Introduction to Queueing Theory with Applications to Air Transportation Systems

John Shortle
George Mason University

November 7, 2011

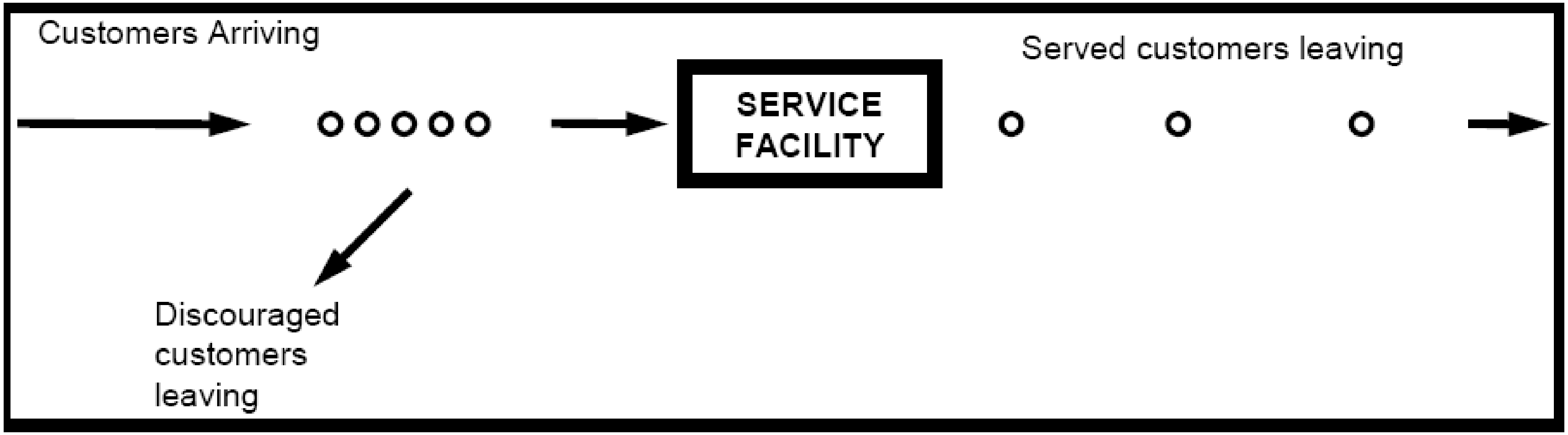
Outline

- **Why stochastic models matter**
- M/M/1 queue
- Little's law
- Priority queues
- Simulation: Lindley's equation
- A simple airport model

Why Stochastic Models Matter

- Often, we simplify models by replacing stochastic values with their averages
- Sometimes this is a reasonable approximation
- But it also has the potential to drastically change the behavior of the model

Typical Queueing Process



<u>Common Notation</u>

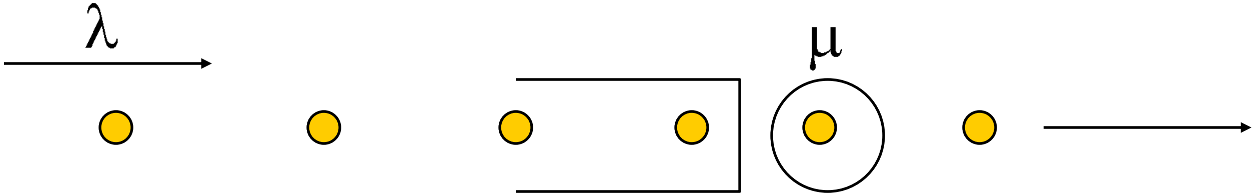
λ: Arrival Rate (e.g., customer arrivals per hour)

μ: Service Rate (e.g., service completions per hour)

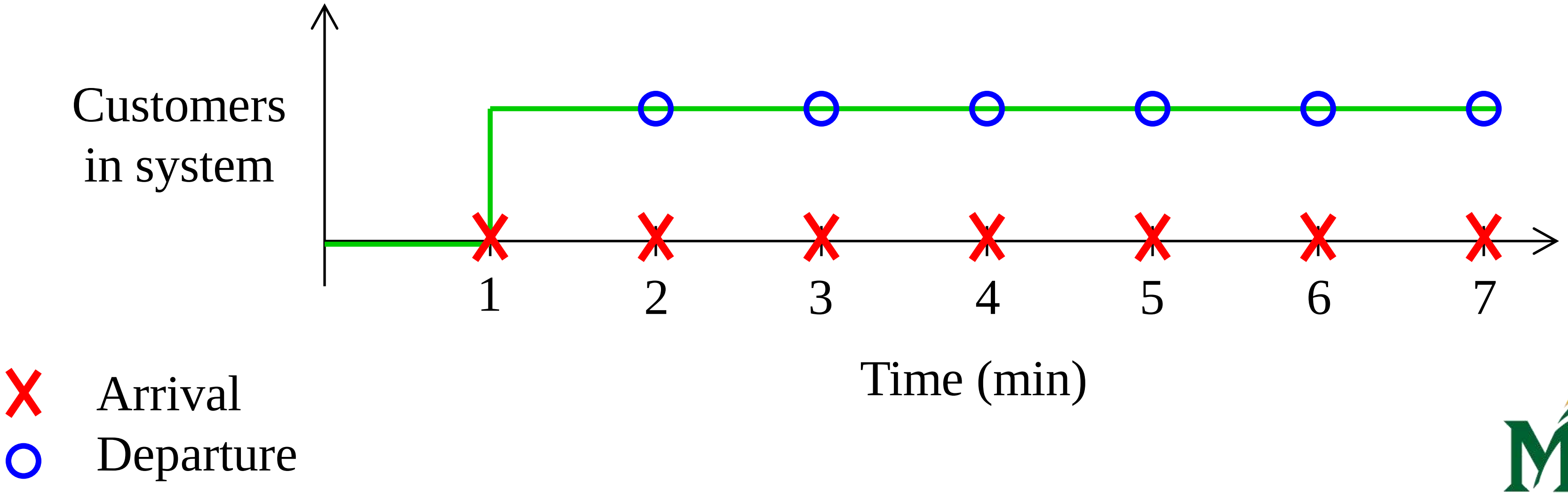
1/μ: Expected time to complete service for one customer

ρ: Utilization: $\rho = \lambda / \mu$

A Simple Deterministic Queue

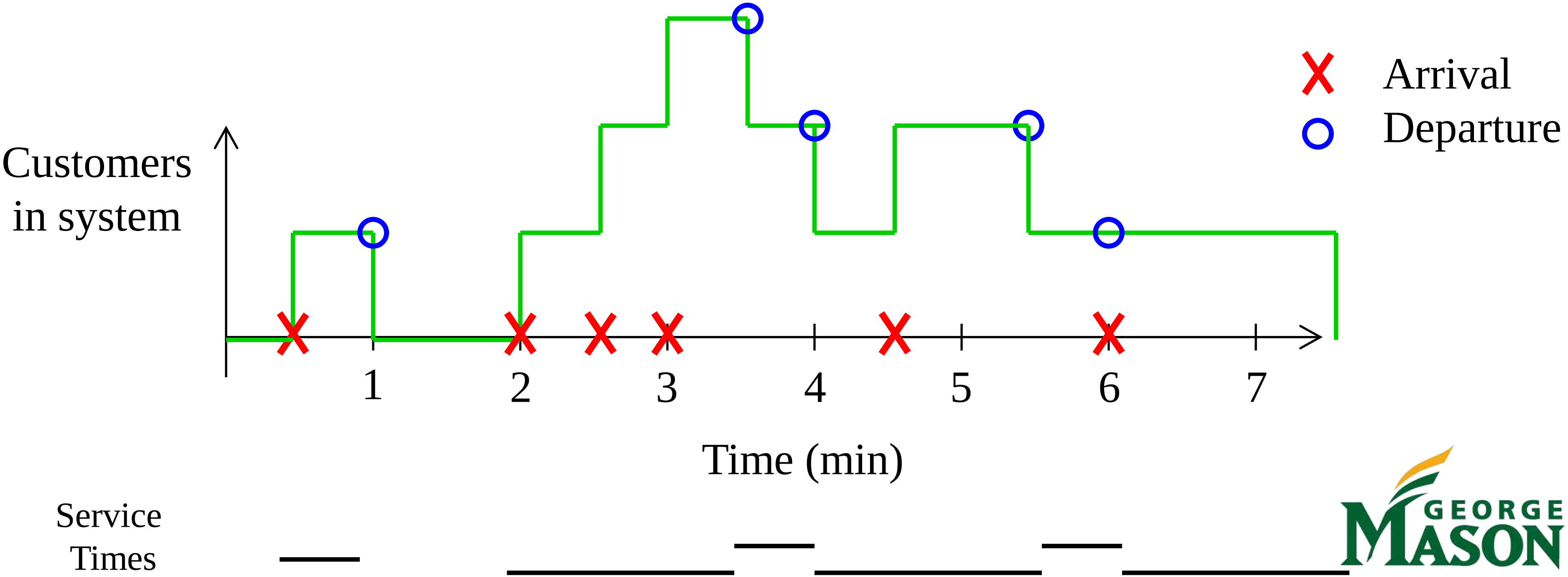


- Customers arrive at 1 min, 2 min, 3 min, etc.
- Service times are exactly 1 minute.
- What happens?

