Self-optimising state-dependent routing in parallel queues

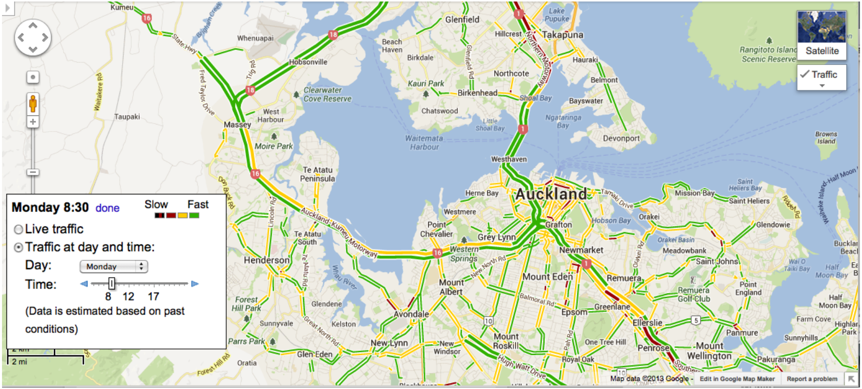
Ilze Ziedins

Joint work with:
Heti Afimeimounga, Lisa Chen, Mark Holmes, Wiremu Solomon,
Niffe Hermansson, Elena Yudovina

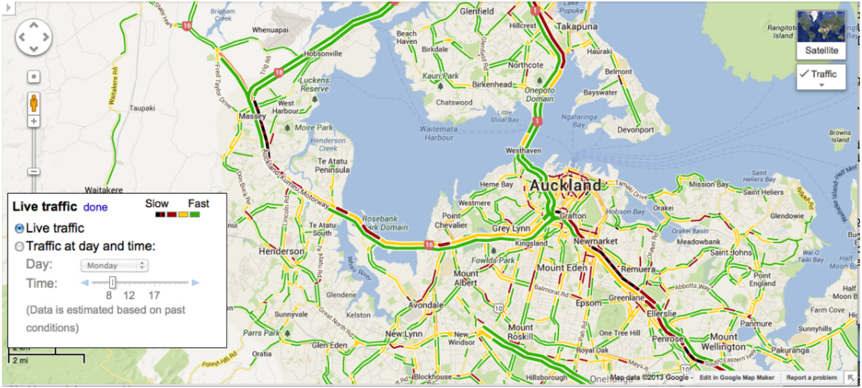
The University of Auckland

MARSDEN FUND
TE PŪTEA RANGAHAU A MARSDEN

THE UNIVERSITY OF AUCKLAND
FACULTY OF SCIENCE
Department of Statistics



Auckland, Monday, 8.30 a.m., predicted traffic
(downloaded 6 July 2013)



Auckland, Monday 10 June, 8.30 a.m., actual traffic

Which route/mode of transport to take?

Individual choice (selfish routing) vs. social optimum

User equilibrium vs. system optimum

Probabilistic routing vs. state-dependent routing.

User equilibrium

Wardrop or user equilibrium

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

Wardrop, J.G. (1952)

Each user has an infinitesimal effect on the system.

Parallel queues

Network with collection R of N routes from A to B.

Probabilistic routing – user optimal/equilibrium policies

p_r = probability of taking route r, with $p_r \geq 0$, $\sum_r p_r = 1$.
$\mathbf{p} = (p_1, p_2, \ldots, p_N)$
$W_r(\mathbf{p})$ = expected transit time via route $r \in R$.

At a user equilibrium, $\mathbf{p^{EQ}}$, there exists c such that

$$\begin{aligned} W_r(\mathbf{p^{EQ}}) &= c \qquad \text{if } p_r^{EQ} > 0 \\ &\geq c \qquad \text{if } p_r^{EQ} = 0. \end{aligned}$$

State dependent routing – user optimal/equilibrium policies

A decision policy $\mathcal{D}$ is a partition of state space, $\mathcal{S}$, into sets D_r, $r \in R$ such that if system is in state $\mathbf{n} \in D_r$ when a user arrives, then they take route r.

