An investigation of proportionally fair ramp metering

Richard Gibbens

Computer Laboratory
University of Cambridge

EURANDOM workshop in honour of Frank Kelly
28–29 April 2011

(Joint work with Frank Kelly)

Congestion and reduced capacity

- ▶ Congestion occurs when demand exceeds available resources and can significantly reduce capacity.

- ▶ Reduced capacity results in additional delays, increased pollution, ...

- ▶ Congestion results in low but highly volatile speeds and more uncertain journey times: flow breakdown or stop-and-go behaviour.
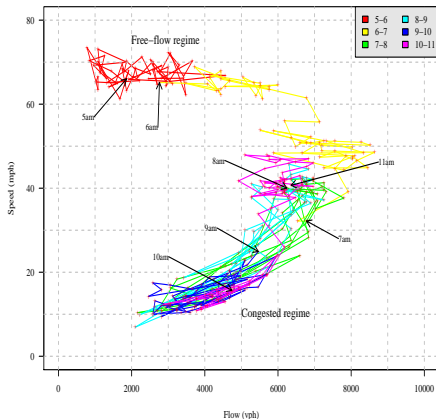
Source: G. & Saatci (2008)

Flow breakdown on the M25

Speeds (mph) on M25 (clockwise) Mon 6 Jan 2003
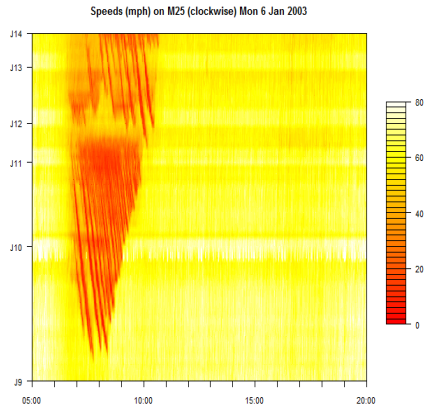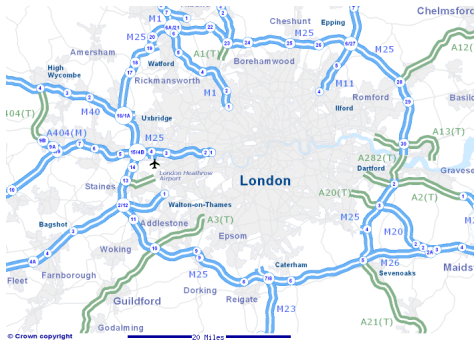
Source: G. & Saatci (2008)

Performance metrics: *VMT*, *VHT*, *VHD*

Measure flow and average speed for locations (cells) of given length and for each time interval.

- ► Vehicle miles travelled

$$VMT = flow \times length$$

- ► Vehicle hours travelled
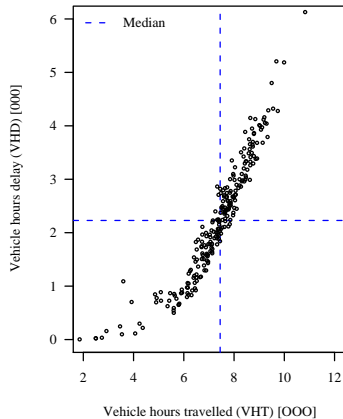
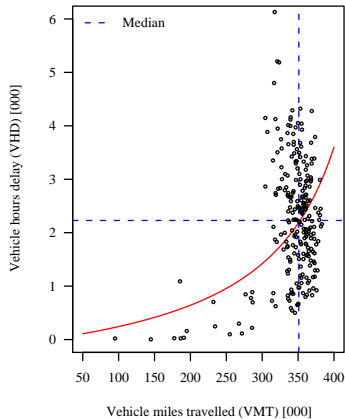$$VHT = VMT/speed$$

