

## Performance indices of an G/G/c queue

A web server receives jobs according to a Poisson process of rate  $\lambda = 10$  j/s. The duration of each job is distributed according to an Hyper-Exponential distribution, of rate  $\mu_1 = 50$  j/s and  $\mu_2 = 5$  j/s and  $p_1 = 0.8$ .

Compute:

1. The utilization of the system
2. The (exact) average response time
3. The (exact) average number of jobs in the system

After a year, the traffic increases and stabilizes: now it can be considered distributed according to a 5 stages Erlang distribution, with  $\lambda = 240$  j/s. To support this new scenario, two extra web servers are added, together with a load-balancer that holds request in a single queue, and dispatches them to the first available server. Assuming the time required by the load balancer to be negligible (i.e., the system can be modelled with a G/G/3 queue),

Compute:

1. The average utilization of the system
2. The approximate average response time
3. The approximate average number of jobs in the system