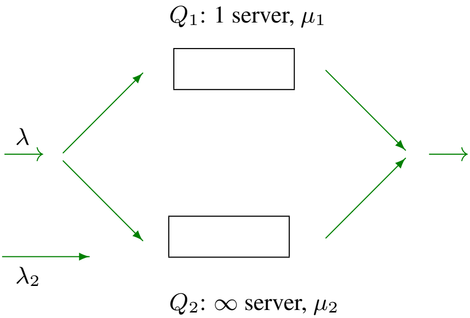
For a policy $D \in \mathcal{D}$ and $\mathbf{n} \in \mathcal{S}$, $z_r^D(\mathbf{n})$ = expected time to reach the destination for a general user, if system is in state $\mathbf{n}$ immediately prior to their arrival, and they choose to take route r.

A policy $D \in \mathcal{D}$ is a user optimal policy or user equilibrium if for each $\mathbf{n} \in \mathcal{S}$

$$\mathbf{n} \in D_r \quad \Longrightarrow \quad z_r^D(\mathbf{n}) \leq z_s^D(\mathbf{n}) \text{ for all } s \neq r, s \in R.$$

Downs-Thomson network

Downs-Thomson network



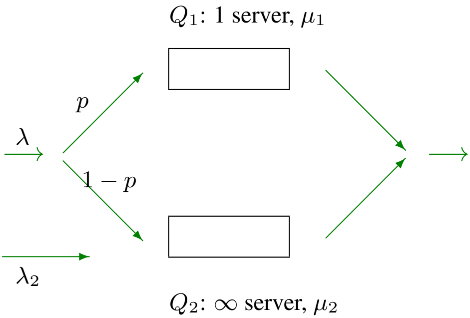
Two Poisson arrival streams	– dedicated users to queue 2 at rate λ_2,
	– general users at rate λ.
General users choose route	– either probabilistic or state-dependent routing.

Q_1 single server queue $(\cdot/M/1)$, exponential service times, mean $1/\mu_1$.
Q_2 batch service ∞ server queue, service times with mean $1/\mu_2$.
Downs(62), Thomson(77), Calvert(97), Afimeimounga,Solomon,Z(05,10)

- Single server queue – private transportation (e.g. cars).
  - delay increases as load increases
- Batch service queue – public transportation (e.g. shuttle bus).
  - delay decreases as load increases
  - frequency of service increases as load increases
- This version of model first proposed by Calvert (1997) as queueing network version of transportation model that gives rise to the Downs Thomson paradox.
- Paradox is that delays for all users can increase when capacity of private transportation (roading) is increased. First observed by Downs (1962) and Thomson (1977).
- Afimeimounga, Solomon, Z (2005, 2010)

Downs-Thomson network –

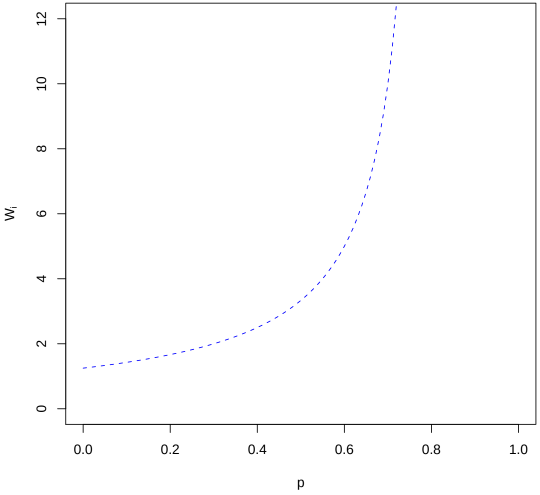
probabilistic routing



Q_1 single server queue $(\cdot/M/1)$. Expected delay $W_1 = \frac{1}{\mu_1 - \lambda p}$

