CS425: Computer Systems Architecture

Homework Problem Set 1
Assignment Date: Friday 31/10/2025
Due Date: Monday 10/11/2025 23:59:59

Instructions: Solve all problems, create a .pdf file and send it via e-mail to HY425 course e-mail (hy425@csd.uoc.gr). Set the e-mail subject: HY425 - Homework 1

Problem 1 (35 points)

Assume that you run four applications (X, Y, Z, W) on a 20-core processor system, given the application characteristics provided in the table below:

Application	X	Y	Z	W
% resources need	40	25	20	15
% parallelizable	50	80	60	90

i. Calculate and plot on a graph the speedup achieved when running each application in isolation on 2, 4, 8, 12, 16, 20 cores compared to running it serially. Show your calculations, the formulas you used, and comment on the results.
ii. Application X needs 40% of the resources, so we assign it 40% of the cores. Find the overall speedup if X runs in parallel but remaining applications run serially. State your assumptions, show your calculations, the formulas you used, and comment on the results.
iii. Propose a core allocation strategy that minimizes completion time when all applications run in parallel. State your assumptions, show your calculations, the formulas you used, and comment on the results.

Problem 2 (35 points)

The servers used by the major cloud providers provide adequate compute capacity to sustain the highest load demand. However, these servers operate at only 40% of their peak while the power comsumption does not scale linearly with the load. One solution would be to turn off the servers, but the problem would be long restart times in response to larger load demands. You propose a new system that does quick restart while waiting in "Active Idle" state.

Go to the SPECpower_ssj2008 website and select a system with Intel Xeon or AMD EPYC processors with published results from Q3'2025. Indicate your specific system selection and present in tables the numbers that you will use. Answer the following questions and make sure that you clearly state your assumptions, show your calculations, the formulas you used, and comment on the results.

i. Calculate the power savings in case of turning off 50% of the servers.
ii. Calculate the power savings when 50% of the servers are in the "quick restart" state.
iii. Calculate the power savings if the voltage is reduced by 20% and frequency is reduced by 40%.
iv. Calculate the power savings if 25% of the servers are in the "quick restart" state and 25% of them are completely off.

Problem 3 (30 points)

Apart from scalability and throughput, availability is a key consideration when designing servers.

i. Your institution has a processor chip which exhibits a Failure In Time of 1000. What is the Mean Time To Failure for this system?

ii. If engineers need two days to fix and get the system running again, find the system's availability.

iii. The institution wants to build a supercomputer with 4096 nodes using these chips, and assigned to you the MTTF estimation of the following two scenarios:

- **a.** To cut down costs, the institution will deploy unreliable computer units to set up the supercomputer, meaning that if one fails, they all fail.
- **b.** The institution will use reliable computer units to build the supercomputer. In this case, the system will experience catastrophic failure only if 128 units fail.

Note: Consider the worst case scenario of simultaneous failures.

Which scenario would you recommend if the extra upfront cost of the second scenario is 3x compared to the first one? State your assumptions, show your calculations, the formulas you used, and comment on the results.