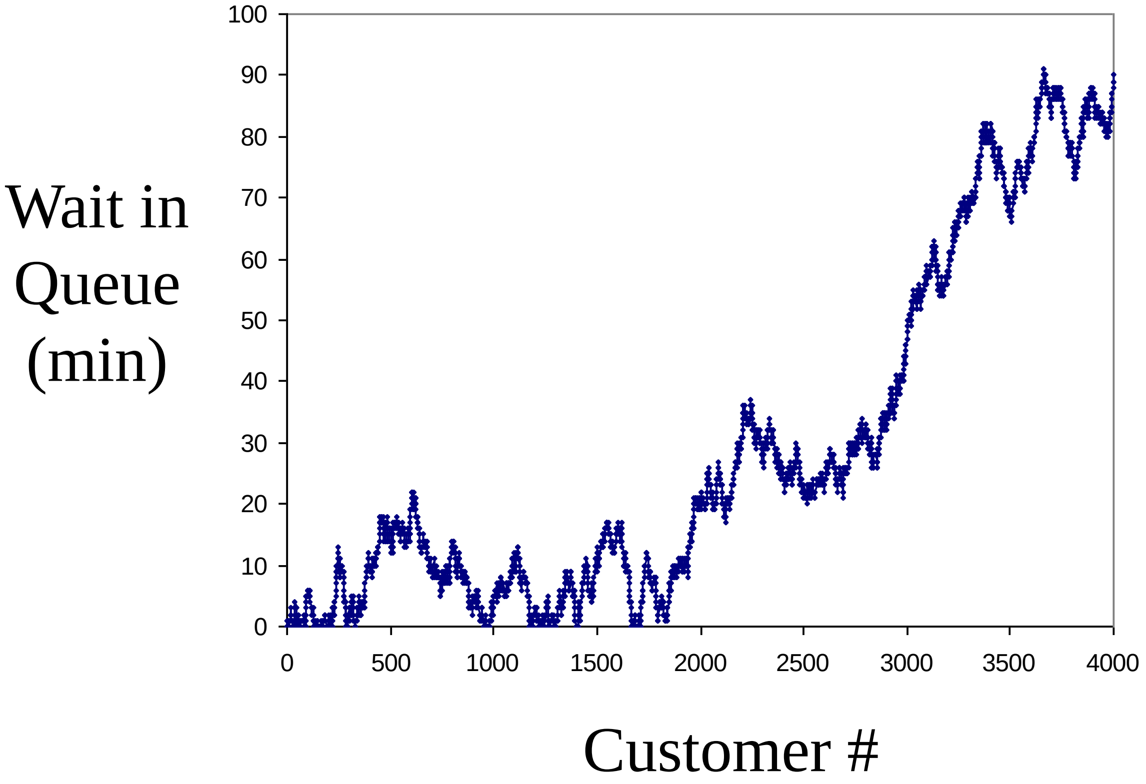


A Stochastic Queue

- Times between arrivals are ½ min. or 1½ min. (50% each)
- Service times are ½ min. or 1½ min. (50% each)
- Average inter-arrival time = 1 minute
- Average service time = 1 minute
- What happens?



Stochastic Queue in the Limit



- Two queues with same <u>average</u> arrival and service rates
  - Deterministic queue: <u>zero</u> wait in queue for <u>every</u> customer
  - Stochastic queue: wait in queue grows <u>without bound</u>
- Variance is an enemy of queueing systems

Outline

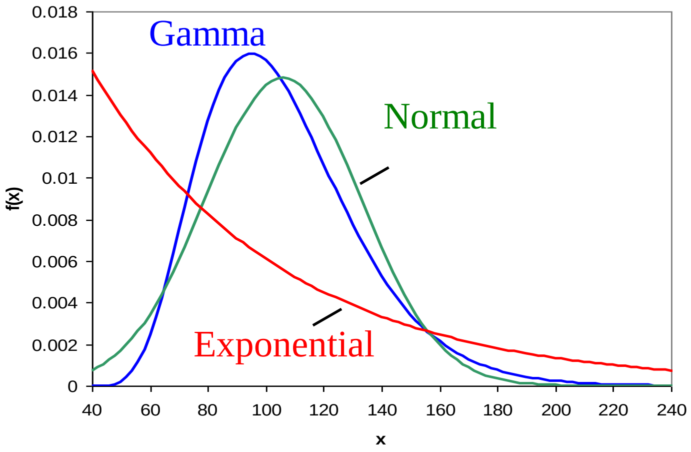
- Why stochastic models matter
- **M/M/1 queue**
- Little’s law
- Priority queues
- Simulation: Lindley’s equation
- A simple airport model

The M/M/1 Queue

Inter-arrival times follow an exponential distribution (or arrival process is Poisson)

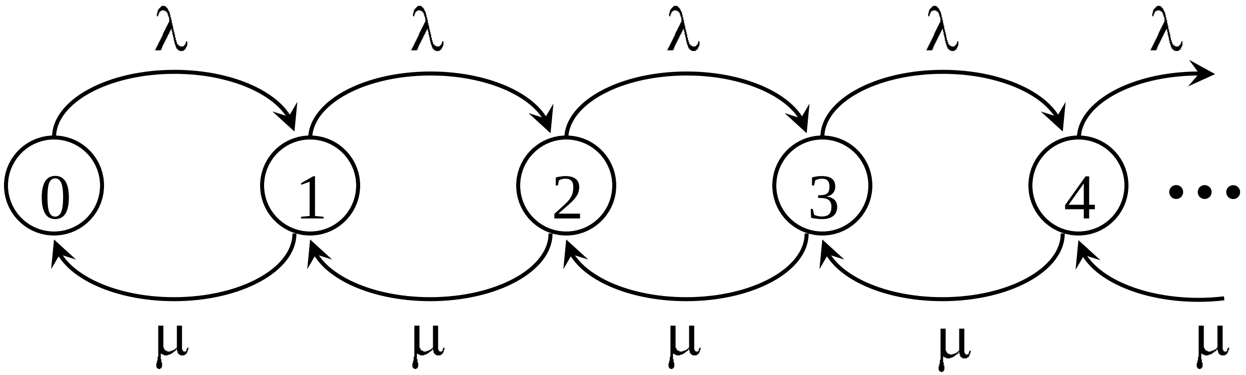
Service times follow an exponential distribution

A single server



M/M/1 Queue

- Inter-arrival times are exponential with rate λ
- Service times are exponential with rate μ
- First-come-first-served discipline
- All random variables are independent
- Infinite queue, no balking, reneging, ...



Number in System

M/M/1 Queue

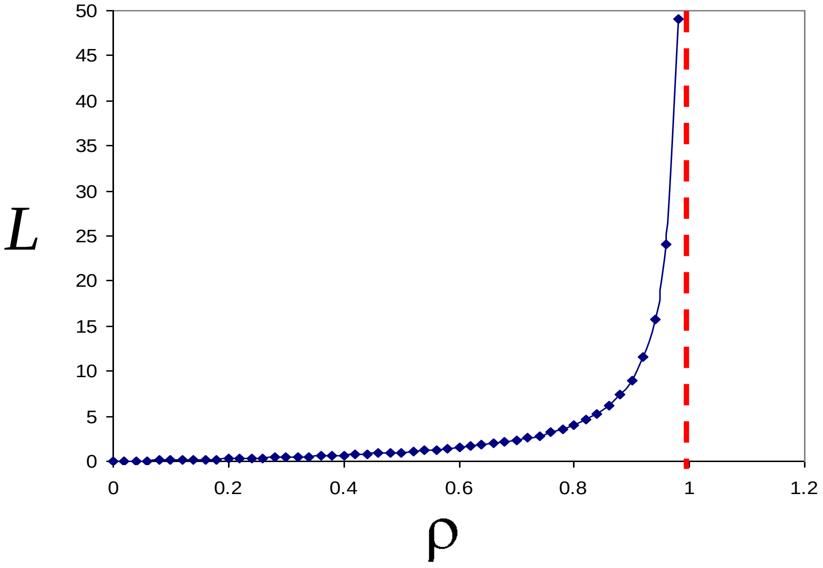
- Can solve system as a *continuous time Markov chain*
- A variety of queueing metrics can be calculated analytically (in steady state)

$L_q = \frac{\rho^2}{1-\rho}$	Average # in queue
$L = \frac{\rho}{1-\rho}$	Average # in system
$W_q = \frac{\rho}{1-\rho} \cdot \frac{1}{\mu}$	Average wait in queue
$W = \frac{1}{\mu - \lambda}$	Average wait in system

The M/M/1 Queue

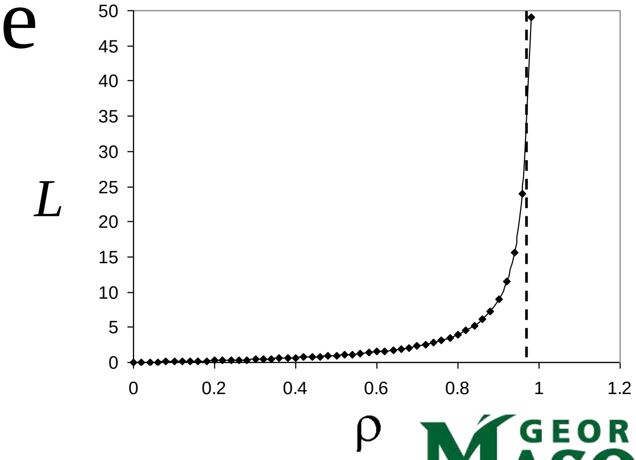
- Observations
  - 100% utilization is not desired
- Limitations
  - Model assumes *steady-state*. Solution does not exist when $\rho > 1$ (arrival rate exceed service rate).
  - Poisson arrivals can be a reasonable assumption
  - Exponential service distribution is usually a bad assumption.

