

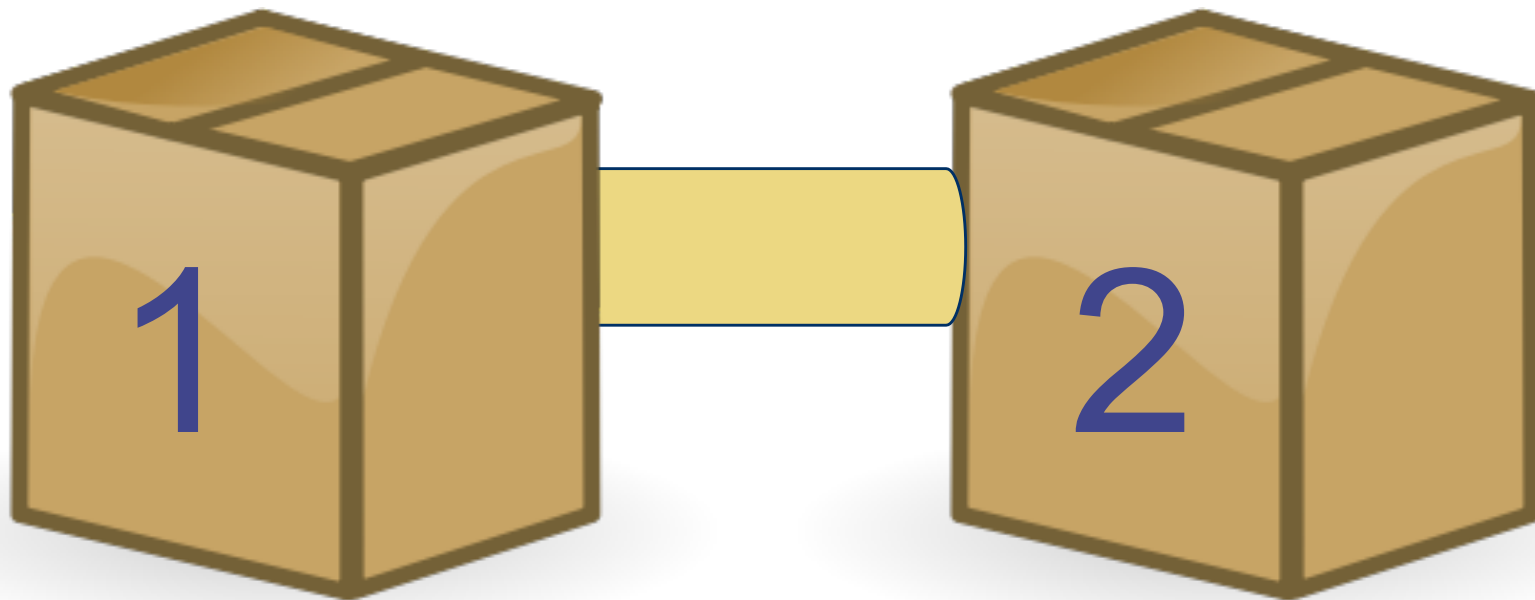
COMP9334

Capacity Planning for Computer Systems and Networks

Week 2B: Queues with Poisson arrivals

Pre-lecture exercise: Where is Felix? (Page 1)

- You have two boxes: Box 1 and Box 2, as well as a cat called Felix
- The two boxes are connected by a tunnel
- Felix likes to hide inside these boxes and travels between them using the tunnel.
- Felix is a very fast cat so the probability of finding him in the tunnel is zero
- You know Felix is in one of the boxes but you don't know which one


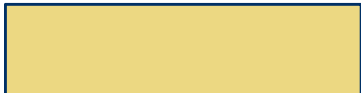


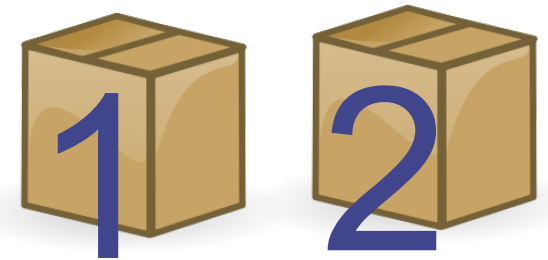
Pre-lecture exercise: Where is Felix? (Page 2)

- Notation:
 - $\text{Prob}[A]$ = probability that event A occurs
 - $\text{Prob}[A \mid B]$ = probability that event A occurs given event B
- You do know
 - Felix is in one of the boxes at times 0 and 1
 - $\text{Prob}[\text{Felix is in Box 1 at time 0}] = 0.3$
 - $\text{Prob}[\text{Felix will be in Box 2 at time 1} \mid \text{Felix is in Box 1 at time 0}] = 0.4$
 - $\text{Prob}[\text{Felix will be in Box 1 at time 1} \mid \text{Felix is in Box 2 at time 0}] = 0.2$
- Calculate
 - $\text{Prob}[\text{Felix is in Box 1 at time 1}]$



