

Stochastic Network Modeling (SNM)

Queuing Theory

Stochastic Network Modeling (SNM)

Llorenç Cerdà-Alabern Universitat Politècnica de Catalunya Departament d'Arquitectura de Computadors llorenc@ac.upc.edu

Parts

- Introduction
- Discrete Time Markov Chains (DTMC)
- Continuous Time Markov Chains (CTMC)
- Queuing Theory



Stochastic Network Modeling (SNM)

Queuing Theory

Part IV

Queuing Theory

Outline

- Introduction
- Kendal Notation
- Little Theorem
- PASTA Theorem
- The M/M/1 Queue
- M/G/1 Oueue

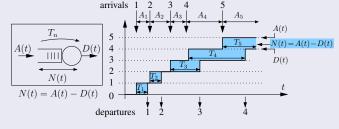
- M/G/1/K Queue
- M/G/1 Busy Period
- M/G/1 Delays
- Oueues in Tandem
- Networks of Oueues
- Matrix Geometric Method



Introduction

Queuing Theory

Introduction



- Queueing theory is the mathematical study of waiting lines, or queues.
- Common notation:
 - A(t): number of arrivals [0, t].
 - A_n : interarrival time between customers n and n+1.
 - T_n : time in the system (response time) for customer n.

3 / 69

• N(t): number in the system at time t.



Kendal Notation

Queuing Theory

Introduction

Kendal

Notation

Little Theore

PASTA Theorem

The M/M/1 Queue

M/G/1 Queu

M/G/1/K

M/G/1 Bus Period

M/G/1 Delay

Queues in Tandem

Kendal Notation

A/S/k[/c/p]

- A: arrival process,
- S: service process,
- k: number of servers,
- c: maximum number in the system (number of servers + queue size). Note: some authors use the queue size.
- p: population.
 If "c" or "p" are missing, they are assumed to be infinite.

arrivals 2 departures k



Kendal Notation

Queuing Theory

Introduction

Kendal

Notation

Little Theorem

PASTA Theorem

The M/M/1 Queue

M/G/1 Queue

M/G/1 Bus

M/G/1 Delays

Queues in Tandem

Common arrivals/service processes

- **G**: general (non specific process is assumed),
- M: Markovian (exponentially or geometrically distributed),
- D: deterministic.
- P: Poisson (discrete RV, *N*, equal to the number of arrivals exponentially dist. in a time *t*):

$$P_p(N = n, t) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}, n \ge 0, t \ge 0.$$

• Er: Erlang (continuous RV equal to the time *t* that last *n* arrivals exponentially dist.):

$$f_e(t) = \lambda P_p(N = n - 1, t) = \frac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!}, t \ge 0, n \ge 1$$

Examples

- M/M/1: M. arr. / M. serv. / 1 server, ∞ queue and population.
- M/G/1: M. arr. / Gen. serv. / 1 server, ∞ queue and population.



Stochastic Network Modeling (SNM)

Queuing Theory

Part IV

Little Theorem

Queuing Theory

Outline

- Little Theorem



Queuing Theory

Introductio:

Little Theorem

Graphical proof

Application to the waiting line and the server

Mean number in the

PASTA Theorem

The M/M

M/G/1 Quei

M/G/1/K Queue

M/G/1 Busy Period

Little Theorem

- Define the stochastic processes:
 - A(t): number of arrivals [0, t].
 - T_n : time in the system (response time) for customer n.
 - N(t): number in the system at time t.
- And the mean values:
 - Mean number of customers in the system:

$$N = \lim_{t \to \infty} \frac{1}{t} \int_0^t N(s) \, \mathrm{d}s$$

- Arrival rate: $\lambda = \lim_{t \to \infty} A(t)/t$
- Mean time in the system: $T = \lim_{t \to \infty} (\sum_n T_n) / A(t)$
- The following relation follows:

$$N = \lambda T$$

Mnemonic: NAT (Number = Arrivals x Time).



Queuing Theory

Introduction

Notation

Graphical proof

Application to the waiting line and the server

Mean number in the Server

Theorem

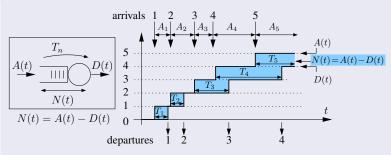
Queue

M/G/1 Queu

M/G/1/K Queue

M/G/1 Bus Period

Graphical proof



• From the graph we have:

$$\frac{1}{t} \int_0^t N(s) \, ds = \frac{1}{t} \sum_{i=1}^{A(t)} T_i = \frac{A(t)}{t} \frac{\sum_{i=1}^{A(t)} T_i}{A(t)}$$

• Taking the limit $t \to \infty$: $N = \lambda T$



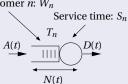
Queuing Theory

Application to the waiting line and the

Application to the waiting line and the server

• We can apply the Little theorem to the waiting line and the server:

> Waiting time in the queue of customer n: W_n



Time in the system:

 $T_n = W_n + S_n$ Expected value:

$$T = W + S$$

where
 $T = E[T_n], W = E[W_n],$
 $S = E[S_n]$

- Mean number of customers in the queue: $N_0 = \lambda W$.
- Mean number of customers in the server: $N_S = \rho = \lambda S$.