$$L = \frac{\rho}{1-\rho}$$



Insights

- Fraction of “wasted time” = 1 – ρ
  - Unused airport capacity, empty seats, etc.
  - Tension between wasted time and delays
- Generally a bad idea to have ρ near 1
  - Higher delays and higher variability
- Results are true in steady state

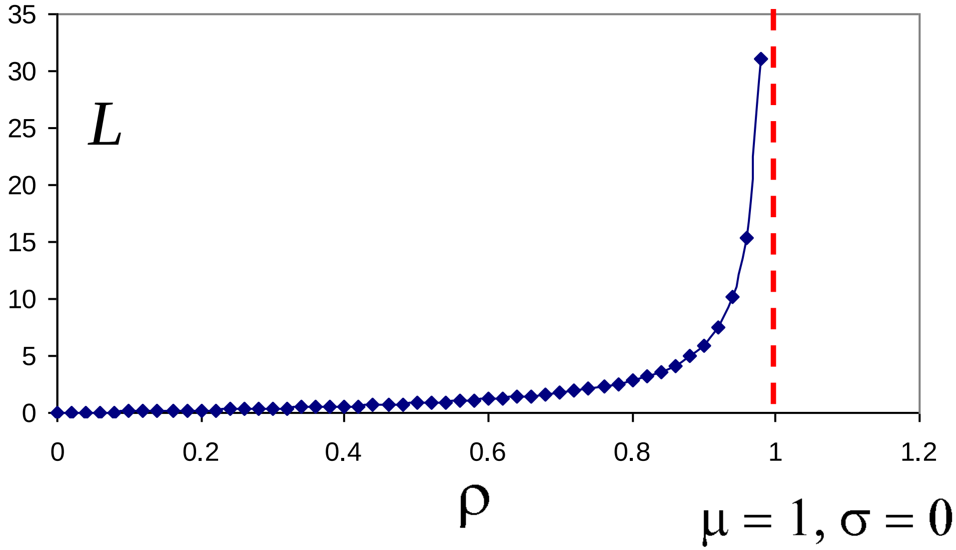


The M/G/1 Queue

Service times follow a *general* distribution

Required inputs:

- λ: arrival rate
- 1/μ: expected service time
- σ: std. dev. of service time

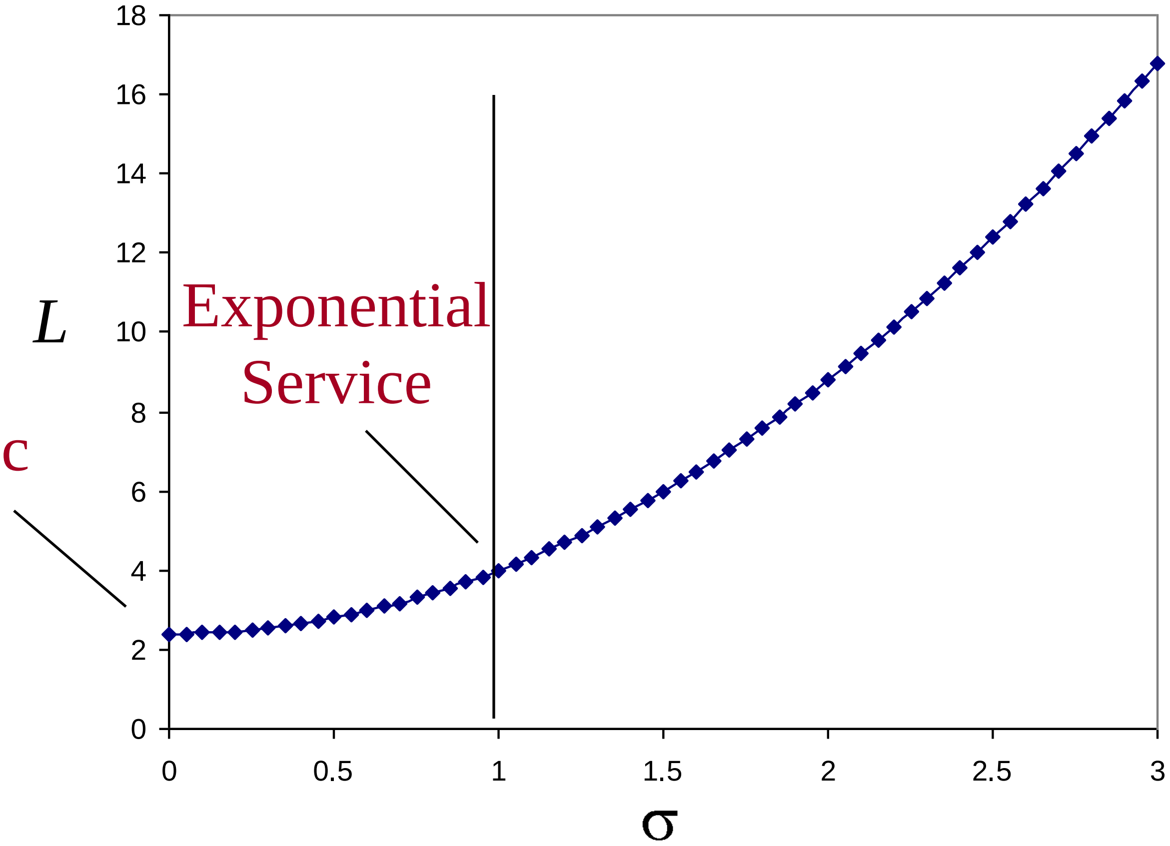


μ = 1, σ = 0.5

$$\boxed{L = \rho + \frac{\rho^2 + \lambda^2\sigma^2}{2(1-\rho)}}$$

Avg. # in System

M/G/1: Effect of Variance



Exponential Service

Deterministic Service

Arrival Rate
Service Rate
Held
Constant

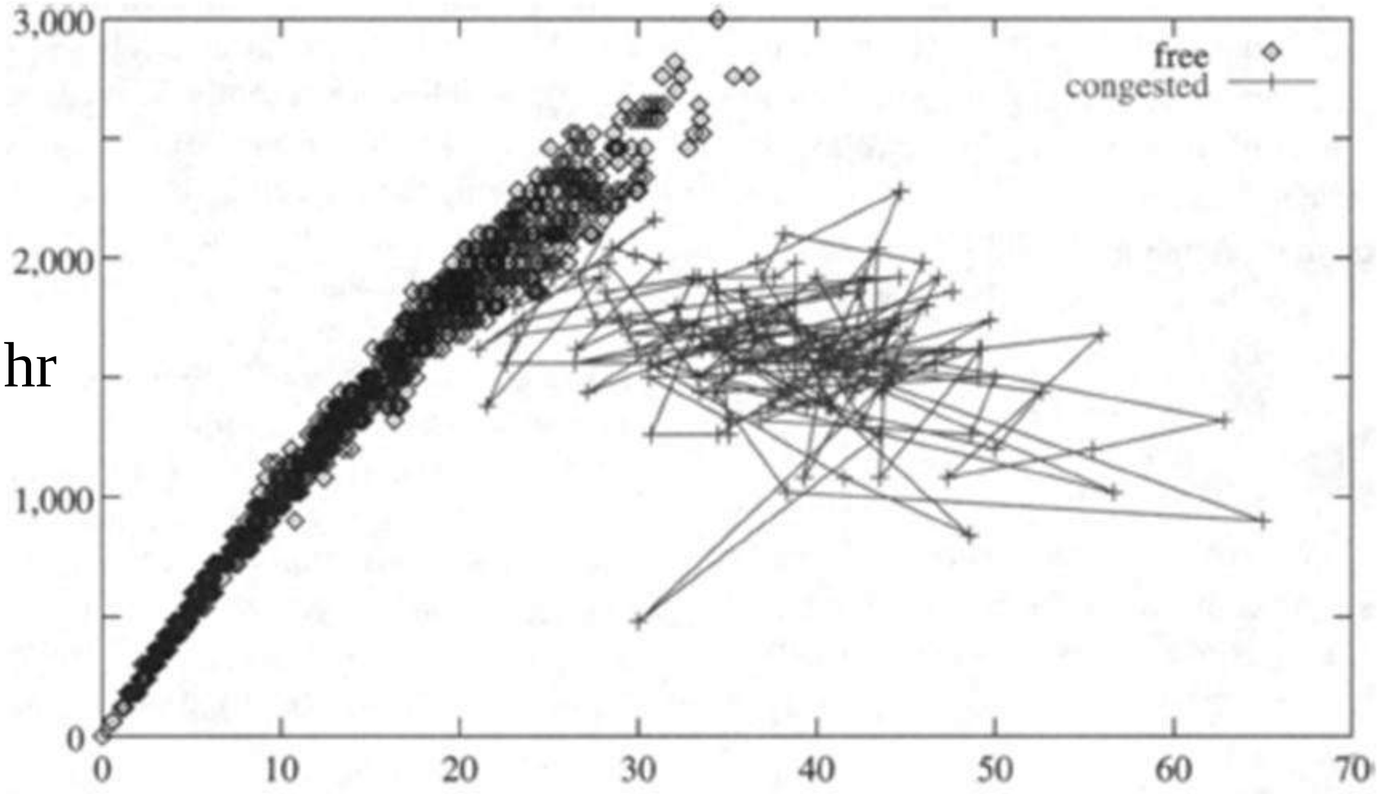
$$L = \rho + \frac{\rho^2 + \lambda^2\sigma^2}{2(1-\rho)}$$

$\lambda = 0.8,\ \mu = 1\ (\rho = 0.8)$

What Happens When ρ > 1?

