# Extensions to Markov Models of One Node Deadlocking Queueing Networks - Baulking & Scheduled Vacations

Cindy Huang, Vincent Knight, Geraint Ian Palmer

## 1 Introduction

This chapter contains collaborative work done jointly by Cindy Huang, Dr. Vincent Knight and myself, during Cindy's Nuffield research placement [2] at the School of Mathematics, Cardiff University. At the time, Cindy was a sixth form student undertaking a four week placement with myself and Dr. Knight. The work involved extending the Markov chain models previously in built in Chapter **??**, in order to model baulking behaviour by the customers, and scheduled vacations of the servers.

## 2 Baulking

Baulking behaviour is when customers arrive, but choose not to join the queue as the queue is too long. This decision to baulk however is probabilistic. This work looks investigates a model of baulking discussed in [1], in which arriving customers baulk with probability $b(m)$ dependent on $m$ the number of customers already at the node. This probability density function is shown in Equation 1.

$$b(m) = \begin{cases} 0 & \text{if } m = 0 \\ 1 - \frac{\beta}{m} & \text{otherwise} \end{cases} \tag{1}$$

where $n$ is the number of customers already at the node, and $0 < \beta \leq 1$ is some measure of willingness to join the queue. Figure 1 shows how the baulking function $b(m)$ behaves for different values of $m$ and $\beta$. The more customers there are at the node already, the more likely a newly arriving customer is to baulk. The higher the value for $\beta$ the less likely a newly arriving customer is to baulk, or the more willing they are to join the queue despite there being customer ahead of them. This justifies the interpretation of $\beta$ being a measure of willingness to join the queue.

### 2.1 Markov Chain Model of One Node Queueing Network with Baulking

Consider the one node queueing network shown in Figure 2. Customers arrive to the system at a rate $\Lambda$, but baulk with probability $1 - b(m)$ when there are $m$ customers already at the node, and so join the queue

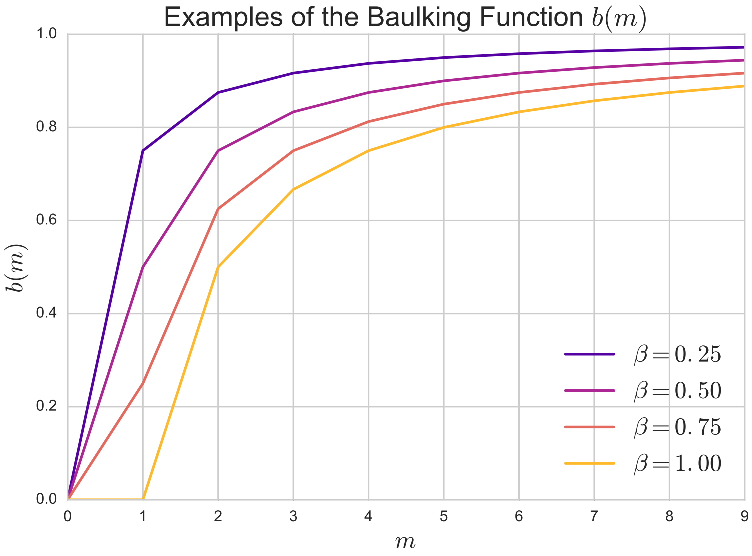


Figure 1: Behaviour of the baulking function $b(m)$ for different values of $\beta$.

with probability $b(m)$. There is one server, who serves customers at a rate of $\mu$, and there is only capacity for $n$ customers to wait in the queue. After service there is a probability $r_{11}$ of rejoining the queue, and so a probability $1 - r_{11}$ of exiting the system after service.

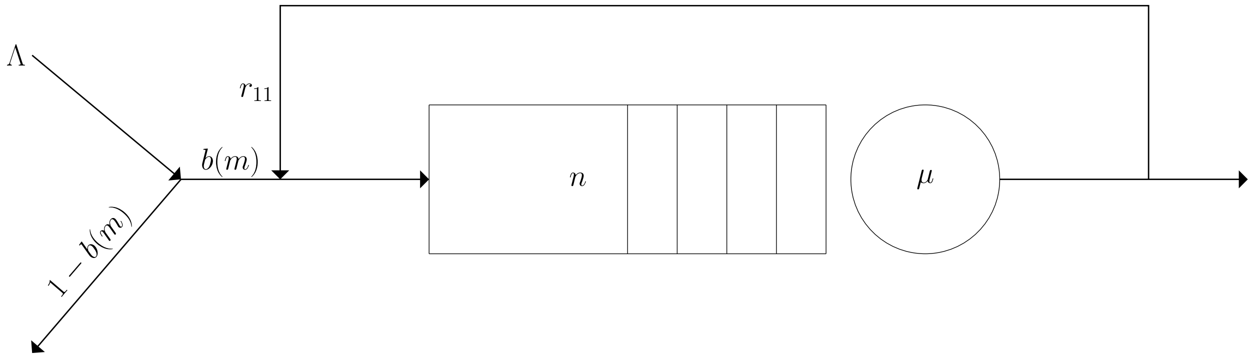


Figure 2: A one node queueing network with baulking.

The state space for the corresponding Markov chain is given by:

$$S = \{i \in \mathbb{N} \mid 0 \leq i \leq n + 1\} \cup \{(-1)\}$$

where $i$ denotes the number of individuals in service or waiting, and $(-1)$ denotes the deadlocked state.

Define $\delta = i_2 - i_1$ for all $i_k \geq 0,\ k \in \{1, 2\}$. The transitions are given by Equations 2, 3 and 4.

$$q_{i_1,i_2} = \begin{cases} \left.\begin{array}{cc} \Lambda & \text{if } i_1 = 0 \\ \frac{\Lambda\beta}{i_1} & \text{if } 0 < i_1 < n+1 \\ 0 & \text{otherwise} \end{array}\right\} & \text{if } \delta = 1 \\ (1-r_{11})\mu & \text{if } \delta = -1 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

$$q_{i,(-1)} = \begin{cases} r_{11}\mu & \text{if } i = n+1 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

$$q_{-1,i} = 0 \tag{4}$$

The Markov chain is shown in Figure 3.

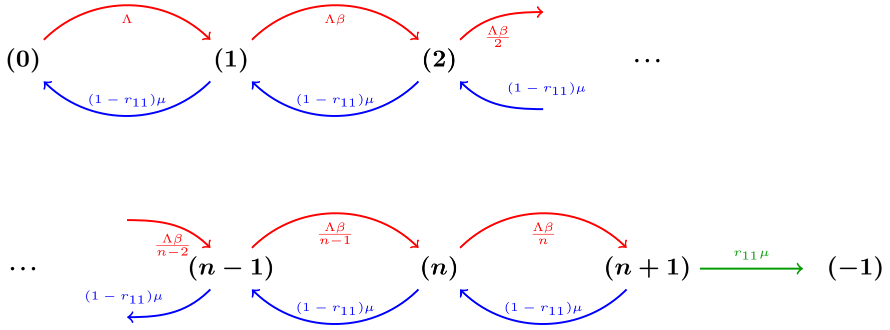


Figure 3: Diagrammatic representation of the Markov chain for the one node system with baulking.

The only difference between this Markov model and the model presented in Chapter **??** is that the rate at which customers enter the queue now varies with the number of customers already in the system. If there are $m$ customers in the system $(m > 0)$ then the rate at which customers join the queue is $\frac{\Lambda\beta}{x}$. Note that for any $0 < \beta \leq 1$, and for any number of customers $m$, there is always a positive probability of a customer joining the queue, and therefore it is always possible to reach the deadlock state. If however $\beta = 0$ was chosen, then the only possibility of a customer joining the queue is when $x = 0$, and so the system will never reach deadlock.

