Homework 1

Problem 1 In this exercise, assume that we are considering enhancing a quad-core machine by adding encryption hardware to it. When computing encryption operations, it is 30 times faster than the normal mode of execution. We will define percentage of encryption as the percentage of time in the original execution that is spent performing encryption operations. The specialized hardware increases power consumption by 4%.

a. Draw a graph that plots the speedup as a percentage of the computation spent performing encryption. Label the y-axis "**Net speedup"** and label the x-axis "**Percent encryption**."

b. With what percentage of encryption will adding encryption hardware result in a speedup of 3?

c. What percentage of time in the new execution will be spent on encryption operations if a speedup of 3 is achieved?

d. Suppose you have measured the percentage of encryption to be 50%. The hardware design group estimates it can speed up the encryption hardware even more with significant additional investment. You wonder whether adding a second unit in order to support parallel encryption operations would be more useful. Imagine that in the original program, 90% of the encryption operations could be performed in parallel. What is the speedup of providing **three or six** encryption units, assuming that the parallelization allowed is limited to the number of encryption units? (Answer **three and six** respectively)

Problem 2 Server farms such as Google and Yahoo! provide enough compute capacity for the highest request rate of the day. Imagine that most of the time these servers operate at only 60% capacity. Assume further that the power does not scale linearly with the load; that is, when the servers are operating at 60% capacity, they consume 90% of maximum power. The servers could be turned off, but they would take too long to restart in response to more load. A new system has been proposed that allows for a quick restart but requires 15% of the maximum power while in this"barely alive"state.

a. How much power savings would be achieved by turning off 50% of the servers?

b. How much power savings would be achieved by placing 50% of the servers in the "barely alive" state?

c. How much power savings would be achieved by reducing the voltage by 30% and frequency by 50%?

d. How much power savings would be achieved by placing 40% of the servers in the "barely alive" state and 20% off?

Problem 3 Availability is the most important consideration for designing servers, followed closely by scalability and throughput.

a. We have a single processor with a failure in time (FIT) of 10. What is the mean time to failure (MTTF) for this system?

b. If it takes two days to get the system running again, what is the availability of the system?

c. Imagine that the government, to cut costs, is going to build a supercomputer out of inexpensive computers rather than expensive, reliable computers. What is the MTTF for a system with 100 processors? Assume that if one fails, they all fail.

Problem 4 They will sell a range of chips from that factory, and they need to decide how much capacity to dedicate to each chip. Imagine that they will sell two chips. Phoenix is a completely new architecture designed with 7 nm technology in mind, whereas RedDragon is the same architecture as their 10 nm Blue Dragon. Imagine that RedDragon will make a profit of $15 per defect-free chip. Phoenix will make a profit of $20 per defect-free chip. Each wafer has a 450 mm diameter.

Chip	Die Size (mm2)	Estimated defect rate (per cm2)	N	Manufacturing size (nm)	Cores
$BlueDragon$	225	0.04	14	10	4
$RedDragon$	150	0.04	14	7	4
$Phoenix^8$	180	0.03	12	7	8

a. How much profit do you make on each wafer of Phoenix chips?

b. How much profit do you make on each wafer of RedDragon chips?

c. If your demand is 30,000 RedDragon chips per month and 40,000 Phoenix chips per month, and your facility can fabricate 70 wafers a month, how many wafers should you make of each chip?

Problem 5 Your company has just bought a new 32-core processor, and you have been tasked with optimizing your software for this processor. You will run four applications on this system, but the resource requirements are not equal. Assume the system and application characteristics listed in Table 1.1.

Table 1.1 Four applications

Application	A	B	C	D
% resources needed	40	25	15	20
% parallelizable	50	70	80	60

The percentage of resources of assuming they are all run in serial. Assume that when you parallelize a portion of the program by X, the speedup for that portion is X.

a. How much speedup would result from running application A on the entire 32-core processor, as compared to running it serially?

b. How much speedup would result from running application D on the entire 32-core processor, as compared to running it serially?

c. Given that application A requires 40% of the resources, if we statically assign it 25% of the cores, what is the overall speedup if A is run parallelized but everything else is run serially?

d. What is the overall speedup if all four applications are statically assigned some of the cores, relative to their percentage of resource needs, and all run parallelized?