Quiz No. 4 [Two Problems, 7 and 3 points each, respectively]

Your Name:

Problem 1 [7 points] Consider an M/G/1 system with vacations, *i.e.,* arrivals are Poisson with rate λ, and service times follow a general distribution with p.d.f. $f_S(t), t \geq 0$, and mean $E[S]$, where the server goes on "vacation" for a *single* period of duration $V = 1$ whenever it becomes idle. Under these assumptions, derive an expression, function of λ and $E[S]$, for $E[B_{V_1}]$, the average duration of a busy period, *i.e.,* the server is busy serving.

Hint: consider separately the cases of 0 and 1 or more arrivals during a vacation period and use that information to first obtain an expression for an average idle period (the server is idle), and use this information to obtain an expression for $E[B_{V_1}]$.

The probability f_0 of 0 arrivals during a vacation (of duration 1) is given by

$$f_0 = \frac{\lambda^0}{0!}e^{-\lambda} = e^{-\lambda}$$

The average duration $E[I]$ of an idle period, *i.e.,* period during which the server is idle, is then of the form

$$E[I] = 1 \cdot (1 - f_0) + \left(1 + \frac{1}{\lambda}\right) f_0 = 1 + \frac{e^{-\lambda}}{\lambda}$$

We also know that the fraction of time $\rho = \lambda E[S]$ (from Little's law applied to the server) that the server is busy is equal to

$$\begin{aligned} \rho &= \frac{E[B_{V_1}]}{E[B_{V_1}] + E[I]} \\ \Rightarrow E[B_{V_1}](1-\rho) &= \rho\left(1 + \frac{e^{-\lambda}}{\lambda}\right) \\ E[B_{V_1}] &= \frac{\rho}{1-\rho}\left(1 + \frac{e^{-\lambda}}{\lambda}\right) \end{aligned}$$

PLEASE TURN TO NEXT PAGE

Problem 2 [3 points] Our goal is to compare a single fast PS server to two slow PS servers. Specifically, assume that jobs arrive according to a Poisson process of rate λ and have service times drawn from a general distribution.

Consider first a scenario where jobs are sent to a single server operating according to a PS service discipline. In this case, the average service time of a job (when running alone on the server) is $E[S] = \frac{1}{\mu}$. Next, consider a scenario where instead of a single, fast server, we have two slower servers, each operating at half the rate, *i.e.,* a job's average service time on either server is $2E[S] = \frac{2}{\mu}$, and where arrivals are split evenly (and randomly) across the two servers.

Which of the two configurations yields the lower response time. Justify your answer.

When using a single, fast PS server with an arrival rate λ and an average service time $\frac{1}{\mu}$, the average response time is of the form:

$$E[T_F] = \frac{1}{\mu - \lambda}$$

Conversely, if we split jobs evenly between two servers running at half the speed, we have

$$E[T_S] = \frac{1}{2} \cdot \frac{1}{\frac{\mu}{2} - \frac{\lambda}{2}} + \frac{1}{2} \cdot \frac{1}{\frac{\mu}{2} - \frac{\lambda}{2}} = \frac{1}{\mu - \lambda} + \frac{1}{\mu - \lambda} = \frac{2}{\mu - \lambda} = 2E[T_F]$$

In other words, a system with a single fast server is twice as fast.