Figure 4 shows the effect of varying the baulking parameter $\beta$ on times to deadlock. Base parameters of $\Lambda = 3$, $n = 2$, $\mu = 3$, and $r_{11} = 0.6$ are used. As $\beta$ increases the time to deadlock decreases. We interpret $\beta$ as the willingness of newly arriving customers to join the queue. Then, as this parameter increases more newly arriving customers join the queue, thus the rate at which occupancy increases, increases. The baulking parameter $\beta$ therefore has a very similar effect on the time to deadlock as the arrival rate $\Lambda$.

Let's investigate the combined effect of $\beta$ along with other parameters on the time to deadlock. Figures 5a and 5b show the effects that the baulking parameter $\beta$ has on the time to deadlock as the arrival rate $\Lambda$ varies, and also how this effect is affected by the arrival rate. Figures 6a and 6b show the effects that the baulking parameter $\beta$ has on the time to deadlock as the transition probability $r_{11}$ varies, and also how this

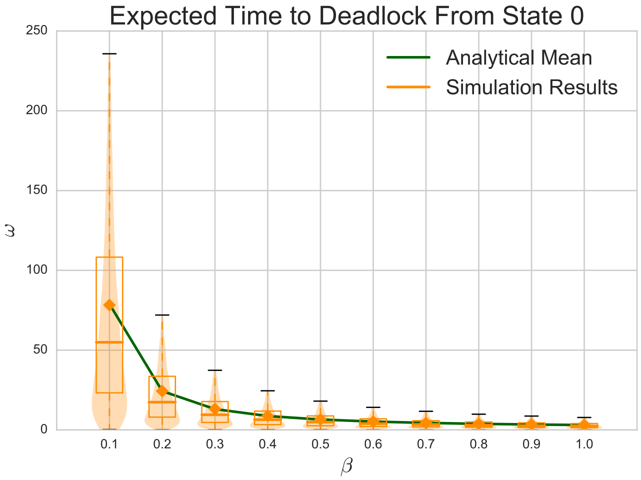
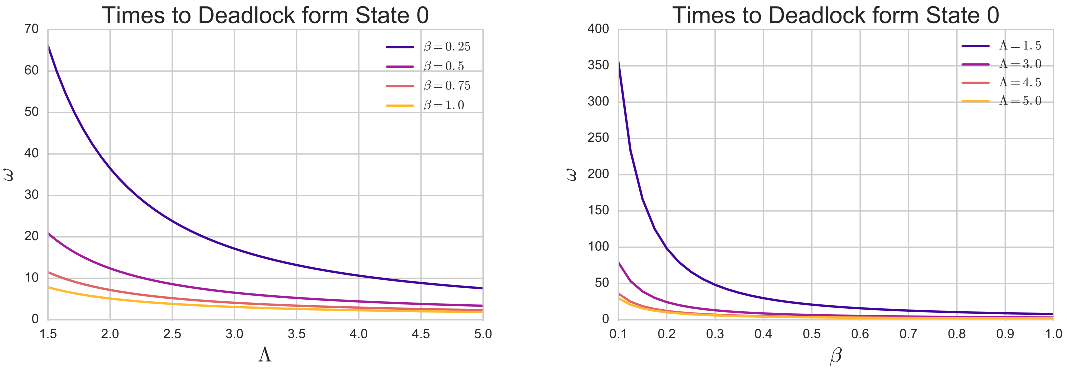


Figure 4: Time to deadlock of the one node baulking network, analytical & simulation results (10,000 iterations), varying $\beta$.

effect is affected by the transition probability. All three parameters, $\Lambda$, $r_{11}$, and $\beta$ have similar effects on the time to deadlock; as these parameters are increased, then the queueing capacity if filled up more quickly, and so we get to a situation where blocking, and deadlock, can occur sooner. Therefore increasing these parameters decreases the time to deadlock. Combining these parameters results in an amplified effect.

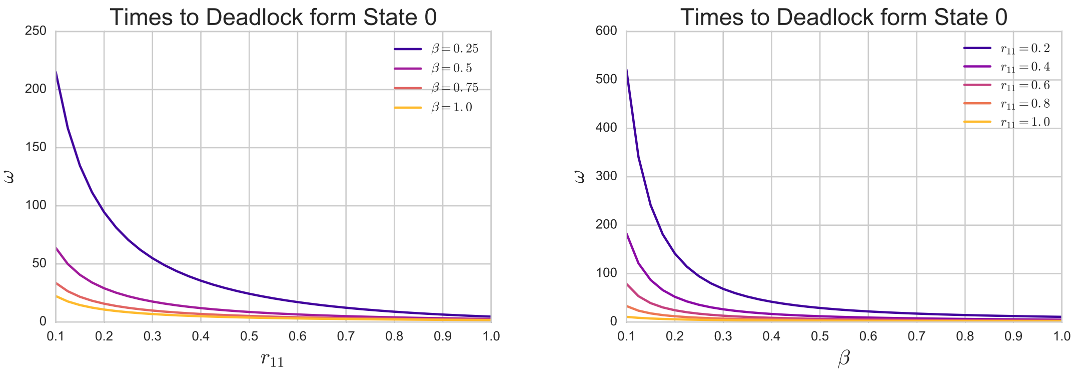


(a) Time to deadlock varying $\Lambda$ by $\beta$.

(b) Time to deadlock varying $\beta$ by $\Lambda$.

Figure 5: Investigating how the combined effect of the baulking parameter $\beta$ and the arrival rate $\Lambda$ on the time to deadlock.

Figures 7a and 7b show the effects that the baulking parameter $\beta$ has on the time to deadlock as the queueing capacity $n$ varies, and also how this effect is affected by the queueing capacity. In Chapter **??** it was shown that, without baulking, as the queueing capacity increased the time to deadlock also increased, as it took longer to fill up a larger queueing capacity. This effect is also seen with baulking, however the baulking exaggerates this effect; the lower the baulking parameter $\beta$ (that is, the more likely customers are to baulk) the more exaggerated this effect. This is because the baulking function $b(m)$ is independent of $n$, and even when customers are most willing to join the queue ($\beta = 1.0$) once there are 4 customers in the system three-

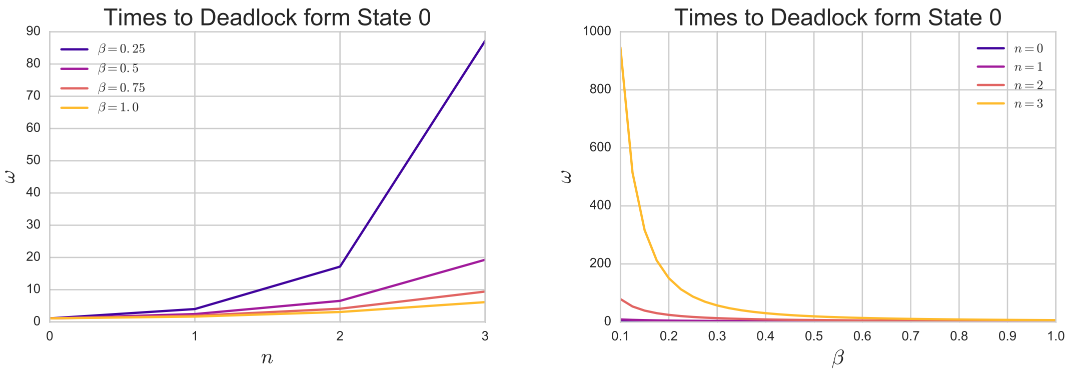


(a) Time to deadlock varying $r_{11}$ by $\beta$.

(b) Time to deadlock varying $\beta$ by $r_{11}$.

Figure 6: Investigating how the combined effect of the baulking parameter $\beta$ and the transition probability $r_{11}$ on the time to deadlock.

quarters of arriving customers baulk. Therefore increasing $n$ makes it less and less likely for a customer to join the queue, and so more and more difficult for occupancy to increase, especially when the number of customers already at the queue approaches the queueing capacity.



