Transaction Performance

COMP755 Advanced OS

Goals

- Understand the process we are using to estimate performance
- Add queuing theory to our performance model

The Process

1. Convert to common time units
2. Calculate the visitation ratios, V_i
3. Calculate the device utilization, ρ_i
4. Calculate the CPU service time
5. Calculate the device time
6. Find the bottleneck device
7. Calculate the maximum transaction rate
8. Calculate the average transaction time

Example

device	data/ hour	lambda	S	V	rho	V*S
web pages	9,000					
CPU					42%	
disk1	108,000		11 ms			
disk2	72,000		16 ms			
network	18,000		23 ms			

Convert to common time units

- Usually seconds is the best time unit
- Divide the reported counts by the number of seconds in the reporting period. This gives you the arrival rate, λ_i, for each device
- Make sure the device service times are in seconds

Example

device	data/ hour	lambda	S	V	rho	V*S
web pages	9,000	2.5				
CPU					42%	
disk1	108,000	30.0	0.0110			
disk2	72,000	20.0	0.0160			
network	18,000	5.0	0.0230			

Calculate the visitation ratios

- The visitation ratio is the number of times a device is used for each transaction
- Divide the device use count by the number of transactions to get the visitation ratio, V_i
- The visitation ratio of the CPU is the sum of all other visitation ratios

Example

device	data/ hour	lambda	S	V	rho	V*S
web pages	9,000	2.5		1		
CPU	207,000	57.5		23	42%	
disk1	108,000	30.0	0.0110	12		
disk2	72,000	20.0	0.0160	8		
network	18,000	5.0	0.0230	2		

Calculate the device utilization

- Remember that $\rho = \lambda * s$
- Utilization, ρ, can be calculated by multiplying the per second arrival rate for each device by its service time

Example

device	data/ hour	lambda	S	V	rho	V*S
web pages	9,000	2.5		1		
CPU	207,000	57.5		23	42%	
disk1	108,000	30.0	0.0110	12	0.33	
disk2	72,000	20.0	0.0160	8	0.32	
network	18,000	5.0	0.0230	2	0.115	

Calculate the CPU service time

- Typically we measure the CPU utilization and not the average service time
- To calculate the service time of the CPU

$$s = \rho / \lambda$$

Example

device	data/ hour	lambda	S	V	rho	V*S
web pages	9,000	2.5		1		
CPU	207,000	57.5	0.0073	23	42%	
disk1	108,000	30.0	0.0110	12	0.33	
disk2	72,000	20.0	0.0160	8	0.32	
network	18,000	5.0	0.0230	2	0.115	

Calculate the device time

- On the average, each transaction uses a device V_i time. Each device access takes S_i time on average.
- The time used by a device for each transaction is $V_i * S_i$

Example

device	data/ hour	lambda	S	V	rho	V*S
web pages	9,000	2.5		1		
CPU	207,000	57.5	0.0073	23	42%	0.168
disk1	108,000	30.0	0.0110	12	0.33	0.132
disk2	72,000	20.0	0.0160	8	0.32	0.128
network	18,000	5.0	0.0230	2	0.115	0.046

Find the bottleneck device

- The bottleneck device is the busiest. As the load increases, this device will limit the maximum throughput
- The device with the largest utilization is the bottleneck device

Example

device	data/ hour	lambda	S	V	rho	V*S
web pages	9,000	2.5		1		
CPU	207,000	57.5	0.0073	23	**42%**	0.168
disk1	108,000	30.0	0.0110	12	0.33	0.132
disk2	72,000	20.0	0.0160	8	0.32	0.128
network	18,000	5.0	0.0230	2	0.115	0.046

Calculate maximum transaction rate

- Using the Visitation Ratio and service time of the bottleneck device, i

$$MaxTrans = \frac{1}{S_i * V_i}$$

Example

device	data/ hour	lambda	S	V	rho	V*S
web pages	9,000	2.5		1		
CPU	207,000	57.5	0.0073	23	**42%**	0.168
disk1	108,000	30.0	0.0110	12	0.33	0.132
disk2	72,000	20.0	0.0160	8	0.32	0.128
network	18,000	5.0	0.0230	2	0.115	0.046