Road Traffic Example

Vehicles / hr



Vehicles / km

Nagel, K., P. Wagner, R. Woesler. 2003. Still flowing: approaches to traffic flow and traffic jam modeling. *Operations Research*, 51(5), 681-710.

Limitations of Analytical Models

- Typically assume *steady state*
- Typically assume *stationary* process
  - No time of day, day of week, ... effects
- Basic models require *exponential distribution*
  - Often this assumption can be relaxes
- Typically assume *independence* of arrival times (and service times)
- Many results exist for *single-server* queues

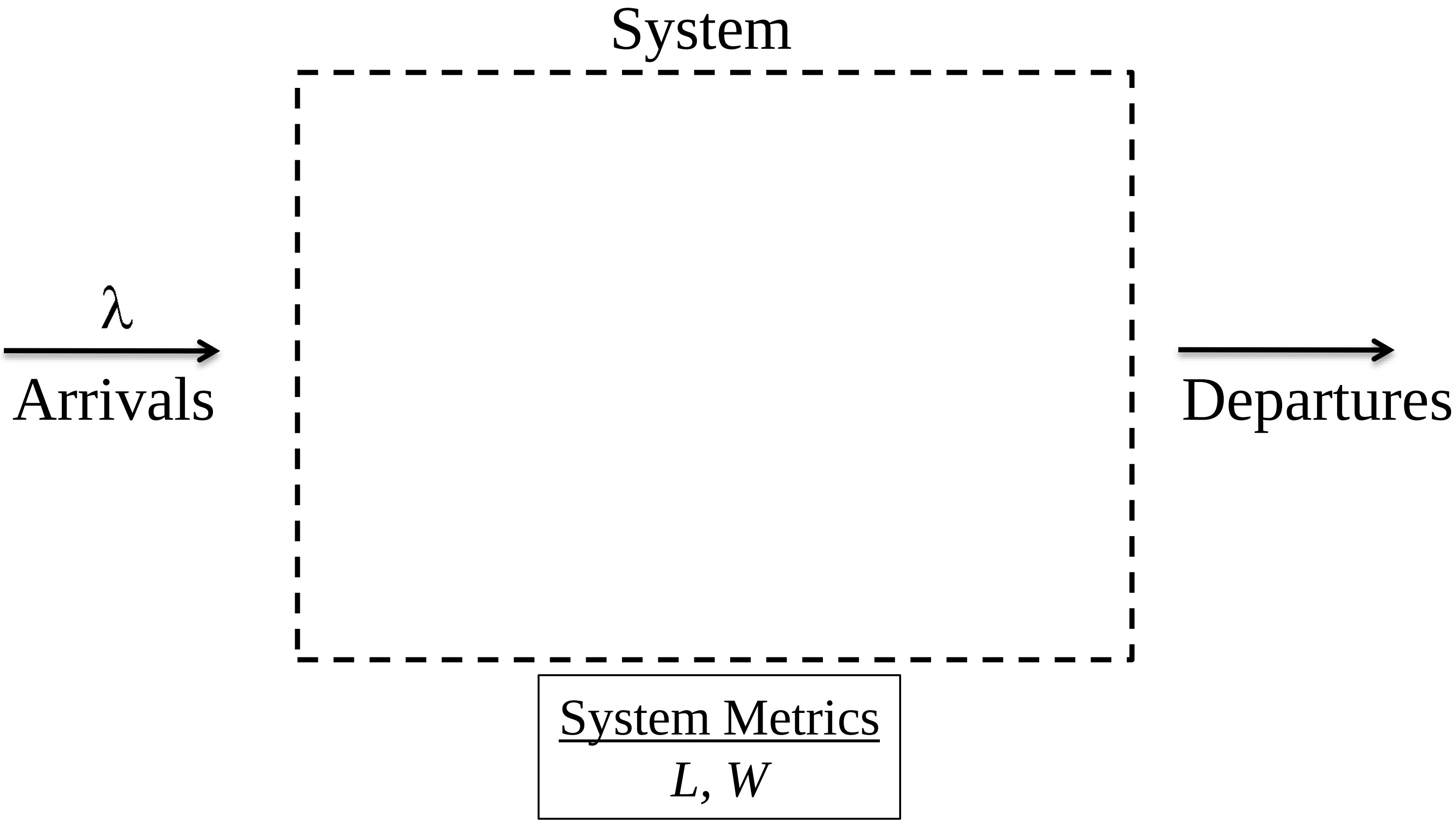
Outline

- Why stochastic models matter
- M/M/1 queue
- **Little's law**
- Priority queues
- Simulation: Lindley's equation
- A simple airport model

Little's Law

- $L = \lambda W$
  - L = average # in system
  - W = average time in system
  - λ = average arrival rate
- Little's law does *not* require
  - Poisson arrivals,
  - Independent arrivals
  - First-come-first-served discipline
  - Exponential service
  - Etc.
- Minimum "physics" required
  - Sequence of arrivals to a system
  - Sequence of departures from a system
  - Stable long-run average behavior

