total

and

$$E[R] = \frac{1}{\mu-\lambda}$$

We use the following two properties of an *M/M/1* queue:

1. From the PASTA property of a Poisson stream, an arriving customer sees the steady-state distribution of the number of customers in system.This arriving customer, with probability p_n, finds n customers already waiting or in service.

2. From the memory less property of the exponential distribution, if the new arrival finds a costumer in service, the remaining service time of that customer is distributed exponentially with mean $1/\mu$, i.e., identical to the service requirements of all waiting customers.

The response time of an arriving customer who finds n customers in the system is therefore the sum of $(n+1)$ exponentially distributed random variables,the n already present plus the n arriving customer itself. Such a sum has an(n+1) stage Erlange density function. Then, if B_k is the service time of customer k, we have

$$Prob\{R > t\} = Prob\left\{\sum_{k=1}^{n+1} B_k > t\right\}.$$

Now, conditioning on the number present when the arrival occurs an using the independence of arrivals and service, we obtain

$$\begin{aligned} Prob\{R > t\} &= \sum_{n=0}^{\infty}\left(Prob\left\{\sum_{k=1}^{n+1} B_k > t\right\}\right)p_n = \sum_{n=0}^{\infty}\left(e^{-\mu t}\sum_{k=0}^{n}\frac{(\mu t)^k}{k!}\right)(1-p)p^n \\ &= \sum_{k=0}^{n}\sum_{n=0}^{\infty}\left(e^{-\mu t}\frac{(\mu t)^k}{k!}\right)(1-p)p^n = \sum_{k=0}^{n}\left(e^{-\mu t}\frac{(\mu t)^k}{k!}\right)\sum_{n=k}^{\infty}(1-p)p^n \\ &= \sum_{k=0}^{n}\left(e^{-\mu t}\frac{(\mu t)^k}{k!}\right)p^k = e^{-\mu(1-p)t}, t \geq 0. \end{aligned}$$

Hence the probability distribution function for the response time in an $M/M/1$ queue is

$$\text{Prob}\{\text{R} \leq t\} = W_r(t) = 1 - e^{(\mu-\lambda)t},$$

i.e., the exponential distribution with mean

$$E[R] = \frac{1}{\mu-\lambda} = \frac{1}{\mu(1-p)}.$$

Notice that is not possible to write this performance measure just in terms of p. It depends on λ and μ but not just trough their ratio p.This means that it is possible to assign values to λ and μ in such s way that the system can be almost saturated with large queues but still have a very short expected response time.

The probability density function for the response time can be immediately found as the density function of the exponential distribution with parameter $\mu(1-p)$or evaluated directly as

$$w_r(t) = \sum_{n=0}^{\infty} P_n g_{n+1}(t) = (1-p)\mu e^{-\mu t}\sum_{n=0}^{\infty}\frac{(p\mu t)^n}{n!} = (\mu-\lambda)e^{-(\mu-\lambda)t},$$

410 Elementary Queueing Theory

and

$$E[R] = \frac{1}{\mu - \lambda}.$$

We use the following two properties of an *M/M/*1 queue:

1. From the PASTA property of a Poisson stream, an arriving customer sees the steady-state distribution of the number of customers in the system. This arriving customer, with probability p_n, finds n customers already waiting or in service.
2. From the memoryless property of the exponential distribution, if the new arrival finds a customer in service, the remaining service time of that customer is distributed exponentially with mean $1/\mu$, i.e., identical to the service requirements of all waiting customers.

The response time of an arriving customer who finds n customers in the system is therefore the sum of $(n+1)$ exponentially distributed random variables, the n already present plus the arriving customer itself. Such a sum has an $(n+1)$ stage Erlang density function. Then, if B_k is the service time of customer k, we have

$$\text{Prob}\{R > t\} = \text{Prob}\left\{\sum_{k=1}^{n+1} B_k > t\right\}.$$

Now, conditioning on the number present when the arrival occurs and using the independence of arrivals and service, we obtain

$$\begin{aligned}
\text{Prob}\{R > t\} &= \sum_{n=0}^{\infty}\left(\text{Prob}\left\{\sum_{k=1}^{n+1} B_k > t\right\}\right) p_n = \sum_{n=0}^{\infty}\left(e^{-\mu t}\sum_{k=0}^{n}\frac{(\mu t)^k}{k!}\right)(1-\rho)\rho^n \\
&= \sum_{k=0}^{n}\sum_{n=k}^{\infty}\left(e^{-\mu t}\frac{(\mu t)^k}{k!}\right)(1-\rho)\rho^n = \sum_{k=0}^{n}\left(e^{-\mu t}\frac{(\mu t)^k}{k!}\right)\sum_{n=k}^{\infty}(1-\rho)\rho^n \\
&= \sum_{k=0}^{n}\left(e^{-\mu t}\frac{(\mu t)^k}{k!}\right)\rho^k = e^{-\mu(1-\rho)t}, \quad t \ge 0.
\end{aligned}$$

Hence the probability distribution function for the response time in an *M/M/*1 queue is

$$\text{Prob}\{R \le t\} = W_r(t) = 1 - e^{-(\mu-\lambda)t},$$

i.e., the exponential distribution with mean

$$E[R] = \frac{1}{\mu - \lambda} = \frac{1}{\mu(1-\rho)}.$$

Notice that it is not possible to write this performance measure just in terms of ρ. It depends on λ and μ but not just through their ratio ρ. This means that it is possible to assign values to λ and μ in such a way that the system can be almost saturated with large queues but still have a very short expected response time.

The probability density function for the response time can be immediately found as the density function of the exponential distribution with parameter $\mu(1-\rho)$ or evaluated directly as

$$w_r(t) = \sum_{n=0}^{\infty} p_n g_{n+1}(t) = (1-\rho)\mu e^{-\mu t}\sum_{n=0}^{\infty}\frac{(\rho\mu t)^n}{n!} = (\mu-\lambda)e^{-(\mu-\lambda)t},$$

Agija Bebrisa REBC03 181REB307