Maximum transaction rate = 1/V*S = 1 / 0.168 = 5.95

Calculate the transaction time

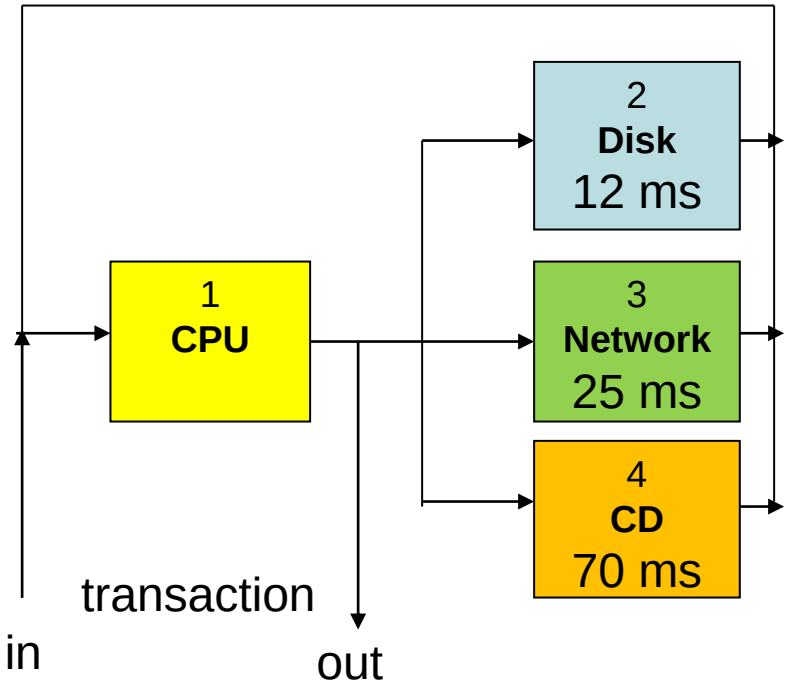
- The time spent by a transaction is the sum of the time spent at each device for that transaction.
- Sum $V_i * S_i$ for all devices

Example

device	data/ hour	lambda	S	V	rho	V*S
web pages	9,000	2.5		1		
CPU	207,000	57.5	0.0073	23	42%	0.168
disk1	108,000	30.0	0.0110	12	0.33	0.132
disk2	72,000	20.0	0.0160	8	0.32	0.128
network	18,000	5.0	0.0230	2	0.115	0.046

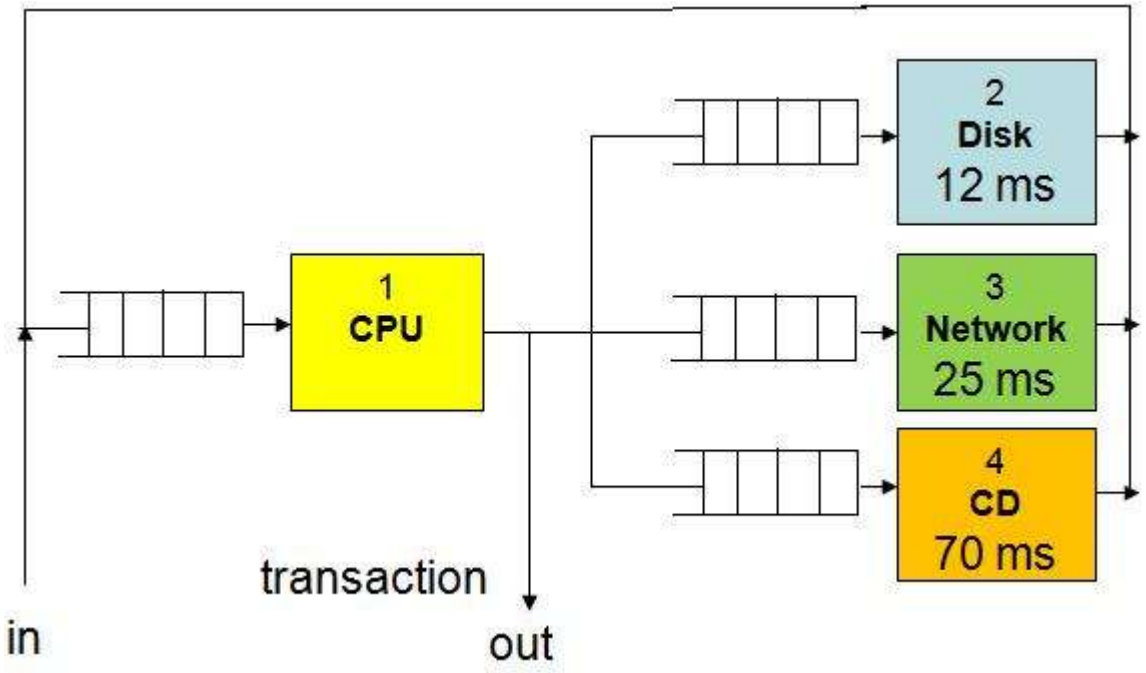
Average transaction time = sum of V*S = 0.474 seconds

Example System



Transactions flow around the system visiting the devices

Example System with Queues



Transactions flow around the system and are queued if the device is busy

Similarities

- The arrival rates, λ, the service times, the visitation ratios and the utilization are all the same.

Device Times

- Assume the arrival rate and the service times are exponentially distributed
- We can calculate Tq for each device, including the CPU
- Tq represents the amount of time it takes for a transaction to use that device, including any wait time

Time Spent at Each Device

- Previously we used $V_i * S_i$ to calculate the total time used by a transaction at a device
- With queuing, we should use $V_i * Tq_i$

Program Saturation

- Using the device that has the highest utilization, the bottleneck device, we can calculate the maximum number of transactions the system can perform.

$$\lambda_0 = \frac{1}{S_i * V_i}$$

Minimum Execution Time

- You can calculate a minimum execution time.

$$\min TransTime = \sum V_i * Tq_i$$