Little's Law



System

λ

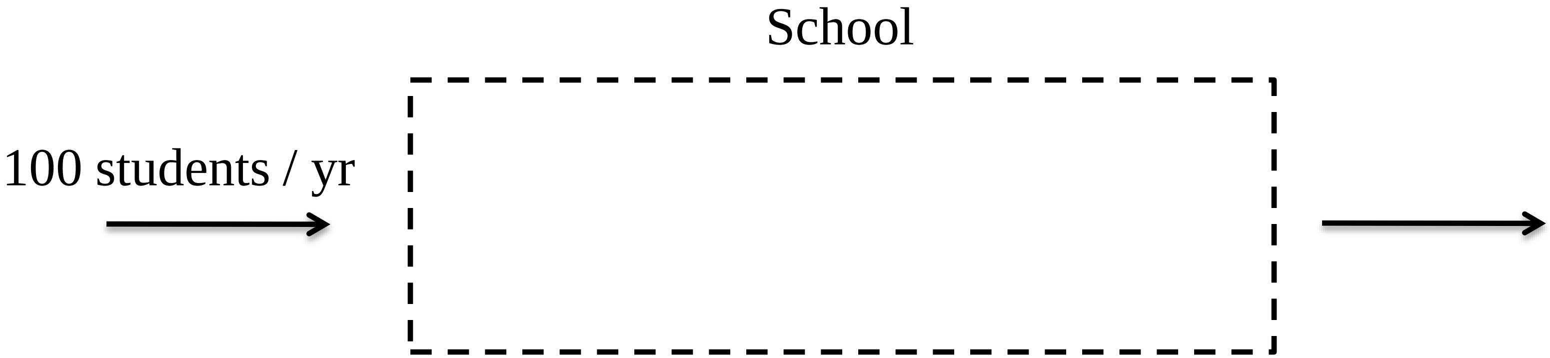
Arrivals

Departures

<u>System Metrics</u>

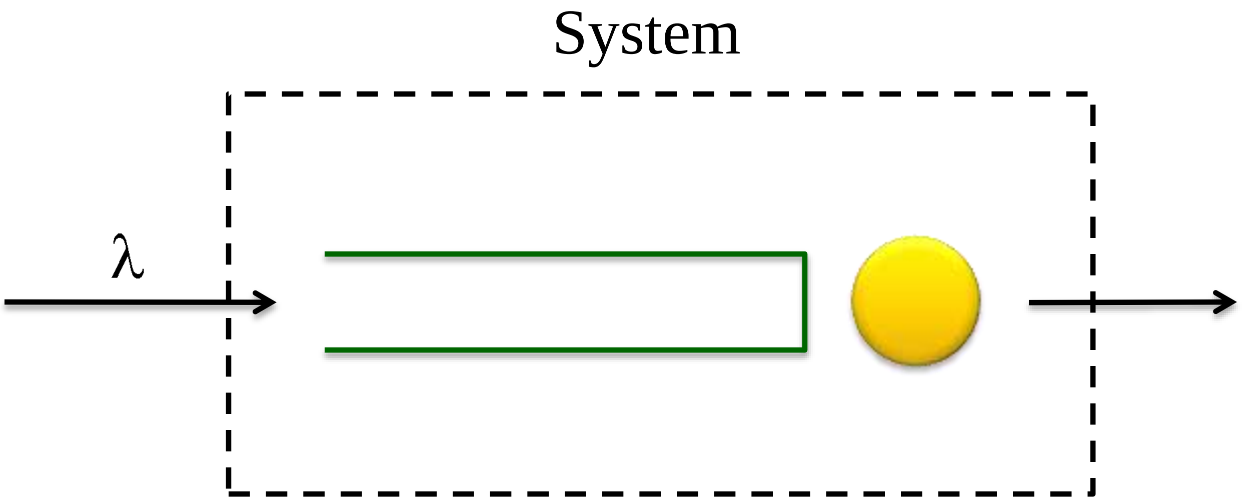
L, W

Example: Non-Queueing



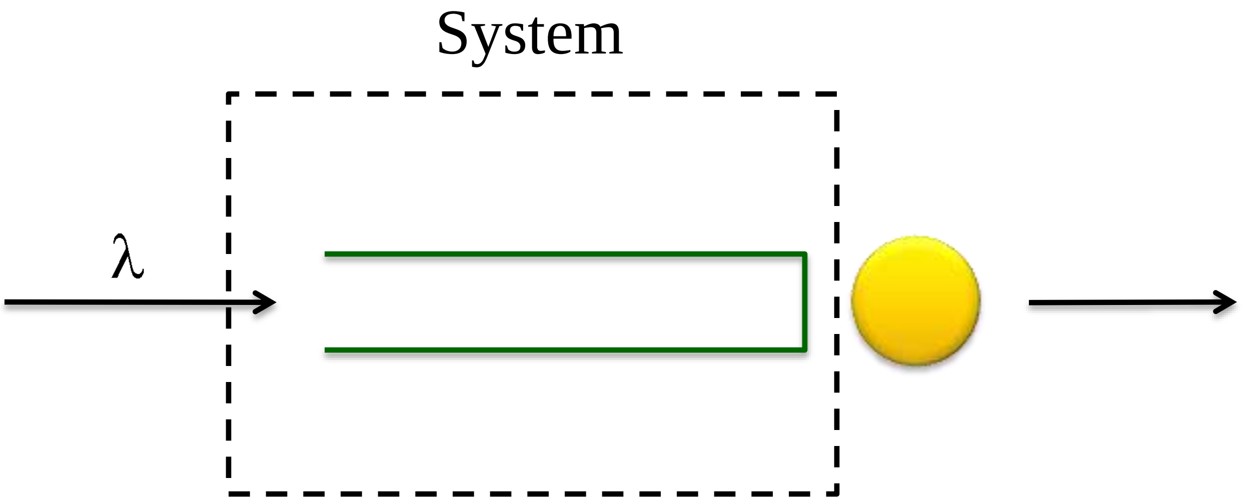
- Four years per student
- How many students in school on average?

Example #1



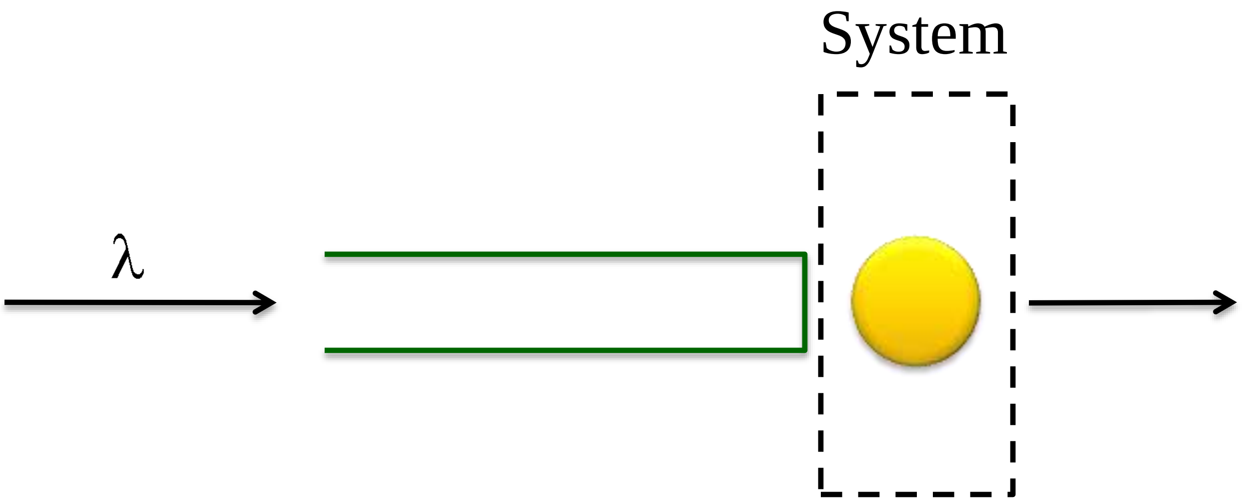
$$L = \lambda W$$

Example #2



$$L_q = \lambda\, W_q$$

Example #3: Single-Server



$$\underset{\substack{\uparrow \\ \text{“}L\text{”}}}{\text{Fraction of time server busy}} = \lambda \underset{\substack{\uparrow \\ \text{“}W\text{”}}}{(1 / \mu)} = \rho$$

Outline

- Why stochastic models matter
- M/M/1 queue
- Little’s law
- **Priority queues**
- Simulation: Lindley’s equation
- A simple airport model

Priority Queues

- Some basic analytical results exist for priority models
- Assumptions
  - n arrival types, each with arrival rate λ_i and service rate μ_i ($\rho_i = \lambda_i / \mu_i$) (exponential distributions)
  - After service completion, customers of type i served before customers of type $j > i$

$$W_q^{(i)} = \frac{\sum_{k=1}^{n} \rho_k / \mu_k}{(1-\sigma_{i-1})(1-\sigma_i)} \qquad \sigma_i \equiv \rho_1 + \rho_2 + \ldots + \rho_i$$

So What?

- Expected wait in queue is minimized by giving priority to the customer class with the shortest average service time
- Generalization: cμ-rule: If each customer class incurs a cost *c* per time and is served with rate μ, the expected cost in queue is minimized by giving priority to the customer class with the lowest value of cμ
  - E.g., to minimize passenger delay (*c* proportional to # of passengers on a plane), land planes with higher numbers of passengers first

Seems Like a Good Idea, But...

- First-come-first-served is "optimal" in the sense that it minimizes higher moments (e.g., variance).
  - Any scheme of priorities not depending on service times makes all higher moments worse than FCFS
- That is, a priority rule may increase the *variance* of delay
  - Priority customers have small delay
  - Non-priority customers have large delay

Outline

- Why stochastic models matter
- M/M/1 queue
- Little’s law
- Priority queues
- **Simulation: Lindley’s equation**
- A simple airport model

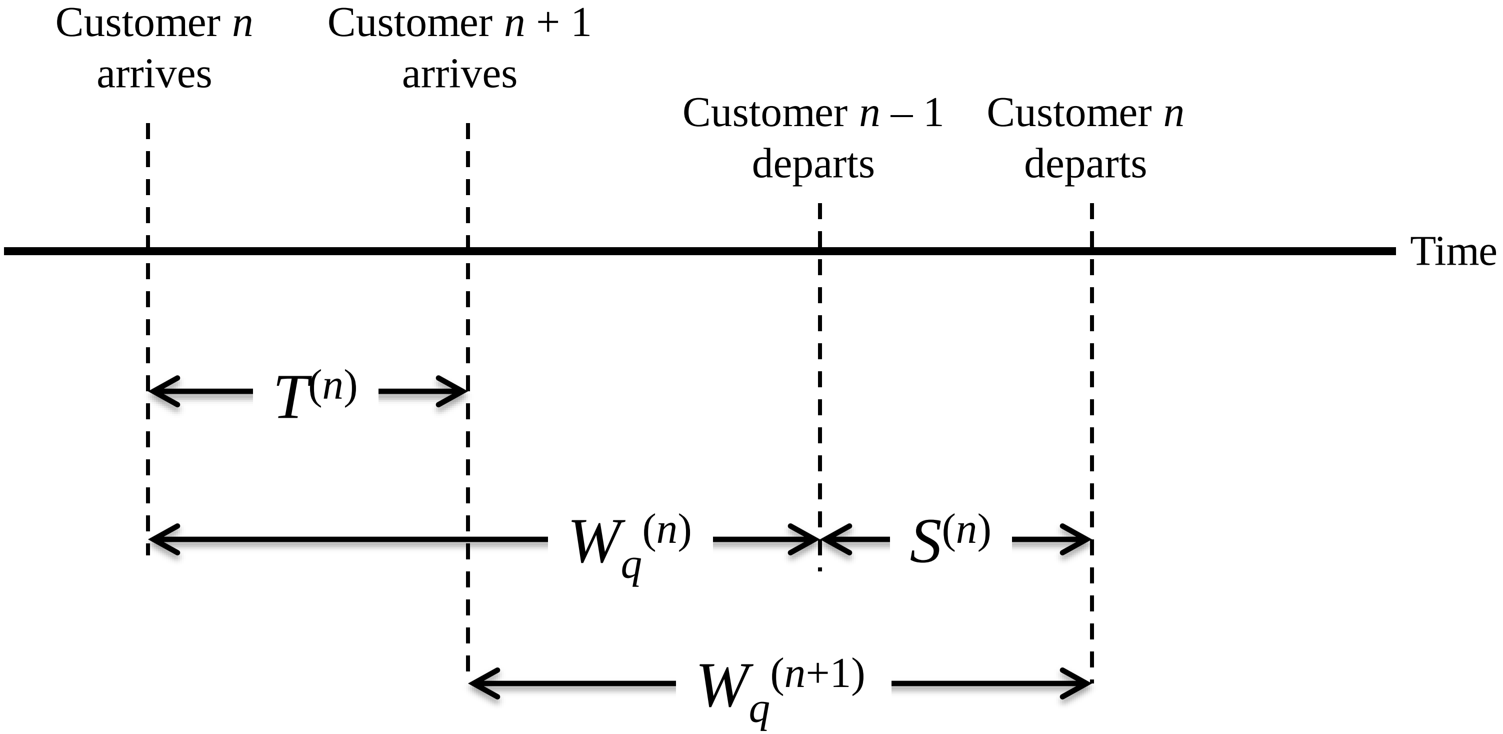
Simulation Models

- Fairly easy to think up scenarios in which analytical results do not exist
- This lecture: Lindley’s equation
  - Can be used as a building block in more complicated models
  - Helps to understand the basic “physics” of queueing models