Queuing Theory

Introduction

Little Theore

Graphical proof

Application to the waiting line and the server

Mean number in the Server

Theorem

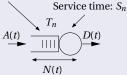
M/G/1 Oner

M/G/1/K Queue

M/G/1 Busy Period

Mean number in the Server

Waiting time in the queue of customer n: W_n



Time in the system:

$$T_n = W_n + S_n$$
 Expected value:

$$T = W + S$$

where
 $T = E[T_n], W = S$

$$T = E[T_n], W = E[W_n],$$

 $S = E[S_n]$

• In a single server queue (even if not Markovian):

$$\rho = N_S = \mathbb{E}[N_S(t)] = \lambda \, \mathbb{E}[S]$$

$$\mathbb{E}[N_S(t)] = 0 \times \pi_0 + 1 \times (1 - \pi_0) = 1 - \pi_0 \Rightarrow \pi_0 = 1 - \rho$$

• $\rho = N_S = \lambda E[S] = 1 - \pi_0$ is the proportion of time the system is busy, in other words, is the server utilization or load.



Stochastic Network Modeling (SNM)

Queuing Theory

ntroduction

Little Tileofer

PASTA Theorem

Example of P

The M/M/

M/G/1 Onem

M/G/1/K Queue

Period

M/G/1 Delays

Queues in Tandem Part IV

Queuing Theory

Outline

- Introduction
 - A Kandal Notation
 - Kendal Notation
- Littl
- PASTA Theorem
 - The M/M/1 Queue
- M/G/1 Oueue

- M/G/1/K Oueue
- M/G/1 Busy Period
- M/G/1 Delays
- Queues in Tandem
- Networks of Queues
- Matrix Geometric Method

Networks of

Llorenç Cerdà-Alabern



PASTA Theorem

Queuing Theory

Introduction

PASTA

Theorem

The M/M/

MICHA

M/G/1/K

Period

M/G/1 Delays

Queues in Tandem

PASTA Theorem: Poisson Arrivals See Time Averages

- The mean time the chain is in state i is $\pi_i \Rightarrow$ using PASTA, the probability that a Markovian arrival see the system in state i is π_i (proof: see [1]).
- The equivalent theorem in discrete time is the arrival theorem, RASTA: Random Arrivals See Time Averages: the probability that a random arrival see the system in state i is π_i .
- [1] Ronald W Wolff. "Poisson arrivals see time averages". In: *Operations Research* 30.2 (1982), pp. 223–231.



Queuing Theory

Introductioi

. . . .

Little Theorer

PASTA Theorem

Example of PASTA

he M/M/1

Queue

MICILIV

Queue M/G/1 Buc

MOUDI

M/G/1 Delay

Queues in Tandem

Example of PASTA

- Assume that a system can have, at most, *N* customers (e.g *N* 1 in the queue and 1 in service).
- Assume that an arrival is lost when the system is full.
- By PASTA the proportion of Poisson arrivals that see the system full, and are lost, is equal to the proportion of time the system has N in the system, π_N .
- Thus, the loss probability is π_N .



Stochastic Network Modeling (SNM)

Queuing Theory

Part IV

Queuing Theory

Outline

The M/M/1 Queue

The M/M/1

Queue



Queuing Theory

Introduction

rz . . . d . l

Little Theorer

PASTA Theorem

The M/M/1

Queue

Q-matrix Stationary Distribution

Stability
Example: Loss
probability in a

M/G/1 Queu

M/G/1/K Queue

The M/M/1 Queue

$$A(t) \xrightarrow{N(t)} T_n = W_n + S_n$$

• Markovian arrivals with rate $\lambda \Rightarrow$ the interarrival time is exponentially distributed with mean $1/\lambda$:

$$P\{A_n \le x\} = 1 - e^{-\lambda x}, x \ge 0$$

 \Rightarrow A(t) is a Poisson process:

$$P(A(t) = i) = \frac{(\lambda t)^i}{i!} e^{-\lambda t}, i \ge 0, t \ge 0$$

• Markovian Services with rate $\mu \Rightarrow$ service time exponentially distributed with mean $1/\mu$:

$$P\{S_n \le x\} = 1 - e^{-\mu x}, x \ge 0$$



Queuing Theory

Introduction

Notation

Little Theorei

PASTA Theorem

Queue

Q-matrix

Stationary Distribution

Stability Example: Loss probability in a

M/G/1 Queu

M/G/1/K Queue

Q-matrix

• The process $N(t) = \{\text{number in the system at time } t \ge 0\}$ is a CTMC.

