CS 536 (Park) **Final** **Dec. 15 (Mon.), 1997**

PROBLEM 1

(a) What is *congestion*? Give a precise definition. Can congestion (in the technical sense) exist in an M/M/1 queue? Explain.

(b) What are the main objectives of ATM networks? What are the features that characterize them and for what purpose? What are the negative aspects of these choices?

(c) What is *Fast Retransmit* and what problem is it intended to solve? What is *Fast Recovery*? In what sense is it "fast"? What is *Slow Start* and in what sense is it supposed to be "slow"?

(d) Explain the philosophical difference between *reactive* congestion control and *proactive* congestion control. What is their principal trade-off relation? Is TCP proactive or reactive? Explain.

(e) Enumerate all the source traffic models we have encountered in class. What are the relative merits (if any) of each? Which ones are considered to be fair models of telephony and data traffic?

PROBLEM 2

(a) What is the "Lunch Date" problem also known as the 2-person consensus problem? How is this related to TCP? Can this problem be solved? How is the problem handled in practice (e.g., TCP)?

(b) Why is the condition $\lambda < \mu$ needed for M/M/1 queues when deriving the equilibrium queue length distribution? Is it still needed for M/M/1/n queues? Explain. What is the role of ergodicity in deriving the equilibrium queue length distribution for M/M/1?

(c) What is the role of VCI in ATM networks? Are the values globally or locally meaningful and what does it mean? What is the role of VPI? What additional optimization does it try to achieve? Based upon VCI and VPI, describe the process that needs to take place at every ATM switch along a path when performing call set-up.

(d) It is said that congestion control in high-bandwidth networks has become an even more difficult problem than before. In what sense is this true? What congestion control *strategy* is suited to handle this problem? Explain. (*Hint: delay-bandwidth product.*)

(e) What is system optimal routing and how does it differ from shortest paths routing (Dijkstra, Bellman-Ford)? What quantity is being optimized? What quantity is being optimized in SPF? Under what scenarios can the SPF perform significantly worse than system optimal routing?

PROBLEM 3

(a) Explain the fundamental dilemma faced by network designers when trying to facilitate both high QoS and high system utilization. Why is this a difficult problem? Under what conditions can this problem be solved?

(b) In what sense does Dijkstra's algorithm for shortest path converge faster than Bellman-Ford? In either cases, why can too frequent updates to the routing table be detrimental? Give two ways of solving this problem. Describe a scenario in which feedback can exercise a negative influence on routing stability. How would you go about fixing this problem?

(c) What is the fragmentation problem in IP? Other things being equal, will the fragmentation problem be more or less severe in ATM networks? Explain. Give two ways of dealing with this problem. What are their costs?

(d) How can weighted fair queueing be used to achieve guaranteed QoS service using resource reservations? Sketch the structure of such a system in terms of the call set-up process and the actions taken at switches. Assume the user specifies that he/she would like to obtain X Mbps of CBR service. (*Hint: Notice that there are two resources that need to be accounted for.*)

PROBLEM 4

(a) Assume you are given a congestion-susceptible network given by a unimodal load-throughput function $\mu = \mu(Q)$ where $\mu(Q)$ achieves maximum at the point (Q^*, μ^*). Let the system be described by the following two differential equations, representing the conservation of flow and a control law, respectively:

$$dQ/dt = \lambda - \mu \quad (1)$$

$$d\lambda/dt = \epsilon(Q^* - Q) \quad (2)$$

Give the intuitive meaning of (1) and (2). Is the optimal operating point (Q^*, μ^*) a fixed point (i.e., stays still in the absence of perturbation) of the system? Explain. Draw the vector field or phase portrait for the two-dimensional system; first, explain how this is done. From your drawing, what can you conclude about the stability of (Q^*, μ^*)? What is the relation between the *linear increase/exponential decrease* heuristic to your conclusion at hand? Construct an *asymmetric control law* of the form "if $Q \leq Q^*$ then do this" and "if $Q > Q^*$ then do that" which possesses better stability properties. On what property of the $\mu(Q)$ function does the success of the "if $Q > Q^*$ then do that"-part of the control depend?

(b) Is TCP a cooperative or noncooperative congestion control protocol? Explain. If there are k sessions running concurrently and everyone is running a cooperative congestion control protocol but one, what is likely to happen to the relative bandwidth being allocated to the k connections? What happens if all k connections behave noncooperatively, in particular, "aggressively"? Can FIFO queues ever hope to resolve this problem, in particular, with respect to fairness? Explain.