Q_2 batch service ∞ server queue. Expected delay $W_2 = \frac{1}{\mu_2} + \frac{N-1}{2(\lambda_2 + \lambda(1-p))}$

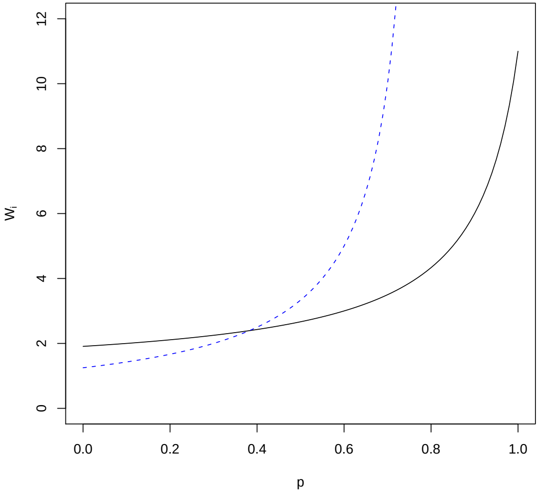
Both W_1 and W_2 are increasing in p.



$$\mu_1 = 0.8$$

$$\lambda = 1, \lambda_2 = .1, \mu_2 = 1, N = 3$$

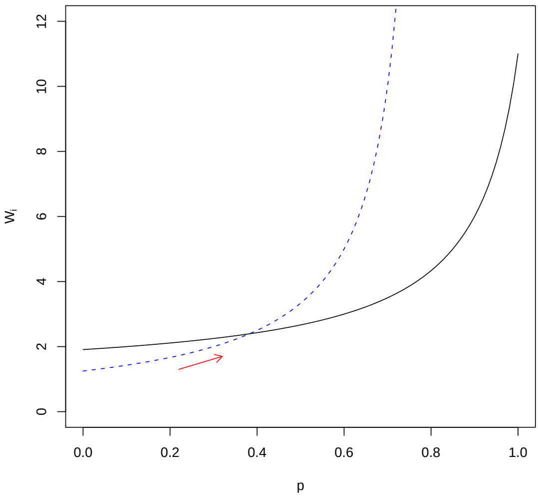
W_1, - - - - - - - -, W_2, ————



$$\mu_1 = 0.8$$

$$\lambda = 1,\ \lambda_2 = .1,\ \mu_2 = 1,\ N = 3$$

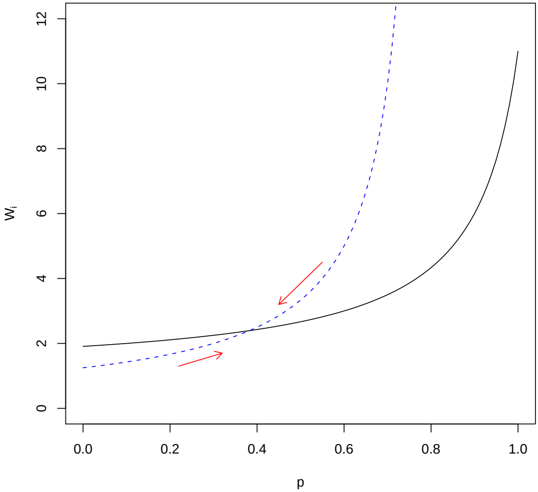
W_1, - - - - - - - -, W_2, ————



$$\mu_1 = 0.8$$

$$\lambda = 1,\, \lambda_2 = .1,\, \mu_2 = 1,\, N = 3$$