OBSERVATION: for a non Markovian service, the process N(t) would not be a MC! State transition diagram:

• Q-matrix:

$$\mathbf{Q} = \begin{bmatrix} -\lambda & \lambda & 0 & 0 & \cdots \\ \mu & -(\mu + \lambda) & \lambda & 0 & \cdots \\ 0 & \mu & -(\mu + \lambda) & \lambda & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$



Queuing Theory

Introduction

Little Theorem

DACTA

Theorem

The M/N Queue

Queue Q-matrix

Stationary Distribution

Stability
Example: Loss
probability in a
telephone switchir

M/G/1 Que

M/G/1/I Queue

Stationary Distribution

• Solving the M/M/1 queue using flux balancing (or the general solution of a reversible chain):

$$\pi_i = (1 - \rho) \rho^i, i = 0, \dots, \infty$$

where $\rho = \frac{\lambda}{\mu}$



Queuing Theory

Introduction

.

Little Theore

PASTA Theorem

The M/M/1

Queue Q-matrix

Stational

Properties Stability

Example: Loss probability in a telephone switching

M/G/1 Q

M/G/1/K Queue

Properties

• Mean customers in the system:

$$N = \lim_{t \to \infty} \frac{1}{t} \int_0^t N(s) \, ds = \sum_{i=0}^{\infty} i \pi_i = \sum_{i=0}^{\infty} i (1 - \rho) \, \rho^i = \frac{\rho}{1 - \rho}$$

• Mean time in the system (response time):

Little:
$$N = \lambda T \Rightarrow T = \frac{N}{\lambda} = \frac{\rho}{\lambda (1 - \rho)} = \frac{1}{\mu - \lambda}$$

- Mean time in the queue: $W = T \frac{1}{\mu} = \frac{\rho}{\mu \lambda}$
- Mean Number in the queue: $N_Q = \lambda W = \frac{\rho^2}{1-\rho}$
- Mean number in the server: $N_s = N N_Q = \rho$ NOTE: $\pi_0 = 1 - \rho$



Queuing Theory

Introductio

Notation

Little Theoren

PASTA Theorem

Oueue

Queue

Q-matrix

Proper

Stability
Example: Loss
probability in a

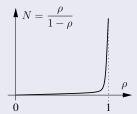
M/G/1 Oueu

M/G/1/K

Queue

Stability

- *N* and *T* are proportional to $1/(1-\rho) \Rightarrow$ when $\rho \to 1 \Rightarrow N, T \to \infty$.
- The process N(t) is positive recurrent, null recurrent or transient according to whether $\rho = \lambda/\mu$ is below, equal or greater than 1, respectively.





Queuing Theory

Introduction

Tital - Th - - - - -

.....

Theoren

The M/M

Queue

Q-matrix

Propertie

Example: Loss probability in a telephone switching

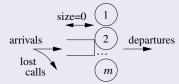
M/G/1 Que

M/G/1/k Queue

MICIA Door

Example: Loss probability in a telephone switching center

• Hypothesis: Switching center with m circuits and "lost call", infinite population, Markovian arrivals with rate λ and exponentially distributed call duration with mean $1/\mu \Rightarrow M/M/m/m$ queue.





Queuing Theory

Introduction

Notation

Little Theorem

PASTA Theorem

The M/N

Queue

Q-matrix

Properti

Example: Loss probability in a telephone switching

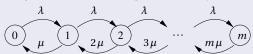
M/G/1 Que

M/G/1/I Queue

M/C/1 Prove

Example: Loss probability in a telephone switching center

• Since the minimum of i independent and identically exponentially distributed RV with parameter service time is exponentially distributed with parameter $i\mu$:





Queuing Theory

Example: Loss probability in a telephone switching

Example: Loss probability in a telephone switching center

- Stationary Distribution of the queue M/M/m/m:
- Solving using the general solution of a reversible chain:

Define
$$\rho_k = \frac{\lambda}{(k+1)\mu}$$
, $k = 0, \dots, m-1$

$$\pi_0 = \frac{1}{G}, \, \pi_i = \frac{1}{G} \prod_{k=0}^{i-1} \rho_k = \frac{1}{G} \left(\frac{\lambda}{\mu}\right)^i \frac{1}{i!}, \, 0 < i \le m \Rightarrow$$

$$\pi_i = \frac{1}{G} \left(\frac{\lambda}{\mu}\right)^i \frac{1}{i!}, 0 \le i \le m. \ G = \sum_{k=0}^m \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!}.$$

• Using PASTA Theorem (Poisson Arrivals See Time Average): the loss call probability is the probability that the queue is in state m: π_m , "Erlang B Formula".