Pre-lecture exercise: Where is Felix? (Page 3)

- You want to calculate: $\text{Prob}[\text{Felix is in Box 1 at time 1}]$
- Hint: There are two ways that Felix can end up in Box 1 at time 1
 - Felix is in Box 1 at time 0 AND he ; OR
 - Felix is in Box 2 at time 0 AND he 
- After you've figured out the above, you will need:



Reminder:

$$\text{Prob}[A \text{ or } B] = \text{Prob}[A] + \text{Prob}[B] - \text{Prob}[A \text{ and } B]$$

$$\text{Prob}[A \text{ and } B] = \text{Prob}[A] \text{Prob}[B | A]$$

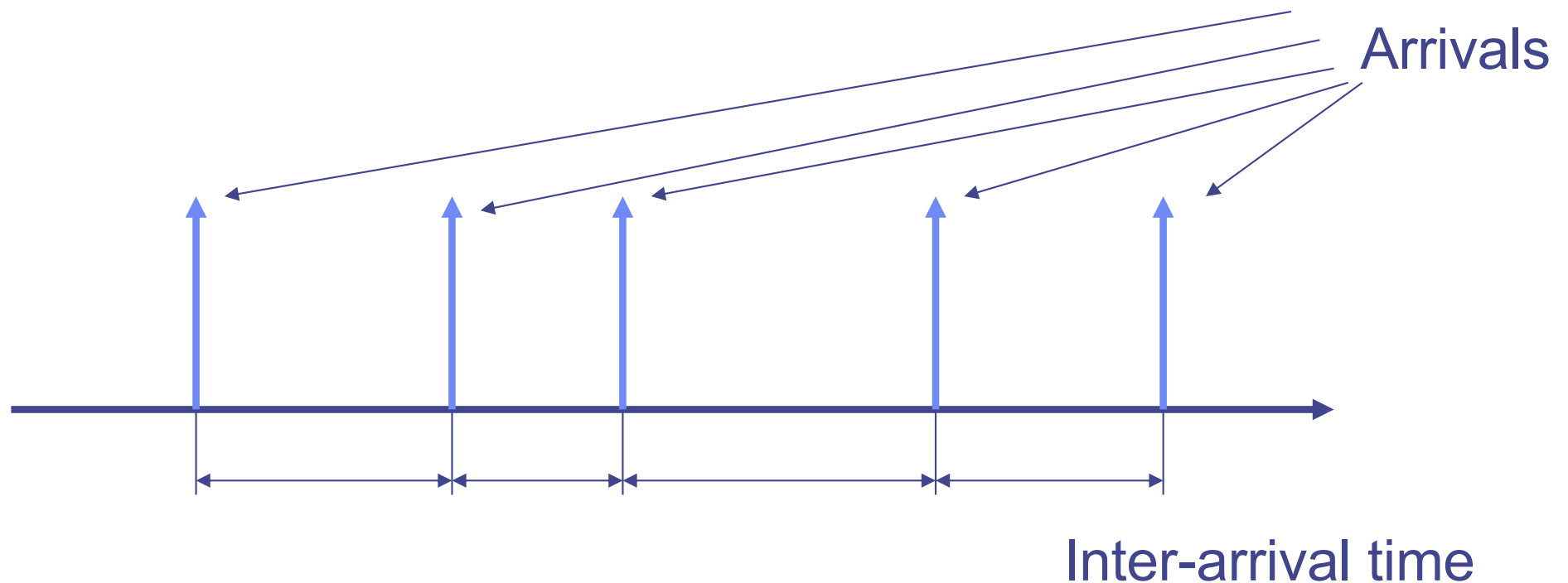
(Special case: A and B are mutually exclusive)

$$\text{Prob}[A \text{ or } B] = \text{Prob}[A] + \text{Prob}[B] \quad \text{if } \text{Prob}[A \text{ and } B] = 0$$

Performance analysis

- Modelling a computer system as a network of queues
- Operational analysis
 - Can be used to find performance bound
- What if you want more exact performance?
 - Need to consider
 - Probability distribution of the arrival process
 - Probability distribution of the service time

Exponential inter-arrival with rate λ



We assume that successive arrivals are independent

Probability that inter-arrival time is between x and $x + \delta x$
 $= \lambda \exp(-\lambda x) \delta x$

Poisson distribution

- The following are equivalent
 - The inter-arrival time is independent and exponentially distributed with parameter λ
 - The number of arrivals in an interval T is a Poisson distribution with parameter λ

$$Pr[k \text{ arrivals in a time interval } T] = \frac{(\lambda T)^k \exp(-\lambda T)}{k!}$$

- Mean inter-arrival time = $1 / \lambda$
- Mean number of arrivals in time interval $T = \lambda T$
- Mean arrival rate = λ

Sample queueing problems

- Consider a call centre
 - Calls are arriving according to Poisson distribution with rate λ
 - The length of each call is exponentially distributed with parameter μ
 - Mean length of a call is $1/\mu$ (in, e.g. seconds)

Call centre:

Arrivals



m operators

If all operators are busy, the centre can put at most n additional calls on hold.

