CS350 Assignment 1

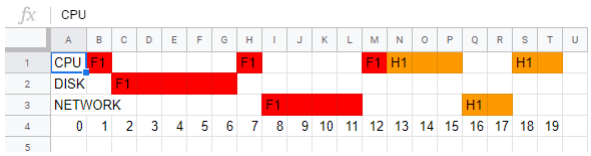
Chris-Emio (chrisr98@bu.edu)

Feb 11, 2020

Written Part

Problem 1

a)

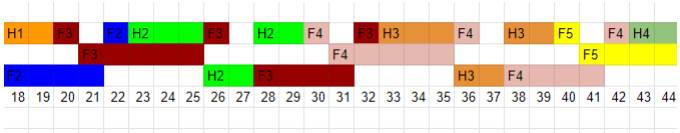


CPU $=\frac{8}{19}$, DISK $=\frac{5}{19}$, NETWORK $=\frac{6}{19}$.

b) 1 request every 19 time units $= 0.053\ \frac{req}{tu}$

c) 1 request every 19 time units $= 0.053\ \frac{req}{tu}$

d)



CPU $=\frac{8}{10}$, DISK $=\frac{1}{2}$, NETWORK $=\frac{6}{10}$.

e) 2 request every 10 time units $= 0.2\ \frac{req}{tu}$

f) The system would reach its max capacity when the CPU is utilized 100%. This happens when the steady state value of T=8s. DISK $= \frac{5}{8}$, and NETWORK$= \frac{6}{8}$. The total throughput of the web server is then $\frac{2}{8}$

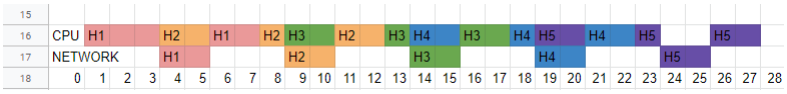
Problem 2

a) Total execution for HTTP request = 7ms. The parallelizable part of the utilization : f $= \frac{3}{7} = 0.4285$

According to Amdahl's Law: Speedup $= 1.4 = \frac{1}{(1-f)+(\frac{f}{x})} = \frac{1}{(1-0.43)+(\frac{0.43}{x})}$. So if we solve for x it equals to x=2.98. Therefore, the minimum number of CPUs required is 3.

b) If x is an arbitrary number, then it goes to infinity. Making $(\frac{f}{x}) = 0$. Therefore speedup would be $\frac{1}{(1-0.43)} = 1.75$. Which means the handling time cannot be halved with an arbitrary number of CPUs.

c)



The response time for H5 is 27ms

Problem 3

Latency of the main memory = 50%

Latenct if cache memory = 20%

a) Cache hit rate is 95%, access time of cache is 1tu, and access time of main memory is 4tu.

Expected access time can be calculated as:

$$\{\text{cache hit rate} \times \text{cache access time} + ((\text{1-cache hit rate}) \times \text{main memory access time})\}$$

$0.95 \times 1tu + ((1 - 0.95) \times 4tu) = 0.95tu + 0.2tu = 1.15tu$

After main memory optimization it would be:

$$\{\text{cache hit rate} \times \text{cache access time} + ((\text{1-cache hit rate}) \times \text{Latency of main memory})\}$$

$0.95 \times 1tu + ((1 - 0.95) \times 0.5tu) = 0.95tu + 0.025tu = 0.975tu$

The speed up would equal $\frac{original\ access\ time}{access\ time\ after\ main\ memory\ optimization}$

Speedup $= \frac{1.15}{0.975} = 1.1794$

After cache memory optimization it would be:

$$\{\text{cache hit rate} \times \text{latency of cache} + ((\text{1-cache hit rate}) \times \text{main memory access time})\}$$

$0.95 \times 0.80tu + ((1 - 0.95) \times 4tu) = 0.76tu + 0.2tu = 0.96tu$

The speed up would equal $\frac{original\ access\ time}{access\ time\ after\ cache\ memory\ optimization}$

Speedup $= \frac{1.15}{0.6} = 1.1979$

b) I would consider to lower the cache hit rate if there is a lot of data being written on the memory to avoid having to access main memory with longer latency. Although the difference between the 2 optimization isn't that high

c)

After both optimization it would be:

$$\{\text{cache hit rate} \times \text{latency of cache} + ((\text{1-cache hit rate}) \times \text{Latency of main memory})\}$$

$0.95 \times 0.80tu + ((1 - 0.95) \times 0.5tu) = 0.76tu + 0.025tu = 0.785tu$

The speed up would equal $\frac{original\ access\ time}{access\ time\ after\ cache\ memory\ optimization}$

Speedup $= \frac{1.15}{0.785} = 1.4649$