The Downs-Thomson Paradox for a parallel queueing system

under state-dependent and probabilistic routing

Rein Nobel
(joined work with Marije Stolwijk)

Symposium Erik van Doorn
September 26, 2014

Taxonomy of true statements (playing with words)

- trivial [i.e. clear for everybody, but everybody who?]

Taxonomy of true statements (playing with words)

- trivial [i.e. clear for everybody, but everybody who?]
- interesting [every true statement which is not trivial]

Taxonomy of true statements (playing with words)

- trivial [i.e. clear for everybody, but everybody who?]
- interesting [every true statement which is not trivial]
- surprising [a collision between an expected and a factual truth]

Taxonomy of true statements (playing with words)

- trivial [i.e. clear for everybody, but everybody who?]
- interesting [every true statement which is not trivial]
- surprising [a collision between an expected and a factual truth]
- paradoxical [so surprising, that initially you hesitate to give up your expected truth;

Taxonomy of true statements (playing with words)

- trivial [i.e. clear for everybody, but everybody who?]
- interesting [every true statement which is not trivial]
- surprising [a collision between an expected and a factual truth]
- paradoxical [so surprising, that initially you hesitate to give up your expected truth; famous example: the French paradox: drink daily a lot of wine and live longer!]

Taxonomy of true statements (playing with words)

- trivial [i.e. clear for everybody, but everybody who?]
- interesting [every true statement which is not trivial]
- surprising [a collision between an expected and a factual truth]
- paradoxical [so surprising, that initially you hesitate to give up your expected truth; famous example: the French paradox: drink daily a lot of wine and live longer!]

We have

$$\{\text{paradoxes}\} \subset \{\text{surprising statements}\} \subset \{\text{interesting statements}\}.$$

Taxonomy of true statements (playing with words)

- trivial [i.e. clear for everybody, but everybody who?]
- interesting [every true statement which is not trivial]
- surprising [a collision between an expected and a factual truth]
- paradoxical [so surprising, that initially you hesitate to give up your expected truth; famous example: the French paradox: drink daily a lot of wine and live longer!]

We have

$$\{\text{paradoxes}\} \subset \{\text{surprising statements}\} \subset \{\text{interesting statements}\}.$$

Theorem

Using the word paradox in the title of a talk keeps the audience awake;

Taxonomy of true statements (playing with words)

- trivial [i.e. clear for everybody, but everybody who?]
- interesting [every true statement which is not trivial]
- surprising [a collision between an expected and a factual truth]
- paradoxical [so surprising, that initially you hesitate to give up your expected truth; famous example: the French paradox: drink daily a lot of wine and live longer!]

We have

$$\{\text{paradoxes}\} \subset \{\text{surprising statements}\} \subset \{\text{interesting statements}\}.$$

Theorem

Using the word paradox in the title of a talk keeps the audience awake; the term suggests that the speaker has to say something interesting which might be even surprising at first sight,

Taxonomy of true statements (playing with words)

- trivial [i.e. clear for everybody, but everybody who?]
- interesting [every true statement which is not trivial]
- surprising [a collision between an expected and a factual truth]
- paradoxical [so surprising, that initially you hesitate to give up your expected truth; famous example: the French paradox: drink daily a lot of wine and live longer!]

We have

$$\{\text{paradoxes}\} \subset \{\text{surprising statements}\} \subset \{\text{interesting statements}\}.$$

Theorem

Using the word paradox in the title of a talk keeps the audience awake; the term suggests that the speaker has to say something interesting which might be even surprising at first sight, but after a second thought the results turn out to be trivial.

Taxonomy of true statements (playing with words)

- trivial [i.e. clear for everybody, but everybody who?]
- interesting [every true statement which is not trivial]
- surprising [a collision between an expected and a factual truth]
- paradoxical [so surprising, that initially you hesitate to give up your expected truth; famous example: the French paradox: drink daily a lot of wine and live longer!]

We have

$$\{\text{paradoxes}\} \subset \{\text{surprising statements}\} \subset \{\text{interesting statements}\}.$$

Theorem

Using the word paradox in the title of a talk keeps the audience awake; the term suggests that the speaker has to say something interesting which might be even surprising at first sight, but after a second thought the results turn out to be trivial.

Proof: The proof will be presented in the Appendix.

The model described for ordinary people, i.e. non-mathematicians

- Three types of passengers arrive at an airport,
  1. business people [rich],
  2. mass tourists [poor],
  3. academic people [neither rich nor poor]

The model described for ordinary people, i.e. non-mathematicians

- Three types of passengers arrive at an airport,
  1. business people [rich],
  2. mass tourists [poor],
  3. academic people [neither rich nor poor]
- To go downtown from the airport there are two options, (i) a taxi or (ii) a shuttle bus

The model described for ordinary people, i.e. non-mathematicians

- Three types of passengers arrive at an airport,
  1. business people [rich],
  2. mass tourists [poor],
  3. academic people [neither rich nor poor]
- To go downtown from the airport there are two options, (i) a taxi or (ii) a shuttle bus
- Business people always take a taxi and mass tourists always take the shuttle bus

The model described for ordinary people, i.e. non-mathematicians

- Three types of passengers arrive at an airport,
  1. business people [rich],
  2. mass tourists [poor],
  3. academic people [neither rich nor poor]
- To go downtown from the airport there are two options, (i) a taxi or (ii) a shuttle bus
- Business people always take a taxi and mass tourists always take the shuttle bus
- Academics are free to choose between a taxi or the shuttle bus

The model described for ordinary people, i.e. non-mathematicians

- Three types of passengers arrive at an airport,
  1. business people [rich],
  2. mass tourists [poor],
  3. academic people [neither rich nor poor]
- To go downtown from the airport there are two options, (i) a taxi or (ii) a shuttle bus
- Business people always take a taxi and mass tourists always take the shuttle bus
- Academics are free to choose between a taxi or the shuttle bus
- The shuttle bus only leaves when it is full (and then immediately a new shuttle bus becomes available)

The model described for ordinary people, i.e. non-mathematicians

- Three types of passengers arrive at an airport,
  1. business people [rich],
  2. mass tourists [poor],
  3. academic people [neither rich nor poor]
- To go downtown from the airport there are two options, (i) a taxi or (ii) a shuttle bus
- Business people always take a taxi and mass tourists always take the shuttle bus
- Academics are free to choose between a taxi or the shuttle bus
- The shuttle bus only leaves when it is full (and then immediately a new shuttle bus becomes available)
- For a taxi (possibly) you have to wait in line for a free taxi.

The question for the academic:

When money is irrelevant what should I do:
- Go to the taxi stand and wait for a taxi or
- Enter the shuttle bus and wait until it is full?

The question for the academic:

When money is irrelevant what should I do:
- Go to the taxi stand and wait for a taxi or
- Enter the shuttle bus and wait until it is full?

The only criterion that counts [for the academic] is expected total transit time [sojourn time], i.e. the sum of his waiting time [in the queue for the taxi stand or in the shuttle bus] and his travel time.

The question for the academic:

When money is irrelevant what should I do:
- Go to the taxi stand and wait for a taxi or
- Enter the shuttle bus and wait until it is full?

The only criterion that counts [for the academic] is expected total transit time [sojourn time], i.e. the sum of his waiting time [in the queue for the taxi stand or in the shuttle bus] and his travel time.

We assume that the academic 'knows' the average arrival intensities of the different types of passengers, the number of taxis, the size of the shuttle bus and the travel times of the taxis and the shuttle bus [at the level of probability distributions].

What does the individual academic see upon arrival?

We distinguish two possible levels of knowledge:

1. He/she has full knowledge of the 'transport situation', i.e.
   - he can observe the number of waiting passengers at the taxi stand and
   - he can see the number of occupied seats in the shuttle bus

What does the individual academic see upon arrival?

We distinguish two possible levels of knowledge:

1. He/she has full knowledge of the 'transport situation', i.e.
   - he can observe the number of waiting passengers at the taxi stand and
   - he can see the number of occupied seats in the shuttle bus
2. He/she is not aware of the queue length at the taxi stand nor does he know the number of occupied places in the shuttle bus, but he knows all parameters involved.

What does the individual academic see upon arrival?

We distinguish two possible levels of knowledge:

1. He/she has full knowledge of the ‘transport situation’, i.e.
   - he can observe the number of waiting passengers at the taxi stand and
   - he can see the number of occupied seats in the shuttle bus
2. He/she is not aware of the queue length at the taxi stand nor does he know the number of occupied places in the shuttle bus, but he knows all parameters involved.

Ad 1 The academic can choose for a selfish strategy or an altruistic strategy

What does the individual academic see upon arrival?