If a call arrives when all operators and holding slots are used, the call is rejected.

- Queueing theory will be able to answer these questions:
 - What is the probability that a call is rejected? (This lecture)
 - What is the mean waiting time for a call? (Next lecture)

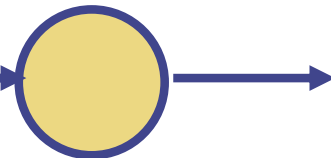
Let us start simple

- We will start by looking at a call centre with one operator and no holding slot
 - This may sound unrealistic but we want to show how we can solve a typical queueing network problem

**Poisson
Arrivals**
→

**Call centre:
1 operator. No holding slot.**

**Poisson
Arrivals**
→



Analysis strategy

- The analysis will consider what happens over a small time interval δ
- This is so that we can consider only two possibilities in each time interval

Poisson distribution


- Consider a small time interval δ
 - This means δ^n (for $n \geq 2$) is negligible
- An interpretation of Poisson arrival:
 - Probability [no arrival in δ] = $1 - \lambda \delta$
 - Probability [1 arrival in δ] = $\lambda \delta$
 - Probability [2 or more arrivals in δ] ≈ 0
- This interpretation can be derived from:

$$Pr[k \text{ arrivals in a time interval } T] = \frac{(\lambda T)^k \exp(-\lambda T)}{k!}$$



Service time distribution

- Service time = the amount of processing time a job requires from the server
- We assume that the service time distribution is exponential with parameter μ
 - The probability that the service time is between t and $t + \delta t$ is:

$$\mu \exp(-\mu t) \delta t$$

- Here: μ = service rate = $1 / \text{mean service time}$
- Another interpretation of exponential service time:
 - Consider a small time interval δ
 - Probability [a job will finish its service in next δ seconds] = $\mu \delta$
 - Probability [a job will **not** finish its service in next δ seconds] = 

Call centre with 1 operator and no holding slots

- Let us see how we can solve the queuing problem for a very simple call centre with 1 operator and no holding slots
- What happens to a call that arrives when the operator is busy?
 - 
- What happens to a call that arrives when the operator is idle?
 - 
- We are interested to find the probability that an arriving call is rejected.

**Poisson
Arrivals**



Call centre:

1 operator. No holding slot.

Solution (1)

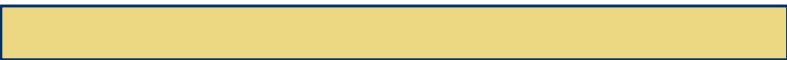
- There are two possibilities for the operator:
 - Busy or
 - Idle
- Let
 - State 0 = Operator is idle (i.e. #calls in the call centre = ?)
 - State 1 = Operator is busy (i.e. #calls in the call centre = ?)

$P_0(t)$ = Prob. 0 call in the call centre at time t

$P_1(t)$ = Prob. 1 call in the call centre at time t

Solution (2)

We try to express $P_0(t + \Delta t)$ in terms of $P_0(t)$ and $P_1(t)$

- No call at call centre at $t + \Delta t$ can be caused by
 - **No call at time t** and **no call arrives in $[t, t + \Delta t]$** , or
 - **1 call at time t** and 

$$P_0(t + \Delta t) = \img alt="yellow box" data-bbox="348 558 652 708" + \img alt="yellow box" data-bbox="698 558 928 708"/>$$

**Question: Why do we NOT have to consider the following possibility:
No customer at time t & 1 customer arrives in $[t, t + \Delta t]$ &
the call finishes within $[t, t + \Delta t]$.**

Solution (3)

- Similarly, we can show that

$$P_1(t + \Delta t) = P_0(t)\lambda\Delta t + P_1(t)(1 - \mu\Delta t)$$

- If we let $\Delta t \rightarrow 0$, we have

$$\frac{dP_0(t)}{dt} = -P_0(t)\lambda + P_1(t)\mu$$

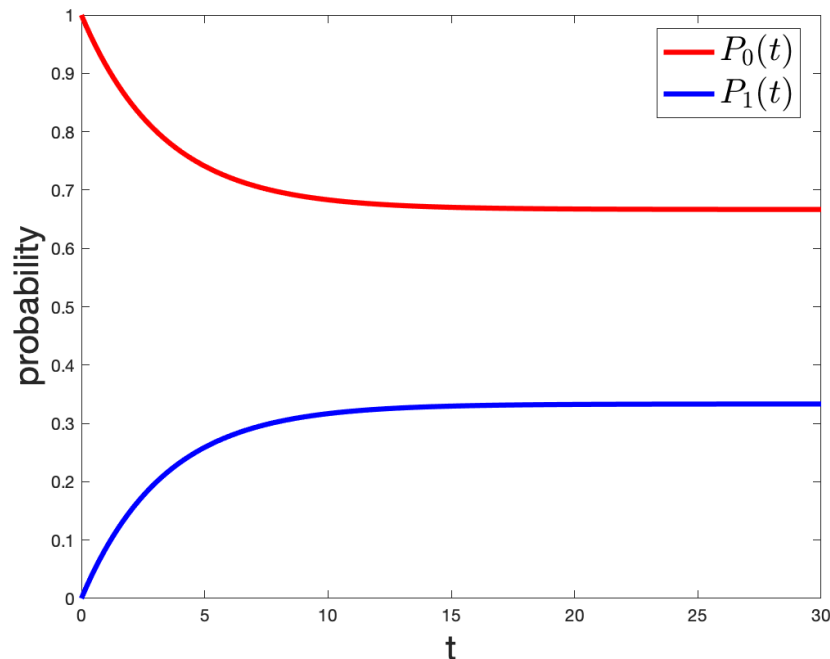
$$\frac{dP_1(t)}{dt} = P_0(t)\lambda - P_1(t)\mu$$