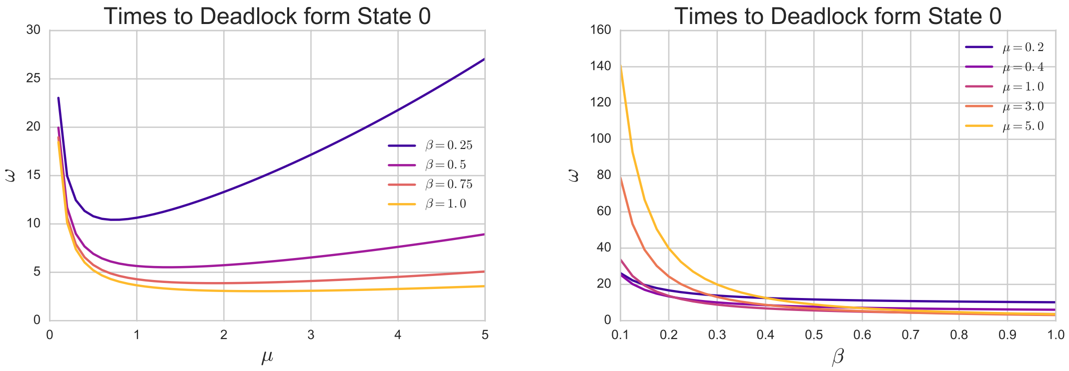
(a) Time to deadlock varying $n$ by $\beta$.

(b) Time to deadlock varying $\beta$ by $n$.

Figure 7: Investigating how the combined effect of the baulking parameter $\beta$ and the queueing capacity $n$ on the time to deadlock.

Figures 8a and 8b show the effects that the baulking parameter $\beta$ has on the time to deadlock as the service rate $\mu$ varies, and also how this effect is affected by the service rate. When looking at the equivalent system without baulking in Chapter **??** a bowl shaped curve was observed in the time to deadlock when varying the service rate, and the presence of some service rate threshold that changes the behaviour of $\omega$. This behaviour is still present when customers exhibit baulking, however the baulking parameter $\beta$ effects the shape of the curve. Note that only the gradient of the slope for values $\mu > \hat{\mu}$, that is service rates greater that the threshold that is effected. This is because at service rates lower than $\hat{\mu}$ we assume that the system is saturated, that is that the queueing capacity is full or is filled up fast enough that the time which it is not full or negligible. Therefore in these situations the arrival rate, and so the baulking rate, cannot affect the time to deadlock, only the service rate and transition probabilities can.

Looking at the inverse plot, the effect that the service rate $\mu$ has on the time to deadlock as $\beta$ varies, Figure 8b, an interesting effect occurs. At low values of $\beta$ the greater the service rate the longer it takes to reach deadlock, however this effect is flipped at high values of $\beta$. At low $\beta$ not many customers join the queue (compared to a low $\beta$), and so a long service time in necessary to allow the queueing capacity to fill up, therefore a lower $\mu$ will yield quicker times to deadlock. At high values of $\beta$ many customers join the queue (compared to a a $\beta$), and so there is no need to wait for the queue to fill up, we can assume this will happen anyway. Therefore a lower service rate $\mu$ will cause longer for a blockage to occur that a higher service rate, and so a longer time to deadlock.



(a) Time to deadlock varying $\mu$ by $\beta$.

(b) Time to deadlock varying $\beta$ by $\mu$.

Figure 8: Investigating how the combined effect of the baulking parameter $\beta$ and the service rate $\mu$ on the time to deadlock.

Figure 9 shows the effect of the baulking parameter $\beta$ on the service rate threshold. As the baulking parameter is closely associated with the arrival rate $\Lambda$, similarly to the effect of $\Lambda$ on $\hat{\mu}$, the threshold increases as the baulking parameter increases. This is due to A high baulking parameter means that more customers join the queue, and so it takes a larger service rate to escape the saturated zone. This relationship is not linear; the general rate of increase of the function decreases as $\beta$ increases. This effect is amplified for smaller values of $n$.

# 3 Scheduled Vacations

This section will extend the Markov chain model of a one node restricted queueing network discussed in Chapter **??**, so that server behaviour may be incorporated. The behaviour we will focus on is scheduled vacations. That is, the server will periodically cycle on and off duty. While the server is on duty the system behaves as usual, however when the server is off duty no services can occur, although arrivals happen as usual. When a server goes off duty, the current service is interrupted, and resumed when the server comes back on duty.

Lets call $s$ the time on duty, and $s_v$ the time on vacation. Therefore the cycle length is $s + s_v$.

As an example, imagine an on-line shop where orders can arrive to a queue at any time of the day or night.

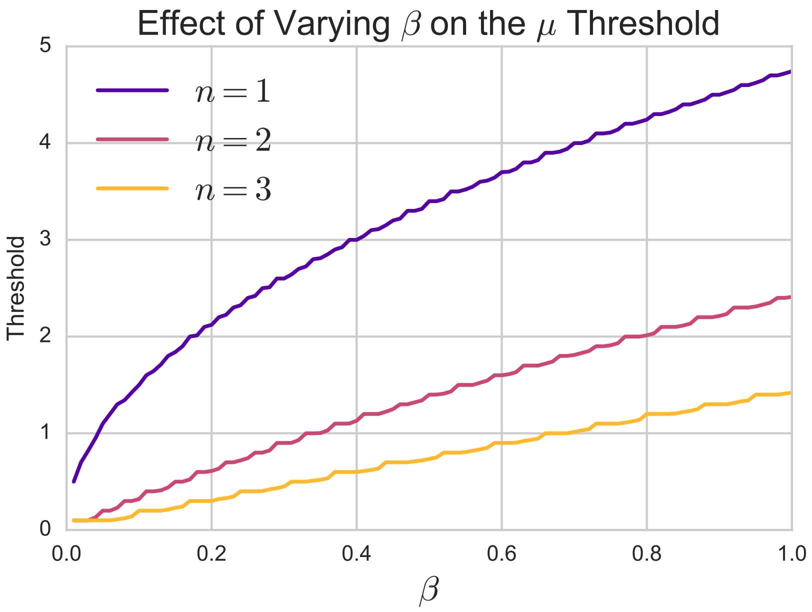


Figure 9: The effect of $\beta$ on the service rate threshold $\hat{\mu}$.

The shop hires one server who processes orders from 9am to 5pm, seven days a week. From 5pm to 9am next day, that server is off duty and no orders can be processed, however orders continue to arrive. That server is on a scheduled vacation. (Note: if the server is mid way through processing an order at 5pm, then he will complete that service before going off duty). Here $s$ is the time on duty (9am - 5pm), and $s_v$ the time on vacation (5pm - 9am). The cycle length is one day. Figure 10 illustrates this.

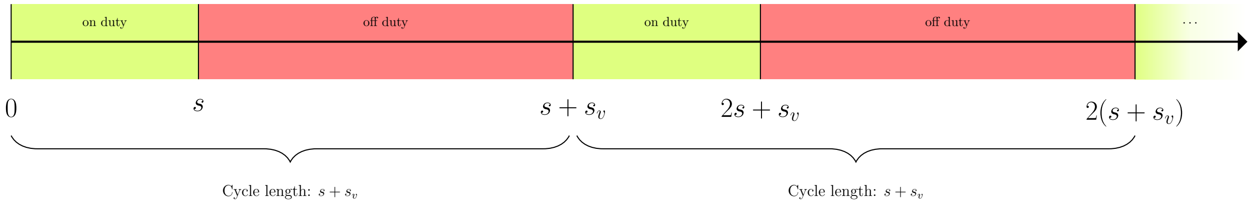


Figure 10: Illustration of the $(s, s_v)$ work cycle.

This will be modelled using time inhomogeneous Markov chains. Before we begin, a result on discretising time inhomogeneous Markov chains is required.

### 3.1 Discretising Time Inhomogeneous Markov Chains

A well known result [3] on discretising continuous Markov chains is that the matrix $P = Q\Delta t + I$ is a stochastic matrix if the following condition is satisfied:

$$\Delta t \leq \frac{1}{\max_i |q_{ii}|} \tag{5}$$

This discrete Markov chain $P$ is equivalent to the continuous Markov chain $Q$ with transitions only occurring a at the discrete time interval $\Delta t$.

