Computer System Fundamentals HW #1

Quan Zhou

Feb 1st, 2016

Problem 1

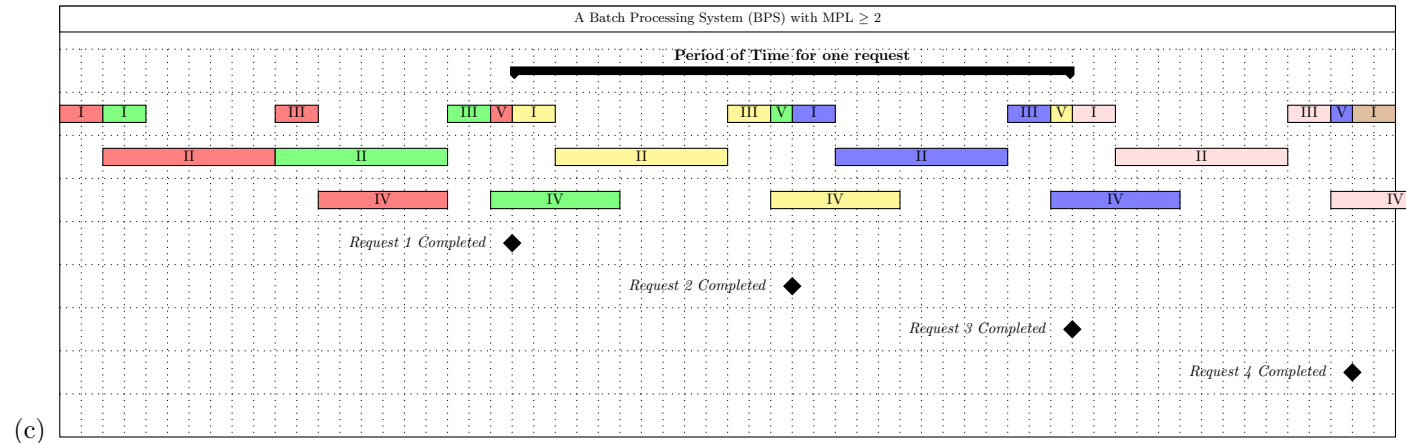A Batch Processing System (BPS) with MPL = 1

(a)
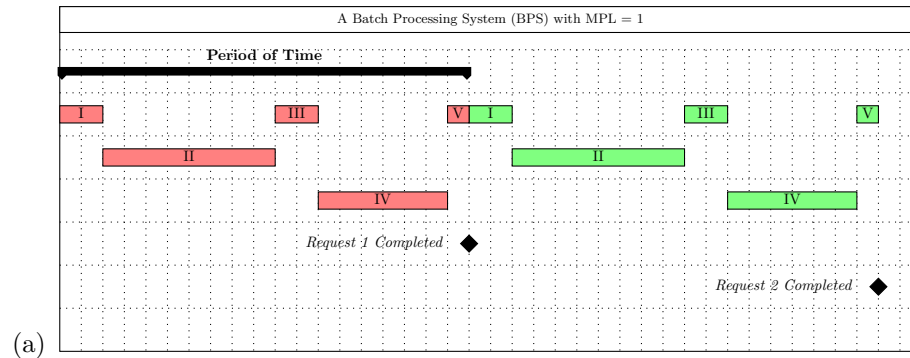
From the above chart, we see that the usage for CPU, disk, and network are 5 msec, 8 msec and 6 msec respectively. The total span of time to process this one request is 19 msec. Therefore, utilization of each resources are $\frac{5}{19}$, $\frac{8}{19}$, $\frac{6}{19}$, or 26.3%, 42.1%, 31.6%.

(b) $Bandwidth = \frac{1}{19}$ or 52.6 requests/sec

A Batch Processing System (BPS) with MPL ≥ 2

(c)

For N = 2: We assume that the more progressed requests are prioritized to the less progressed requests. From the above chart, we see that the usage for CPU, disk, and network are 10 msec, 16 msec and 12 msec respectively. The period of time to complete one process when N = 2 is 26 msec. Utilization of each resources can be found by following the definition and we get $\frac{10}{26}$, $\frac{16}{26}$, $\frac{12}{26}$, or 38.5%, 61.5%, 46.2%.