Stochastic Network Modeling (SNM)

Queuing Theory

Part IV

Gendal

Queuing Theory

.....

Theorem

The M/M/I Queue

M/G/1 Queue

M/G/1 Queue

Matrix

Properties of the stationary distribution $(\pi = \pi P, \pi e = 1)$

Proof of the Level Crossing Law Theorem

M/G/1/K Queue

M/G/1 Busy

Outline

- Introduction
- Kendal Notation
- Little Theorer
- PASTA Theorem
- The M/M/1 Queue
- M/G/1 Queue

- M/G/1/K
- M/G/1 Rusy Period
- M/G/1 Delays
- Oueues in Tandem
- Networks of Queues
- Matrix Geometric Method



Queuing Theory

Introduction

Kendal

Little Theorer

PASTA Theorem

The M/M/ Queue

M/G/1 Queue

Matrix
Properties of the stationary distribution $(\pi = \pi P, \pi e = 1)$ Proof of the Level

M/G/1

M/G/1 Busy Period

M/G/1 Queue

- The process $N(t) = \{\text{number in the system at time } t \ge 0\}$ in general it is not a MC (it is so only if G is Markovian).
- We can build a semi-Markov process observing the system at departure times t_n (note that t_n are also the service completion times). Define the discrete time process:

 $X(n) = \{\text{number in the system at time } t_n \ge 0, n = 0, 1, \dots \}$

- Theorem: The process X(n) is a DTMC.
- Proof: X(n) only depends on the number of arrivals in non overlapping intervals. Since arrivals are Markovian, this is a memoryless process.
- NOTE: Looking at departure times the chain may have self transitions (in contrast to observing at transition times): we can have the same number in the system after a departure.



Queuing Theory

Introductio

Notation

Little Theorem

PASTA Theorem

The M/M/I Queue

Queue

M/G/1 Queue
Transition Probability

Properties of the stationary distribution $(\pi = \pi P, \pi e = 1)$

Proof of the Level Crossing Law Theore

M/G/1 Bus

M/G/1 Bus Period

Transition Probability Matrix

- Let $f_S(x)$, $x \ge 0$ be the service time density function.
- Define the RV V = {number of arrivals during a service time}, and the probabilities: v_i = P{V = i}.
- Conditioning on the service duration:

$$v_i = \int_{x=0}^{\infty} P\{i \text{ arrivals in time } x \mid S = x\} f_S(x) dx \Rightarrow$$

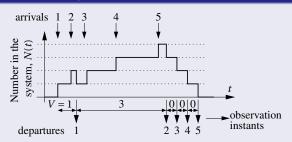
$$v_i = \int_{x=0}^{\infty} \frac{(\lambda x)^i}{i!} e^{-\lambda x} f_S(x) dx$$



Queuing Theory

Transition Probability

Transition Probability Matrix



• $v_i = P\{\text{number of arrivals during a service time} = i\} \Rightarrow$

$$p_{ij} = \begin{cases} 0, & j < i-1 \quad (N(t) \text{ can only be decreased by 1}) \\ v_j, & i = 0, j \ge 0 \quad (i = 0 \to \text{the queue was empty}) \\ v_{j-i+1}, & i > 0, j \ge i-1 \quad (i > 0 \to \text{the queue was busy}) \end{cases}$$



Queuing Theory

Introduction

Little Theorem

PASTA

The M/M/

M/G/1 Queue Transition Probability

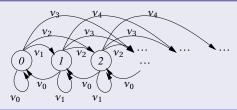
Matrix $\begin{aligned} & \text{Properties of the} \\ & \text{stationary distribution} \\ & (\pi = \pi \, P, \pi \, e = 1) \end{aligned}$

Proof of the Level Crossing Law Theore

Queue

M/G/1 Busy Period

Transition Probability Matrix



$$p_{ij} = \begin{cases} 0, & j < i-1 \\ v_j, & i = 0, j \ge 0 \\ v_{j-i+1}, & i > 0, j \ge i-1 \end{cases} \Rightarrow \mathbf{P} = \begin{bmatrix} v_0 & v_1 & v_2 & v_3 & \cdots \\ v_0 & v_1 & v_2 & v_3 & \cdots \\ 0 & v_0 & v_1 & v_2 & \cdots \\ 0 & 0 & v_0 & v_1 & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

• Stationary distribution: $\pi = \pi P$, $\pi e = 1$.