We distinguish two possible levels of knowledge:

1. He/she has full knowledge of the ‘transport situation’, i.e.
   - he can observe the number of waiting passengers at the taxi stand and
   - he can see the number of occupied seats in the shuttle bus
2. He/she is not aware of the queue length at the taxi stand nor does he know the number of occupied places in the shuttle bus, but he knows all parameters involved.

Ad 1 The academic can choose for a selfish strategy or an altruistic strategy

Ad 2 All academics together can choose for a user equilibrium or for a social equilibrium.

Selfish versus altruistic strategies

When the academic upon arrival has full knowledge of the system

- he/she can choose the transport [taxi/shuttle] for which his/her own expected transit time is shorter [selfish strategy] or

Selfish versus altruistic strategies

When the academic upon arrival has full knowledge of the system

- he/she can choose the transport [taxi/shuttle] for which his/her own expected transit time is shorter [selfish strategy] or
- he/she can possibly sacrifice him/herself to guarantee a minimal long-run average transit time seen over all academics [social or altruistic strategy].

User equilibrium versus social equilibrium

When the academic upon arrival cannot observe the state of the system

- all academics can choose the taxi with a fixed probability such that the long-run average transit times at the taxi stand and at the shuttle bus are equal [user equilibrium]

User equilibrium versus social equilibrium

When the academic upon arrival cannot observe the state of the system

- all academics can choose the taxi with a fixed probability such that the long-run average transit times at the taxi stand and at the shuttle bus are equal [user equilibrium]
- all academics can choose the taxi with a fixed probability such that the long-run average transit times of all academics is minimal [social equilibrium]

User equilibrium versus social equilibrium

When the academic upon arrival cannot observe the state of the system

- all academics can choose the taxi with a fixed probability such that the long-run average transit times at the taxi stand and at the shuttle bus are equal [user equilibrium]
- all academics can choose the taxi with a fixed probability such that the long-run average transit times of all academics is minimal [social equilibrium]

To compare the different strategies our criterion of interest is this long-run average transit time of the academics.

Main question: What happens when the capacity of the taxi stand is increased?

The capacity of the taxi stand can be increased by

- faster taxis, i.e. shorter travel times

Main question: What happens when the capacity of the taxi stand is increased?

The capacity of the taxi stand can be increased by

- faster taxis, i.e. shorter travel times
- increasing the number of taxis

Main question: What happens when the capacity of the taxi stand is increased?

The capacity of the taxi stand can be increased by

- faster taxis, i.e. shorter travel times
- increasing the number of taxis
- (for queueing people only!) decreasing the variance of the travel time, *ceteris paribus*.

Main question: What happens when the capacity of the taxi stand is increased?

The capacity of the taxi stand can be increased by

- faster taxis, i.e. shorter travel times
- increasing the number of taxis
- (for queueing people only!) decreasing the variance of the travel time, *ceteris paribus*.

Ordinary people expect that the long-run average transit time of the academics will decrease when the capacity of the taxi stand will be increased.

This turns out not to be the case. For that reason we are faced with a paradox: Increasing the capacity of the taxi stand sometimes leads to longer average transit times!

Main question: What happens when the capacity of the taxi stand is increased?

The capacity of the taxi stand can be increased by

- faster taxis, i.e. shorter travel times
- increasing the number of taxis
- (for queueing people only!) decreasing the variance of the travel time, *ceteris paribus*.

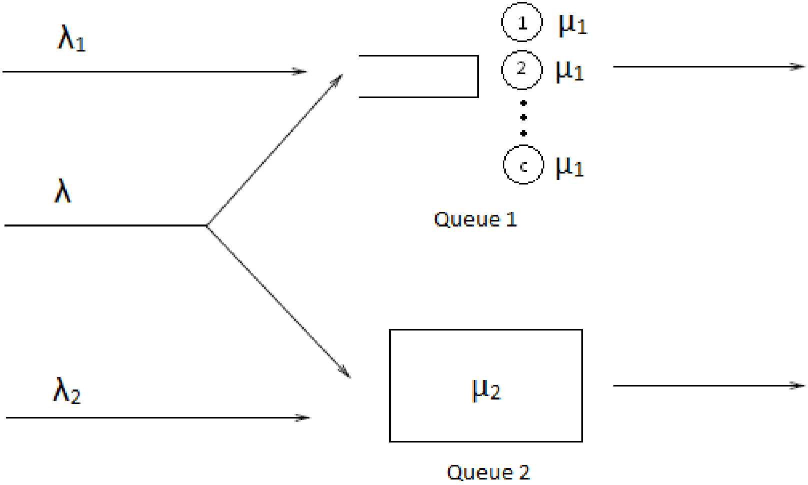
Ordinary people expect that the long-run average transit time of the academics will decrease when the capacity of the taxi stand will be increased.
This turns out not to be the case. For that reason we are faced with a paradox: Increasing the capacity of the taxi stand sometimes leads to longer average transit times! This phenomenon is called the Downs-Thomson paradox.

Model description

- Two parallel queues
  1. a standard $M/G/c$ queue with individual service in FIFO order
  2. an $M/G^{[N]}/\infty$ batch service queue: customers are served simultaneously in batches of size N
- Two Poisson streams of dedicated customers: type i arrives at queue i with rate λ_i $[i = 1, 2]$
- A third Poisson stream of general customers with rate λ
- The mean service time at queue i is $\frac{1}{\mu_i}$ $[i = 1, 2]$
- Upon arrival the general customers have to decide which queue to join.

Quantity of interest: the steady-state average transit time [sojourn time] of the general customers for different arrival strategies.

Model



What is the problem?

To study the sensitivity of the average transit time for several system parameters, given that the general customers act according to one of the following type of strategies:

- **Probabilistic routing**: with a fixed probability p general customers choose to join queue 1
- **State-dependent selfish routing**: upon arrival the general customer chooses the queue with the smaller expected transit time, given full knowledge of the state of the system
- **State-dependent social routing**: the strategy for which the overall expected transit time is minimal
- **Heuristic state-dependent routing**: upon arrival the general customer chooses the queue with the smaller estimated transit time based on incomplete knowledge of the state of the system.

Probabilistic Routing

- General customers only have knowledge of steady-state expected delay in each queue
- They choose queue 1 with probability p and queue 2 with probability $1 - p$, resulting in a steady-state average transit time $W_i(p)$ at queue i $[i = 1, 2]$
- General customers choose an optimal p according to Wardrop principle: $W_1(p) = W_2(p)$

Wardrop principle

The journey times on all routes actually used are equal, and less than those which would be experienced by a single vehicle on any unused route.

Definition

A user equilibrium is any value $p^* \in [0,1]$ which satisfies at least one of the following conditions,

1. $W_1(0) \geq W_2(0)$. Then $p^* = 0$ is a user equilibrium.
2. $W_1(1) \leq W_2(1)$. Then $p^* = 1$ is a user equilibrium.
3. For some $p^* \in (0,1) : W_1(p^*) = W_2(p^*)$. Then p^* is called a mixed user equilibrium.

We are mainly interested in so-called stable mixed user equilibria, i.e. values $p^* \in (0,1)$ with the following two properties,

1. For some $\varepsilon > 0$ and for all $p \in (p^*, \min\{p^* + \varepsilon, 1\})$: $W_1(p) > W_2(p)$
2. For some $\varepsilon > 0$ and for all $p \in (\max\{p^* - \varepsilon, 0\}, p^*)$: $W_2(p) > W_1(p)$.

User equilibria for the single-server case $[c = 1]$

$$\begin{aligned} W_1(p) &= \text{steady-state transit time for a customer who joins queue 1} \\ W_2(p) &= \text{steady-state transit time for a customer who joins queue 2} \\ W(p) &= pW_1(p) + (1-p)W_2(p) = \\ &\quad \text{the average transit time for all general customers.} \end{aligned}$$

The Pollaczek-Khintchine formula gives

$$W_1(p) = \frac{1}{\mu_1} + \frac{\lambda_1 + \lambda p}{2\mu_1(\mu_1 - \lambda_1 - \lambda p)}[1 + c_S^2]$$

A simple steady-state analysis gives

$$W_2(p) = \frac{1}{\mu_2} + \frac{N-1}{2(\lambda_2 + (1-p)\lambda)}.$$

Solve the quadratic equation $W_1(p) = W_2(p)$ for p and check whether the found equilibrium is stable.