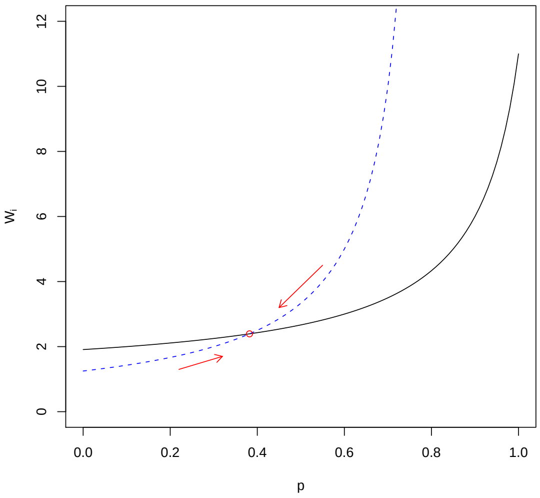
W_1, - - - - - - - -, W_2, ———



$$\mu_1 = 0.8$$

$$\lambda = 1,\ \lambda_2 = .1,\ \mu_2 = 1,\ N = 3$$

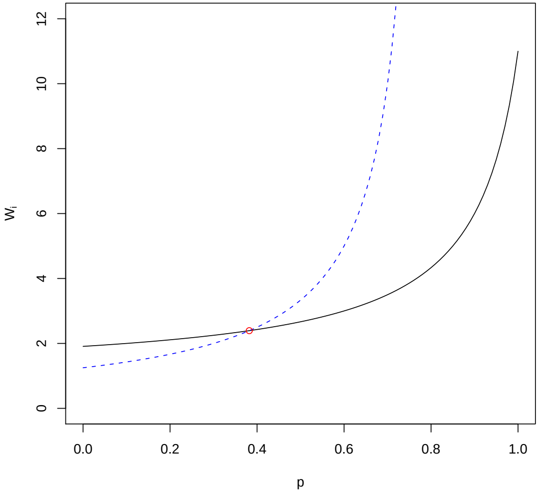
W_1, - - - - - - - - -, W_2, ———



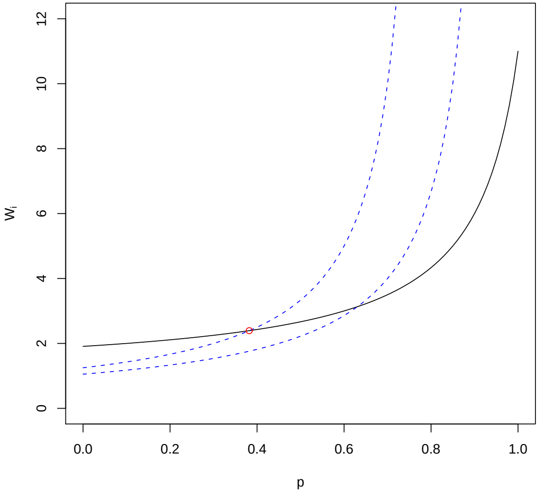
$$\mu_1 = 0.8$$

$$\lambda = 1,\ \lambda_2 = .1,\ \mu_2 = 1,\ N = 3$$

W_1, - - - - - - - - -, W_2, ———————



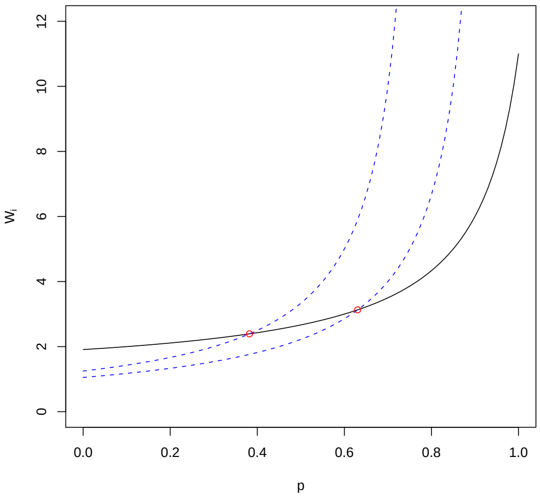
$$\mu_1 = 0.8$$
$$\lambda = 1,\ \lambda_2 = .1,\ \mu_2 = 1,\ N = 3$$
W_1, - - - - - - - -, W_2, ———



