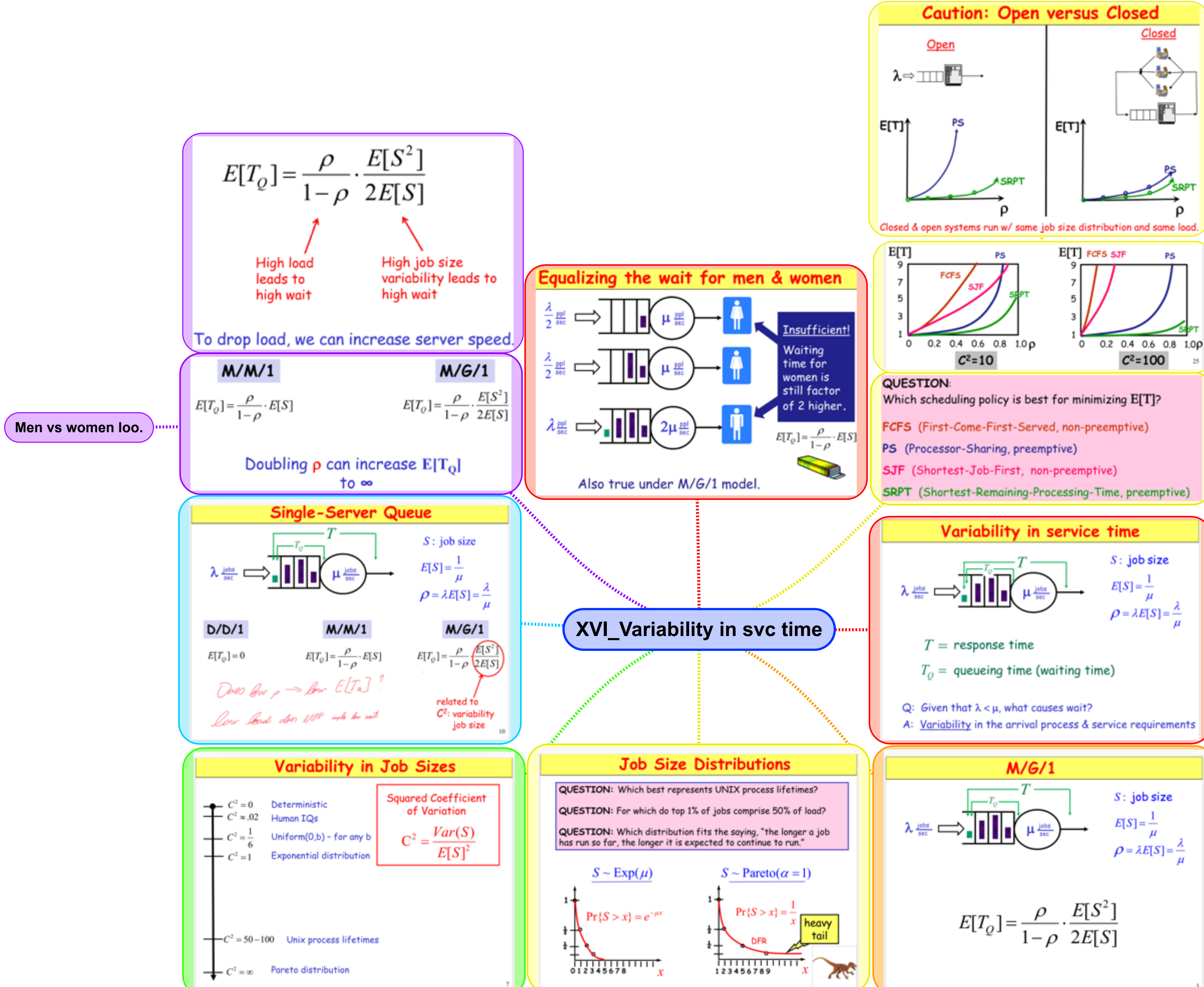
XVI_Variability in svc time
Men vs women loo.
$E[T_Q] = \frac{\rho}{1-\rho} \cdot \frac{E[S^2]}{2E[S]}$
High load leads to high wait
High job size variability leads to high wait
To drop load, we can increase server speed.
M/M/1
M/G/1
$E[T_Q] = \frac{\rho}{1-\rho} \cdot E[S]$
$E[T_Q] = \frac{\rho}{1-\rho} \cdot \frac{E[S^2]}{2E[S]}$
Doubling ρ can increase $E[T_Q]$ to ∞
Equalizing the wait for men & women
Insufficient!
Waiting time for women is still factor of 2 higher.
Also true under M/G/1 model.
Caution: Open versus Closed
Open
Closed
E[T]
PS
SRPT
ρ
Closed & open systems run w/ same job size distribution and same load.
FCFS
SJF
$C^2=10$
$C^2=100$
QUESTION:
Which scheduling policy is best for minimizing E[T]?
FCFS (First-Come-First-Served, non-preemptive)
PS (Processor-Sharing, preemptive)
SJF (Shortest-Job-First, non-preemptive)
SRPT (Shortest-Remaining-Processing-Time, preemptive)
Single-Server Queue
S: job size
$E[S] = \frac{1}{\mu}$
$\rho = \lambda E[S] = \frac{\lambda}{\mu}$
D/D/1
M/M/1
M/G/1
$E[T_Q] = 0$
related to C^2: variability job size
Variability in service time
T = response time
T_Q = queueing time (waiting time)
Q: Given that λ < μ, what causes wait?
A: Variability in the arrival process & service requirements
Variability in Job Sizes
$C^2 = 0$ Deterministic
$C^2 \approx .02$ Human IQs
$C^2 = \frac{1}{6}$ Uniform(0,b) - for any b
$C^2 = 1$ Exponential distribution
$C^2 = 50-100$ Unix process lifetimes
$C^2 = \infty$ Pareto distribution
Squared Coefficient of Variation
$C^2 = \frac{Var(S)}{E[S]^2}$
Job Size Distributions
QUESTION: Which best represents UNIX process lifetimes?
QUESTION: For which do top 1% of jobs comprise 50% of load?
QUESTION: Which distribution fits the saying, "the longer a job has run so far, the longer it is expected to continue to run."
$S \sim \mathrm{Exp}(\mu)$
$\Pr\{S > x\} = e^{-\mu x}$
$S \sim \mathrm{Pareto}(\alpha = 1)$
$\Pr\{S > x\} = \frac{1}{x}$
heavy tail
DFR
M/G/1