Solution (4)

- We can solve these equations to get

$$P_0(t) = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda + \mu)t}$$

$$P_1(t) = \frac{\lambda}{\lambda + \mu} - \frac{\lambda}{\lambda + \mu} e^{-(\lambda + \mu)t}$$



Solution (5)

- We can solve these equations to get

$$P_0(t) = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda + \mu)t}$$

$$P_1(t) = \frac{\lambda}{\lambda + \mu} - \frac{\lambda}{\lambda + \mu} e^{-(\lambda + \mu)t}$$

- This is too complicated, let us look at **steady state** solution

$$P_0 = P_0(\infty) = \frac{\mu}{\lambda + \mu}$$

$$P_1 = P_1(\infty) = \frac{\lambda}{\lambda + \mu}$$

Solution (5)

- From the steady state solution, we have
 - The probability that an arriving call is rejected
 - = The probability that the operator is busy

- =
$$P_1 = \frac{\lambda}{\lambda + \mu}$$

- Let us check whether it makes sense
 - For a constant μ , if the arrival rate λ increases, will the probability that the operator is busy go up or down?
 - Does the formula give the same prediction?

An alternative interpretation

- We have derived the following equation:

$$P_0(t + \Delta t) = P_0(t)(1 - \lambda\Delta t) + P_1(t)\mu\Delta t$$

- Which can be rewritten as:

$$P_0(t + \Delta t) - P_0(t) = -P_0(t)\lambda\Delta t + P_1(t)\mu\Delta t$$

- At steady state:

Change in Prob in State 0 = 0

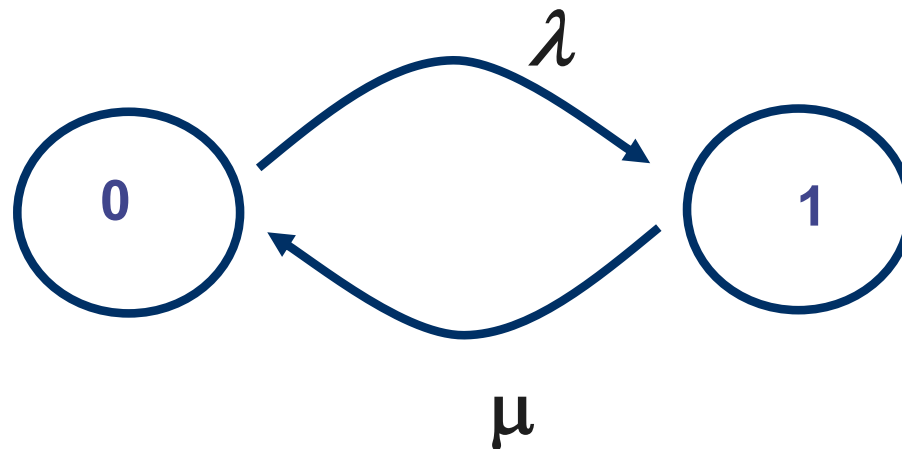
$$\Rightarrow 0 = -\boxed{P_0\lambda}\Delta t + \boxed{P_1\mu}\Delta t$$

Rate of leaving state 0

Rate of entering state 0

Faster way to obtain steady state solution (1)