$$\mu_1 = 0.8, 0.95$$

$$\lambda = 1,\ \lambda_2 = .1,\ \mu_2 = 1,\ N = 3$$

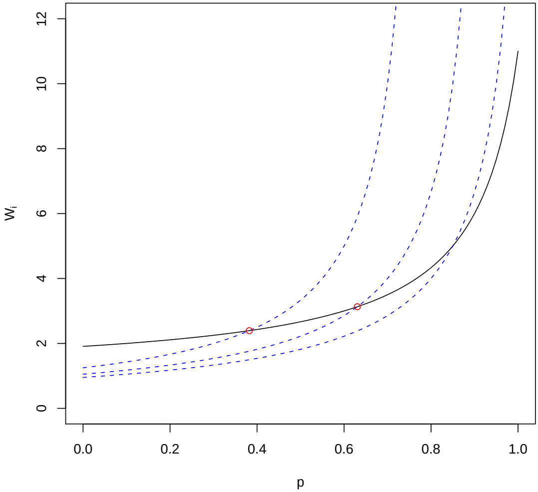
W_1, - - - - - - - -, W_2, ————



$$\mu_1 = 0.8, 0.95$$

$$\lambda = 1,\ \lambda_2 = .1,\ \mu_2 = 1,\ N = 3$$

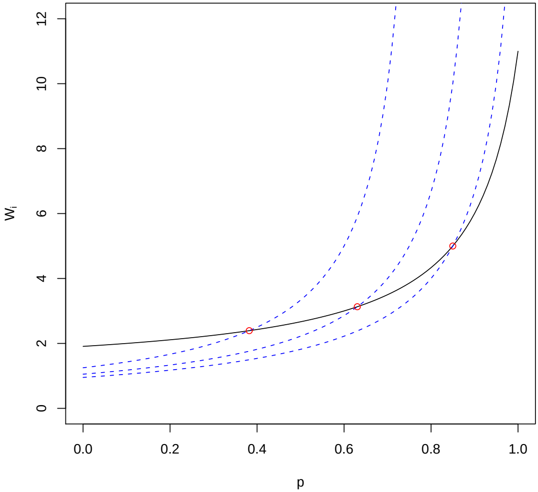
W_1, - - - - - - - -, W_2, ————



$$\mu_1 = 0.8, 0.95, 1.05$$

$$\lambda = 1,\ \lambda_2 = .1,\ \mu_2 = 1,\ N = 3$$

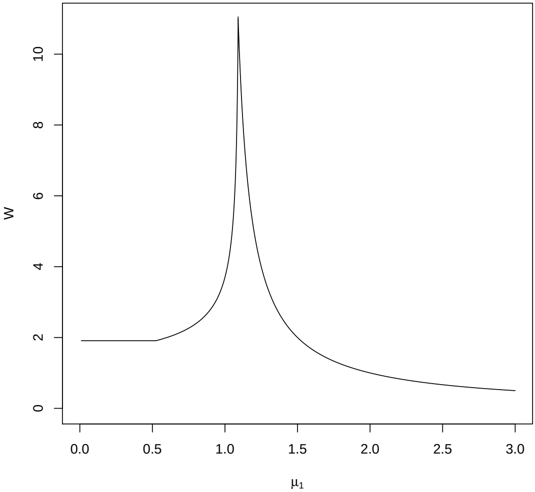
W_1, - - - - - - - -, W_2, ———



$$\mu_1 = 0.8, 0.95, 1.05$$

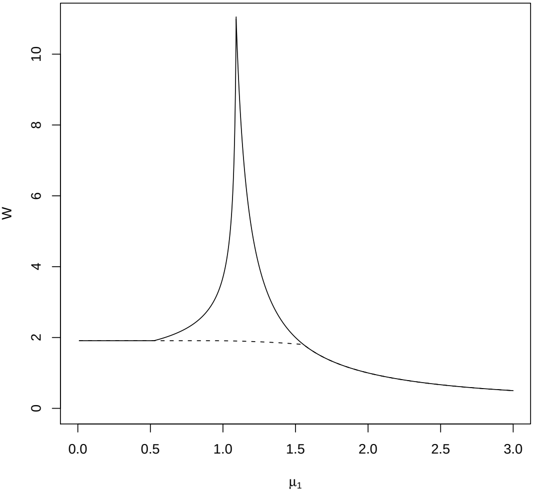
$$\lambda = 1,\ \lambda_2 = .1,\ \mu_2 = 1,\ N = 3$$

W_1, - - - - - - - -, W_2, ———————



$$\lambda = 1, \lambda_2 = .1, \mu_2 = 1, N = 3$$

$$W = p^{EQ} W_1 + (1 - p^{EQ}) W_2$$



$\lambda = 1, \lambda_2 = .1, \mu_2 = 1, N = 3$

$W = p^{EQ} W_1 + (1 - p^{EQ}) W_2$ ————————

$W = \min_p pW_1 + (1 - p)W_2$ - - - - - - - - - - - - - -

Consequences of individual choice

- Network performance may be poorer than expected

- Adding capacity may lead to worse performance

Downs-Thomson network –

state dependent routing

State dependent policies