Queuing Theory

Properties of the

<u>Properties of the stationary distribution ($\pi = \pi P$, $\pi e = 1$)</u>

• Using the "Level Crossing Law" theorem: a queue with unitary arrivals and departures satisfies:

> $P\{\text{an arriving customer finds } i \text{ in the system}\} =$ $P\{a \text{ departing customer leaves } i \text{ in the system}\} \Rightarrow$

 $\pi_i = P\{\text{an arriving customer find } i \text{ in the system}\}\$

Using PASTA:

 $\pi_i = P\{\text{there are } i \text{ customers in the } \}$ system at an arbitrary time}

So, in an M/G/1 the stationary distribution of the EMC obtained observing the departures, is the stationary distribution of the continuous time process.



Queuing Theory

Introduction

Theorem

The M/M/1 Queue

M/G/1 Queue
Transition Probabilit

Matrix

Properties of the stationary distribution $(\pi = \pi P, \pi e = 1)$ Proof of the Level Crossing Law Theorem

M/G/1/

M/G/1 Busy

Proof of the Level Crossing Law Theorem

- Define:
 - $A_i(t) = \{\text{number of arrivals finding } i \text{ in the system at } t \ge 0\}$
 - D_i(t) ={number of departures leaving i in the system at t ≥ 0}
 - $P\{\text{a customer finds } i \text{ in the system}\} = \lim_{t\to\infty} A_i(t)/A(t)$
 - $P\{\text{a customer leave } i \text{ in the system}\} = \lim_{t\to\infty} D_i(t)/D(t)$
- An arriving customer that finds i in the system produce a transition $i \rightarrow i+1$. A customer leaving i in the system produce a transition $i+1 \rightarrow i$.
- Since arrivals and departures are unitary, the number of transitions $i \to i+1$ and $i+1 \to i$ can only differ in 1: $|A_i(t) D_i(t)| \le 1$. Note that N(t) = A(t) D(t).
- For a stable queue: $A(t) D(t) < \infty$



Queuing Theory

Proof of the Level Crossing Law Theorem

Proof of the Level Crossing Law Theorem

- We have:
 - $A_i(t) = \{\text{number of arrivals finding } i \text{ customer in the system}\}$
 - $D_i(t) = \{\text{number of departures leaving } i \text{ customers in the } i \text{ cu$ system}
 - $P\{\text{a customer finds } i \text{ in the system}\} = \lim_{t\to\infty} A_i(t)/A(t)$
 - $P\{a \text{ customer leave } i \text{ in the system}\} = \lim_{t\to\infty} D_i(t)/D(t)$
 - $A_i(t) D_i(t) \in \{0, 1\}, N(t) = A(t) D(t) < \infty.$
 - $\lim_{t\to\infty} A(t) = \infty$, $\lim_{t\to\infty} D(t) = \infty$.
- Thus:

$$\lim_{t \to \infty} \left\{ \frac{A_i(t)}{A(t)} - \frac{D_i(t)}{D(t)} \right\} = \lim_{t \to \infty} \left\{ \frac{A_i(t)}{A(t)} - \frac{D_i(t)}{A(t)} - \left(\frac{D_i(t)}{D(t)} - \frac{D_i(t)}{A(t)} \right) \right\} = \lim_{t \to \infty} \left\{ \frac{A_i(t) - D_i(t)}{A(t)} - \frac{D_i(t)}{D(t)} \frac{A(t) - D(t)}{A(t)} \right\} = 0$$



Stochastic Network Modeling (SNM)

Queuing Theory

M/G/1/K

Queue

Outline

Queuing Theory

Part IV

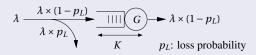
M/G/1/K Queue



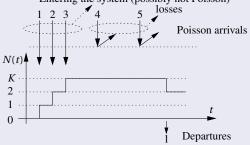
Queuing Theory

Problem Formulation

Problem Formulation



Entering the system (possibly not Poisson)





Queuing Theory

Stationary Distribution

- Using the general solution of an M/G/1/K we obtain the stationary distribution of the number in the system left by a departing customer: π_i^d .
- By the Level Crossing Law this is the stationary distribution of the number in the system found by the successful arrivals:

$$\pi_i^s = \pi_i^d, i = 0, 1, \dots K - 1.$$

and

$$\pi_i^s = P(\text{a customer entering the system finds } i)$$

 NOTE: a departing customer cannot leave the system full (nor an arrival can enter the system when it is full).