(d) $Bandwidth = \frac{2}{26}$ or 76.9 requests/sec

1

(e)
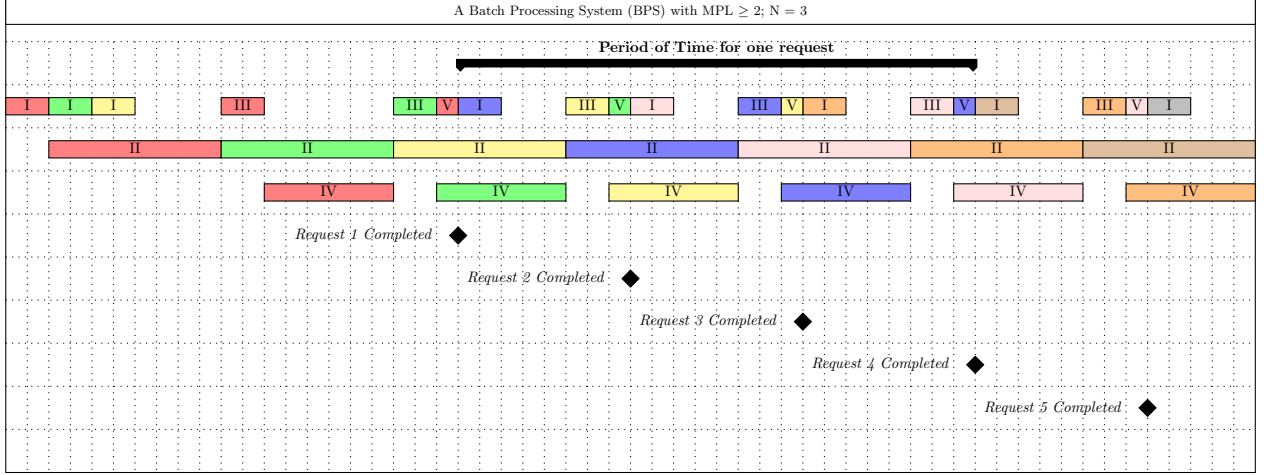
From the above chart, we see that the usage for CPU, disk, and network are 15 msec, 24 msec and 18 msec respectively in a total span of 24 msec. Utilization of each resources can be found by following the definition and we get $\frac{15}{24}$, $\frac{24}{24}$, $\frac{18}{24}$, or 62.5%, 100%, 75%.

$Bandwidth = \frac{3}{24}$ or 125 requests/sec

(f) We know that the disk has hit 100% utilization, then the maxx capacity is equal to the bandwidth found in part (e) which is : 125 requests/sec

(g) The bottleneck resource is the disk: as N increases, the utilization of the disk would first approach to 100%. This limits the throughput of this system.

Problem 2

(a) $F_{CPU} = \frac{5msec}{19msec} = \frac{5}{19}$, and $S = r$
Using Amdahl's Law, we have:
System Overall Speedup $= \frac{1}{1 - \frac{5}{19} + \frac{5}{19*r}} = \frac{1}{\frac{14}{19} + \frac{5}{19*r}}$
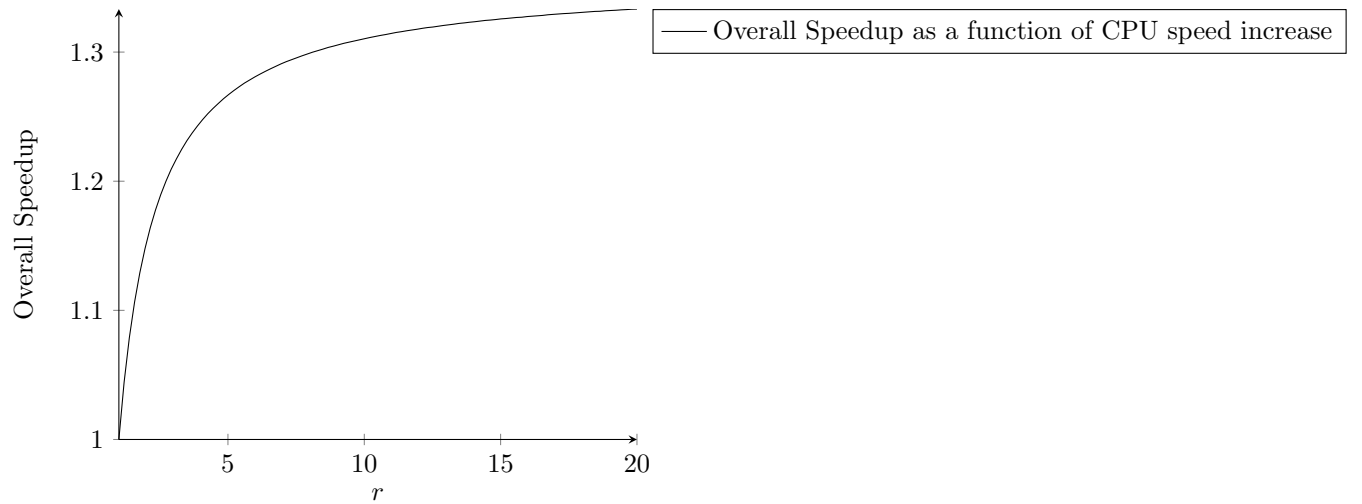
(b) Assume CPU can infinitely fast, then the value of r increases to ∞. From the function from part (a), we are left with Overall Speedup $= \frac{1}{\frac{14}{19}} = \frac{19}{14}$ or 1.36. And that is our theoretical limit.