$X_1(t)$ = number of customers in queue 1 (including customer in service)

$X_2(t)$ = number of customers waiting for service in queue 2 (not including those in service)

State space $\mathcal{S} = Z_+ \times \{0, 1, 2, \ldots, N-1\}$.

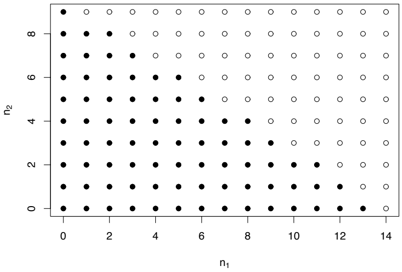
Process X_D operating under decision policy D has transition rates:-

$$\mathbf{n} \longrightarrow \begin{cases} \mathbf{n} - e_1 & \text{at rate } \mu_1 I_{\{n_1 > 0\}} \\ \mathbf{n} + e_1 & \text{at rate } \lambda I_{\{\mathbf{n} \in D_1\}} \\ (n_1, (n_2 + 1) \bmod N) & \text{at rate } \lambda_2 + \lambda I_{\{\mathbf{n} \in D_2\}} \end{cases}$$

where $I_A = 1$ if A occurs, and $I_A = 0$ otherwise.

A policy $D \in \mathcal{D}$ is a user optimal policy or user equilibrium if

$$\mathbf{n} \in D_1 \Longleftrightarrow z_1^D(\mathbf{n}) < z_2^D(\mathbf{n}) \qquad \text{for all } \mathbf{n} \in \mathcal{S}.$$



Points in D_1 – $\bullet$. Points in D_2 – $\circ$.
Unique user optimal policy for
$N = 10, \lambda = 1.5, \lambda_2 = 0.5, \mu_1 = 2, \mu_2 = 1.$

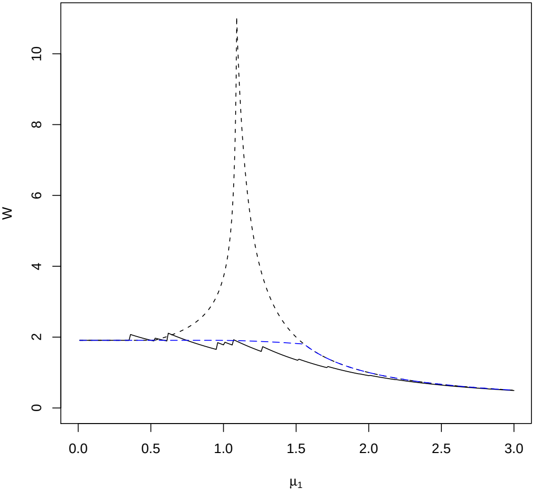
A policy $D \in \mathcal{D}$ is monotone if D satisfies