Queuing Theory

Introduction

ivotation

Little Theoren

PASTA Theoren

The M/M/1 Queue

M/G/1 Queue

M/G/1/K

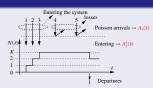
Problem Formulation
Stationary

Distribution Loss Probability

M/G/1 Delay

M/G/1 Delays

Loss Probability



Define:

- $A_i^a(t)$: Number of arrivals (lost or not) finding i in the system.
- $A_i^{\dot{s}}(t)$: Number of successful arrivals finding i in the system.
- π_i^a , π_i^s the stationary distribution of the embedded Markov chains $A_i^a(t)$, $A_i^s(t)$. By PASTA π_i^a is also the stationary distribution of the continuous time process. Thus,

 $\pi_i^s = P(\text{a customer entering the system finds } i), i = 0, 1, \dots K - 1 \Rightarrow$

$$\pi_i^s = \lim_{t \to \infty} \frac{A_i^s(t)}{\sum_{k=0}^{K-1} A_k^s(t)} \frac{\sum_{k=0}^K A_k^a(t)}{\sum_{k=0}^K A_k^a(t)} = \frac{\pi_i^a}{\sum_{k=0}^{K-1} \pi_i^a} = \frac{\pi_i^a}{1 - \pi_K^a} = \frac{\pi_i^a}{1 - p_L}, \Rightarrow$$

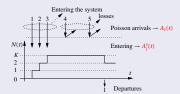
$$\pi_i^a = \pi_i^s (1 - p_L) = \pi_i^d (1 - p_L), i = 0, 1, \dots \frac{K - 1}{I}$$



Queuing Theory

Loss Probability

Loss Probability



- Applying Little: $\rho_S = E[N_S] = 1 \pi_0 = \lambda (1 p_L) E[S] = \rho (1 p_L)$. Where $\rho = \lambda E[S]$ and π_0 is the proportion of time the server is empty.
- Using PASTA: $\pi_0 = \pi_0^a$ (Poisson arrivals). Using $\pi_i^a = \pi_i^d (1 p_I)$:

$$\begin{vmatrix} 1 - \pi_0 = 1 - \pi_0^a = 1 - \pi_0^d (1 - p_L) \\ 1 - \pi_0 = \rho (1 - p_L) \end{vmatrix} \Rightarrow p_L = \frac{\rho + \pi_0^d - 1}{\rho + \pi_0^d}, \rho = \lambda E[S]$$

• Where π_0^d is computed using the general solution of an M/G/1/K.



Stochastic Network Modeling (SNM)

Queuing Theory

Introduction

....

PASTA Theorem

The M/M/1

M/G/1 Oueur

mrori Queuc

M/G/1 Busy

Period Period

M/G/1 Delay

Queues in Tandem

Network

Queues

Part IV

Queuing Theory

Outline

- Introduction
- Kendal Notation
- Little Theorem
- PASTA Theorem
- The M/M/1 Oueue
- M/G/1 Queue

- M/G/1/K Onene
- M/G/1 Busy Period
- M/G/1 Delays
- Queues in Tandem
- Networks of Queues
- Matrix Geometric Method



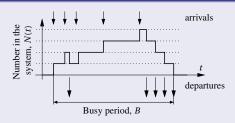
M/G/1 Busy Period

Queuing Theory

M/G/1 Busy

Period

Expected Length of a Busy Period



- Define the RV:
 - Busy period, B.
 - Idle period, I. Poisson arrivals with rate $\lambda \Rightarrow E[I] = 1/\lambda$
- Clearly:

System load
$$\rho = \lambda E[S] = \frac{E[B]}{E[I] + E[B]} \Rightarrow E[B] = \frac{1}{\lambda} \frac{\rho}{1 - \rho}$$



M/G/1 Busy Period

Queuing Theory

Introduction

introductio

Little Theorem

_ . . _ .

Theorem

The M/M/I Queue

M/G/1 Oueu

M/G/1 Busy Period

M/G/1 Delay

Queues in Tandem

Networks of

M/G/1/K Busy Period

$$\lambda \xrightarrow{\lambda \times (1-p_L)} \overline{\qquad \qquad } \lambda \times p_L$$

$$\downarrow \qquad \qquad \downarrow \qquad \qquad$$

- Busy period, *B*.
- Idle period, I. Poisson arrivals with rate $\lambda \Rightarrow E[I] = 1/\lambda$
- Clearly:

System load
$$\rho_s = \lambda (1 - p_L) E[S] = \frac{E[B]}{E[I] + E[B]} \Rightarrow$$

$$E[B] = \frac{1}{\lambda} \frac{\rho (1 - p_L)}{1 - \rho (1 - p_L)}, \rho = \lambda E[S]$$

• Or, in terms of $\pi_0 = \pi_0^d (1 - p_L)$: System load $\rho_s = 1 - \pi_0 = \frac{E[B]}{E[I] + E[B]} \Rightarrow E[B] = \frac{1}{\lambda} \frac{1 - \pi_0}{\pi_0}$



Master in Innovation and Research in Informatics (MIRI) Computer Networks and Distributed Systems

Stochastic Network Modeling (SNM)

