NTNU
Norwegian University of
Science and Technology

Week 43: Lecture 1

Queueing theory

Geir-Arne Fuglstad

October 19, 2020

Information

— Reference group meeting will be on October 22.

— Let me or the reference group know if you have feedback.

— We plan to end new material on November 11 (or maybe November 9).

Theorem (Little's law (Part 2))

$$L_0 = \lambda W_0$$

L_0: *Average number of customers waiting in the system.*

λ: *Rate of arrival to the system.*

W_0: *Average waiting time before service starts.*

Example

On average, 10 customers arrive to the store per hour. The average time spent in the store is 0.5 hours. What is the average number of customers in the store?

Example

There are on average 2 customers queuing and being served. If 10 customers arrive per hour, on average, how long do customers spend checking out on average?

Section 9.1.2

Notation

We write “A/B/c queue” for a queueing system with

1) Interarrival distribution “A”
2) Individual service time distribution “B”
3) Number of servers “c”

Section 9.2.1

Definition

A **M/M/1 queue** with arrival rate $\lambda > 0$ and expected service time $1/\mu > 0$ has

1) Interarrival times are independent and identically distributed as Exp(λ).
2) Service times are independent and identically distributed as Exp(μ).
3) One server, and service times are independent of arrival process.

Example

Assume requests to a web server is described by a M/M/1 queue with arrival rate $\lambda = 5$ (per second) and expected service time $1/\mu$. What is the behaviour for $\mu > 5$, $\mu = 5$ and $\mu < 5$.

Example

Assume requests to a web server is described by a M/M/1 queue with arrival rate $\lambda > 0$ and expected service time $1/\mu > 0$. Derive

1) Limiting distribution
2) Expected number of requests in the system
3) Server utilization
4) Expected time in system

Section 9.2.2

Definition

A **M/M/∞ queue** with arrival rate $\lambda > 0$ and expected service time $1/\mu > 0$ has

1) Interarrival times are independent and identically distributed as Exp(λ).
2) Service times are independent and identically distributed as Exp(μ).
3) Infinitely many servers, and service times are independent of arrival process.

Example

Assume requests to a web server is described by a M/M/∞ queue with arrival rate λ and expected service time $1/\mu$. Derive

1) Limiting distribution
2) Expected number of requests in the system
3) Server utilization
4) Expected time in system

Section 9.2.3

Definition

A **M/M/**s **queue** with arrival rate $\lambda > 0$ and expected service time $1/\mu > 0$ has

1) Interarrival times are independent and identically distributed as Exp(λ).
2) Service times are independent and identically distributed as Exp(μ).
3) s servers, and service times are independent of arrival process.

Example

Assume requests to a web server is described by a M/M/*s* queue with arrival rate λ and expected service time $1/\mu$. Derive

1) Limiting distribution
2) Expected number of requests in the system
3) Server utilization
4) Expected time in system

Example

Requests to a web server is modelled as a M/M/s queue with arrival rate $\lambda = 5$, expected service time $1/\mu = 1$, and number of servers $s \in \{1, 5, 6\}$. How many servers are needed? How does the limiting distribution compare to those of a M/M/1 queue and a M/M/∞ queue.