- ► Vehicle hours delay

$$VHD = (VHT - VMT/speed_{\mathrm{ref}})^+$$

Aggregate these metrics over locations and times.
For M25, take $speed_{\mathrm{ref}} = 67$ mph (PSA1 target)
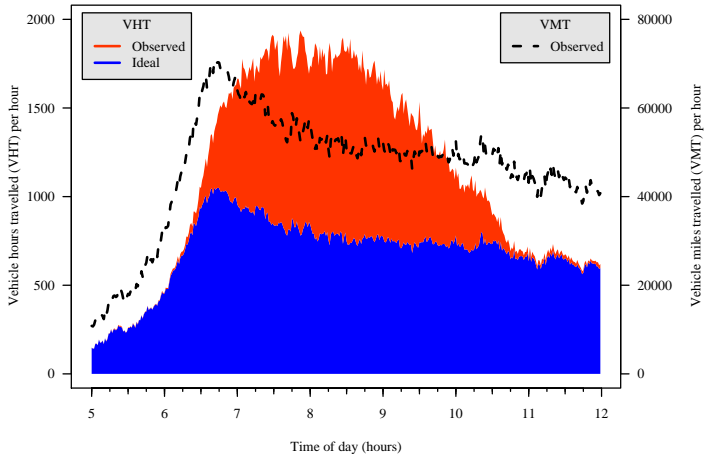
Daily performance metrics

M25 on weekdays in 2003 for 10 miles clockwise within J9 – J14

Daily performance profile

Monday, 6 Jan 2003

Ramp metering

- Ramp metering intends to control the entry of new flow so as to maintain steady flow and avoid the flow breakdown associated with congestion.

- The rate of flow entry is set by the choice of ramp metering strategy.

- A key issue for the design of ramp metering strategies is the trade-off between efficiency and fair use of resources.

- This trade-off has been much studied in the context of communication networks.

Source: DfT

Ramp metering: a form of distributed access control

- ▶ Access control is a common problem in networks, including communication networks as well as road networks.
- ▶ View ramp metering systems as part of a larger network: drivers generate demand and select their routes in ways that are responsive to delays incurred or expected, which depend on the controls applied in the road network.
- ▶ As mobile devices and Internet applications improve we might expect drivers' responses to be more immediate.

Ramp metering: signals and incentives

- ▶ We seek to understand the interactions between the ramp metering system and the larger network and investigate the signals such as delay provided to the larger network.

- ▶ In the communication network context fairness of the control scheme has emerged as an effective means by which the appropriate information and incentives are provided to the larger network by flow control and routing strategies.

- ▶ Kelly & Williams (2010) introduced the proportionally fair ramp metering strategy motivated by transfering some of these ideas from communication networks to road networks and we explore this further here.

A linear network
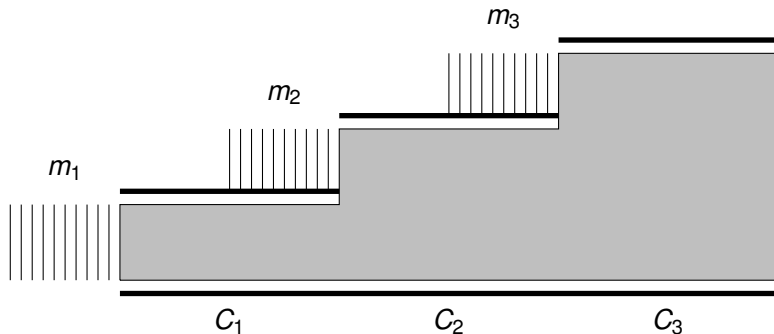
- ▶ Traffic entering at upstream on-ramps roads may all pass through the same downstream bottleneck, and if more traffic is admitted at one junction it will reduce the amount of traffic that can be admitted at later junctions.

Queue size processes

- We suppose that the queue sizes, $m_j(t)$, evolve according to the following dynamics which take account of vehicle arrivals and on-ramp metered rates at the entry points

$$m_j(t + \delta t) = m_j(t) + e_j(t) - L_j(t)\delta t \,.$$

- Here, $e_j(t)$ is the (random) number of arrivals in a short interval of time $[t, t + \delta t)$ and $L_j(t)$ is the realized metered rate of flow.