Queuing Theory

Part IV

Queuing Theory

M/G/1 Delays

Outline

M/G/1 Delays



Queuing Theory

Introductio

Notation

Little Theorei

PASTA Theoren

The M/M/1 Oueue

M/G/1 Onem

M/G/1 Queut

M/G/1 Bus

M/G/1 Delays M/G/1 Mean Time in

M/G/1 Mean Time i the Queue

M/G/1 Mean Time in the Queue

• Method of the moments: Using PASTA, the mean time in the queue (W) for an arriving customer, is the mean time to finish the current service (mean residual time, R) plus the mean time to service the customers in the queue ($E[S]N_O$):

$$W = R + E[S] N_Q$$

• Using Little for the queue length:

$$N_Q = \lambda \: W \Rightarrow W = R + E[S] \: \lambda \: W \Rightarrow W = \frac{R}{1-\rho}, \: \rho = \lambda \: E[S].$$



Queuing Theory

Introduction

Little Theore

PASTA Theoren

The M/M/I

M/G/1 Oueur

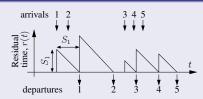
M/G/1 Queut

M/G/1 Busy

M/G/1 Delays

the Queue Queues in

M/G/1 Mean Time in the Queue



• From the figure (note the right triangles with two equal cathetus), we have:

$$\mathbf{R} = \frac{1}{t} \int_0^t r(\tau) d\tau = \frac{1}{t} \sum_{i=1}^{A(t)} \frac{S_i^2}{2} = \frac{1}{2} \frac{A(t)}{t} \sum_{i=1}^{A(t)} \frac{S_i^2}{A(t)} = \frac{1}{2} \lambda \operatorname{E}[S^2]$$



Queuing Theory

Introduction

Introductio

Little Theore

....

PASTA Theorem

The M/M/1 Queue

M/G/1 Queue

mi di i Queuc

M/G/1 Bus

M/G/1 Delays
M/G/1 Mean Time in
the Queue

the Queue Queues in

M/G/1 Mean Time in the Queue

• For instance, for an M/M/1

$$E[S^2] = Var(S) + E[S]^2 = \frac{1}{\mu^2} + \left(\frac{1}{\mu}\right)^2 = \frac{2}{\mu^2},$$

thus, the residual time is:

$$R = \frac{1}{2}\lambda \operatorname{E}[S^2] = \frac{\lambda}{\mu^2} = \frac{\rho}{\mu}, \, \rho = \frac{\lambda}{\mu}.$$

• We can check that E[R|S idle] = 0 and $E[R|S \text{ busy}] = 1/\mu$, thus

$$R = \mathrm{E}[R|\mathrm{S}\ \mathrm{idle}]\ \pi_0 + \mathrm{E}[R|\mathrm{S}\ \mathrm{busy}]\ (1-\pi_0) = \frac{\rho}{\mu},\ \rho = 1-\pi_0,$$
 as expected.



Queuing Theory

Introduction

endal

Little Theorer

......

Theorem

Queue

M/G/1 Queue

WI/G/I Queue

M/G/1 Bus

Period

M/G/1 Mean Time in the Queue

M/G/1 Mean Time in the Queue

• We have:

$$W = \frac{R}{1 - \rho}, \rho = \lambda E[S]$$

 $R = \frac{1}{2}\lambda \,\mathrm{E}[S^2]$

Substituting we get the Pollaczek-Khinchin, P-K formula:

$$W = \frac{\lambda E[S^2]}{2(1-\rho)}, \rho = \lambda E[S]$$



Queuing Theory

Introduction

PASTA Theorem

The M/M/1 Queue

M/G/1 Queue

mr dr i Queu

M/G/1 Bus

M/G/1 Delays M/G/1 Mean Time in the Queue

the Queue Queues in Fandem

M/G/1 Mean Time in the Queue

• Mean time in the system (response time):

$$T = \mathrm{E}[S] + W = \mathrm{E}[S] + \frac{\lambda \, \mathrm{E}[S^2]}{2 \, (1-\rho)}$$

- For an M/M/1 queue: $E[S^2] = \frac{2}{\mu^2} \Rightarrow W = \frac{\rho}{\mu(1-\rho)}$
- For an M/D/1 queue: $E[S^2] = \frac{1}{\mu^2} \Rightarrow W = \frac{\rho}{2\mu(1-\rho)}$
- Observation: The M/D/1 has the minimum value of $E[S^2] \Rightarrow$ is a lower bound of W, T, N_Q and N for an M/G/1.