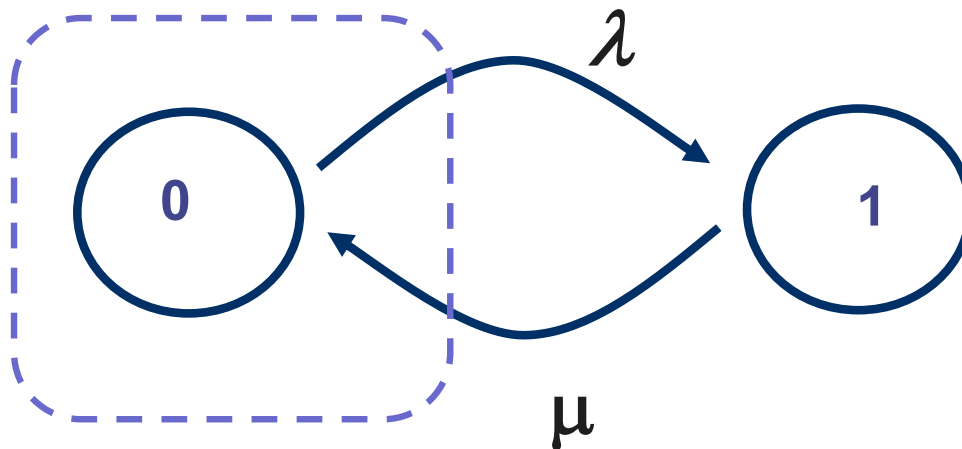
- Transition from State 0 to State 1
 - Caused by an arrival, the rate is λ
- Transition from State 1 to State 0
 - Caused by a completed service, the rate is μ
- State diagram representation
 - *Each circle is a state*
 - *Label the arc between the states with transition rate*



Faster way to obtain steady state solution (2)

- Steady state means
 - **rate of transition out of a state = Rate of transition into a state**
- We have for state 0:

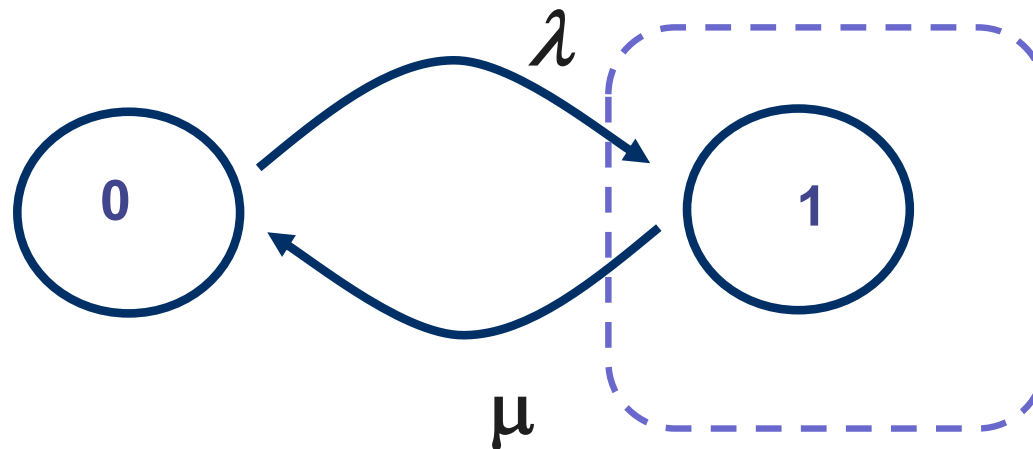
$$\underline{\lambda P_0} = \underline{\mu P_1}$$



Faster way to obtain steady state solution (3)

- We can do the same for State 1:
- Steady state means
 - **Rate of transition into a state** = **rate of transition out of a state**
- We have for state 1:

$$\underline{\lambda P_0} = \underline{\mu P_1}$$



Faster way to obtain steady state solution (4)

- We have one equation $\lambda P_0 = \mu P_1$
- We have 2 unknowns and we need one more equation.
- Since we must be either one of the two states:

$$P_0 + P_1 = 1$$

- Solving these two equations, we get the same steady state solution as before

$$P_0 = \frac{\mu}{\lambda + \mu} \quad P_1 = \frac{\lambda}{\lambda + \mu}$$

Summary

- Solving a queueing problem is not simple
- It is harder to find how a queue evolves with time
- It is simpler to find how a queue behaves at steady state
 - Procedure:
 - Draw a diagram with the states
 - Add arcs between states with transition rates
 - Derive flow balance equation for each state, i.e.
 - Rate of entering a state = Rate of leaving a state
 - Solve the equation for steady state probability

Don't forget the probabilistic interpretation

- Change in probability in State 0

$$P_0(t + \Delta t) - P_0(t) = -P_0(t)\lambda\Delta t + P_1(t)\mu\Delta t$$

$$\Rightarrow 0 = -\boxed{P_0\lambda}\Delta t + \boxed{P_1\mu}\Delta t$$

Rate of leaving state 0

Rate of entering state 0

$$\Rightarrow 0 = -\boxed{P_0\lambda\Delta t} + \boxed{P_1\mu\Delta t}$$

**Prob[Leaving State 0 |
State 0]**

**Prob[Entering State 0 |
State 1]**

A call centre with 1 operator and 1 holding slot

- We want to determine the probability that an arriving call will be rejected

**Poisson
Arrivals**



Call centre:

1 operators. 1 holding slot.