- For example, $e_j(t)$ may be given by Poisson random variables with mean parameters $\rho_j \delta t$ corresponding to independent Poisson processes of arrivals with rates ρ_j.

Greedy strategy

- ▶ Realized metered rates, $L_j(t)$, are updated as follows

  $L_1(t) \leftarrow \text{ifelse}(m_1(t) > 0, C_1, 0)$
  $L_2(t) \leftarrow \text{ifelse}(m_2(t) > 0, C_2 - L_1(t - \tau_1 + \tau_2), 0)$
  $L_3(t) \leftarrow \text{ifelse}(m_3(t) > 0, C_3 - L_1(t - \tau_1 + \tau_3) - L_2(t - \tau_2 + \tau_3), 0)$

- ▶ Optimality property: this strategy minimizes, for all times T, the sum of the line sizes at time T, $\sum_{j=1}^{3} m_j(T)$.

- ▶ This is a compelling property if arrival patterns of traffic are exogenously determined.

- ▶ However, the strategy will concentrate delay upon flows entering at the more downstream entry points.

- ▶ This seems intuitively unfair since such flows use fewer system resources and may well have perverse and suboptimal consequences if driver behaviour is influenced by delays.

Fairness

- Suppose that given queue sizes $m = (m_r, r \in R)$, a rate $\lambda_r(m)$ is allocated to route r, for each $r \in R$. The allocation $\lambda(m) = (\lambda_r(m), r \in R)$ is proportionally fair if, for each $m \in \mathbb{R}_+^R$, $\lambda(m)$ solves

$$\text{maximize} \qquad \sum_{r \in R: m_r > 0} m_r \log \lambda_r \qquad (1)$$

$$\text{subject to} \qquad \sum_{r \in R} A_{jr} \lambda_r \leqslant C_j \qquad j \in J, \qquad (2)$$

$$\text{over} \qquad \lambda_r \geqslant 0 \qquad r \in R. \qquad (3)$$

for all $m \in \mathbb{R}_+^R$.

- Note that the constraint (2) captures the limited capacity of resource j where A_{jr} is the resource-route incidence matrix.

- The problem (1–3) is a straightforward convex optimization problem, and a vector $\lambda \in \mathbb{R}^R$ is a solution if and only if there exists a vector $p \in \mathbb{R}^J$ satisfying

$$p \geqslant 0; \quad \lambda \geqslant 0, \quad A\lambda \leqslant C \qquad (4)$$

$$p \cdot (C - A\lambda) = 0 \qquad (5)$$

$$m_r = \lambda_r \sum_{j \in J} A_{jr} p_j, \quad r \in R. \qquad (6)$$

- The variables $p = (p_j, j \in J)$ are Lagrange multipliers (or shadow prices) for the capacity constraints (2).

Fairness (3)

- Given queue sizes, the ratio $m_j/\lambda_j(m)$ is the time it would take to empty the workload in queue j at the current metered rate for queue j. Thus, for the linear network

$$d_j = \sum_{i=j}^{|J|} p_i, \qquad j \in J$$

  give estimates of queueing delay in each queue.

- These estimates do not take into account any change in the queue sizes over the time taken for traffic to move through the queue, but are a reasonable prediction of queueing delay at the time of arrival to the queue.

