

M/M/1/K and M/M/2/K systems

A web server receives requests arriving according to a Poisson process of rate $\lambda = 150 \text{ req/min}$, and serves them with an average service time $D = 350 \text{ ms}$. The web server can handle a maximum of 32 requests at a time: if the system is full, new arrivals are discarded.

Determine:

1. The utilization of the system
2. The loss probability
3. The average number of jobs in the system
4. The drop rate
6. The average response time
7. The average time spent in the queue (waiting for service)

After 1 year, the load has increased to $\lambda = 250 \text{ req/min}$, making the current solution no longer applicable. The system administrator adds a second web server and a load balancer: requests enqueue at the load balancer, and then are sent to the first available webserver. Considering the communication time between load balancers and servers to be negligible compared to the service times, determine for this new configuration:

1. The total and average utilization of the system
2. The loss probability
3. The average number of jobs in the system
4. The drop rate
6. The average response time
7. The average time spent in the queue (waiting for service)