Project Summary: Queueing Analysis of a Hospital Emergency Room

Koorosh Asil Gharehbaghi
Department of Computer Science, Faculty of Mathematics
K. N. Toosi University of Technology

Instructor: Dr. Razieh Khodsiani

June 3, 2025

Introduction

This project applies queueing theory and discrete-event simulation to analyze patient flow in a hospital Emergency Room (ER). We model arrivals as a Poisson process and service times as exponential, constructing both M/M/1 (single-server) and M/M/c (multi-server) queues. The objectives are to derive analytical performance measures, validate them via simulation, and determine an optimal staffing level.

Objectives

- **Analytical Derivation:**
  - For M/M/1 and M/M/c systems, compute:
    - Utilization ρ,
    - Average number in system L,
    - Average queue length L_q,
    - Average time in system W,
    - Average waiting time W_q,
    - Probability of waiting P_{wait}.
- **Discrete-Event Simulation (DES):**
  - Implement a simulator that:
    - Generates arrivals using $\text{Exp}(\lambda)$,
    - Generates service times using $\text{Exp}(\mu)$,
    - Maintains an event list for `ARRIVAL` and `DEPARTURE`,
    - Tracks time-weighted counters to estimate $\hat{L}, \hat{L}_q, \hat{W}, \hat{W}_q, \hat{P}_{\text{wait}}, \hat{\rho}$.

- Compare simulation outputs to analytical results.

- **Sensitivity Analysis:**
  - Vary arrival rate λ, service rate μ, and number of servers c.
  - Observe how L, L_q, W, W_q, and P_{wait} change.

- **Staffing Optimization:**
  - Define a target waiting time, for example $W_q \leq 0.5$ hours.
  - Use the Erlang C formula to find the smallest c meeting that target.
  - Introduce a cost model:

    $$C_{\text{total}}(c) \;=\; 100\,c \;+\; 20\,\lambda\,W_q(c),$$

    where each server costs \$100/hour and patient waiting costs \$20 per patient-hour.
  - Determine the staffing level c^* that minimizes $C_{\text{total}}(c)$.

Methodology

Analytical Component

- **M/M/1 Model:**

$$\rho = \frac{\lambda}{\mu}, \quad \pi_n = (1-\rho)\,\rho^n, \quad n \geq 0,$$

$$L = \frac{\rho}{1-\rho}, \quad L_q = \frac{\rho^2}{1-\rho}, \quad W = \frac{1}{\mu-\lambda}, \quad W_q = \frac{\rho}{\mu-\lambda}, \quad P_{\text{wait}} = \rho.$$

- **M/M/c Model:**

$$\rho = \frac{\lambda}{c\mu} < 1, \quad \pi_0 = \left[\sum_{n=0}^{c-1} \frac{(\lambda/\mu)^n}{n!} \;+\; \frac{(\lambda/\mu)^c}{c!\,(1-\rho)}\right]^{-1}.$$

$$P_{\text{wait}} = \frac{\frac{(\lambda/\mu)^c}{c!}\,\frac{1}{1-\rho}\,\pi_0}{\sum_{n=0}^{c-1}\frac{(\lambda/\mu)^n}{n!} + \frac{(\lambda/\mu)^c}{c!}\,\frac{1}{1-\rho}},$$

$$L_q = \frac{(\lambda/\mu)^c\,\rho}{c!\,(1-\rho)^2}\,\pi_0, \quad W_q = \frac{L_q}{\lambda}, \quad L = L_q + \frac{\lambda}{\mu}, \quad W = W_q + \frac{1}{\mu}.$$

Simulation Component

- **Event Types:** `ARRIVAL` and `DEPARTURE`.

- **Data Structures:**
  - *Event list:* Priority queue of $\langle type, time, patientID \rangle$.

  - *FIFO queue:* Holds waiting patient IDs.
  - *Counters:* BusyServers, QueueLength, AreaUnderQ, AreaInSystem.

- **Pseudocode Outline:**

  1. Initialize time = 0, BusyServers = 0, QueueLength = 0.
  2. Schedule first `ARRIVAL` at time $\mathrm{Exp}(\lambda)$.
  3. While time $< T_{\max}$:
     - Pop next event $\langle type, t_{\mathrm{ev}}, ID \rangle$.
     - Update time-weighted counts:

       $$\text{AreaUnderQ} \mathrel{+}= (t_{\mathrm{ev}} - t_{\mathrm{last}}) \times \text{QueueLength}, \quad \text{AreaInSystem} \mathrel{+}= (t_{\mathrm{ev}} - t_{\mathrm{last}}) \times (\text{BusyServe}$$

       Set $t_{\mathrm{last}} = t_{\mathrm{ev}}$.
     - If `ARRIVAL`:
       * If BusyServers $< c$, increment BusyServers, draw service time $\mathrm{Exp}(\mu)$, schedule `DEPARTURE`.
       * Else, enqueue this patient (increment QueueLength).
       * Schedule next `ARRIVAL` at $t_{\mathrm{ev}} + \mathrm{Exp}(\lambda)$.
     - If `DEPARTURE`:
       * Decrement BusyServers. If QueueLength > 0, dequeue one patient, increment BusyServers, draw service time $\mathrm{Exp}(\mu)$, schedule next `DEPARTURE`.
  4. After $T_{\max}$, compute:

     $$\hat{L} = \frac{\text{AreaInSystem}}{T_{\max}}, \quad \hat{L}_q = \frac{\text{AreaUnderQ}}{T_{\max}},$$

     $$\hat{W} = \frac{1}{N_{\text{served}}} \sum (\text{DepartureTime} - \text{ArrivalTime}), \quad \hat{W}_q = \frac{1}{N_{\text{served}}} \sum \text{WaitingTime}.$$
  5. Run $N = 30$ replications to obtain mean and confidence intervals.

Expected Outcomes

- Validation that DES outputs closely match analytical formulas.
- Quantification of how L, L_q, W, W_q, and P_{wait} vary with λ, μ, and c.
- Identification of the minimum c to achieve a waiting-time target.
- Determination of cost-optimal staffing c^* via

  $$C_{\text{total}}(c) = 100\,c + 20\,\lambda\,W_q(c).$$

- Presentation of comparative tables and plots (theory vs. simulation, cost vs. servers).