Some of the answers to these questions can be found using the lecture slides, the recommended textbooks or other sources (question marked with *). Some questions may have more than one possible answer, or be more or less open for discussion. Note that no answers (solutions) will be given to these questions, but if help is needed the assistants will be available to answer questions. The concepts marked with yellow are important and should be fully understood.

OS Scheduling

1. Name and explain several scheduling optimization criteria.
2. What is the average waiting time and how can this be computed? Give an example using a Gantt chart as a visual aid.
3. Consider the following set of processes, with the length of the CPU burst given in milliseconds:

Process	**Burst time**	**Priority**
P_1	10	3
P_2	1	1
P_3	2	3
P_4	1	4
P_5	5	2

   The processes are assumed to have arrived in the order P_1, P_2, P_3, P_4, P_5 all at time 0.

   (a) Draw four Gantt charts that illustrate the execution of these processes using the following scheduling algorithms: FCFS, SJF, non-preemptive priority (smaller priority number implies a higher priority), and RR (quantum=1).

   (b) What is the turnaround time of each process for each of the scheduling algorithms in the previous subproblem?

   (c) What is the waiting time of each process for each of these scheduling algorithms?

   (d) Which of the algorithms results in the minimum average waiting time (over all processes)?

4. A preemptive version of shortest job first is

   (a) Shortest remaining time next

   (b) Three-level scheduling

   (c) First-come first served

   (d) Round-robin scheduling

(e) None of the above

5. Discuss how the following pairs of scheduling criteria conflict in certain settings.

   (a) CPU utilization and response time

   (b) Average turnaround time and waiting time

   (c) I/O device utilization and CPU utilization

6. Consider the exponential average formula used to predict the length of the next CPU burst. What are the implications of assigning the following values to the parameters used by the algorithm?

   (a) $\alpha = 0$ and $\tau_0 = 100$ miliseconds

   (b) $\alpha = 0.99$ and $\tau_0 = 10$ miliseconds

7. Consider the fairness metric proposed in `http://www.cs.wustl.edu/~jain/papers/ftp/fairness.pdf`:

$$f(x) = \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n \sum_{i=1}^{n} x_i^2}, x_i \geq 0 \tag{1}$$

   The formula given in equation 1 describes the fairness index. It is applicable to resource sharing or allocation problems among n requesting entities. The index computed by $f(x)$ is in the range $[0, 1]$. A fairness index of 0.1 implies unfairness to 90% of the users in a given context.

   (a) Imagine there are 10 resources given to 10 processes. We assume that each process is assigned to exactly one resource. Calculate the fairness index for this example.

   (b) Now, imagine there is a priority concept, i.e., the first two out of ten processes get 5 resources each and the remaining processes get 0 resources. Calculate the fairness index for this situation.

   (c) Discuss the results. How could such a fairness metric help to improve an OS?