It is also known that if a discrete time inhomogeneous Markov chain can be described my transition matrix $P_1$ between time steps 1 and 2; and by transition matrix $P_2$ between time steps 2 and 3, where $P_1$ and $P_2$ have the same states, then that system's state can be described by the transition matrix $P_1 P_2$ between time steps 1 and 3. This is equivalent to a discrete time homogeneous Markov chain described by $P$, can be described by $P^n$ for the first $n$ time steps.

How about a continuous time inhomogeneous Markov chain that is described by $Q_1$ for the first $s_1$ time units, and by $Q_2$ for the next $s_2$ time units? How can we discretise these in order to obtain transient solutions to this discretised time inhomogeneous Markov chain? Theorem 1 discusses this discretisation.

**Theorem 1.** *Given a time inhomogeneous Markov chain that is described sequentially by the continuous transition matrices $Q_j$ for $s_j$ time units, $s_j \in \mathbb{Z}$, for $j \in J = \{1, 2, ..., m\}$, then the discretised system requires the discretised transition matrices $P_j = Q_j \Delta t + I$, where*

$$\Delta t = \frac{\gcd_j(s_j)}{\left\lceil \frac{\gcd_j(s_j)}{\min_j \bar{\delta}_j} \right\rceil}$$

*where $\bar{\delta}_j = \frac{1}{\max_i |q_{ii}|}$, with $q_{ii}$ being the diagonal entries in the transition matrix $Q_j$.*

*Proof.* The discretised system requires a single time step, $\Delta t$. $\Delta t$ must be chosen such that each continuous transition matrix is discretised using $P_j = Q_j \Delta t + I$, and such that $\Delta t$ can be multiplied an integer amount of times to give all $s_j$, for $j \in J$.

First we require that $Q_j \Delta t + I$ is a stochastic matrix for each $j \in J$. In order for this to be true, we must have

$$\Delta t \leq \frac{1}{\max_j |q_{ii}|} \; \forall \, j \in J$$

where $q_{ii}$ are the diagonal entries of $Q_j$. Defining $\bar{\delta}_j = \frac{1}{\max_j |q_{ii}|} \; \forall \, j \in J$, we have that

$$\Delta t \leq \min_j \bar{\delta}_j \tag{6}$$

.

Now $\Delta t \in \mathbb{R}$, but as we require that $\Delta t$ can be multiplied by integers to give each of $s_j$ for $j \in J$, then we force $\Delta t \in \mathbb{Q}$. Therefore $\exists \, k, m \in \mathbb{Z}$ such that $k\Delta t = m$. Now we have an integer $m$ that must divide each $s_j$ for $j \in J$. For efficiency, we take the largest such $m$, and so $m = \gcd_j(s_j)$. Therefore, we have $\Delta t$ of form

$$\Delta t = \frac{\gcd_j(s_j)}{k} \tag{7}$$

.

Combining the conditions in Equations 6 and 7, we get $k$ such that

$$\frac{\text{gcd}_j(s_j)}{k} \leq \min_j(\bar{\delta}_j)$$

and so we take $k = \left\lceil \frac{\text{gcd}_j(s_j)}{\min_j(\bar{\delta}_j)} \right\rceil$.

Finally, substituting in for $k$, we get

$$\Delta t = \frac{\text{gcd}_j(s_j)}{\left\lceil \frac{\text{gcd}_j(s_j)}{\min_j \bar{\delta}_j} \right\rceil}$$

□

## 3.2 Modelling Deadlock with Servers with Scheduled Vacations

Consider the one node restricted queueing network shown in Figure 11. Here customers arrive at a rate of $\Lambda$ per time unit, there is only enough room for $n$ customers to queue at a time, and after service customers rejoin the queue with probability $r_{11}$. The service rate now alternates between having $s$ time units as $\mu$, and $s_v$ time units as 0. If a customer is in service as the service rate changes, their service is interrupted, and is resumed once the server come back on duty after $s_v$ time units on vacation.

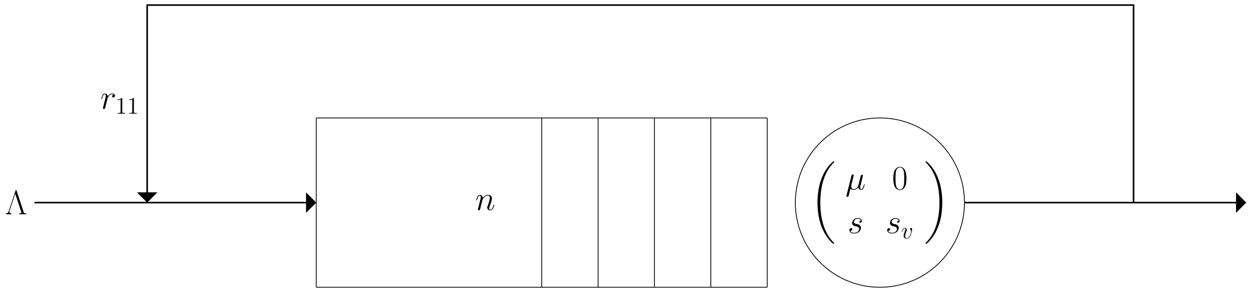


Figure 11: A one node queueing network with scheduled vacations.

This system is described by two transition matrices, $Q$ while the server is on duty, and $Q_v$ when the server is off duty.

The state space for both is given by:

$$S = \{i \in \mathbb{N} \mid 0 \leq i \leq n+1\} \cup \{(-1)\}$$

where $i$ denotes the number of individuals in service or waiting, and $(-1)$ denotes the deadlocked state.

Define $\delta = i_2 - i_1$ for all $i_k \geq 0$. The transitions for $Q$ are given by Equations 8, 9 and 10.

$$q_{i_1,i_2} = \begin{cases} \left.\begin{array}{cc} \Lambda & \text{if } i_1 < n+1 \\ 0 & \text{otherwise} \end{array}\right\} & \text{if } \delta = 1 \\ (1-r_{11})\mu & \text{if } \delta = -1 \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

$$q_{i,(-1)} = \begin{cases} r_{11}\mu & \text{if } i = n+1 \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

$$q_{-1,i} = 0 \tag{10}$$

The transitions for $Q_v$ are given by Equations 11 and 12.

$$q_{i_1,i_2} = \begin{cases} \Lambda & \text{if } i_1 < n \text{ and } \delta = 1 \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

$$q_{i,-1} = q_{-1,i} = 0 \tag{12}$$

These two cases are visualised in Figure 12.

We will use two different methods for getting the CDF and expeced time to deadlock from these Markov models. Method 1, described in Section 3.3, discretises the transition matrices, and Method 2, described in Section 3.4 uses matrix exponentials of the continuous transition matrices.

## 3.3 Method 1

Using Theorem 1, the two transition matrices described in Section 3.2 can be discretised appropriately. As we know the structure of these Markov chains, we can simplify the theorem for this purpose.