(c) $F_{disk} = \frac{8msec}{19msec} = \frac{8}{19}$, and $S = r$

Using Amdahl's Law, we have:

System Overall Speedup (disk) $= \frac{1}{1 - \frac{8}{19} + \frac{8}{19*r}} = \frac{1}{\frac{11}{19} + \frac{8}{19*r}}$

Assume the disk can infinitely fast, then the value of r increases to ∞. From the equation above, we can find the theoretical limit for the speedup is $= \frac{1}{\frac{11}{19}} = \frac{19}{11}$ or 1.73. And that is our theoretical limit.

$F_{network} = \frac{6msec}{19msec} = \frac{5}{19}$, and $S = r$

Using Amdahl's Law, we have:

System Overall Speedup (network) $= \frac{1}{1 - \frac{6}{19} + \frac{6}{19*r}} = \frac{1}{\frac{13}{19} + \frac{6}{19*r}}$

Assume the network can infinitely fast, then the value of r increases to ∞. From the equation above, we can find the theoretical limit for the speedup is $= \frac{1}{\frac{13}{19}} = \frac{19}{13}$ or 1.46. And that is our theoretical limit.
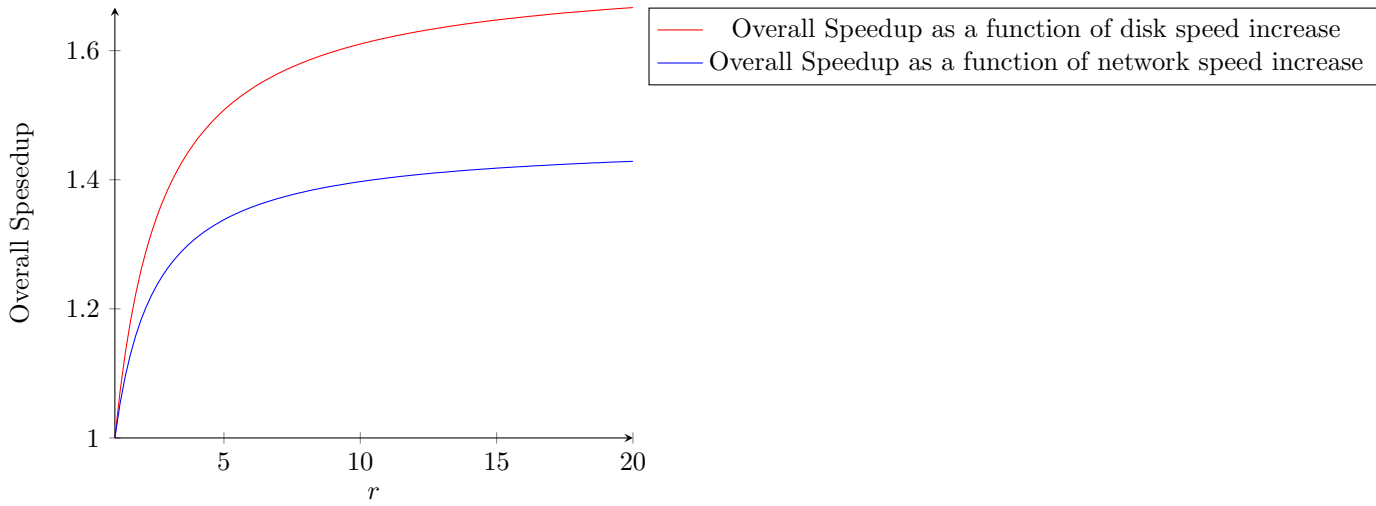
(d) The minimum speedup for the various device to achieve equal utilization of all resources is to increase the speed of the disk, which has the higher utilization among the three resources of this web server system. Current utilization of the disk is 8 msec for every request; if we speed up the disk to 5 msec for every request, and likewise speed up the network also to 5 msec so the utilization for all three resources are equal, then we could get an overall speedup of 1.19 and 1.05 respectively.