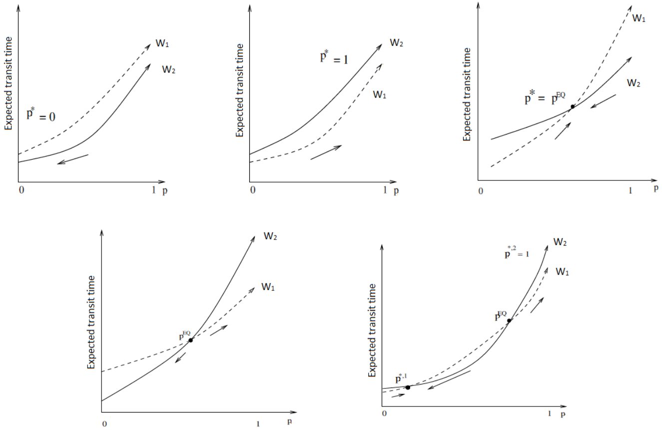
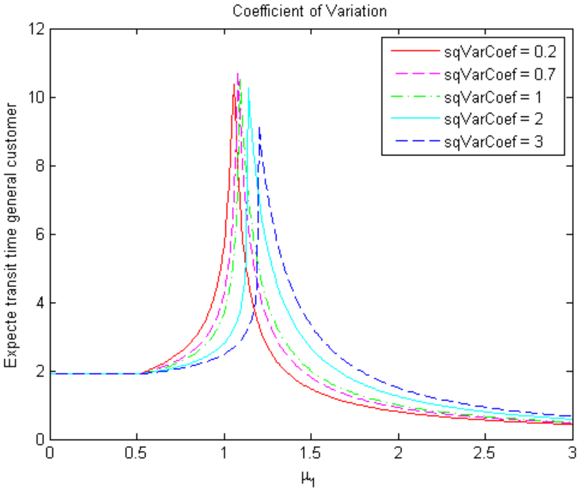


Figure: Possible user equilibria

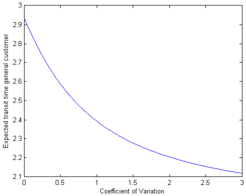
The Downs-Thomson paradox varying μ_1

1 server, $N = 3,\ \lambda = 1,\ \lambda_1 = 0,\ \lambda_2 = 0.1,\ \mu_2 = 1,\ 0 \le \mu_1 \le 3$

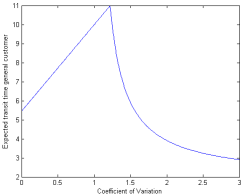


The Downs-Thomson paradox varying c_S^2

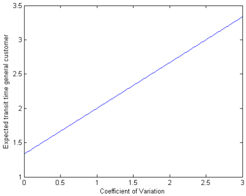
$\mu_1 = 0.8$



$\mu_1 = 1.1$



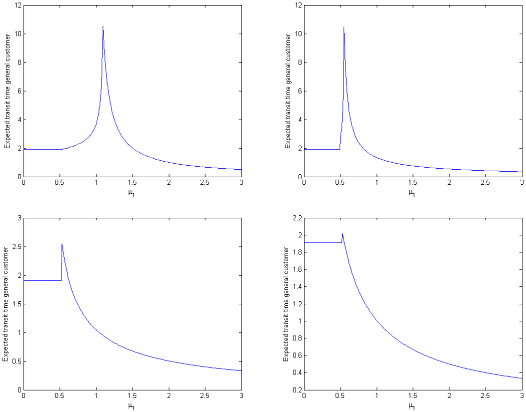
$\mu_1 = 1.5$



Paradox for c_S^2

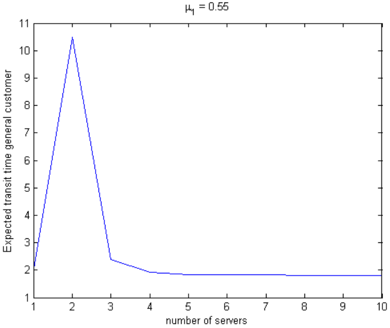
For values of μ_1 where we observe a paradox, there is also a paradox for the squared coefficient of variation

The Downs-Thomson paradox for more servers, varying μ_1



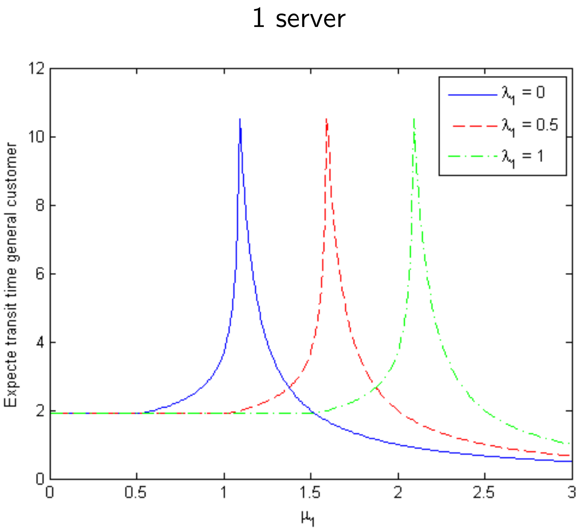
- As the number of servers increases, the interval in which there is a mixed equilibrium decreases
- Size of the paradox also decreases in the number of servers

The Downs-Thomson paradox varying the number of servers c



- μ_1 fixed at 0.55
- vary the number of servers
- Paradox found: expected transit time increases in the number of servers
- 1 server: $p^* = 0.034$
- 2 servers: $p^* = 1$

Example including λ_1 - probabilistic routing



State-dependent routing

- General customers have full knowledge of the state of the system upon arrival,
- Based on their knowledge they choose the queue with the smaller expected transit time.

For exponential service times the state space is

$$\mathcal{S} = \{(i,j) | i = 0,1,2,\ldots;\ j = 0,1,2,\ldots,N-1\}.$$

A policy or strategy for the general customers is a partition of $\mathcal{S}$ into two disjoint subsets $\mathcal{S}_1$ and $\mathcal{S}_2$ such that

$(i,j) \in \mathcal{S}_1 \iff$ the customer who sees state (i,j) chooses queue 1.

Notation $\mathcal{D} := (\mathcal{S}_1, \mathcal{S}_2)$. Define for every state $(i,j) \in \mathcal{S}$ seen by a customer upon arrival

$y_{\mathcal{D}}(i,j) =$ the expected transit time when the customer joins queue 1,

$z_{\mathcal{D}}(i,j) =$ the expected transit time when the customer joins queue 2.

Under the assumption of exponential service times we get

$$y_{\mathcal{D}}(i,j) = \frac{1}{\mu_1} + \mathbf{I}_{\{i \geq c\}} \frac{i-c+1}{c\mu_1}. \tag{1}$$

Of course,

$$z_{\mathcal{D}}(i, N-1) = \frac{1}{\mu_2} \quad \text{for} \quad i = 0, 1, 2, \ldots.$$

Further, if $(i, j+1) \in \mathcal{S}_1$ then

$$\begin{aligned} z_{\mathcal{D}}(i,j) &= \frac{1}{\lambda_1 + \lambda_2 + \lambda + \min\{i,c\}\mu_1} \times [1 + (\lambda_1 + \lambda) z_{\mathcal{D}}(i+1,j) \\ &\quad + \lambda_2 z_{\mathcal{D}}(i,j+1) + \min\{i,c\}\mu_1 z_{\mathcal{D}}(i-1,j)]. \end{aligned}$$

If on the other hand $(i, j+1) \in \mathcal{S}_2$ then

$$\begin{aligned} z_{\mathcal{D}}(i,j) &= \frac{1}{\lambda_1 + \lambda_2 + \lambda + \min\{i,c\}\mu_1} \times [1 + \lambda_1 z_{\mathcal{D}}(i+1,j) \\ &\quad + (\lambda_2 + \lambda) z_{\mathcal{D}}(i,j+1) + \min\{i,c\}\mu_1 z_{\mathcal{D}}(i-1,j)]. \end{aligned}$$

How to determine the selfish policy $\mathcal{D}^* = (\mathcal{S}_1^*, \mathcal{S}_2^*)$ for which

$$(i,j) \in \mathcal{S}_1^* \iff y_{\mathcal{D}^*}(i,j) < z_{\mathcal{D}^*}(i.j)? \tag{2}$$

We build up this policy $\mathcal{D}^*$ gradually as follows $[\lambda_1 = 0]$

1. Start with $z_{\mathcal{D}^*}(i, N-1) = \frac{1}{\mu_2}$ and compare these quantities with $y_{\mathcal{D}^*}(i, N-1)$ for $i = 0, 1, 2, \ldots$
2. Then $(i, N-1) \in \mathcal{S}_1^* \iff y_{\mathcal{D}^*}(i, N-1) < z_{\mathcal{D}^*}(i, N-1)$
3. Suppose we find $(i, N-1) \in \mathcal{S}_1^*$ for $i = 0, 1, \ldots, i_{N-1}$ and $(i, N-1) \in \mathcal{S}_2^*$ for $i = i_{N-1}+1, i_{N-1}+2, \ldots$
4. Then using the recursion scheme, set up a system of $i_{N-1}+2$ linear equations to calculate $z_{\mathcal{D}^*}(i, N-2)$ for $i = 0, 1, \ldots, i_{N-1}+1$
5. Now for $i = i_{N-1}+2, i_{N-1}+3, \ldots$ the $z_{\mathcal{D}^*}(i, N-2)$ can be calculated directly from the recursion scheme
6. Then $(i, N-2) \in \mathcal{S}_1^* \iff y_{\mathcal{D}^*}(i, N-2) < z_{\mathcal{D}^*}(i, N-2)$ for $i = 0, 1, 2, \ldots$.
7. Continue the above procedure for $j = N-3, \ldots, 0$.