**Theorem 2.** *For the one node restricted network described above, described by continuous transition matrices $Q$ for $s$ time units and $Q_v$ for $s_v$ time units, the appropriate time step to discretise the transition matrices is $\Delta t = \text{gcd}(s, s_v) / \left\lceil \text{gcd}(s, s_v) \left(\Lambda + (1 + r_{11}\mu\right) \right\rceil$.*

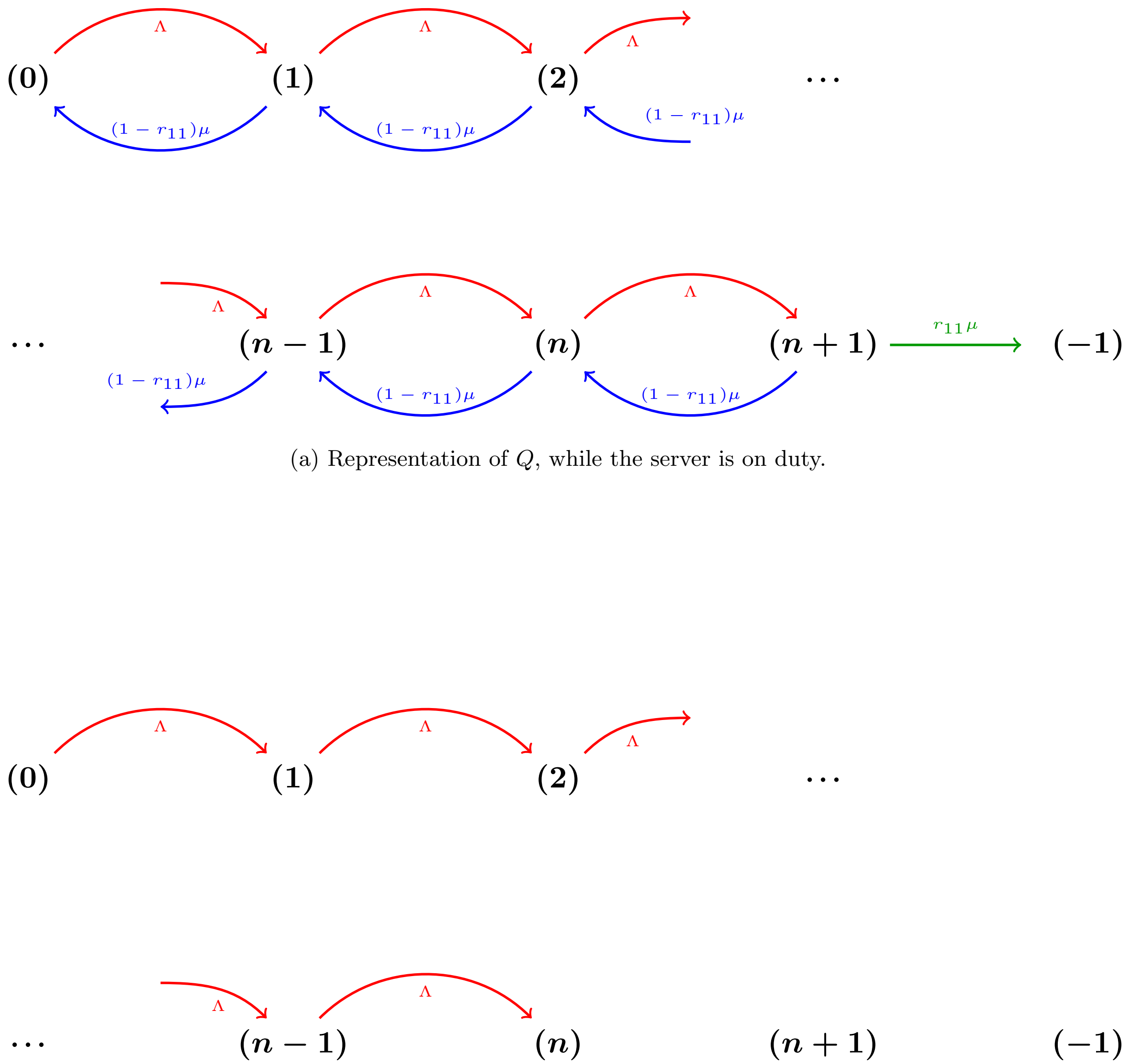
*Proof.* From Theorem 1 we require

$$\Delta t = \frac{\text{gcd}_j(s_j)}{\left\lceil \frac{\text{gcd}_j(s_j)}{\min_j \bar{\delta}_j} \right\rceil}$$

Now $\text{gcd}_j(s_j) = \text{gcd}(s, s_v)$.

For $Q$, $\bar{\delta} = 1/\Lambda$. For $Q_v$, $\bar{\delta} = 1/(\Lambda + (1 - r_{11})\mu)$. As $1/(\Lambda + (1 - r_{11})\mu) \leq 1/\Lambda$, for any $\Lambda \geq 0$, $0 < r_{11} \leq 1$, $\mu \geq 0$, then for any valid parameters $\min_j \bar{\delta}_j = 1/(\Lambda + (1 - r_{11})\mu)$.

Substituting these into the original theorem yields the required result. □



(a) Representation of $Q$, while the server is on duty.

(b) Representation of $Q_v$, while the server is off duty.

Figure 12: Diagrammatic representation of $Q$ and $Q_v$, the continuous time Markov chains describing a system with scheduled vacations while the server is on and off duty respectively.

Using this time step $Q$ and $Q_v$ can be discretised to $P$ and $P_v$ respectively. We can now find the transient solutions, and approximate stationary solutions for this system using Equation 13.

$$\pi_{t+1} = \pi_t \hat{P} \tag{13}$$

Where $\hat{P}$ is defined by Equation 14.

$$\hat{P} = \begin{cases} P & \text{if } t \bmod (s + s_v) \leq s \\ P_v & \text{if } t \bmod (s + s_v) > s \end{cases} \tag{14}$$

In this way the probability density function, and the cumulative density function over the time steps can be found. A stepping through methods was used, described in Algorithm 1 to find the entire CDF up to some tolerance $\epsilon$.

**Algorithm 1:** A stepping through method to find the CDF of a system with scheduled vacations; Method 1.

```
input  : $P$, $P_v$, $s$, $s_v$, $\Delta t$, $\epsilon$.
output: A cumulative density function of reaching deadlock from an empty system.
$t, p = 0, 0$
$\hat{P} = P$
$T_{Steps}, CDF = [0, \Delta t], [0, \hat{P}_{0,-1}]$
while $p < 1 - \epsilon$ do
    $t \leftarrow t + \Delta t$
    if $t \bmod (s + s_v) \leq s$ then
        $\hat{P} \leftarrow \hat{P}P$
    end
    else
        $\hat{P} \leftarrow \hat{P}P_v$
    end
    $p = \hat{P}_{0,-1}$
    Append $p$ to $CDF$
    Append $t$ to $T_{Steps}$
end
$CDF \leftarrow \frac{1}{p} CDF$
return $CDF$, $T_{Steps}$
```

## 3.4 Method 2

An alternative method of finding the state distributions at time $t$ is to use the following matrix exponential formula [3]:

$$P(t) = e^{Qt} \tag{15}$$

Therefore $e^{Qt}$ accounts for all the probability of the system described by the continuous transition matrix $Q$ for $t$ time units. Transient solutions, and approximate stationary solutions for this system can be found using using Equation 16.

$$\pi_{t+u} = \pi_t e^{\hat{Q}u} \tag{16}$$

Where $\hat{Q}$ is defined by Equation 17.

$$\hat{Q} = \begin{cases} Q & \text{if } t \bmod (s + s_v) \leq s \\ Q_v & \text{if } t \bmod (s + s_v) > s \end{cases} \tag{17}$$

Now $e^{Qs}$ accounts for all the probability during the period where the server is on duty, and $e^{Qs_v}$ accounts for all the probability during the vacation period. Figure 13 compares the transient solutions of both methods by plotting the CDF of reaching deadlock from an empty system.

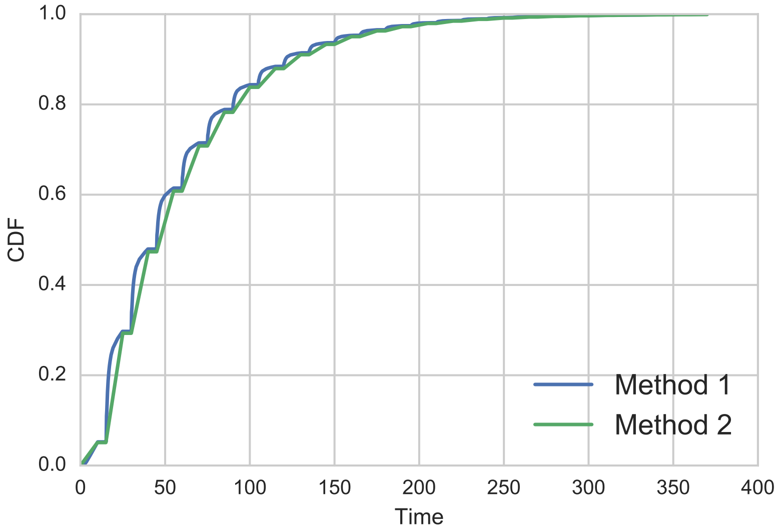


Figure 13: Comparing the CDF of the time to deadlock using Method 1 and Method 2.