(e) Since all three resources/devices have equal utilization, the resulting capacity is the max throughput of any device and therefore equals to $\frac{1}{5}$ or 200 requests/sec

Problem 3

(a) I. cut the latency of the main memory by 1/2
II. cut the latency of the cache by 1/5

The expected execution time without improvement is:

$$\mathbb{E}(T) = \mathbb{P}(X = cachehit) \cdot (T_{hit}) + \mathbb{P}(X = cachemiss) \cdot (T_{miss})$$
$$= (95\%) \cdot (1) + (5\%) \cdot (8)$$
$$= 1.35 \text{ units of time}$$

Now, with improvement I, the execution time is:

$$\mathbb{E}(T_I) = (95\%) \cdot (1) + (5\%) \cdot (4)$$
$$= 0.95 + 0.20$$
$$= 1.15 \text{ units of time}$$

therefore achieving a speedup of $\frac{1.35}{1.15} = 1.17$
versus, with improvement II:

$$\mathbb{E}(T_{II}) = (95\%) \cdot (0.8) + (5\%) \cdot (8)$$
$$= 0.76 + 0.40$$
$$= 1.16 \text{ units of time}$$

therefore achieving a speedup of $\frac{1.35}{1.16} = 1.16$

(b) Since the choice to optimized the cache or main memory may change as the cache hit rate varies, we have the follow equation:
Given hit rate = h, and therefore miss rate = 1 - h, we have: $h \cdot (1) + (1 - h) \cdot (4) = h \cdot (0.8) + (1 - h) \cdot (8)$
$4.2 \cdot h = 4$ so h $= \frac{20}{21}$ this number means that if the cache hit rate is less than $\frac{20}{21}$, cutting the latency of the main memory (improvement I) would be a better choice whereas if the cache hit rate is more than $\frac{20}{21}$, improving the cache latency (improvement II) would be a better choice.

(c)

$$Speedup = \frac{\mathbb{E}(T)}{\mathbb{E}(T_{I+II})}$$
$$= \frac{h \cdot (1) + (1 - h) \cdot (8)}{h \cdot (0.8) + (1 - h) \cdot (4)}$$
$$= \frac{8 - 7h}{4 - 3.2h}$$

Problem 4

(a) Since the percentage of payload for each packet is $\frac{150}{1500}$ or 10%, the we have:
Throughput = Capacity $\times (1 - 10\%) \times (1 - P)^H = (10 \text{ Mbps}) \cdot (90\%) \cdot (1 - P)^H = 9(1 - P)^H$ Mbps

(b) For H = 1 and P = 0.01, we substitue H and P into the equation above and we obtain:
$9(1 - P)^H = 9 * 0.99^1 = 8.91$ Mbps

(c) To achieve a bandwidth of more than 6 Mbps:
$9(1 - P)^H > 6$
$(1 - P)^H > \frac{2}{3}$
$1 - P > (\frac{2}{3})^{\frac{1}{H}}$ so $P < 1 - (\frac{2}{3})^{\frac{1}{H}}$