The overall average transit time

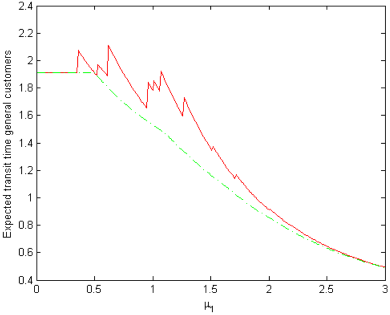
Once we have determined the optimal selfish policy $\mathcal{D}^*$ we can calculate the steady-state distribution $\{\pi_{\mathcal{D}^*}(i,j)\}_{(i,j)\in\mathcal{S}}$ of the continuous-time Markov chain [CTMC] which describes the probabilistic evolution when the system is controlled by policy $\mathcal{D}^*$.

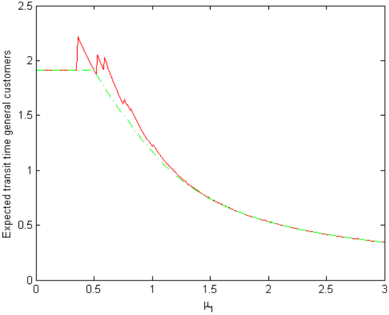
The overall mean transit time for the general customers, say $W_{\mathcal{D}^*}$, can then be calculated as

$$W_{\mathcal{D}^*} = \sum_{(i,j)\in\mathcal{S}} \pi_{\mathcal{D}^*}(i,j) \left[y_{\mathcal{D}^*}(i,j) \mathbf{I}_{\{(i,j)\in\mathcal{S}_1^*\}} + z_{\mathcal{D}^*}(i,j) \mathbf{I}_{\{(i,j)\in\mathcal{S}_2^*\}} \right]. \tag{3}$$

Numerical example for different servers I

$$N = 3, \; M = 20, \; \lambda = 1, \; \lambda_1 = 0, \; \lambda_2 = 0.1, \; \mu_2 = 1, \; 0 \leq \mu_1 \leq 3$$





Numerical example for different servers II

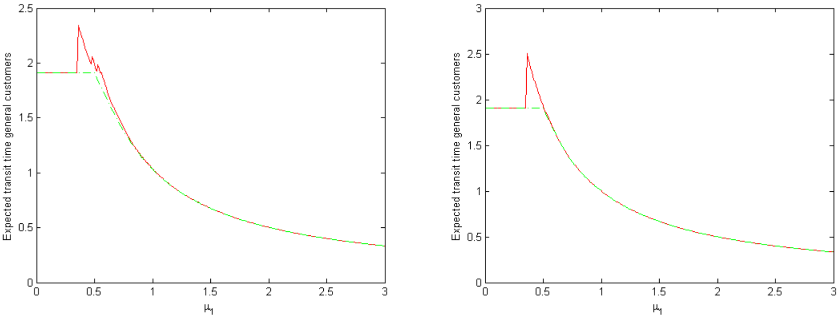


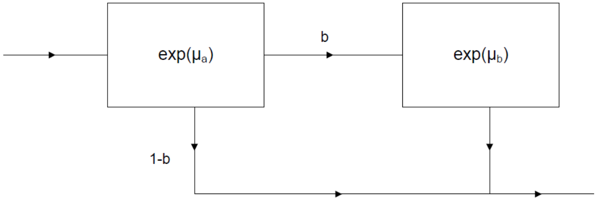
Figure: The expected transit times of general customers under state-dependent routing for the selfish policy (red solid line) and for the social optimal policy (green dotted line) for 1, 2, 3 and 5 servers.

Coxian-2 service time

A random variable X is Coxian-2 distributed if S can be represented as:

$$X = \begin{cases} A + B & \text{with probability } b \\ A & \text{with probability } 1 - b \end{cases}$$

where $A \sim \exp(\mu_a)$ and $B \sim \exp(\mu_b)$, A, B independent random variables.



For Coxian-2 distributed service times the state space is

$$\mathcal{S} = \{(i,j,k) | i = 0,1,2,\ldots;\ j = 0,1,2,\ldots,N-1;\ k = 0,1,\ldots\},$$

where $i = \#$customers in queue 1, $j = \#$customers in queue 2 and $k = \#$customers in service-phase 1.
Now for a policy $\mathcal{D} = (\mathcal{S}_1, \mathcal{S}_2)$ we have

$$(i,j,k) \in \mathcal{S}_1 \iff \text{the customer who sees state } (i,j,k) \text{ chooses queue 1,}$$

Define again for every state $(i,j,k) \in \mathcal{S}$ seen by a customer upon arrival

$$y_{\mathcal{D}}(i,j,k) = \text{the expected transit time when the customer joins queue 1,}$$

$$z_{\mathcal{D}}(i,j,k) = \text{the expected transit time when the customer joins queue 2.}$$

Recursion scheme for the $y_{\mathcal{D}}(i,j,k)$

$$
y_{\mathcal{D}}(i,j,k) = \begin{cases}
\dfrac{1}{k\mu_a + (c-k)\mu_b}\Big(1 + bk\mu_a y_{\mathcal{D}}(i,j,k-1) \\
+(1-b)k\mu_a y_{\mathcal{D}}(i-1,j,k) \\
+(c-k)\mu_b y_{\mathcal{D}}(i-1,j,k+1)\Big), & i > c, \\
\dfrac{1}{k\mu_a + (c-k)\mu_b}\Big(1 + bk\mu_a y_{\mathcal{D}}(i,j,k-1) \\
+(1-b)k\mu_a y_{\mathcal{D}}(i-1,j,k-1) \\
+(c-k)\mu_b y_{\mathcal{D}}(i-1,j,k)\Big), & i = c, \\
\dfrac{1}{\mu_a} + \dfrac{b}{\mu_b}, & i < c.
\end{cases} \quad (4)
$$

Recursion scheme for the $z_{\mathcal{D}}(i, j, k)$

$$z_{\mathcal{D}}(i, N-1, k) = \frac{1}{\mu_2}, \quad \text{for } i = 0, 1, \ldots; \; k = 0, 1, \ldots \min\{i, c\}.$$

Let $\Lambda(i, k) = \lambda + \lambda_1 + \lambda_2 + i\mu_a + (\min\{i, c\} - k)\mu_b$.
If $(i, j+1, k) \in \mathcal{S}_1$:

$$z_{\mathcal{D}}(i, j, k) = \begin{cases} \frac{1}{\Lambda(i,k)}\Big(1 + (\lambda + \lambda_1) z_{\mathcal{D}}(i+1, j, k + \mathbf{I}_{\{i<c\}}) + \lambda_2 z_{\mathcal{D}}(i, j+1, k) \\ + bk\mu_a z_{\mathcal{D}}(i, j, k-1) + (1-b)k\mu_a z_{\mathcal{D}}(i-1, j, k-1) \\ + (i-k)\mu_b z_{\mathcal{D}}(i-1, j, k)\Big), \quad i = 0, 1, \ldots, c \\ \frac{1}{\Lambda(i,k)}\Big(1 + (\lambda + \lambda_1) z_{\mathcal{D}}(i+1, j, k) + \lambda_2 z_{\mathcal{D}}(i, j+1, k) \\ + bk\mu_a z_{\mathcal{D}}(i, j, k-1) + (1-b)k\mu_a z_{\mathcal{D}}(i-1, j, k) \\ + (c-k)\mu_b z_{\mathcal{D}}(i-1, j, k+1)\Big), \quad i = c+1, c+2, \ldots \end{cases} \quad (5)$$

If $(i, j+1, k) \in \mathcal{S}_2$:

$$z_{\mathcal{D}}(i,j,k) = \begin{cases} \frac{1}{\Lambda(i,k)}\Big(1 + \lambda_1 z_{\mathcal{D}}(i+1, j, k + \mathbf{I}_{\{i<c\}}) + (\lambda+\lambda_2) z_{\mathcal{D}}(i, j+1, k) \\ + bk\mu_a z_{\mathcal{D}}(i,j,k-1) + (1-b)k\mu_a z_{\mathcal{D}}(i-1,j,k-1) \\ + (i-k)\mu_b z_{\mathcal{D}}(i-1,j,k)\Big), \quad i = 0,1,\ldots,c \\ \frac{1}{\Lambda(i,k)}\Big(1 + \lambda_1 z_{\mathcal{D}}(i + \mathbf{I}_{\{n_1<c\}}, j, k) + (\lambda+\lambda_2) z_{\mathcal{D}}(i, j+1, k) \\ + bk\mu_a z_{\mathcal{D}}(i,j,k-1) + (1-b)k\mu_a z_{\mathcal{D}}(i-1,j,k) \\ + (c-k)\mu_b z_{\mathcal{D}}(i-1,j,k+1)\Big), \quad i = c+1, c+2, \ldots. \end{cases} \tag{6}$$

