1. (20 points.) Consider a system with 2 CPUs and 3 memory modules that requires at least 1 CPU and 1 memory module to be functioning for the system to be functioning. Use SHARPE to calculate the system reliability $R(t)$ at $t=1500$ hours. Assume that the MTTFs of a CPU and a memory module are 1000 and 2000 hours, respectively. The MTTRs for a CPU and a memory module are 5 and 10 hours, respectively. Show the SHARPE program listing and output. Consider the following three cases separately. For each case, you need to compute $R(t)$ at $t=1500$ hours. Use any model you like.

(a) Each component has an independent repair facility.
(b) Each subsystem (cpu or memory) has an independent repair facility that can repair failed components within the subsystem one at a time.
(c) The whole system shares a repair facility which repairs failed components one at a time with the repair priority of memory modules over CPUs.

(Hint: define Markov models that allow repairs to occur only when the system is still alive.)

2. (20 points.) Suppose that a network switch center has $n=3$ slots to accommodate incoming high and low priority clients, with arrival rates of $\lambda_h$ and $\lambda_l$ and departure rates of $\mu_h$ and $\mu_l$, respectively. A high priority client must always occupy one full slot. A low priority client, on the other hand, can lower its quality of service (QoS) by occupying only one half of a slot, if necessary. When a low priority client occupies a full slot, we call it a low priority, high QoS client; when it occupies one half of a slot, we call it a low priority, low QoS client. Draw a Markov state transition diagram for modeling the following resource control policy:

(i) If there is at least one full slot available, an incoming client always occupies an empty full slot, regardless of its priority class.
(ii) If case (i) is not true, an incoming high priority client can lower the QoS of two low priority, high QoS clients, if they are available, after which the high priority client occupies one full slot and the two low priority, high QoS clients both become low priority, low QoS clients, with each occupying one half slot.
(iii) If case (i) is not true, an incoming low priority client can lower the QoS of one low priority, high QoS client, if it is available, after which each occupies one half slot.
(iv) A client will turn away if none of the above cases is applicable.
(v) A low priority, low QoS client occupying one half slot can immediately become a low priority, high QoS client occupying one full slot upon a client’s departure.

Use the representation $(a, b, c)$ where $a$ stands for the number of low-priority, low QoS clients, $b$ stands for the number of low-priority, high QoS clients, and $c$ stands for the number of high-priority clients. Organize the Markov model so that when a high priority client arrives, the transition goes right, and when a low priority client arrives, the transition goes down. Label the transition rate of each transition clearly. Assume $\lambda_h = \mu_h = 2$ and $\lambda_l = \mu_l = 3$. Use Sharpe and assign rewards to states of the Markov model to obtain:

(a) the average number of low-priority clients (regardless of their QoS levels) in the system;
(b) the throughput of low-priority, high QoS clients.

3. (20 points.) Use the MVA solution technique to hand-calculate the average response time per terminal user once it enters the central system for $N=4$ for the QNM shown on page 144 of the lecture slide. (Hint: treat the terminal center as a center with infinite servers in the closed QNM and calculate the visit ratios first based on routing probabilities. Compare the average response time per user hand-calculated with that generated by Sharpe code. Note: the terminal service time per user is 25 sec, not 25 msec.)