Analysing the queueing problem

- The system can be in one of the following three states
 - State 0 = 0 call in the system (= the operator is idle)
 - State 1 = 1 call in the system (= Operator busy. Holding slot empty.)
 - State 2 = 2 calls in the system (= Operator busy. Holding slot occupied.)
- Define the probability that a certain state occurs

P_0 = Probability in State 0

P_1 = Probability in State 1

P_2 = Probability in State 2

The transition probabilities

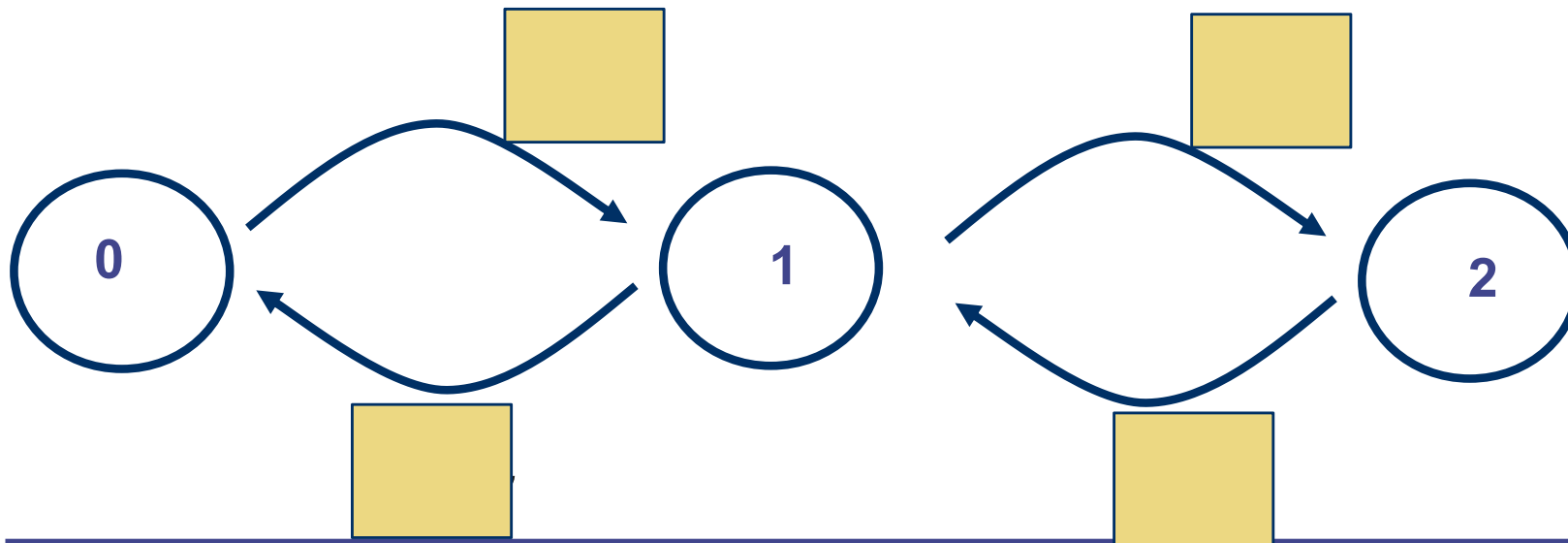
- Consider a small time interval δ
 - Given the system is in State 1
 - What is the probability that it will move to State 0?
 - What is the probability that it will move to State 2?
- Transiting from *State 1* \rightarrow *State 0*
 - This can only occur when
 - Conditional probability for this to occur =
- Transiting from *State 1* \rightarrow *State 2*
 - This can only occur when
 - Conditional probability for this to occur =
- Prob [*State 1* \rightarrow *State 0* | *State 1*] =
- Prob [*State 1* \rightarrow *State 2* | *State 1*] =

Exercise: The transition probabilities

- Can you work out the following transition probabilities
 - Prob [*State 0* → *State 1* | *State 0*] =
 - Prob [*State 0* → *State 2* | *State 0*] =
 - Prob [*State 2* → *State 0* | *State 2*] =
 - Prob [*State 2* → *State 1* | *State 2*] =