We want to find $\mathcal{D}^* = (\mathcal{S}_1^*, \mathcal{S}_2^*)$ for which

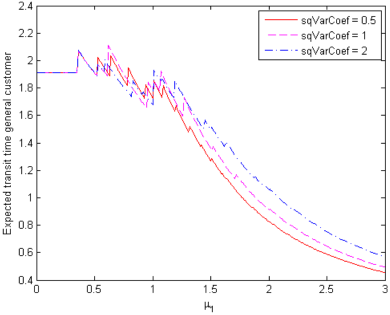
$$(i,j,k) \in \mathcal{S}_1^* \iff y_{\mathcal{D}^*}(i,j,k) < z_{\mathcal{D}^*}(i.j,k) \tag{7}$$

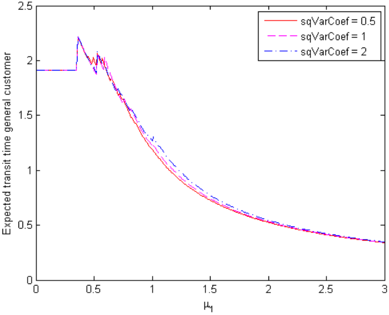
Again we build up this policy $\mathcal{D}^*$ gradually, but now we introduce truncation in queue 1:
When M customers are present in queue 1, no other customers will be accepted. Then we can proceed as before

1. Start with $z_{\mathcal{D}^*}(i, N-1, k) = \frac{1}{\mu_2}$ and compare these quantities with $y_{\mathcal{D}^*}(i, N-1, k)$ for $i = 0, 1, 2, \ldots$ and $k = 0, 1, \ldots, c$
2. Then
$(i, N-1, k) \in \mathcal{S}_1^* \iff y_{\mathcal{D}^*}(i, N-1, k) < z_{\mathcal{D}^*}(i, N-1, k)$
3. Using the recursion scheme, set up a system of linear equations to calculate $z_{\mathcal{D}^*}(i, N-2,)$ for $i = 0, 1, \ldots, M$ and $k = 0, 1, \ldots, c$
4. Then
$(i, N-2, k) \in \mathcal{S}_1^* \iff y_{\mathcal{D}^*}(i, N-2, k) < z_{\mathcal{D}^*}(i, N-2, k)$ for $i = 0, 1, 2, \ldots, M$ and $k = 0, 1, \ldots, c$
5. Continue the above procedure for $j = N-3, \ldots, 0$.

State-dependent routing - User optimum I

$$N = 3,\ M = 20,\ \lambda = 1,\ \lambda_1 = 0,\ \lambda_2 = 0.1,\ \mu_2 = 1,\ 0 \le \mu_1 \le 3$$





State-dependent routing - User optimum II

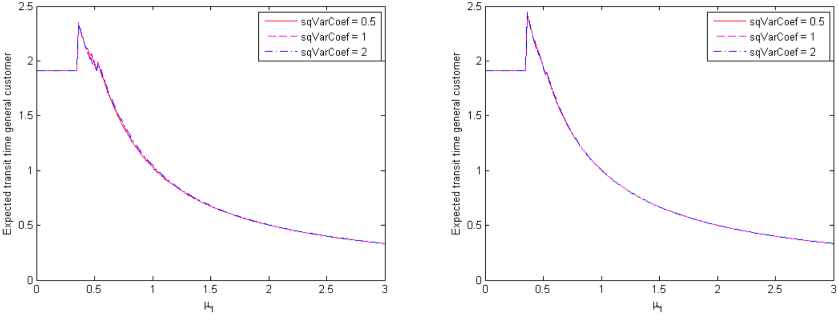
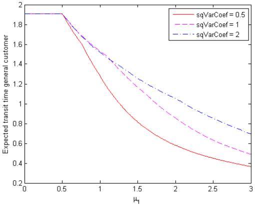


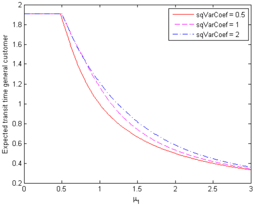
Figure: The expected transit times of general customers under state-dependent routing for the selfish policy for 1, 2, 3 and 4 servers and different values of the squared coefficient of variation.

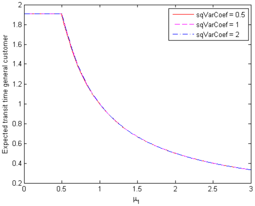
State-dependent routing - Social optimum

The optimal social policy can be calculated using Markov Decision Theory (no details here). Then of course, no paradox shows up!

$$N = 3,\ M = 20,\ \lambda = 1,\ \lambda_1 = 0,\ \lambda_2 = 0.1,\ \mu_2 = 1,\ 0 \leq \mu_1 \leq 3$$

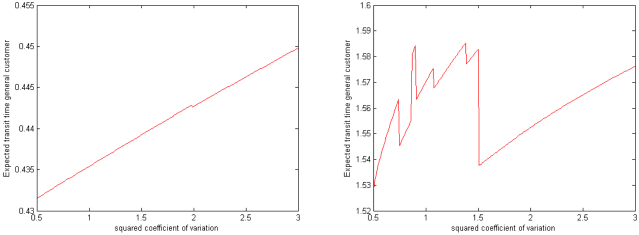






Two servers, selfish policy

When varying the squared coefficient of variation and fixing the value of μ_1 at 2.4 [not in the D-T interval!] and 0.8:

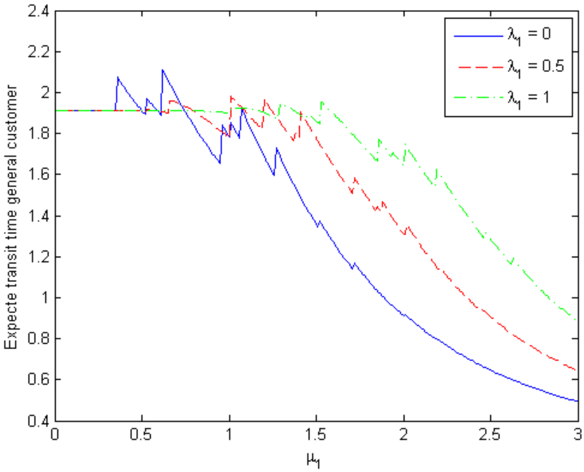


Paradox for c_S^2

A paradox is observed for the squared coefficient of variation, which, just as the paradox for μ_1, shows multiple small jumps.

Example including λ_1 - state-dependent routing

1 server



Heuristic estimates for queue 1 and queue 2

In practice customers who want to decide which queue to join have no knowledge of $k = \#$customers in service-phase 1. So, they cannot calculate

$y_{\mathcal{D}}(i, j, k) =$ the expected transit time when the customer joins queue 1,

let alone

$z_{\mathcal{D}}(i, j, k) =$ the expected transit time when the customer joins queue 2,

which also depends on future decisions.

So, we propose that customers only have knowledge of the number of customers present in queue 1 and 2: i and j and we define

$$y_{\mathcal{D}}^{H} = \frac{1}{\mu_1} + \mathbf{I}_{\{i \geq c\}} \frac{i+1-c}{c\mu_1},$$

$$z_{\mathcal{D}}^{H}(i,j) = w_1 \left(\frac{N-i-1}{\lambda_2} \right) + w_2 \left(\frac{N-i-1}{\lambda+\lambda_2} \right) + \frac{1}{\mu_2}, \quad \text{with } w_1 + w_2 = 1.$$

We introduce a heuristic state-dependent policy $\mathcal{D}^H = (\mathcal{S}_1^H, \mathcal{S}_2^H)$ by

$$(i,j) \in \mathcal{S}_1^H \iff y_{\mathcal{D}^H}^H(i,j) < z_{\mathcal{D}^H}^H(i,j). \tag{8}$$

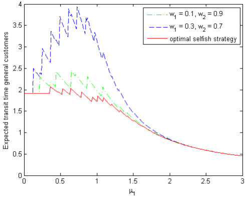
For this policy $\mathcal{D}^H$ calculate the overall average transit time $W_{\mathcal{D}^H}$ by considering the CTMC induced by policy $\mathcal{D}^H$, $W_{\mathcal{D}^H} =$

$$\sum_{(i,j,k) \in \mathcal{S}} \pi_{\mathcal{D}^H}(i,j,k) \left[y_{\mathcal{D}^H}(i,j,k) \mathbf{I}_{\{(i,j,k) \in \mathcal{S}_1^H\}} + z_{\mathcal{D}^H}(i,j,k) \mathbf{I}_{\{(i,j,k) \in \mathcal{S}_2^H\}} \right].$$

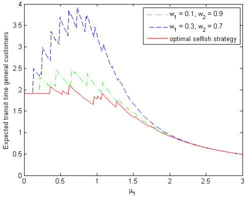
Examples of heuristic policies I

The expected transit times of general customers under state-dependent routing for the optimal selfish policy and two heuristics for one, two and three servers and different values of the squared coefficient of variation.

