LTAT.06.007 Distributed Systems

Seminar 4 - Processes I

Mohan Liyanage, PhD

Tartu, Estonia May 12, 2020
Quiz 4

Questions are based on the third lecture
Short answers

Link to quiz - https://tinyurl.com/swdqfko
Quiz will be available until March-06-2020:23.59
Three reasons to use threads in programs.

Avoid needless blocking
   - block when doing I/O

Exploit parallelism
   - process can be scheduled to run in parallel

Avoid process switching
   - easy to swatch context between threads
Advantages of using the user-level thread...

Cheap to create and destroy threads
Easy to switching thread context
What is defined by the service demand in QN modelling?

The total average service time of a transaction at a resource is called its service demand.

Ex. If the average I/O time is 0.01 seconds

Service demand = average number of I/O x average time per I/O

<table>
<thead>
<tr>
<th>Class (r)</th>
<th>Avg. Number of I/Os</th>
<th>Service demand at disk (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Trivial)</td>
<td>5.5</td>
<td>0.055</td>
</tr>
<tr>
<td>2 (Medium)</td>
<td>28.9</td>
<td>0.289</td>
</tr>
<tr>
<td>3 (Complex)</td>
<td>85.0</td>
<td>0.850</td>
</tr>
</tbody>
</table>
If 1200 bytes of data packets are transmitted over the 64 Kbps line. The throughput of the link is 4 packets/sec. What is the utilization of the link?

Service demand = 1200x8 bits / 64x1024bits = 9600/65536=0.14s
Name three characteristics of the multiclass QN model

Heterogeneous service demands
- requests that form the workload can be clustered into groups
- exhibit significantly different service demands on the system resources

Different types of workloads
- workload are different (short online transactions, batch workload)

Different service level objectives
- different classes of requests have different service level objectives
  (acceptable limit on the average response time)
What are the types of resources referred by 1, 2, and 3
Load independent (LI)
- resources have a constant service rate that does not dependent on the load (CPU, disk, not depend on the number of requests in the queue)

Load dependent (LD)
- service rate is a function of the number of requests in the queue
  (the service rate increases as the number of requests grows or also decrease as the number of requests grows, LAN)

Delay (D)
- There is no waiting line
- Resources not shared with other requests
- A request that arrives is served immediately (e.g., client)
Thank You !!