At the points where the all probabilities were accounted for directly ($s$, $s_v$, $s + s_v$, $2s + s_v$, $\dots$) the solutions are similar. However, between these points this method is inaccurate. This is problematic as we require an accurate CDF in order to find the mean time to deadlock. In order to increase the number of points where the CDF is accurate, we adapt Algorithm **??** to use Method 2. This new algorithm is shown in Algorithm 2

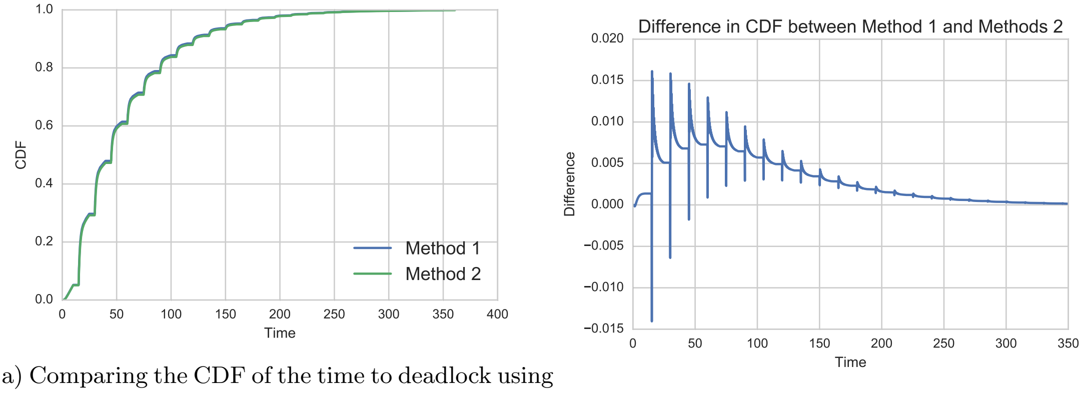
This adapted method, taking $\Delta t$ as that yielded by Theorem 2, ensures that the CDF is accurate for every multiple of $\Delta t$, that includes multiples of both $s$ and $s_v$, the times in which servers go on and off duty. Figure 14 compares the results of the algorithms for Method 1 and Method 2, both using $\Delta t$ as a time step.

**Algorithm 2:** A stepping through method to find the CDF of a system with scheduled vacations; Method 2.

**input** : $Q$, $Q_v$, $s$, $s_v$, $\Delta t$, $\epsilon$.

**output**: A cumulative density function of reaching deadlock from an empty system.

```
$t, p = 0, 0$
$\hat{Q} = e^{Q0}$
$T_{Steps}, CDF = [0], [0]$
while $p < 1 - \epsilon$ do
    $t \leftarrow t + \Delta t$
    if $t \bmod (s + s_v) \leq s$ then
        $\hat{Q} \leftarrow \hat{Q} e^{Q \Delta t}$
    end
    else
        $\hat{Q} \leftarrow \hat{Q} e^{Q_v \Delta t}$
    end
    Append $\hat{Q}_{0,-1}$ to $CDF$
    Append $t$ to $T_{Steps}$
end
$CDF \leftarrow \frac{1}{p} CDF$
return $CDF$, $T_{Steps}$
```



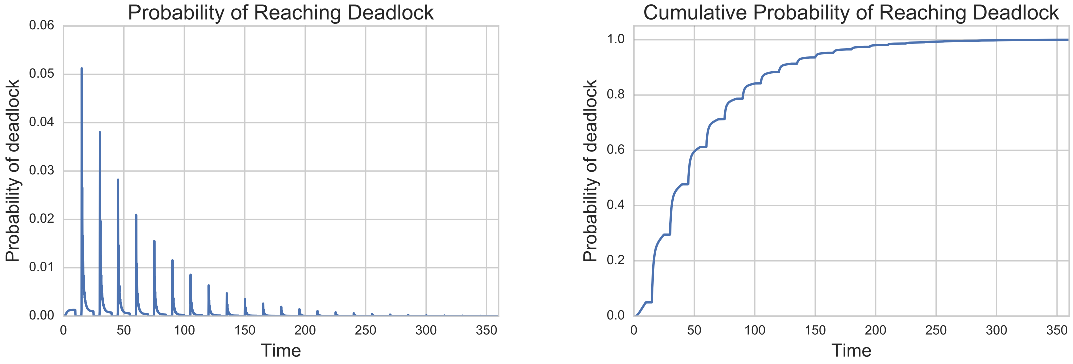
(a) Comparing the CDF of the time to deadlock using Method 1 and Method 2, same $\Delta t$.

(b) Discrepencies between Method 1 and Method 2.

Figure 14: Comparing the CDF of the time to deadlock using Method 1 and Method 2, same $\Delta t$, including discrepencies.

## 3.5 Results

Figures 15a and 15b show the PDF and CDF respectively of the system with parameters $\Lambda = 2.0$, $\mu = 4.0$, $n = 6$, $r_{11} = 0.2$, $s = 10$, $s_v = 5$. In the time periods where the server is on vacation, no customer can finish service, and so deadlock cannot be reached. Therefore the probability of reaching deadlock during these periods is 0. This is observed in the PDF where the probabilities drop to 0, and in the CDF where the cumulative probability does not increase for the duration of the vacation, at each vacation. During this vacation time, arrivals occur, but no customer leaves the queue, and so by the end of the vacation time the queue has had chance to fill up, but not empty. Therefore, immediately after a vacation period ends, the probability of deadlock shoots up as the queue will likely be in a fuller state than at the beginning of the vacation. At this point, the server returns, and so the queue begins to empty, and so the probability of deadlock decreases (although not to 0, as the server can now send customers to rejoin the queue and cause deadlock). This results in the curved step-wise shape in the CDF, and oscillatory behaviour in the PDF that is observed.



(a) Probability density function of reaching deadlock with vacations.

(b) Cumulative density function of reaching deadlock with vacations.

Figure 15: PDF & CDF of reaching deadlock in a system with scheduled vacations.

To show that the Markov chain models agree with the simulation models, Figure 16 shows the time to deadlock results of system with parameters $\Lambda = 5.0$, $\mu = 2.0$, $r_{11} = 0.5$, $s = 10$, and $s_v = 5$, for various values of the queueing capacity $n$. Violin and box plots show the distribution of the times to deadlock obtained using the simulation model (using Ciw), and the green line plot shows the corresponding solution from the Markov chain formulation.

## 3.6 Effect of Parameters of The Time to Deadlock with Vacations

In order to enable further discussion of the CDF of the time to deadlock, let's define three periods during the schedule cycle of length $s + s_v$:

- 'Steady Period': during this period arrivals and services occur as usual, and the CDF increases steadily. Note that this does not imply a steady state.