(A) $\mathbf{n} \in D_2 \Rightarrow \mathbf{n} + e_1 \in D_2$ for all $\mathbf{n} \in \mathcal{S}$ and

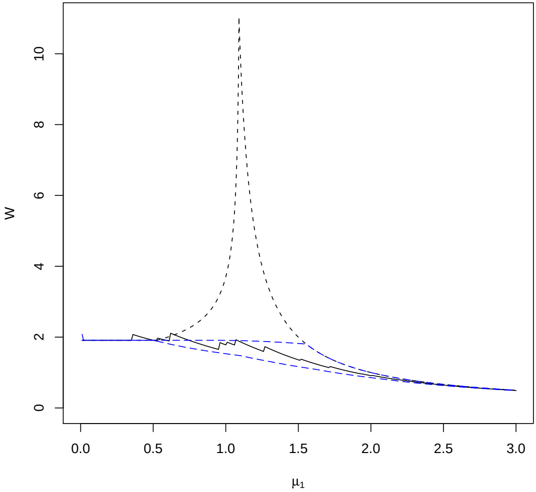
(B) $\mathbf{n} \in D_2 \Rightarrow \mathbf{n} + e_2 \in D_2$ for all $\mathbf{n} \in \mathcal{S}$

Properties

- A user optimal policy exists and is unique (no randomization needed).
- The user optimal policy is monotone.
- The user optimal policy is monotone in the parameters λ, λ_2, μ_1, μ_2 in the following sense. Let $X^{(1)}$ and $X^{(2)}$ be two processes, with common batch size N and user optimal policies $D^*(1)$, $D^*(2)$ respectively. If $\lambda^{(1)} \geq \lambda^{(2)}$, $\mu_1^{(1)} \leq \mu_1^{(2)}$, $\lambda_2^{(1)} \geq \lambda_2^{(2)}$ and $\mu_2^{(1)} \geq \mu_2^{(2)}$, then $D_1^*(1) \subset D_1^*(2)$.
- Proof uses a coupling argument.
- As part of the proof show monotonicity of $z_2^D(\mathbf{n})$ in λ, λ_2, μ_1, μ_2; and in the decision policy.
- Afimeimounga, Solomon, Z (2010), Calvert (1997), Ho (2003), Altman and Shimkin (1998), Ben-Shahar, Orda and Shimkin (2000), Brooms (2005), Hassin and Haviv (2003).



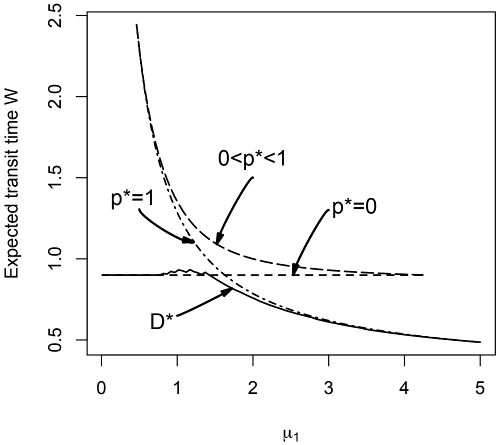
Expected transit times under user optimal policy for
state-dependent routing (————), and probabilistic routing $(-\,-\,-\,-\,-\,-\,-)$
$\lambda = 1$, $\lambda_2 = 0.1$, $\mu_2 = 1$, $N = 3$ for $0 \leq \mu_1 \leq 3$.



Expected transit times under user optimal policy for
state-dependent routing (———), and probabilistic routing (− − − − − − − −)
$\lambda = 1$, $\lambda_2 = 0.1$, $\mu_2 = 1$, $N = 3$ for $0 \leq \mu_1 \leq 3$.