Proportionally fair strategy

- First, at each time epoch, solve the optimization problem to construct metered rates $\lambda_1, \lambda_2, \lambda_3$ given queue sizes m_1, m_2, m_3.
- (The appendix to the paper gives calculations for constructing this solution.)
- Realized metered rates, $L_j(t)$, are updated as follows

$$L_1(t) \leftarrow \min\{C_1, \lambda_1\}$$
$$L_2(t) \leftarrow \min\{C_2 - L_1(t - \tau_1 + \tau_2), \lambda_2\}$$
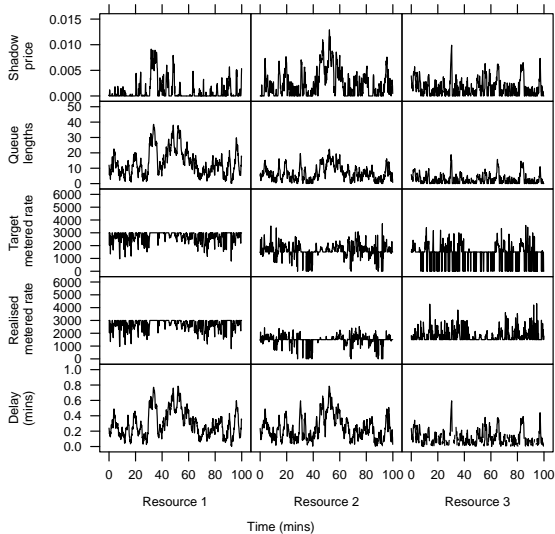$$L_3(t) \leftarrow C_3 - L_1(t - \tau_1 + \tau_3) - L_2(t - \tau_2 + \tau_3).$$

Simulation results

- Capacities: $C_1 = 3000$, $C_2 = 4500$ and $C_3 = 6000$ vehicles per hour.
- Travel times: $\tau_1 = 9$, $\tau_2 = 6$ and $\tau_3 = 3$ minutes.
- Arrival rates: $\rho_1 = 0.45 C_3 = 2700$, $\rho_2 = 0.25 C_3 = 1500$, $\rho_3 = 0.25 C_3 = 1500$ vehicles per hour.

|	Greedy			Proportionally fair	
	Mean	Standard error		Mean	Standard error
m_1	5.6	0.1	m_1	18.0	0.4
m_2	5.3	0.1	m_2	10.0	0.3
m_3	5.7	0.1	m_3	4.0	0.1

Simulation results (2)

Responsive traffic

▶ Suppose that the traffic arriving at entry point j is a time-varying Poisson process, with rate

$$\rho_j(t) = \frac{\kappa_j}{(d_j(t) + \tau_j)^{\eta}} \tag{7}$$

where $d_j(t)$ is again given by $d_j(t) = \sum_{i=j}^{|J|} p_i(t)$. Thus the arrival rate is related inversely to the estimated journey time, that is the sum of the estimated delay in the queue plus the free-flow travel time along the motorway.

▶ The isoelastic demand function (7) is such that the elasticity of demand with respect to estimated journey time is η.

▶ For example, if $\eta = 0.3$ a 10% increase in journey time will reduce the arrival rate of traffic by 3%.

▶ Simulations used $\kappa_1 = 1550$, $\kappa_2 = 770$ and $\kappa_3 = 640$.

Responsive traffic: simulation results
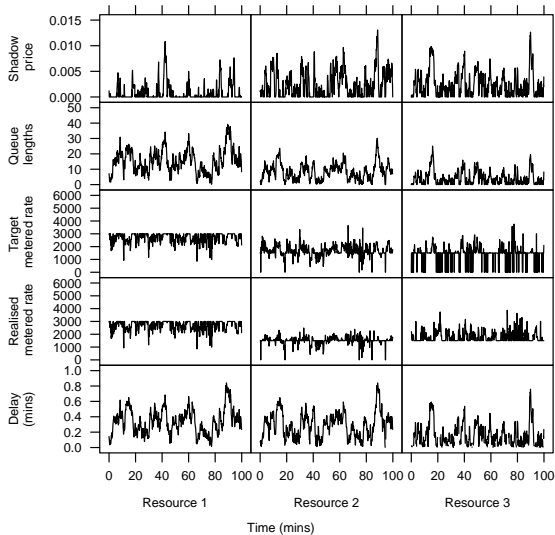
Route choices

- Consider a situation where vehicles have access to several parallel roads with a common destination.
- Assume that traffic arriving with access to more than one road distributes itself in an attempt to minimize its queueing delay.
- An alternative scenario is a priority access scheme, for example high occupancy vehicles may have a larger set of routes to choose from.