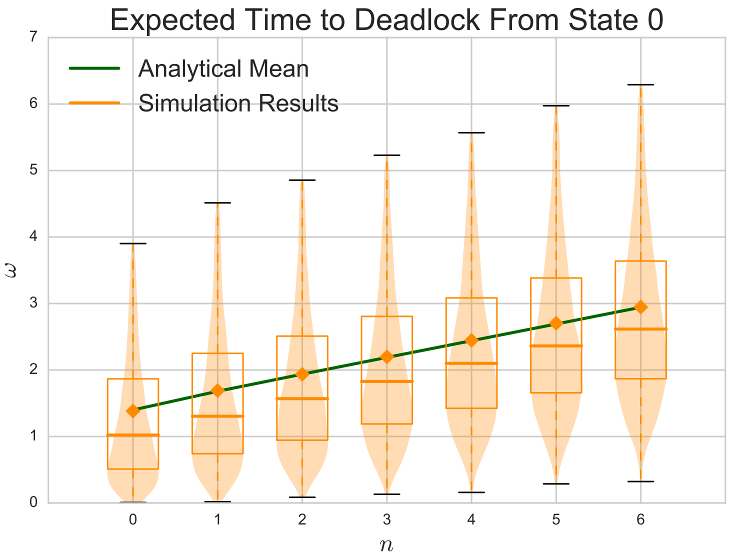
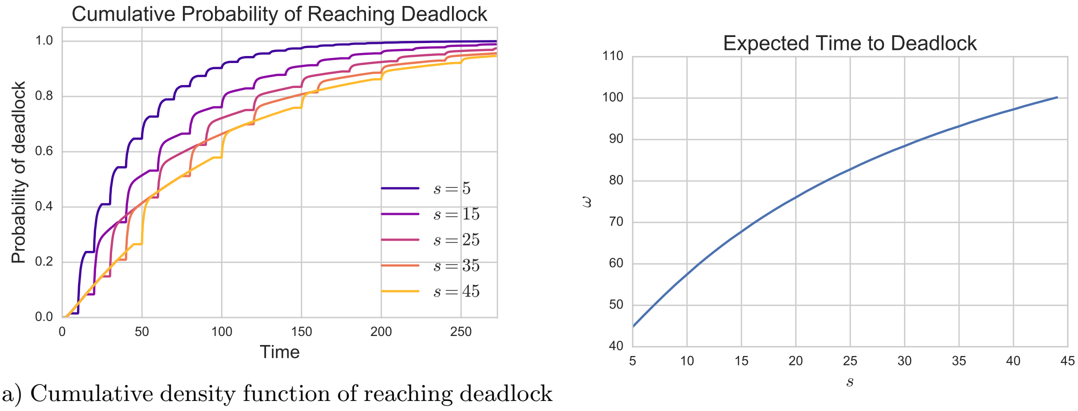


Figure 16: Plot showing the time to deadlock of a system with scheduled vacations, comparing simulation and analytical results.

- 'Vacation Period': this is the period, lasting $s_v$ time units, where the server is on vacation. During these periods, the CDF remains constant.
- 'Catch-up Period': this period occurs immediately after the server returns from vacation, after the system has had a chance to accept arrivals over the Vacation period, but not serve them. During this time the CDF shoots up, and then tails off into the Steady Period. The Steady period and the Catch-up Period lasts $s$ time units.

The height difference of the CDF during the Catch-up Period is effected by the amount the system fills up during the Vacation period. The length of the Catch-up Period is effected by how well the server can deal with the build up of customers caused by the vacation. If the server can deal with them well (for example because there is a high service rate) then the Catch-up Period will be short and will swiftly progress to a Steady Period, however if the server fails to cope, then the Catch-up Period will be long.

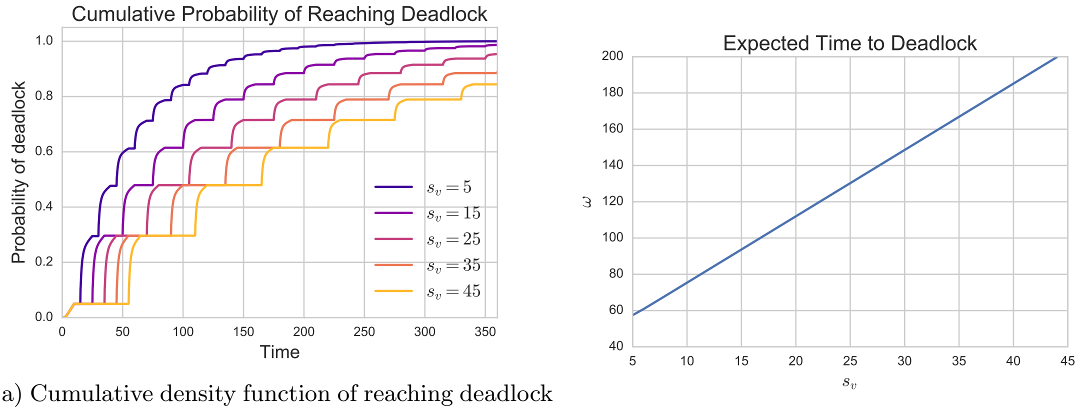
Figures 17a and 17b show the effect of the time on duty, $s$ on the cdf of reaching deadlock and the expected time to deadlock. Figures 18a and 18b show the effect of the vacation time, $s_v$ on the cdf of reaching deadlock and the expected time to deadlock. Increasing both $s$ and $s_v$ result in a fairly linear increase in the time to deadlock, however the behaviour of the system differs. Increasing $s$ allows a longer Steady Period in which the CDF increases slowly. Therefore at low values of $s$ the entire time the server is on duty is a Catch-up Period, however at high values of $s$ each period where the server is on duty is made up from a Catch-up Period and a Steady Period that increases as $s$ increases. As the CDF increases faster during the Catch-up Period, and at lower values of $s$ the system has a greater frequency of Catch-up Periods, then the CDF will increase more quickly for lower values of $s$, and so the expected time to deadlock is lower for lower values of $s$. On the other hand, increasing $s_v$ does not seem to have an effect on the height difference of the CDF during the Catch-up Period, and so increasing $s_v$ simply deplays the CDF from increasing. Therefore higher values of $s_v$ simply delay the effects of the system reaching deadlock, and so result in a longer time to deadlock.



(a) Cumulative density function of reaching deadlock with vacations, varying $s$.

(b) Effect of varying $s$ on the time to deadlock.

Figure 17: The effect of varying the time on duty, $s$, on the CDF and time to deadlock.



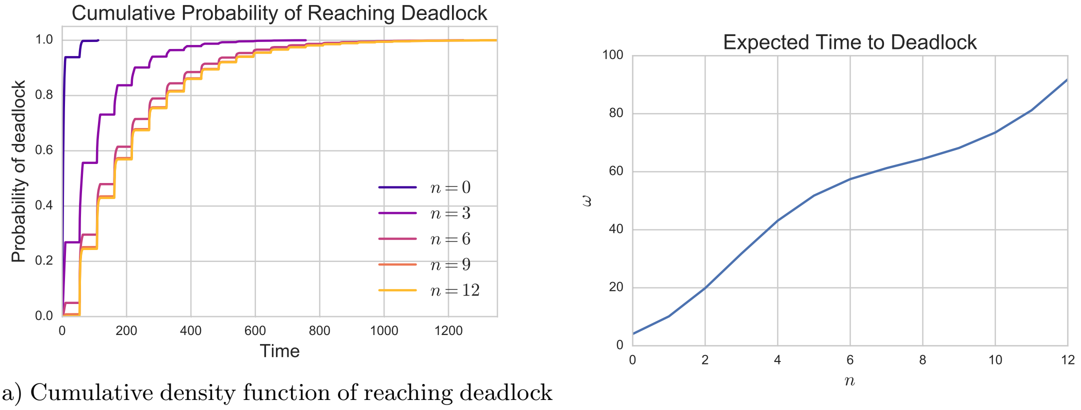
(a) Cumulative density function of reaching deadlock with vacations, varying $s_v$.

(b) Effect of varying $s_v$ on the time to deadlock.

Figure 18: The effect of varying the vacation time, $s_v$, on the CDF and time to deadlock.