Problem 5

(a) There are four distinctive paths between the source "A" and the destination "B" (see Figure 1).
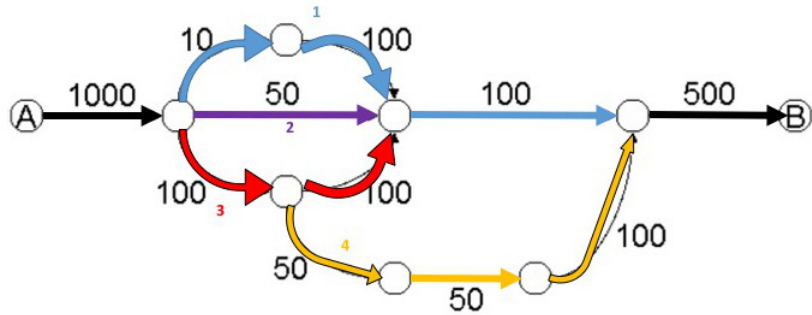
Figure 1. Feasible Paths between A & B in a general Network

By following the algorithm given in the problem we first examine the path colored in blue (see Figure 2)
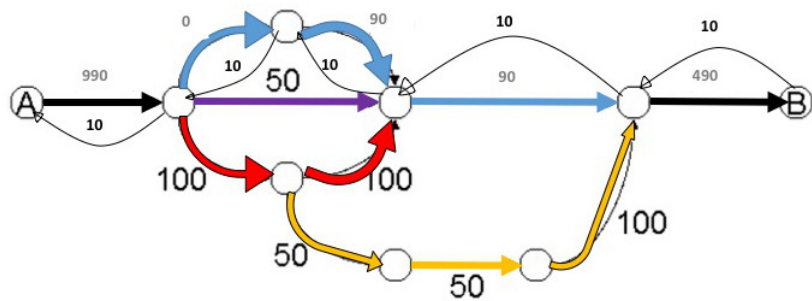

Figure 2. Blue Path from A & B in a residual graph

The bottleneck flow in the blue path is 10 and therefore the flow capacity is 10. Since there are still augmenting paths to be added to the residual graph, we go to the next path.
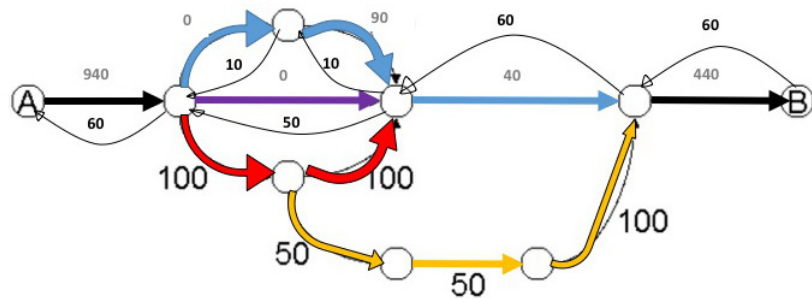

Figure 3. Blue Path adding Purple Path from A & B in a residual graph

Since the minimum flow in the purple path is 50, so the flow capacity now becomes 60 (after adding 50 from the purple path). Since there are still augmenting paths to be added to the residual graph, we go to the next path.
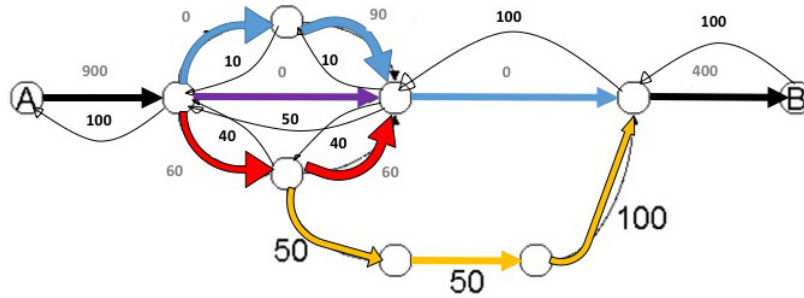
Figure 4. Blue Path adding Purple Path adding Red Path a from A & B in a residual graph

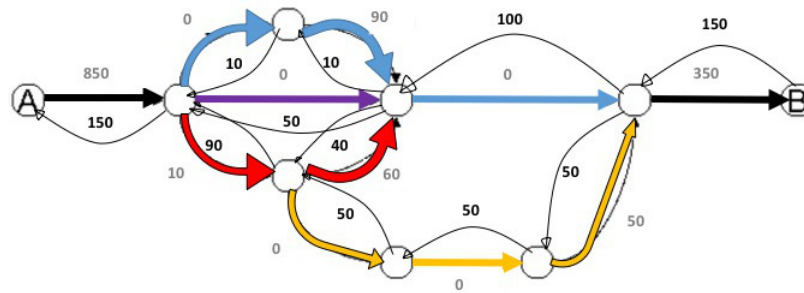We can add another 40 to the flow capacity from the red path.