Single-Server Queue

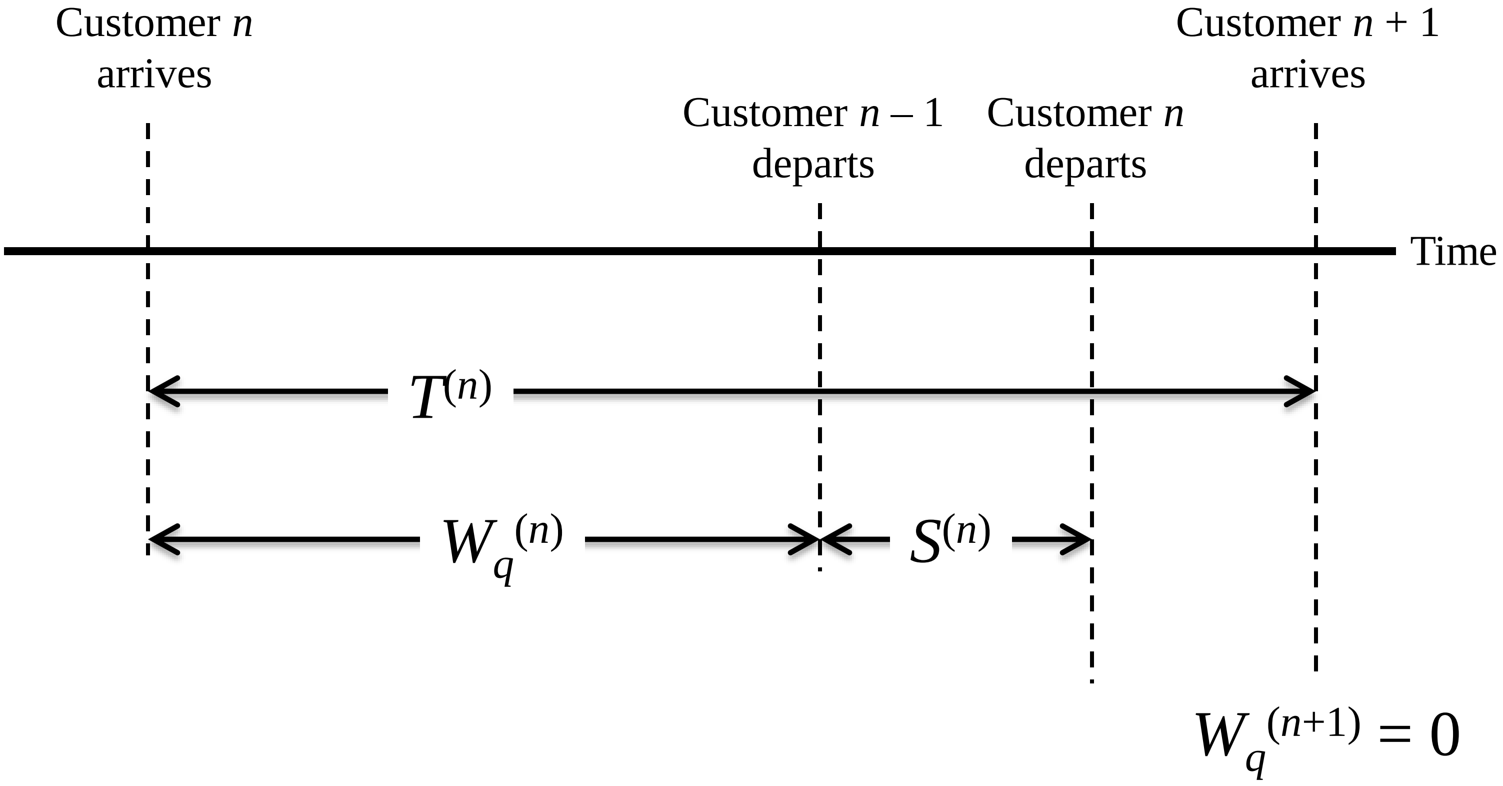
- FCFS single-server queue is a fundamental unit in many queueing models
- Only requires a sequence of arrival times and service times
  - Times do not have to follow an exponential distribution
  - Nor do they have to follow *any* distribution
  - Stationarity, independence, etc. not required
- Definitions
  - $T^{(n)}$: Time between arrivals of customers n and $n+1$
  - $S^{(n)}$: Service time of customer n
  - $W_q^{(n)}$: Waiting time in queue of customer n