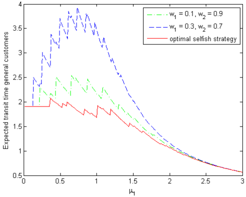
$c_S^2 = 0.5$



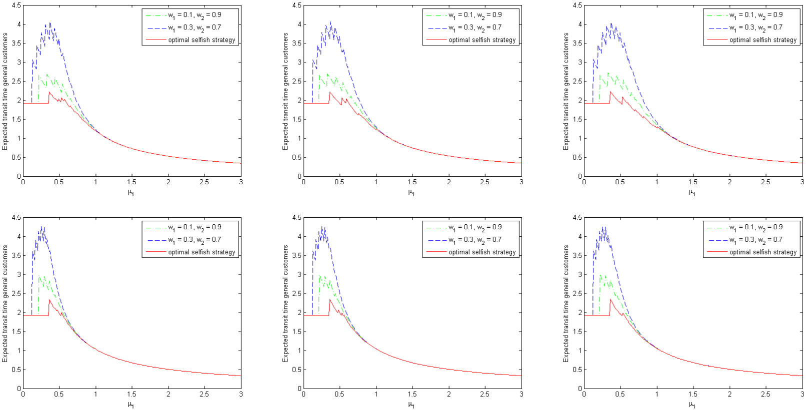
$c_S^2 = 1$



$c_S^2 = 2$

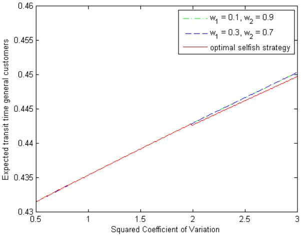


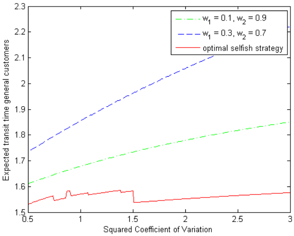
Examples of heuristic policies II



Two servers, heuristic policy

When varying the squared coefficient of variation and fixing the value of μ_1 at 2.4 [not in the D-T interval!] and 0.8:





No paradox for c_S^2

A paradox is not observed for the squared coefficient of variation, due to the fact that the policy does not change for different values of the squared coefficient of variation.

Summary

- The Downs-Thomson paradox shows up for user optimal policies under both probabilistic and state-dependent routing for the service rate μ_1, the squared coefficient of variation c_S^2 and the number of servers c
- For almost all values of μ_1, c_S^2 and c having full knowledge of the system mitigates the effect of the paradox
- For natural intuitively appealing strategies based on incomplete knowledge of the system the effects of the Downs-Thomson paradox can be dramatic
- The paradox only shows up when changing a parameter results in a different policy
- For optimal social state-dependent strategies [calculated by Markov decision theory] no paradoxes show up.

Appendix

Theorem

Using the word paradox in the title of a talk keeps the audience awake; the term suggests that the speaker has to say something interesting which might be even surprising at first sight, but after a second thought the results turn out to be trivial.

Appendix

Theorem

Using the word paradox in the title of a talk keeps the audience awake; the term suggests that the speaker has to say something interesting which might be even surprising at first sight, but after a second thought the results turn out to be trivial.

Proof: Trivial!

State-dependent routing - Social optimum

- $X_1(t)$ = the number of customers in queue 1, including any customer in service, at time t
- $X_2(t)$ = the number of general customers waiting for service in queue 2, not including those customers already in service, at time t
- $X_3(t)$ = the number of dedicated customers to queue 2 waiting for service in queue 2, not including those customers already in service, at time t.
- State space: $S = \{(n_1, n_2, n_3) : n_1 \in \{0, 1, 2, \dots, C\}, n_2, n_3 \in \{0, 1, 2, \dots, N-1\}, n_2 + n_3 \leq N-1\}$

Let $\Lambda = \lambda_1 + \lambda_2 + \lambda + \mu_1$.

State-dependent routing - Social optimum

Transition probabilities: [for simplicity we take $\lambda_1 = 0$]

$$p(\mathbf{n}, \mathbf{n}'; 1) = \begin{cases} \frac{\mu_1}{\Lambda} & \mathbf{n}' = \mathbf{n} - \mathbf{e}_1 \mathbf{I}_{\{n_1 > 0\}} \\ \frac{\lambda}{\Lambda} & \mathbf{n}' = \mathbf{n} + \mathbf{e}_1 \mathbf{I}_{\{n_1 < C\}} \\ \frac{\lambda_2}{\Lambda} & \mathbf{n}' = \mathbf{n} + \mathbf{e}_3 \text{ if } n_2 + n_3 < N - 1 \\ \frac{\lambda_2}{\Lambda} & \mathbf{n}' = (n_1, 0, 0) \text{ if } n_2 + n_3 = N - 1 \\ 0 & \text{otherwise} \end{cases}$$

$$p(\mathbf{n}, \mathbf{n}'; 2) = \begin{cases} \frac{\mu_1}{\Lambda} & \mathbf{n}' = \mathbf{n} - \mathbf{e}_1 \mathbf{I}_{\{n_1 > 0\}} \\ \frac{\lambda}{\Lambda} & \mathbf{n}' = \mathbf{n} + \mathbf{e}_2 \text{ if } n_2 + n_3 < N - 1 \\ \frac{\lambda_2}{\Lambda} & \mathbf{n}' = \mathbf{n} + \mathbf{e}_3 \text{ if } n_2 + n_3 < N - 1 \\ \frac{\lambda + \lambda_2}{\Lambda} & \mathbf{n}' = (n_1, 0, 0) \text{ if } n_2 + n_3 = N - 1 \\ 0 & \text{otherwise.} \end{cases}$$

State-dependent routing - Social optimum

Cost function:

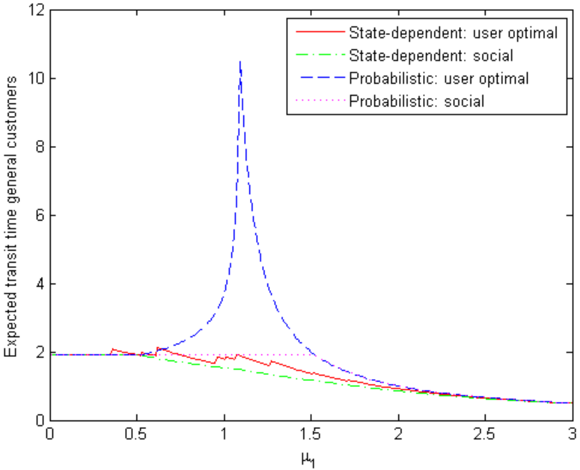
$$
\begin{aligned}
c(\mathbf{n};1) &= (n_1+n_2) && 0 \le n_1 \le C-1,\ 0 \le n_2 \le N-1,\ 0 \le n_3 \le N-1 \\
c(\mathbf{n};2) &= (n_1+n_2)+\frac{\lambda}{\mu_2} && 0 \le n_1 \le C,\ 0 \le n_2 \le N-1,\ 0 \le n_3 \le N-1.
\end{aligned}
$$

Recursion:

$$
\begin{aligned}
V_{k+1}(\mathbf{n}) \quad = \quad & n_1+n_2+\frac{\mu_1}{\Lambda}\left(V_k(\mathbf{n}-\mathbf{e}_1)\mathbf{I}_{\{n_1>0\}}+V_k(\mathbf{n})\mathbf{I}_{\{n_1=0\}}\right) \\
&+\frac{\lambda_2}{\Lambda}\left(V_k(\mathbf{n}+\mathbf{e}_3)\mathbf{I}_{\{n_2+n_3<N-1\}}+V_k(n_1,0,0)\mathbf{I}_{\{n_2+n_3=N-1\}}\right) \\
&+\frac{\lambda}{\Lambda}\left(\left\{\frac{\Lambda}{\mu_2}+V_k(\mathbf{n}+\mathbf{e}_2)\mathbf{I}_{\{n_2+n_3<N-1\}}+V_k(n_1,0,0)\mathbf{I}_{\{n_2+n_3=N-1\}}\right\}\mathbf{I}_{\{n_1=C\}}\right. \\
&+\min\left\{V_k(\mathbf{n}+\mathbf{e}_1);\ \frac{\Lambda}{\mu_2}+V_k(\mathbf{n}+\mathbf{e}_2)\mathbf{I}_{\{n_2+n_3<N-1\}}\right. \\
&\left.\left.+V_k(n_1,0,0)\mathbf{I}_{\{n_2+n_3=N-1\}}\right\}\mathbf{I}_{\{n_1<C\}}\right).
\end{aligned}
$$