Figures 19a and 19b show the effect of the queueing capacity, $n$ on the cdf of reaching deadlock and the expected time to deadlock. We see immediately that the height difference of the CDF during the Catch-up Period for smaller values of $n$ is much greater than for larger values. This is intuitive, as a queue with a smaller capacity will have greater chance of filling up during the Vacation Periods than a queue with lareger capacity. In addition the increase in CDF during the Steady Periods is steeper for smaller capacities, also intuitive as smaller capacities can fill up quicker than larger capacities, as mentioned in Chapter **??**.

Figures 20a and 20b show the effect of the service rate, $\mu$ on the cdf of reaching deadlock and the expected time to deadlock. As with the case without scheduled vacations, as $\lim_{\mu \to 0} \omega = \infty$, and $\lim_{\mu \to \infty} \omega = \infty$, with a service time threshold in between where the time to deadlock is minimised. What is interesting is the difference in behaviour of the CDF of a system the $\mu_-$ and $\mu_+$, service rates either side of the threshold that yield similar expected times to deadlock. At $\mu_-$, the service rate to the left of the threshold, the service time

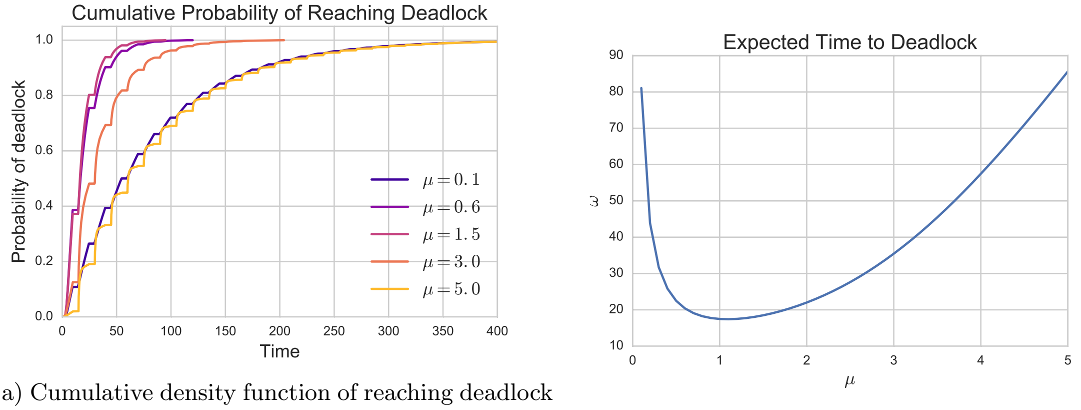


(a) Cumulative density function of reaching deadlock with vacations, varying $n$.

(b) Effect of varying $n$ on the time to deadlock.

Figure 19: The effect of varying the queueing capacity, $n$, on the CDF and time to deadlock.

is very long. Therefore it takes a long time for the first customer to finish service after the Vacation Period, and so the Catch-up Period takes a long time and increases slowly. At $\mu_+$ however, the service rate to the right of the threshold, the service rate is short, customers can leave the system quicker, and so the Catch-up Period is shorter and steeper, followed by a Steady Period in which the CDF does not increase much due to customers begin served so quickly.

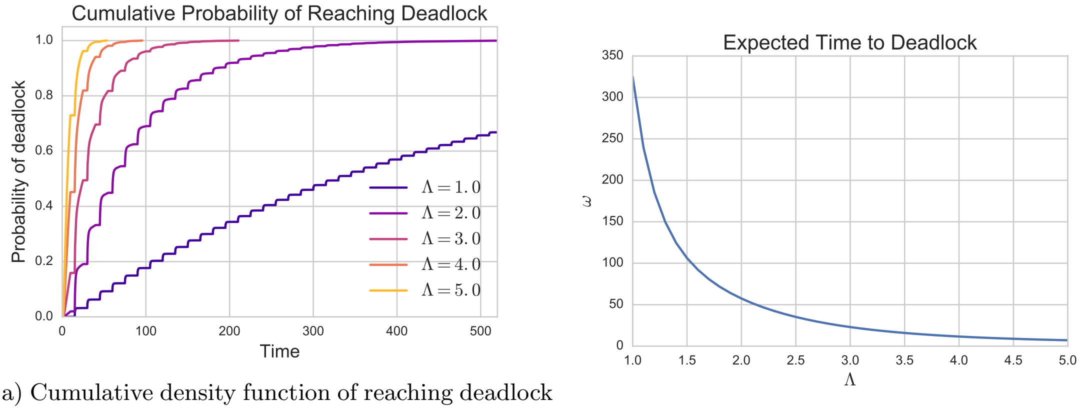


(a) Cumulative density function of reaching deadlock with vacations, varying $\mu$.

(b) Effect of varying $\mu$ on the time to deadlock.

Figure 20: The effect of varying the service rate, $\mu$, on the CDF and time to deadlock.

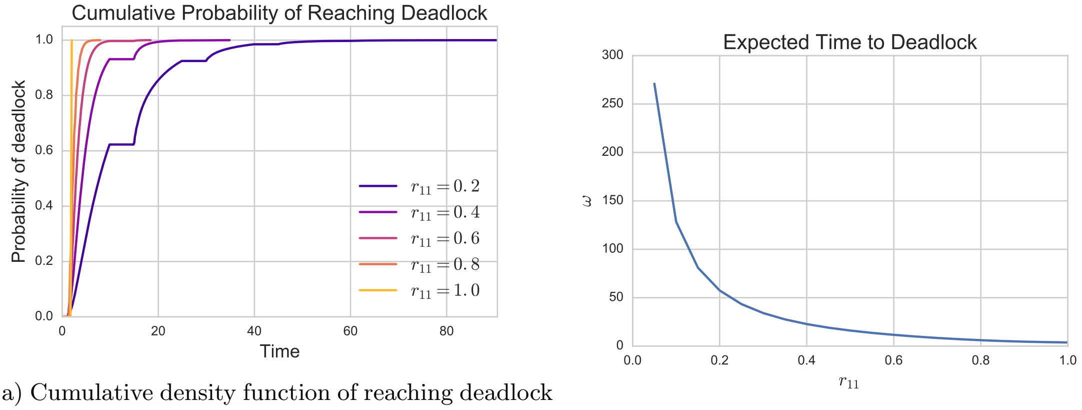
Figures 21a and 21b show the effect of the arrival rate, $\Lambda$ on the cdf of reaching deadlock and the expected time to deadlock, while Figures 22a and 22b show the effect of the transition probability, $r_{11}$ on the cdf and time to deadlock. As discussed previously, increasing the arrival rate $\Lambda$ and the transition probability $r_{11}$ decreases the time to deadlock. Looking at the CDF, higher arrival rates and transition probabilities lead to a greater jump in the CFD during the Catch-up Periods. For increasing the arrival rate, this is due to the queue filling up it's capacity quicker during the Vacation Period, and there is a higher probability of a fuller queue at the end of the server's vacation. For increasing transition probability however, the arrivals during the Vacation Period stay constant (as there are no services to cause rejoins during this time); the increase in CDF during the Catch-up Period is due to the probability of a rejoin, and hence a blockage and then deadlock, immediately after the Vacation Period once the queue has had chance to fill is increases as the transition probability increases.



(a) Cumulative density function of reaching deadlock with vacations, varying $\Lambda$.

(b) Effect of varying $\Lambda$ on the time to deadlock.

Figure 21: The effect of varying the arrival rate, $\Lambda$, on the CDF and time to deadlock.



(a) Cumulative density function of reaching deadlock with vacations, varying $r_{11}$.

(b) Effect of varying $r_{11}$ on the time to deadlock.

Figure 22: The effect of varying the transition probability, $r_{11}$, on the CDF and time to deadlock.

# References

[1] CJ Ancker Jr and AV Gafarian. Some queuing problems with balking and reneging-ii. *Operations Research*, 11(6):928–937, 1963.

[2] Nuffield Foundation. Nuffield research placements. `http://www.nuffieldfoundation.org/nuffield-research-placements`.

[3] W. Stewart. *Probability, markov chains, queues, and simulation.* Princeton university press, 2009.