Queuing Theory

M/G/1 Mean Time in the Queue

P-K Formula Does Not Apply to an M/G/1/K Queue

- P-K formula is not applicable to an M/G/1/K queue because the customers entering the system might not be Poisson. Thus, they does not observe the mean residual time.
- Example: Customers entering an M/G/1/1 queue (0 queue size) observe the system always empty. Thus, in an M/G/1/1 queue the expected time in the queue is W = 0 (P-K formula does not apply), and the expected time in the system is T = E[S] (mean service time).
- With an M/G/1/K we can compute $N = \sum_{n=1}^{K} n \pi_n^a$, and use Little: $N = \lambda (1 - p_L) T$. For instance, for an M/G/1/1 we have $\pi_0^d = 1$, and N = 0 $\pi_0^a + 1$ $\pi_1^a = \pi_1^a = p_L$. Thus, $p_L = \frac{\rho + \pi_0^d - 1}{\rho + \pi_0^d} = \frac{\rho}{\rho + 1}$, and $T = \frac{N}{\lambda(1-n_I)} = \frac{p_L}{\lambda(1-n_I)} = \frac{\rho}{\lambda} = E[S]$, as expected.



Master in Innovation and Research in Informatics (MIRI) Computer Networks and Distributed Systems

Stochastic Network Modeling (SNM)

Queuing Theory

Part IV

Queuing Theory

Outline

- **Oueues in Tandem**

Oueues in Tandem



Queuing Theory

Introduction

v 1...1 mm

PASTA

The M/M/

M/G/1 Queue

M/G/1 Bus

M/G/1 Delay

Queues in Tandem Burke theorem

Burke theorem

- The departure process in an M/M/m queue, $1 \le m \le \infty$, is a Poisson process with the same parameter than the arrival process.
- At each time t, the number of customers in the system is independent of the sequence of departures previous to t.



Queuing Theory

Introduction

Little Theorem

PASTA Theorem

The M/M/1 Queue

M/G/1 Queu

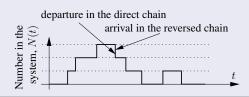
M/G/1 Bus

M/G/1 Delay

Queues in Tandem Burke theorem

Burke theorem. Proof (1)

- Relation between the arrival and departure process:
 The departure process in a reversible queue has the same joint distribution than the arrival process.
- Proof:
 - If the queue is reversible: q_{ij} = q^r_{ij} ⇒ the arrival process in the reversed chain has the same distribution than the arrival process in the direct chain,
 - but:





Queuing Theory

Introduction

Little Theorem

PASTA

The M/M/1

M/G/1 Queu

M/G/1 Bus

M/G/1 Delay

Queues in Tandem Burke theorem

Burke theorem. Proof (2)

- The queue M/M/m is reversible ⇒ The departures are Poisson with the same parameter than the arrivals.
- The arrivals in the reversed chain previous to *t* are
 Markovian, thus, independent of the number of customers
 in the system after *t*. This implies that the departures in the
 direct chain are independent of the number in the system
 before *t*. □



Queuing Theory

Introductio

muoductio

Little Theorem

PASTA Theoren

The M/M/I Queue

M/G/1 Queu

M/G/1 Bus

M/G/1 Delays

Queues in Tandem Burke theorem

Tandem M/M/m Queues

Define the chain:

 $X(n,m) = \{n \text{ in the system 1}, m \text{ in the system 2}\}\$

• The stationary distribution is the product of the stationary distributions of the isolated queues:

$$\pi_{nm} = (1 - \rho_1) \rho_1^n (1 - \rho_2) \rho_2^m, \rho_1 = \lambda/\mu_1, \rho_2 = \lambda/\mu_2$$

Proof: Using Burke, the departures of system 1 are Poisson and the number in the system 1 is independent of the previous departures (arrivals to system 2), thus, independent from the number of customers in system 2.



Queuing Theory

Introduction

Little Theore

Theorem

The M/M/I Queue

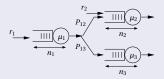
M/G/1 Queu

M/G/1 Bus

M/G/1 Delay

Queues in Tandem Burke theorem

Feed Forward Queues



- Suppose M/M/1 queues with outside arrivals with rate r_i randomly forwarded with probabilities P_{ii} (see figure).
- The network has the following product form solution:

$$\pi(n_1, n_2, \dots n_K) = (1 - \rho_1) \, \rho_1^{n_1} (1 - \rho_2) \, \rho_2^{n_2} \dots (1 - \rho_k) \rho_k^{n_k},$$
$$\rho_i = \lambda_i / \mu_i.$$

- The rates λ_i are computed solving: $\lambda_i = r_i + \sum_j \lambda_j P_{ji}$.
- Stability condition: $\rho_i < 1$.



Queuing Theory

Introductior

Proof (draft)

Little Theorem

PASTA Theorem

The M/M/1 Queue

M/G/1 Queuϵ

MIC/I Buo

Period

M/G/1 Delay

Queues in Tandem

Burke theorem

Feed Forward Queues

_ . .

- Burke theorem.
- Superposition of Poisson processes with rates λ_i is Poisson with rate $\sum_i \lambda_i$.
- A Poisson process with rate λ randomly split with probabilities p_i , $\sum_i p_i = 1$, produce Poisson processes with rates $p_i \lambda$.