Lindley's Equation: Case 1



$$\boxed{W_q^{(n+1)} = W_q^{(n)} + S^{(n)} - T^{(n)}}$$

Lindley's Equation: Case 2



Lindley’s Equation

$$\boxed{W_q^{(n+1)} = \max(W_q^{(n)} + S^{(n)} - T^{(n)}, 0)}$$

Wait of previous
customer

Inter-arrival
time

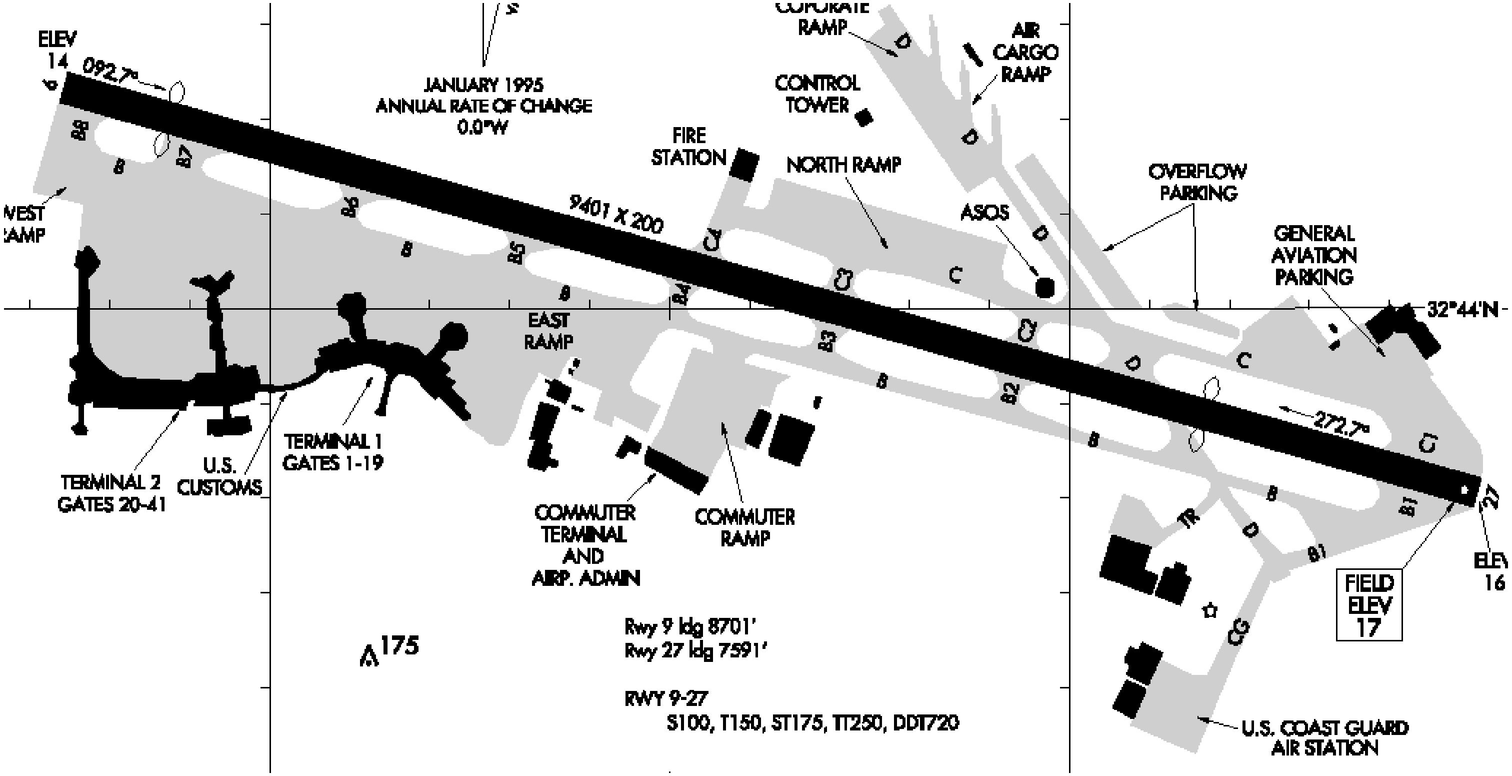
Service time
Of that customer

No “negative”
wait

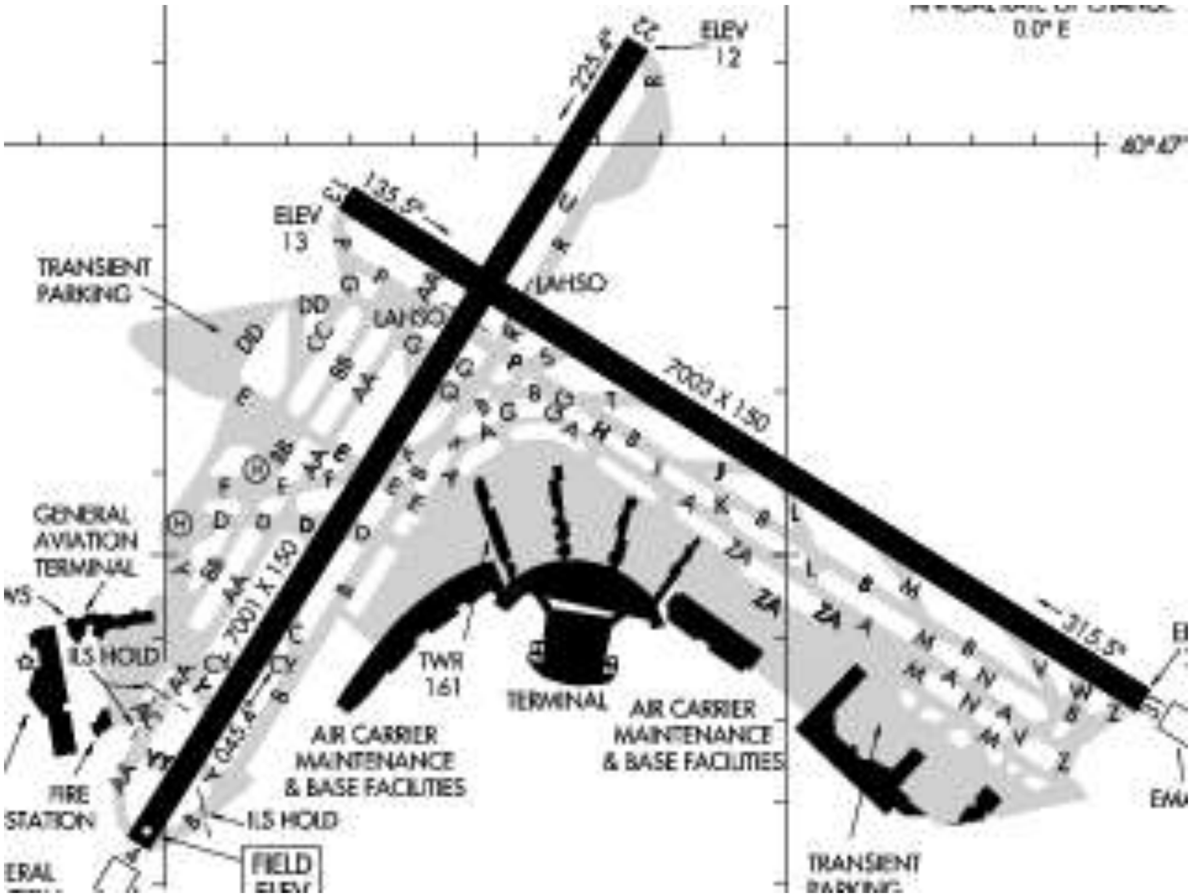
Outline

- Why stochastic models matter
- M/M/1 queue
- Little’s law
- Priority queues
- Simulation: Lindley’s equation
- **A simple airport model**

SAN Runway Layout

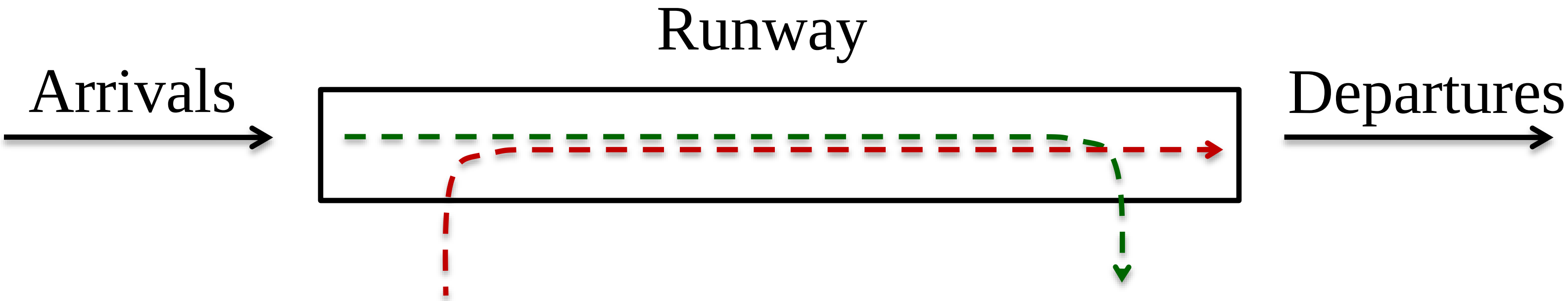


LGA Runway Layout



ORD Runway Layout

Single Runway Resource



Single-Runway Separation Rules

Lead	**Trail**	**Separation Requirement**
Arrival	Arrival	Trailing aircraft cannot cross threshold… until leading aircraft has landed and is clear of the runway
Arrival	Departure	Trailing aircraft cannot begin take-off roll… until leading aircraft has landed and is clear of the runway
Departure	Departure	Trailing aircraft cannot begin take-off roll… until leading aircraft has departed, crossed the runway end
Departure	Arrival	Trailing aircraft cannot cross threshold… until leading aircraft has departed, crossed the runway end

This simplifies to…

Operation	**Occupies Runway**
Arrival	Once aircraft crosses threshold until it has landed and is clear of the runway
Departure	Once aircraft begins take-off roll until aircraft has departed, crossed the runway end

FAA 7110.65R 3-9-6, 3-10-3. There are also exceptions to these rules

Assumptions

- An arrival occupies the runway for a time S_A
- A departure occupies the runway for a time S_D
- Then use Lindley's equation

$$W_q^{(n+1)} = \max(W_q^{(n)} + S^{(n)} - T^{(n)}, 0)$$

Note: In this problem, "arrival" refers in a generic sense to an operation (either an arrival or departure) that "arrives" to use the runway resource

Spreadsheet Calculations

- Inter-arrival time = –LN(RAND()) / λ
- Next arrival time = arrival time + inter-arrival time
- Operation = IF(RAND() < p, “A”, “D”)
- Runway time = IF(operation = “A”, S_A, S_D)
- Next queue time = max(queue time + runway time – interarrival time, 0)
- Departure time = queue time + runway time
- Arrival count = IF(operation = “A”, 1, 0)

Operation #	Inter-arr. Time	Arr. Time	Operation	Rwy Time	Queue Time	Dep. Time	Arr	Dep
1	0.356	0.000	D	2	0.000	2	0	1

Radar/Wake Separation Requirements

- In addition to previous separation requirements…

Lead	**Trail**	**Separation Requirement**
Arrival	Arrival	Trailing aircraft must be at least *x* nm behind leading aircraft at runway threshold (*x* = 3, 4, 5, 6 nm, depending on aircraft weights)
Arrival	Departure	No additional separation requirements
Departure	Departure	Trailing aircraft must satisfy radar/wake separation requirements once airborne (*x* = 3, 4, 5, 6 nm, depending on aircraft weights) and at least 2 minutes behind Heavy/B757
Departure	Arrival	While airborne, aircraft must satisfy standard radar separation

FAA 7110.65R 3-9-6, 5-5-4. There are also exceptions to these rules

Extended Model

- Time that an operation holds the runway is…

Following Operation

Leading Operation

	Arrival	**Departure**
Arrival	S_{AA}	S_{AD}
Departure	S_{DA}	S_{DD}

$$W_q^{(n+1)} = \begin{cases} (W_q^{(n)} - T^{(n)} + S_{AA})^+ & \text{if} \quad n = \text{Arrival},\ n+1 = \text{Arrival} \\ (W_q^{(n)} - T^{(n)} + S_{AD})^+ & \text{if} \quad n = \text{Arrival},\ n+1 = \text{Departure} \\ (W_q^{(n)} - T^{(n)} + S_{DA})^+ & \text{if} \quad n = \text{Departure},\ n+1 = \text{Arrival} \\ (W_q^{(n)} - T^{(n)} + S_{DD})^+ & \text{if} \quad n = \text{Departure},\ n+1 = \text{Departure} \end{cases}$$