Route choice example network
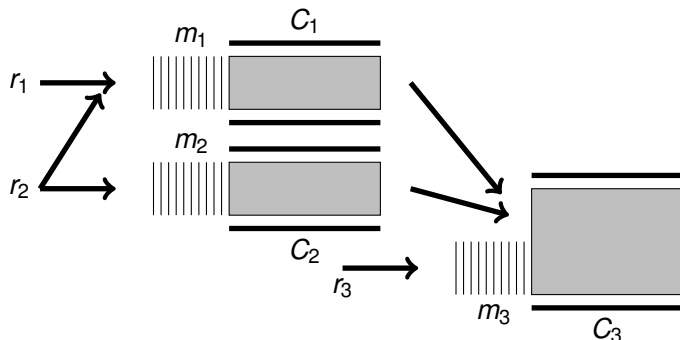
- Capacities: $C_1 = 3000$, $C_2 = 1500$, $C_3 = 6000$ vehicles per hour.
- Travel times: $\tau_1 = \tau_2 = 6$, $\tau_3 = 3$ minutes.
- Arrival rates: $\rho_1 = 0.45 C_3 = 2700$, $\rho_2 = 1500$, $\rho_3 = 1500$ vehicles per hour.

Route choices: simulation results
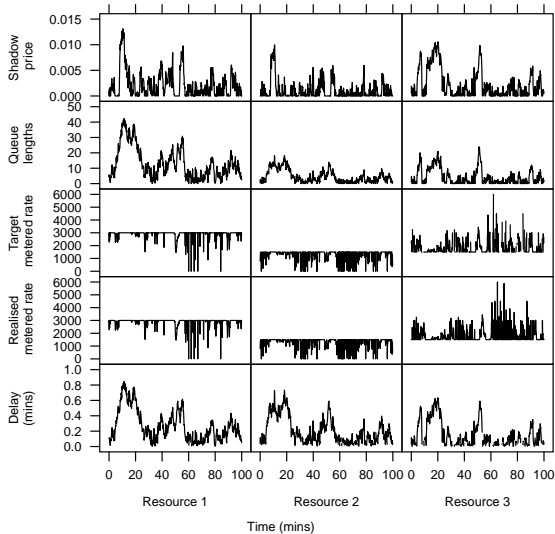
Conclusions

- ► We have explored some of the network aspects to ramp metering, especially whether delay at the entry points to a controlled motorway can provide incentives to drivers that are aligned with efficient use of the scarce resource.

- ► Specifically we looked at properties of the proportionally fair ramp metering strategy for two simple network topologies.

- ► The proportionally fair ramp metering strategy is inspired by rate control algorithms developed for the Internet, and attempts to set delays in proportion to shadow prices for the scarce resources.

Acknowledgements

- Computer Laboratory, University of Cambridge TIME (Transport Information Management Environment) EPSRC-funded project (EP/C547632/1).
- Isaac Newton Institute for Mathematical Sciences, Stochastic Processes in Communication Sciences programme, Jan–Jul 2010.
- Alje van den Bosch (intern) for computer animation software.

References

G and Saatci, Y. (2008).
Data, modelling and inference in road traffic networks.
Phil. Trans. R. Soc. **A366**, 1907–1919.

Kelly, F.P. and Williams, R.J. (2010).
Heavy traffic on a controlled motorway.
In *Probability and Mathematical Genetics: Papers in Honour of Sir John Kingman*, Eds Bingham, N.H. and Goldie, C.M., Cambridge University Press.

G and Kelly, F.P.
An investigation of proportionally fair ramp metering
Submitted.