The state transition diagram

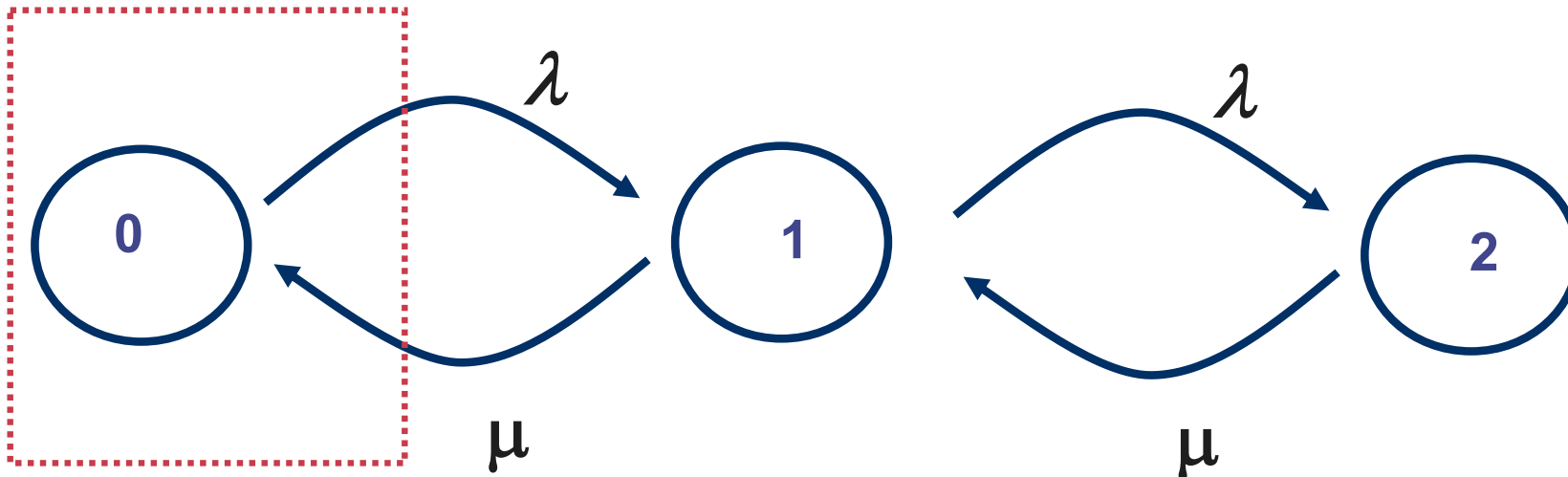
- Given the following transition probabilities (over a small time interval δ)
 - Prob [*State 0* \rightarrow *State 1* | *State 0*] =
 - Prob [*State 0* \rightarrow *State 2* | *State 0*] =
 - Prob [*State 1* \rightarrow *State 0* | *State 1*] =
 - Prob [*State 1* \rightarrow *State 2* | *State 1*] =
 - Prob [*State 2* \rightarrow *State 0* | *State 2*] =
 - Prob [*State 2* \rightarrow *State 1* | *State 2*] =
- We draw the following state transition diagram
 - Note 1: We label the arc with transition rate = transition probability / δ
 - Note 2: Arcs with zero rate are not drawn



Setting up the balance equations (1)

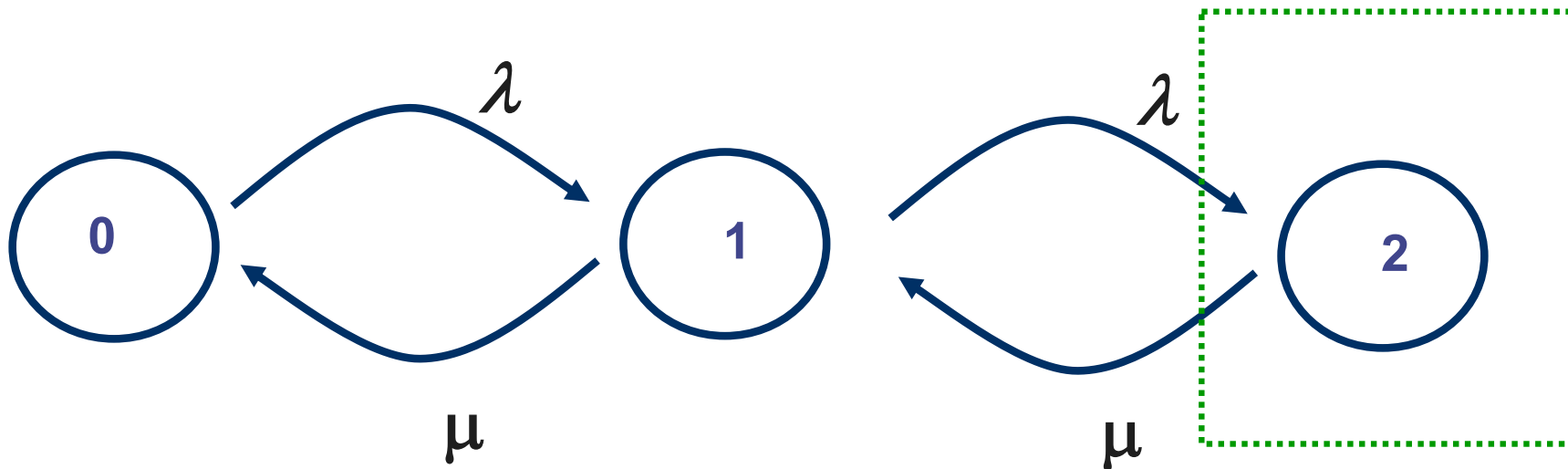
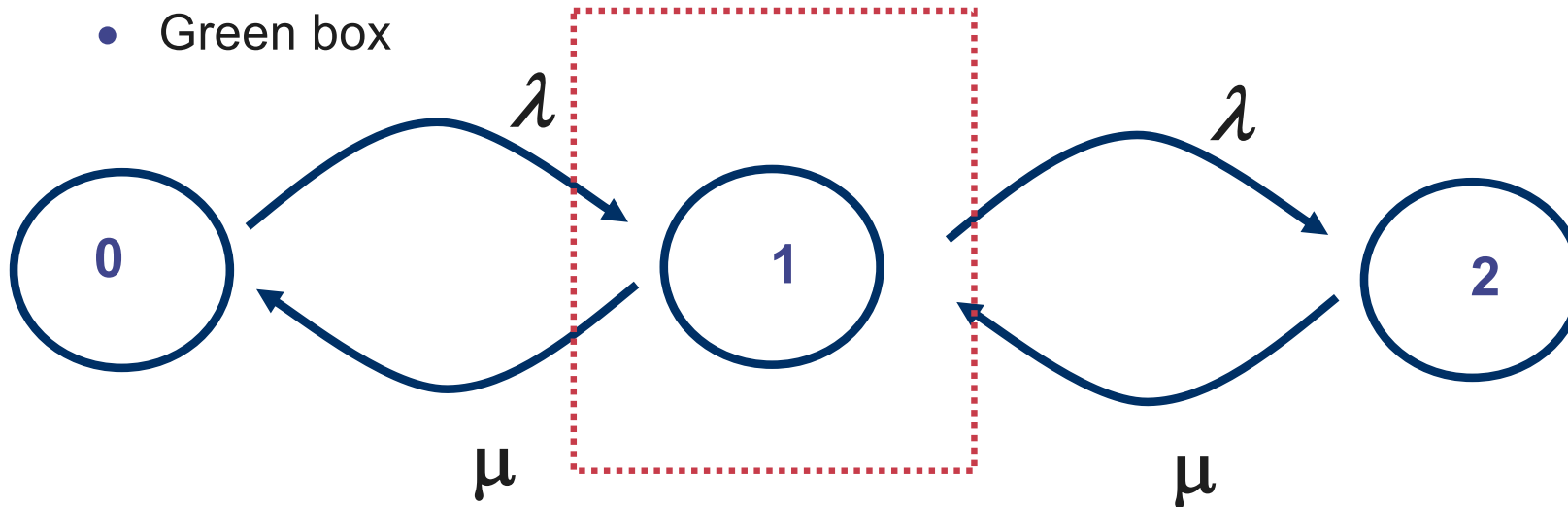
- For steady state, we have
 - Prob of transiting into a “box” = Prob of transiting out of a “box”
 - Rate of transiting into a “box” = Rate of transiting out of a “box”
- Note a “box” can include one or more state
- The “box” is the dotted square shown below