Numerical example

$$N = 3,\ M = 20,\ \lambda = 1,\ \lambda_1 = 0,\ \lambda_2 = 0.1,\ 0 \le \mu_1 \le 3$$



Strategy for $\mu_1 = 1$

- Strategy for the social optimum:

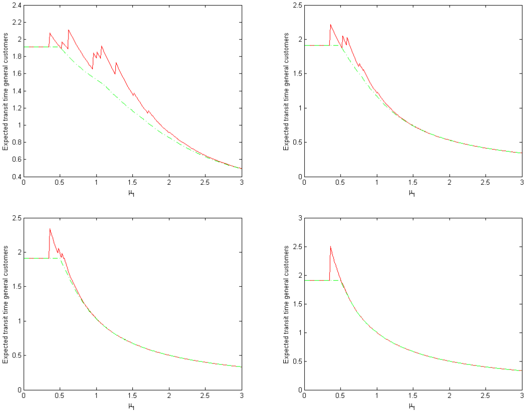
$$s_S(n_1, n_2, 0) = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \\ \vdots & \vdots & \vdots \\ 2 & 2 & 2 \end{bmatrix}, \quad s_S(n_1, n_2, 1) = \begin{bmatrix} 1 & 2 & - \\ 2 & 2 & - \\ 2 & 2 & - \\ \vdots & \vdots & \vdots \\ 2 & 2 & - \end{bmatrix}, \quad s_S(n_1, n_2, 2) = \begin{bmatrix} 1 & - & - \\ 2 & - & - \\ 2 & - & - \\ \vdots & \vdots & \vdots \\ 2 & - & - \end{bmatrix}$$

- Strategy for the user optimum:

$$s_U(n_1, n_2) = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 2 \\ 1 & 2 & 2 \\ 2 & 2 & 2 \\ \vdots & \vdots & \vdots \\ 2 & 2 & 2 \end{bmatrix}$$

- Strategy for probabilistic routing: $p^* = 0.7298$

State-dependent routing



- Social optimum: expected transit time decreases in the number of servers
- User optimum:
  - Fewer paradoxes observed for more servers
  - Increase in paradox for $\mu_1 = 0.36$

The $M/M/1{-}$R retrial queue versus the $M/G^{[N]}/\infty$ batch-service queue

- Two parallel queues
  1. a standard $M/M/1$ retrial queue with individual service
  2. an $M/G^{[N]}/\infty$ batch service queue: customers are served simultaneously in batches of size N
- Two Poisson streams of dedicated customers: type i arrives at queue i with rate λ_i [$i = 1, 2$]
- A third Poisson stream of general customers with rate λ
- The mean service time at queue i is $\frac{1}{\mu_i}$ [$i = 1, 2$]
- The mean retrial time at queue 1 is $\frac{1}{\nu}$ [exponential]
- Upon arrival the general customers have to decide which queue to join.

Quantity of interest: the steady-state average transit time [sojourn time] of the general customers for different arrival strategies.

Intermezzo (joint work with Dinard van der Laan)

Consider the $M/M/1-R$ retrial queue modelled as a CTMC $\{(C_t, Q_t),\ t \geq 0\}$ with its state-space

$$\mathcal{S} = \{(k, j)\, | k = 0, 1;\ j = 0, 1, 2, \ldots\}$$

Here C_t describes the state of the server (0=idle, 1=busy) and Q_t the number of customers in the orbit at time t.
Introduce a tagged customer [in the orbit] and define

$y^*(j, k) \quad = \quad$ the expected (residual) delay of the tagged customer in the orbit, given that j other customers are present in the orbit and the server state is k.

We have the following recursions

$$y^*(j,0) = \frac{1+\lambda y^*(j,1)+j\nu y^*(j-1,1)}{\lambda+(j+1)\nu}, \quad (9)$$

$$y^*(j,1) = \frac{1+\lambda y^*(j+1,1)+(j+1)\nu y^*(j,1)+\mu y^*(j,0)}{\lambda+(j+1)\nu+\mu}. \quad (10)$$

Substituting (9) in (10) gives after some manipulations

$$(\lambda^2+(j+1)\nu\lambda)[y^*(j+1,1)-y^*(j,1)]+\lambda+\mu+(j+1)\nu = (j+1)\nu\mu[y^*(j,1)-y^*(j-1,1)]+\nu\mu y^*(j-1,1). \quad (11)$$

With $y^*(-1,1)=0$ and the conjecture $y^*(j+1,1)-y^*(j,1)=C=\frac{1}{2\mu-\lambda}$ we find from (11) and (9)

$$y^*(j,1) = \frac{\lambda+2\mu+(j+2)\nu}{\nu(2\mu-\lambda)}, \quad (12)$$

$$y^*(j,0) = \frac{\lambda+2\mu+j\nu}{\nu(2\mu-\lambda)}. \quad (13)$$

The conjecture

$$y^*(j+1,1) - y^*(j,1) = \text{CONSTANT} = \frac{1}{2\mu - \lambda}$$

has been checked using the well-known results for the $M/M/1$-R queue ($\rho = \lambda/\mu$):

$$\overline{D} = \frac{\lambda(\mu+\nu)}{\mu\nu(\mu-\lambda)} \quad \text{and} \quad p_{1j} = \frac{\rho^{j+1}}{j!\nu^j} \prod_{i=1}^{j} (\lambda + i\nu)(1-\rho)^{\frac{\lambda}{\nu}+1}$$

where $\overline{D}$ is the long-run average delay in the orbit and $p_{kj} = \lim_{t\to\infty} \Pr(C_t = k;\ Q_t = j)$ the limiting distribution of the CTMC $\{(C_t, Q_t),\ t \geq 0\}$. Then we find $\overline{D} = \sum_{j=0}^{\infty} y^*(j,1) p_{1j} \stackrel{?}{=}$

$$\sum_{j=0}^{\infty} \frac{\lambda + 2\mu + (j+2)\nu}{\nu(2\mu - \lambda)} \frac{\rho^{j+1}}{j!\nu^j} \prod_{i=1}^{j} (\lambda + i\nu)(1-\rho)^{\frac{\lambda}{\nu}+1} = \cdots = \frac{\lambda(\mu+\nu)}{\mu\nu(\mu-\lambda)}.$$

This is encouraging, but not a formal proof of the conjecture!!

WANTED: a probabilistic proof for the conjecture

$$\forall j: \quad y^*(j+1,1) - y^*(j,1) = \text{CONSTANT} = \frac{1}{2\mu - \lambda}.$$

SIMULATION RESULTS SHOW A PERFECT CORRESPONDENCE EVEN FOR $\mu < \lambda < 2\mu$!!

A retrial queue parallel with a batch-service queue

(joint work with Jacqueline Heinerman)

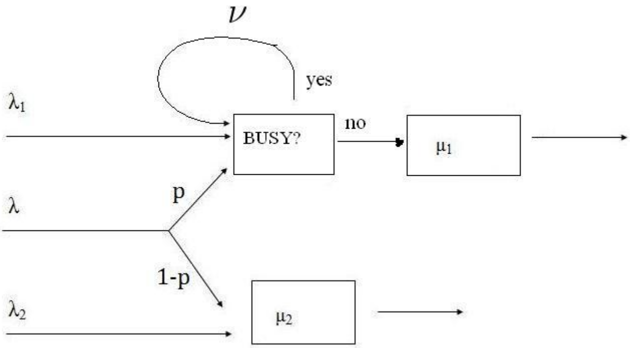


Figure: System with retrials and probabilistic routing

User equilibria for the retrial model

$$\begin{aligned} W_1(p) &= \text{steady-state transit time for a customer who joins queue 1} \\ W_2(p) &= \text{steady-state transit time for a customer who joins queue 2} \\ W(p) &= pW_1(p) + (1-p)W_2(p) = \\ &\quad \text{the average transit time for all general customers.} \end{aligned}$$

The formula for the $M/M/1{-}\mathrm{R}$ queue gives

$$W_1(p) = \frac{1}{\mu_1} + \frac{(p\lambda + \lambda_1)(\mu_1 + \nu)}{\mu_1 \nu (\mu_1 - p\lambda - \lambda_1)}.$$

As before a simple steady-state analysis gives

$$W_2(p) = \frac{1}{\mu_2} + \frac{N-1}{2(\lambda_2 + (1-p)\lambda)}.$$

Solve the quadratic equation $W_1(p) = W_2(p)$ for p and check whether the found equilibrium is stable.

