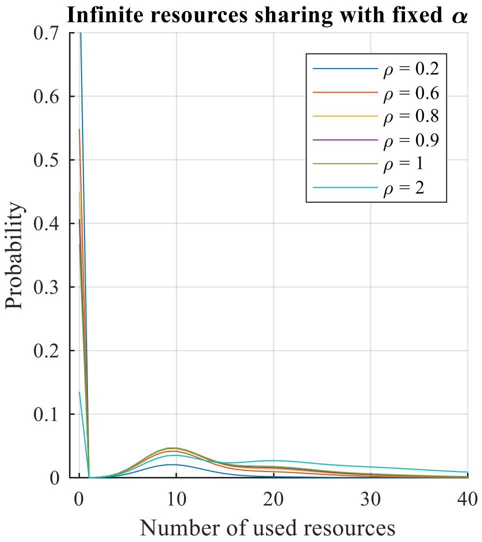

Fig. 4. Distribution of occupied resources for fixed α in an infinite resources sharing

There is a peak near the average number of resources requested by customers, and then the curves spread to the left and to the right. It indicates that the number of busy resources is concentrated around the average requested by the customers. Other small peaks with low probability at multiples of this mean are also observed, explaining the fact that several customers are saved at the same time by the server.

The probability of finding the server empty (in terms of the number of customers) decreases as ρ increases. The increase of ρ assumes that many customers are standing in the queueing area, either due to high service duration compared to inter-arrival, or to arrivals that are too frequent compared to service times. For infinite resources, ρ can exceed 1, and the number of occupied resources also tends towards a high quantity amount as shown by the cyan curve ($\rho = 2$) in Fig. 4.

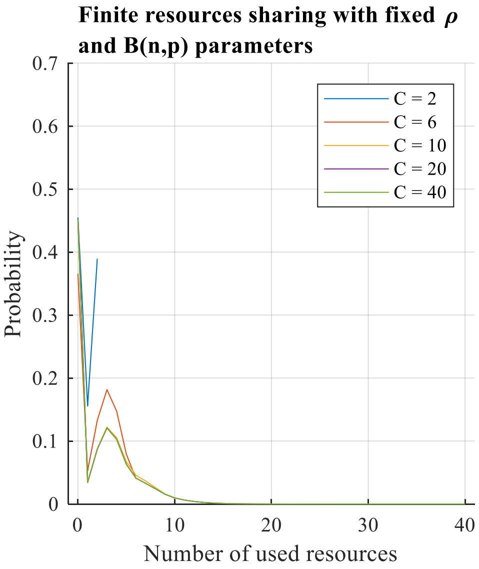

Fig. 5. Distribution of occupied resources for fixed ρ in a finite resources sharing

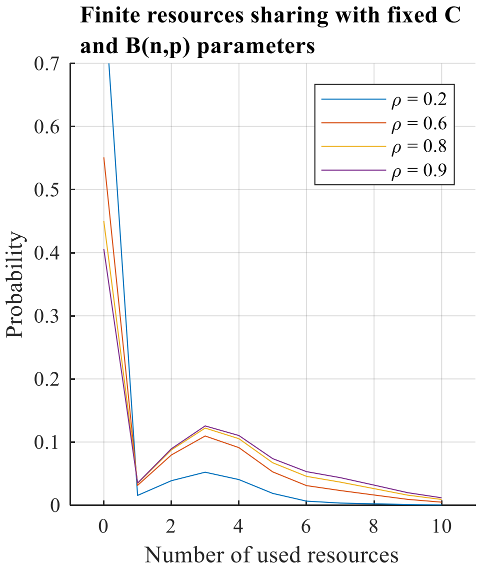

Fig. 6. Distribution of occupied resources for fixed C in a finite resources sharing

The higher the capacity, the more the system tends towards the previous infinite resource scenarios as we can see in Fig. 5. The probability of finding the system free (0 customer, 0 utilization) decreases as the number of resources is reduced (Fig. 5) or as the charging factor ρ increases (Fig. 6.). The notion of resource dimensioning begins to intervene from this step. We do not want to deploy resources that are not going to be used, or that will be under-used.

V. CONCLUSION

This paper focuses upon the study of occupied resources in queueing system whereby the customer’s arrival is Poisson process, they request service of duration exponential, and it requires a random amount of server resources. The server system has discrete resources and can share them to customers that request services to him. We proposed analytical model of the amount of the occupied resources by their probability distribution. We examined two cases of resources sharing: one with infinite amount of resources, and one with finite amount.

Special cases are also discussed regarding the individual resource usage following Poisson distribution and Binomial distribution. We validated our model using simulations on Matlab-Simulink.

The abacus presented in this paper can be used for resources dimensioning in case of Poisson or Binomial individual usage of discrete resource, but we can build more another abacus based on the formula that we validated.

In these proposed models, the server resources are discrete. In the future, we plan to build an analytical model of continuous resources sharing that we can also find in major cases of communication systems.

REFERENCES

[1] E. Daru, “Méthodes de résolution pour quelques problèmes de files d’attente comportant des serveurs d’efficacités différentes,” in Revue de la Recherche Opérationnelle, vol. 2, n°8, 3è trimestre 1958.

[2] L. Green, “A queueing system in which customers require a random number of servers”, in Operations Research, vol. 28, n°6, pp. 1335-1346, Nov-Dec 1980.

[3] E. L. Romm, V.V. Skitovitch, “On certain generalization of problem of Erlang”, in Automation and Remote Control, vol. 32, n°5, pp. 1000-1003, 1971.

[4] O.M. Tikhonenko, “Generalized Erlang problem for service systems with finite total capacity”, in Problems of Information Transmission, vol. 41, n°3, pp. 243-253, 2005.

[5] O.M. Tikhonenko, “Destricted Capacity Queueing Systems: Determination of their Characteristics”, in Automation and Remote Control, vol. 58, n°6, pp. 969-973, 1997.

[6] O.M. Tikhonenko, K.G. Klimovich, “Analysis of queuing systems for random-length arrivals with limited cumulative volume”, in Problems of Information Transmission, vol. 37, n°1, pp. 77-79, 2001.

[7] O.M. Tikhonenko, “Determination of Loss Characteristics in Queueing Systems with Demands of Random Space Requirement”, A. Dudin, A. Nazarov, R. Yakupov. (eds) Information Technologies and Mathematical Modelling – Queueing Theory and Applications, in Communications in Computer and Information Science, vol. 564, pp. 209-215, 2015.

[8] V. Naumov, K. Samouylov, “Analysis of multi-resource loss system with state dependent arrival and service rates”, in Probability in the Engineering and Informational Sciences, vol. 31, n°4, pp. 413-419, 2017.

[9] G.Y. Fletcher, H.G. Perros, W.J. Stewart, “A queueing system where customers require a random number of servers simultaneously”, in Computer Science Department, Center for Communications and Signal Processing North Carolina State University, Raleigh, NC 27695-8206, U.S.A., 2011.

[10] V.A. Naumov, K.E. Samuilov, A.K. Samuilov, “On the total amount of resources occupied by serviced customers”, in Automation and Remote Control, vol. 77, n°8, pp. 1419-1427, 2016.