Variations

Two batch-service queues



Expected transit times under user optimal policy for
state-dependent routing (————), and probabilistic routing ($- - - - - - - -$)
$\lambda = 4$, $\lambda_1 = 3$, $\lambda_2 = 1$, $\mu_2 = 2$, $N_1 = N_2 = 5$ for $0 \leq \mu_1 \leq 6$.
Chen, Holmes, Z(2011)

Other variations

Processor-sharing queues

- Iterative procedure may converge to periodic orbit
- User equilibrium doesn't always possess monotonicity properties
- Randomization needed

Braess's paradox

- State dependent routing mitigates worst effects here as well
  Cohen, Kelly (1990), Calvert, Solomon, Z (1997)

Some final comments

- Do user equilibria exist more generally under state dependent routing, and if yes, when are they unique?
- How to overcome poor performance at user equilibria?
- Does more information lead to shorter delays in general?
  Effects of partial information
- Add monetary and other costs to the problem, as well as delays
- Convergence issues – effect of delayed information.
- Differing information and/or policies for different customer classes
  Argument for investment in public transport, using public transport
  ....

- Afimeimounga, H., Solomon, W. and Ziedins, I. (2005) The Downs-Thomson paradox: Existence, uniqueness and stability of user equilibria. *Queueing Systems* **49**, 321-334.
- Afimeimounga, H., Solomon, W. and Ziedins, I. (2010) User equilibria for a parallel queueing system with state dependent routing. *Queueing Systems* **66**, 169-193.
- Altman, E. and Wynter, L.(2004) Equilibrium games and pricing in transportation and telecommunication networks. *Networks and Spatial Economics* **4**, 7–21.
- Altman, E. and Shimkin, N. (1998) Individual equilibrium and learning in processor sharing systems. *Operations Research* **46**, 776–784.
- Bell, C.E. and Stidham, S., Jr. (1983) Individual versus social optimization in the allocation of customers to alternative servers. *Management Science* **29**, 831–839.
- Ben-Shahar, I., Orda, A. and Shimkin, N. (2000) Dynamic service sharing with heterogeneous preferences. *Queueing Systems* **35**, 1572–9443.
- Brooms, A.C. (2005) On the Nash equilibria for the FCFS queueing system with load-increasing service rate. *Adv. Appl. Prob.* **37**, 461–481.
- Calvert, B. (1997) The Downs-Thomson effect in a Markov process. *Probability in the Engineering and Information Sciences* **11**, 327–340.

- Calvert, B., Solomon, W. and Ziedins, I. (1997) Braess's paradox in a queueing network with state-dependent routing. *Journal of Applied Probability* **34**, 134–154.
- Cohen, J.E. and Kelly, F.P. (1990) A paradox of congestion in a queueing network. *Journal of Applied Probability* **27**, 730–734.
- Downs, A. (1962) The law of peak-hour expressway congestion. *Traffic Quarterly* **16**, 393-409.
- Hassin, R. and Haviv, M. (2003) *To Queue or not to Queue: Equilibrium Behavior in Queueing Systems*. Kluwer.
- Ho, B. (2003) *Existence, Uniqueness and Monotonicity of the State-Dependent User Optimal Policy for a Simple Markov Transport Network.* MSc Thesis. The University of Auckland.
- Naor, P. (1969) The regulation of queue size by levying tolls. *Econometrica* **37**, 15 – 24.
- Parlaktürk, A.K. and Kumar, S. (2004) Self-interested routing in queueing networks. *Management Science* **50**, 949–966.
- Roughgarden, T. and Tardos, E. (2002) How bad is selfish routing? *Journal of the ACM* **49**, 236–259.
- Wardrop, J.G. (1952) Some theoretical aspects of road traffic research. *Proceedings, Institution of Civil Engineers, II* **1**, 325–378.