Figure 5. Complete Residual Graph for the Generic Network

Lastly, the yellow path could contribute another 50, tallying up to 150 as the max-flow in this particular network. By following the algorithm, there are two other possible residual graphs and they are shown below. In Figure 6, the blue path was not utilized at all because there is no other augmenting path available.
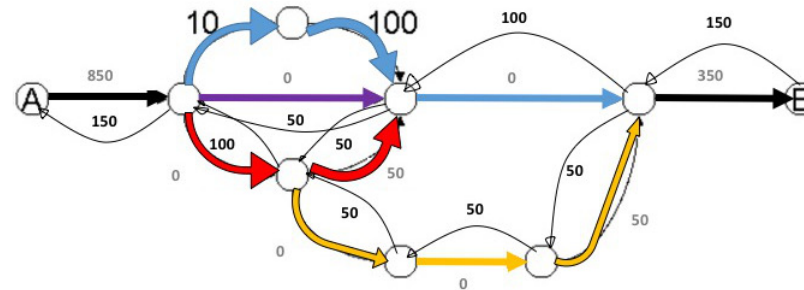

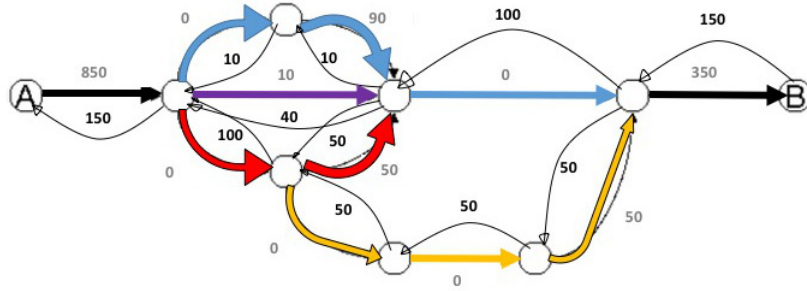Figure 6. Residual Graph B for the Generic Network

Figure 7. Residual Graph C for the Generic Network

All of the residual graphs indicate that the maximum capacity is 150 for this network.

(b) Given a residual graph (for example in Figure 5), we look at each feasible path that cannot be augmented: For each edge (u,v) across the path that has residual flow $f' = 0$, the maximum amount of extra flow that is allowed is given by the minimum residual flow in each path from s to u and the path from v to t. In case of the minimum residual flow being zero, then there is no single edge change that could increase the max-flow in the network. In Figure 5, we first look at the blue path: there are two edges that have 0 residual flow. Unless both edges flow capacities are increased, there is no change in the max flow in the blue path (as the max stays at the level of 10); Then the purple path also has two edges that have 0 residual flow. Same logic and we move on to the red path. Notice in this path, only one edge has zero residual flow and the minimum residual flow before and after this edge is 10. So we could increase the edge (where 3 different paths merges) by 10 so as to increase the max flow of this network. Lastly, the yellow path has three different edges that have 0 residual flow and therefore no single edge increase would improve the max flow of the network.
Therefore, by upgrading the edge (where 3 different paths merges) by 10 we would increase the capacity of the network by 10.

(c) 10 units higher. So the capacity of the network after the update is 160, which makes sense because the flow capacities of the three input edges (after A) add up to $160 = 10 + 100 + 50$.

(d) I would choose the link/edge found in part (b)-(c) because it is the bottleneck edge in this network. Without the upgrade, 100 out of 150 capacity is disabled by removing this particular edge. The other edge which is two nodes away from B and has a capacity of 50 is the other bottleneck edge in this network and removing it does not produced as much damage on the performance of the network as this edge.