$$\begin{array}{l} \text{Prob out of "box"} = P_0 \lambda \delta \\ \text{Prob into "box"} = P_1 \mu \delta \end{array} \quad \begin{array}{l} \nearrow \\ \nearrow \end{array} \quad \lambda P_0 = \mu P_1$$



Exercise: Setting up the balance equations (2)

- Set up the balance equations for the
 - Red box
 - Green box



The balance equations

- There are three balance equations



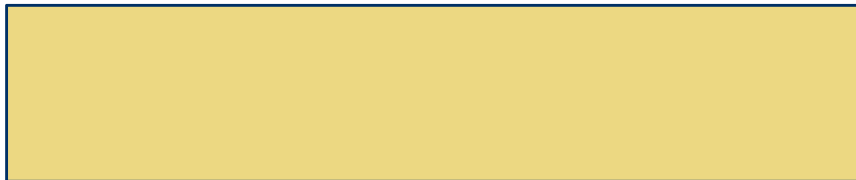
- Note that these three equations are not linearly independent
 - First equation + Third equation = Second equation
- There are 3 unknowns (P_0 , P_1 , P_2) but we have only 2 equations
- We need 1 more equation. What is it?

Solving for the steady state probabilities

- An addition equation: $\text{Sum(Probabilities)} = 1$
- Solve the following equations for the steady state probabilities P_0, P_1, P_2 :

$$\lambda P_0 = \mu P_1$$

$$\mu P_2 = \lambda P_1$$



- By solving these 3 equations, we have

Steady state probabilities

- By solving the equations on the previous slide, we have the steady state probabilities are:

$$P_0 = \frac{1}{1 + \frac{\lambda}{\mu} + \left(\frac{\lambda}{\mu}\right)^2}$$

$$P_1 = \frac{\frac{\lambda}{\mu}}{1 + \frac{\lambda}{\mu} + \left(\frac{\lambda}{\mu}\right)^2}$$

$$P_2 = \frac{\left(\frac{\lambda}{\mu}\right)^2}{1 + \frac{\lambda}{\mu} + \left(\frac{\lambda}{\mu}\right)^2}$$

- If we know the values of λ and μ , we can find the numerical values of these probabilities
- Do the expressions make sense?

Summary and References

- Summary
 - Poisson queues with 1 server + (0 or 1) holding slot
 - How to solve the steady state solution
- Recommended reading
 - Queues with Poisson arrival are discussed in
 - Bertsekas and Gallager, *Data Networks*, Sections 3.3 to 3.4.3
 - Note: I derived the formulas here using continuous Markov chain but Bertsekas and Gallager used discrete Markov chain