probabilistic routing, varying μ_1

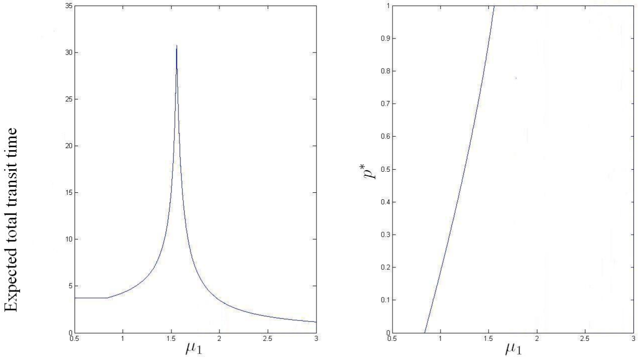


Figure: Left: Expected equilibrium transit times of the general customers for the parameters $\lambda = 1, \lambda_1 = 0.5, \lambda_2 = 0.1, \mu_2 = 1, \nu = 2, N = 7$.
Right: Optimal p^* as a function of μ_1

probabilistic routing, varying ν; $\mu_1 = 1.25$

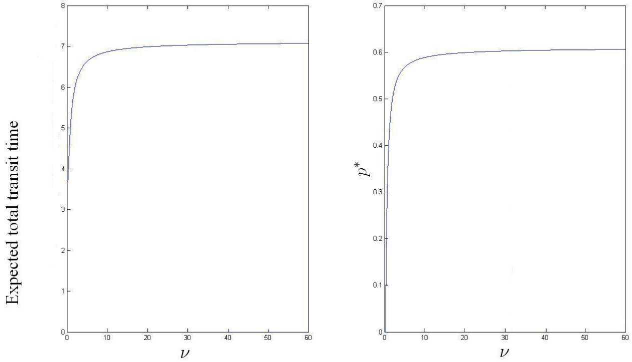


Figure: Left: Expected equilibrium transit times of the general customers for parameters $\lambda = 1, \lambda_1 = 0.5, \lambda_2 = 0.1, \mu_2 = 1, \mu_1 = 1.25, N = 7$.
Right: Optimal p^* as a function of ν

probabilistic routing, varying ν; $\mu_1 = 2$

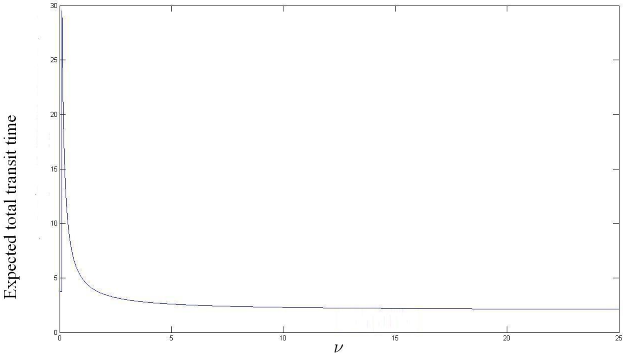


Figure: Expected equilibrium transit times of the general customers for parameters $\lambda = 1, \lambda_1 = 0.5, \lambda_2 = 0.1, \mu_2 = 1, \mu_1 = 2, N = 7$.

State-dependent routing

- General customers have full knowledge of the state of the system upon arrival (irrevocably),
- Based on their knowledge they choose the queue with the smaller expected transit time.

The state space is

$$\mathcal{S} = \{(i,k,j) | i = 0,1,2,\ldots;\ k = 0,1;\ j = 0,1,2,\ldots,N-1\}.$$

A policy or strategy for the general customers is a partition of $\mathcal{S}$ into two disjoint subsets $\mathcal{S}_1$ and $\mathcal{S}_2$ such that

$(i,k,j) \in \mathcal{S}_1 \Longleftrightarrow$ the customer who sees state (i,k,j) chooses queue 1.

Notation $\mathcal{D} := (\mathcal{S}_1, \mathcal{S}_2)$. Define for every state $(i,k,j) \in \mathcal{S}$

$y_{\mathcal{D}}(i,k,j) =$ the expected transit time of a tagged customer in the orbit when i other customers are in the orbit, the server is in state k, and j customers are in the batch.

For $(i+1,k,j) \in \mathcal{S}_1$ we get

$$\begin{aligned} y_{\mathcal{D}}(i,k,j) &= \frac{1}{\lambda_1+\lambda_2+\lambda+(i+1)\nu+k\mu} \\ &\times \Big[1+(\lambda_1+\lambda)y_{\mathcal{D}}(i+k,1,j)+(i+k)\nu y_{\mathcal{D}}(i-1+k,1,j) \\ &+ (1-k)\nu\frac{1}{\mu_1}+\lambda_2 y_{\mathcal{D}}(i,k,j+1)+k\mu y_{\mathcal{D}}(i,0,j)\Big], \end{aligned} \tag{14}$$

and if $(i+1,k,j) \in \mathcal{S}_2$ then we have

$$\begin{aligned} y_{\mathcal{D}}(i,k,j) &= \frac{1}{\lambda_1+\lambda_2+\lambda+(i+1)\nu+k\mu} \\ &\times \Big[1+\lambda_1 y_{\mathcal{D}}(i+k,1,j)+(i+k)\nu y_{\mathcal{D}}(i-1+k,1,j) \\ &+ (1-k)\nu\frac{1}{\mu_1}+(\lambda_2+\lambda)y_{\mathcal{D}}(i,k,j+1)+k\mu y_{\mathcal{D}}(i,0,j)\Big], \end{aligned} \tag{15}$$

$z_{\mathcal{D}}(i,k,j) =$ the expected transit time when the customer joins queue 2.

Of course,

$$z_{\mathcal{D}}(i,k,N-1) = \frac{1}{\mu_2} \quad \text{for} \quad i = 0,1,2,\ldots;\ k = 0.1.$$

Further, if $(i,k,j+1) \in \mathcal{S}_1$ then

$$\begin{aligned} z_{\mathcal{D}}(i,k,j) &= \frac{1}{\lambda_1+\lambda_2+\lambda+i\nu+k\mu_1}\left[1+(\lambda_1+\lambda)z_{\mathcal{D}}(i+k,1,j)\right. \quad (16) \\ &+ \left.\lambda_2 z_{\mathcal{D}}(i,k,j+1) + i\nu z_{\mathcal{D}}(i-1+k,1,j) + k\mu_1 z_{\mathcal{D}}(i,0,j)\right]. \end{aligned}$$

If on the other hand $(i,k,j+1) \in \mathcal{S}_2$ then

$$\begin{aligned} z_{\mathcal{D}}(i,k,j) &= \frac{1}{\lambda_1+\lambda_2+\lambda+i\nu+k\mu_1}\left[1+(\lambda_2+\lambda)z_{\mathcal{D}}(i,k,j+1)\right. \quad (17) \\ &+ \left.\lambda_1 z_{\mathcal{D}}(i+k,1,j) + i\nu z_{\mathcal{D}}(i-1+k,1,j) + k\mu_1 z_{\mathcal{D}}(i,0.j)\right]. \end{aligned}$$

The optimal selfish policy

The problem is to find the optimal selfish policy $\mathcal{D}^*$ for which

$$(i,0,j) \in \mathcal{S}_1^* \iff \frac{1}{\mu_1} < z_{\mathcal{D}^*}(i,0,j) \tag{18}$$

and

$$(i,1,j) \in \mathcal{S}_1^* \iff y_{\mathcal{D}^*}(i,1,j) < z_{\mathcal{D}^*}(i.1,j) \tag{19}$$

- Determine smallest numbers L_j such that

  $$y^*(L_j,1) + \frac{1}{\mu_1} := \frac{\lambda_1 + 2\mu_1 + (i+2)\nu}{\nu(2\mu_1 - \lambda_1)} + \frac{1}{\mu_1} \geq \frac{1}{\lambda_2}(N-j+1).$$

  For all $i \geq L_j$, put $(i,1,j) \in \mathcal{S}_2^*$ and for $i \geq L_j$ set

  $$\begin{aligned} y_{\mathcal{D}^*}(i,0,j) &:= \frac{\lambda_1 + 2\mu_1 + i\nu}{\nu(2\mu_1 - \lambda_1)} + \frac{1}{\mu_1} \\ y_{\mathcal{D}^*}(i,1,j) &:= \frac{\lambda_1 + 2\mu_1 + (i+2)\nu}{\nu(2\mu_1 - \lambda_1)} + \frac{1}{\mu_1}, \end{aligned}$$

The optimal selfish policy (cont'd)

- Solve the system of equations (14) and (15).
- Using the solution of (14) and (15), determine the policy $\mathcal{D}^*$ recursively from the equations (16) and (17).