CSC470 Fall, 2000

Computer System Modelling And Analysis

Laboratory #5
November 14th/16th & 21st/23rd, 2000

Assignment

Given the layered software environment described in the preparation handout, you are asked to apply various techniques, involving both analysis and simulation, to size and configure systems for different workload combinations and variations on the system. Answer the questions below using suitable combinations of three analytical techniques:

- Bounding or utilization analysis (applied to software as well as hardware servers),
- Method of Layers analysis (to take into account queueing effects for both hardware and software servers), and
- Simulation in Csim.

1. Bounding and Utilization Analysis

Assume for now that the system has **six client nodes**, and **two disks at each client node**.

(a). At a single client node, what are the arrival rates (in transactions per second) for each of the three types of client requests?

(b). At a single client node, what will the utilizations be of the client processor and each client disk?

Assume for now that the system has **six server nodes**, and **two disks per server node**.

(c). At a single server node, what are the arrival rates (in transactions per second) for the four types of services?

(d). At a single server node, what will the utilizations be of the server processor and each server disk?

Assume that the total arrival rate of client requests can change, but that the ratio among the types (i.e., 20 : 5 : 1) remains unchanged.

(e). If multithreading is allowed at the server nodes (so that the software servers are not bottlenecks themselves), what is an upper bound on the highest throughput of client requests attainable with a single server node? (It might be helpful to first determine an upper bound on the throughput of server requests fulfilled.)

(f). With unlimited multithreading, what is the minimum number of server nodes that would be required to satisfy the given workload volume?

(g). If multithreading is **not allowed** at the server nodes (so that each transaction must run to completion before the next one can be started), what is an upper bound on the highest throughput attainable with a single server node?

(h). With multithreading disallowed, what is the minimum number of server nodes that would be required to satisfy the given workload volume?

2. System Sizing

(a). Assuming that a **multithreading level of three** is allowed at each server node, choose a number of server nodes to use and a number of disks to be used at each server node such that transactions spend significantly less time (say about half as much time) queueing than they do in service. (In the MAP "performance" report, this corresponds to obtaining expansion factors" around 1.5 to 1.7.)

(b). Next, assume that multithreading is disallowed (i.e., the **multithreading level is one**), and again choose a configuration of server nodes satisfying the same criteria indicated above.

(c). Making use of the knowledge of the response times to server requests with the configuration chosen in part (a) for multithreading level three, choose a number of client nodes to acquire for the system and choose the number of disks to place at each client node. The intent should be to approximately meet the response time goals for the client requests:

Inquiries	15 seconds
Purchases	60 seconds
Returns	40 seconds

(d). For multithreading level one and the configuration chosen in part (b), again choose a number of client nodes to acquire for the system and choose the number of disks to place at each client node.

(e). If your answers to parts (c) and (d) differ, briefly explain why.

3. Software Improvements

Assume that, with some amount of programming effort, you could re-write the software supporting the server processes so that the average instruction path length is halved for each of the four services.

(a). Assuming the case of multithreading level of three, how much would response times to client requests be reduced by such improvements to the software?

(b). After improving the server based software as indicated above, would you suggest any changes in the system configuration (i.e., the numbers of each type of node or the number of disks at each node)?

Assume that, with that same amount of programming effort, you could have instead halved the average instruction path lengths of the three processes that execute on the client nodes.

(c). Which do you believe would be the better investment of the available programming time? ... improving the server processes, or improving the client processes? Justify your answer based on the relative impacts of the two choices on overall performance of the system.

4. Evaluation of Approaches

Comment briefly on the *strengths* and *weaknesses* of each of the three evaluation techniques proposed for use in this assignment (with respect to how helpful they were in doing this assignment):

- Bounding and utilization analysis
- Method of Layers analytical model
- Simulation with Csim

(Your full answer to question 4 should not exceed about 30 lines.)