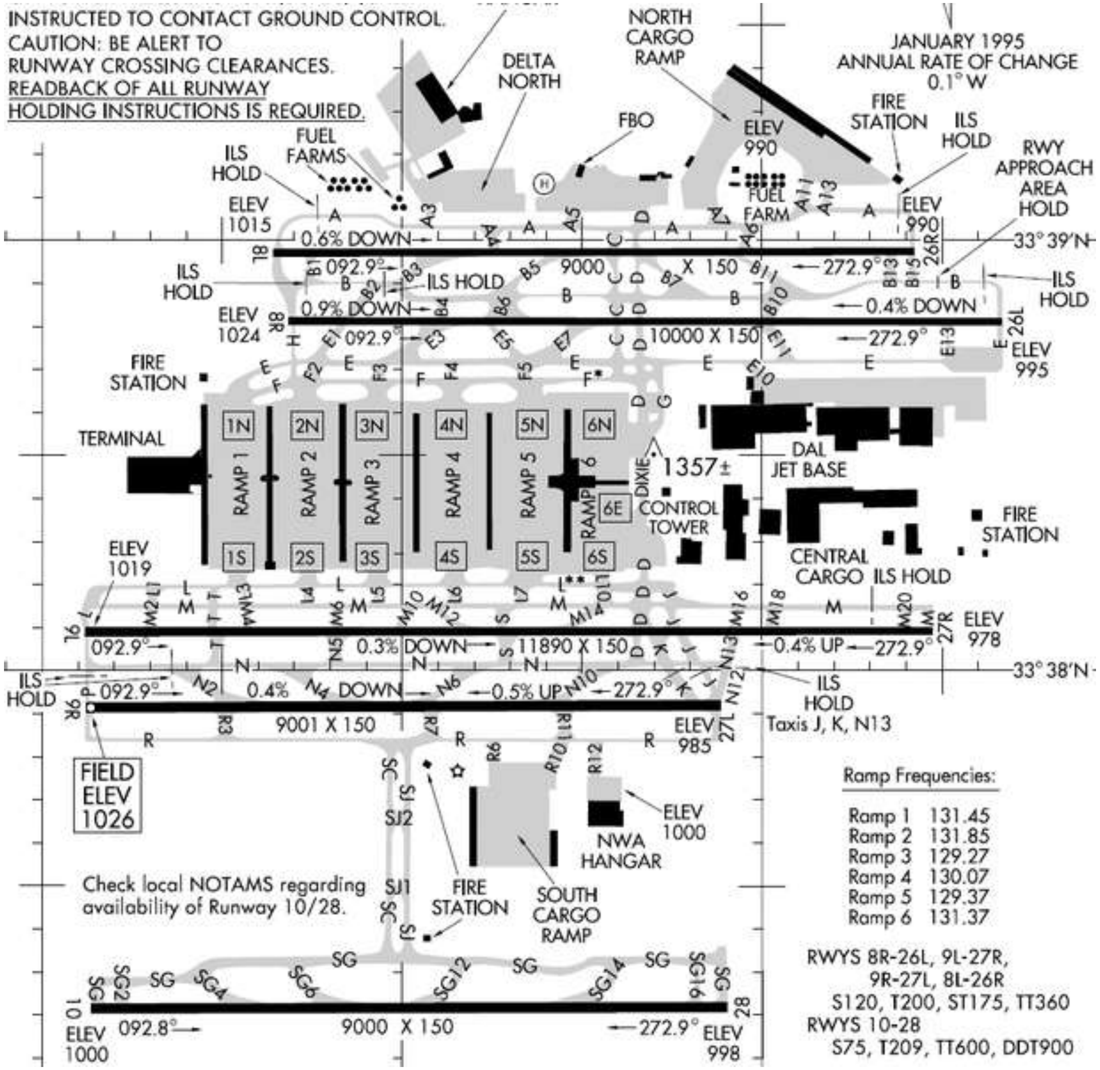
Queueing Theory Summary

- Analytical queueing theory demonstrates fundamental insights
  - Stochastic variability matters
  - Delays generally increase with increased variability
  - Avoid getting utilization too close to 1
  - Priority rules can reduce delay but increase variance
- Potential abuses
  - Only simple models are analytically tractable
  - Analytical formulas generally assume steady-state
  - Theoretical models can predict exceptionally high delays
  - Correlation in arrival process often ignored
- Simulation can be used to overcome limitations

Other Queues

- *G*/*G*/1
  - No simple analytical formulas
  - Approximations exist
- *G*/*G*/∞
  - Infinite number of servers – no wait in queue
  - Time in system = time in service
- *M*(*t*)/*M*(*t*)/1
  - Arrival rate and service rate vary in time
  - Arrival rate can be temporarily bigger than service rate

ATL Runway Layout



INSTRUCTED TO CONTACT GROUND CONTROL.
CAUTION: BE ALERT TO RUNWAY CROSSING CLEARANCES. READBACK OF ALL RUNWAY HOLDING INSTRUCTIONS IS REQUIRED.

JANUARY 1995
ANNUAL RATE OF CHANGE
0.1° W

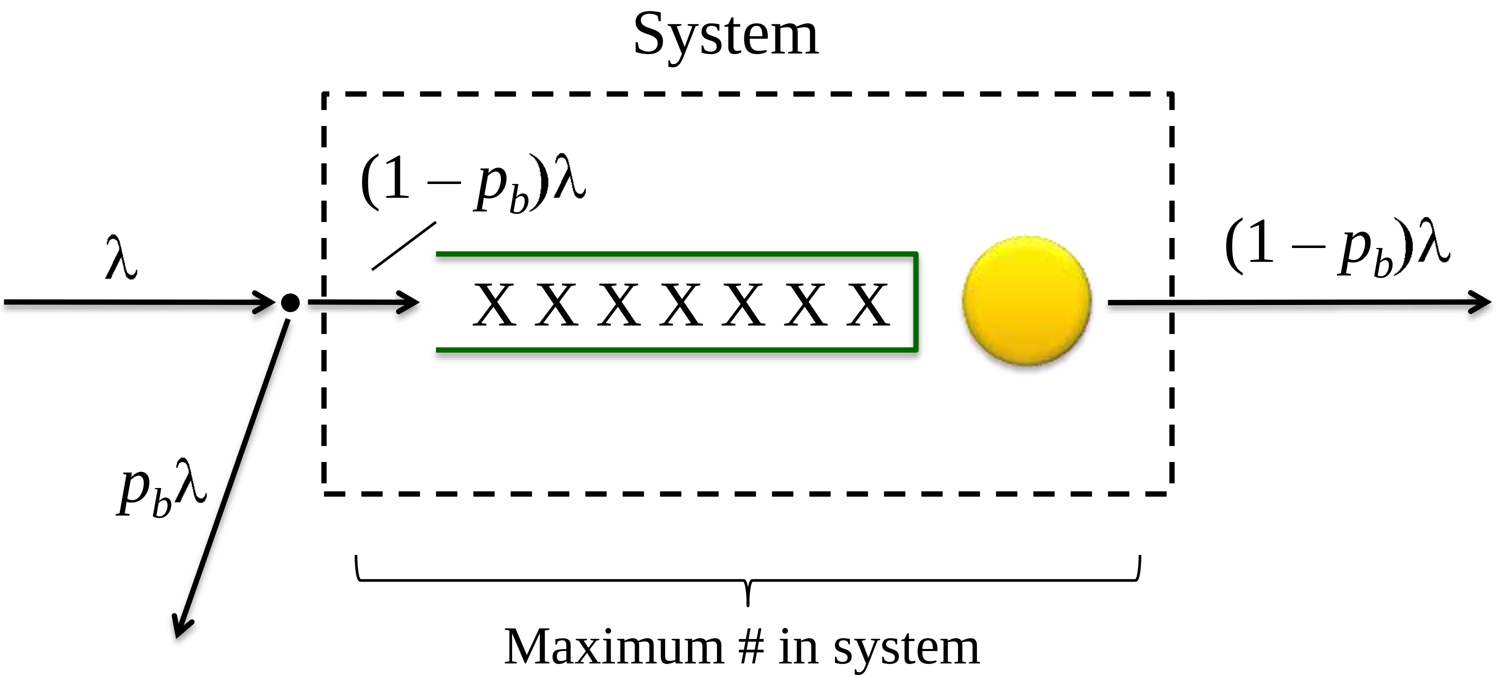
Check local NOTAMS regarding availability of Runway 10/28.

Ramp Frequencies:

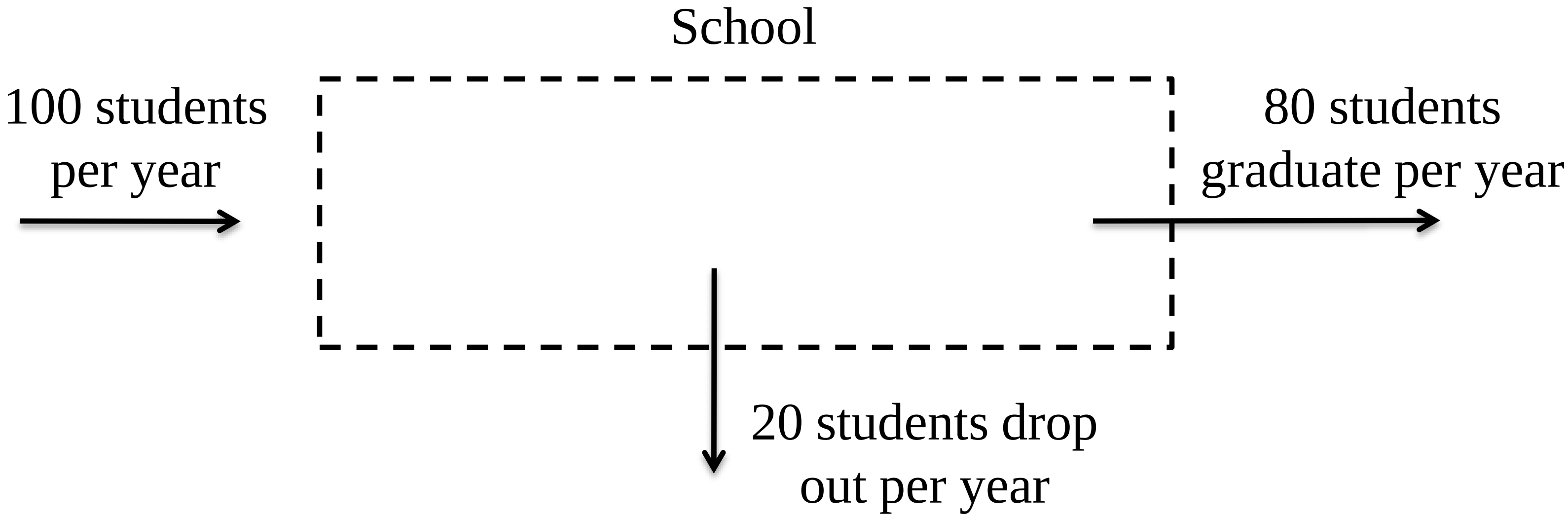
Ramp 1	131.45
Ramp 2	131.85
Ramp 3	129.27
Ramp 4	130.07
Ramp 5	129.37
Ramp 6	131.37

RWYS 8R-26L, 9L-27R,
9R-27L, 8L-26R
S120, T200, ST175, TT360
RWYS 10-28
S75, T209, TT600, DDT900

Example #4: Loss System



Example 5b: Variation



School

100 students per year

80 students graduate per year

20 students drop out per year

Little's Law