Faculty of Information Engineering, Informatics and Statistics
Department of Computer Science

Computer Network Performance

Author:
Simone Lidonnici

26 september 2025

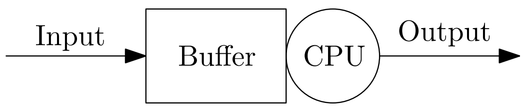
Contents

1

Introduction

1.1 Basic system

We start using a basic system composed by only one CPU and a queue to buffer the jobs that the CPU has to perform.



We define a bunch of variables:

- **CPU Capacity** (μ): expressed as how many jobs can the CPU perform per second on average.
- **Arrival rate** (λ): expressed as how many jobs arrive to the buffer per second on average.
- **CPU Service time (S)**: the mean time a job spends inside the CPU.
- **Waiting time (W)**: the mean time a job spends in the buffer.
- **Response time (R)**: the sum of service time and waiting time.
- **Throughput (X)**: expressed as how many jobs are performed by the system per second on average, measured in the output. The throughput of a system has value $0 < X < \mu$ and we have:
  - $X = \mu$ if the CPU is always running.
  - $X = \lambda$ if the rate is precise (very unlikely).
  - $X < \lambda$ if the buffer is full and some jobs are dropped.

Example:
If we have a system with $\lambda = 3$ j/s and $\mu = 5$ j/s, the service time will be:

$$S = \frac{1}{\mu} = \frac{1}{5} \text{ s} = 0.2 \text{ s}$$

The response time if the buffer is infinite is:

$$R = \frac{1}{\mu - \lambda} = \frac{1}{5 - 1} \text{ s} = 0.25 \text{ s}$$

1.2 System with multiple CPU

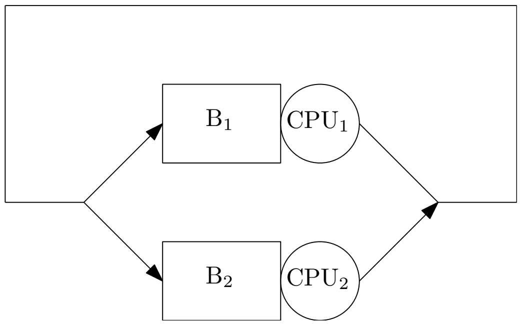
In the case of a system with multiple CPU (each with a personal buffer) we define the probability of a job going in a specific CPU i with p_i. The total response time of the system is calculated by the formula:

$$R = \sum_{i=1}^{n} p_i R_i$$

In which n is the number of CPU.

Example:

We have a closed system with 2 CPU like the following and a total of 6 jobs running in the system:



If $S_1 = S_2 = 3$ s the average number of jobs in every buffer is 2 (plus 1 in every CPU), so the response time of the system will be:

$$R = p_1 R_1 + p_2 R_2 = \frac{1}{2}(3 \cdot 3) + \frac{1}{2}(3 \cdot 3) = 9 \text{ s}$$

If we change $S_1 = 1.5$ s the average number of jobs in buffer 1 is 0 (1 in the CPU) and in buffer 2 is 4.5 (plus 1 in the CPU). This is caused by the fact that every job going to CPU 1 is completed faster than CPU 2 and the other jobs are stuck in buffer 2. So the average response time in this case is:

$$R = \frac{1}{2}(1 \cdot 1.5) + \frac{1}{2}(5.5 \cdot 3) = \frac{3}{4} + \frac{33}{4} = 9 \text{ s}$$

So we can see that upgrading only one CPU on a multi-CPU system doesn't change the response time. To lower the response time the routing probabilities must be changed as well.