Journal of Industrial and Systems Engineering
Vol. 4, No. 1, pp 1-17
Spring 2010

Congestion Pricing: A Parking Queue Model

Richard C. Larson[1*], Katsunobu Sasanuma[2]

[1]Engineering Systems Division, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 USA
rclarson@mit.edu

[2]Heinz College, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213 USA

ABSTRACT

Drivers in urban neighborhoods who patrol streets, seeking inexpensive on-street parking create a significant fraction of measured traffic congestion. The pool of drivers patrolling at any time can be modeled as a queue, where 'queue service' is the act of parking in a recently vacated parking space and queue discipline is SIRO – Service In Random Order. We develop a queueing model of such driver behavior, allowing impatient drivers to abandon the queue and to settle for more expensive off-street parking. We then relate the model to the economic theory of congestion pricing, arguing that price differentials between on-street and off-street parking should be reduced in order to reduce traffic congestion. Reducing the number of "patrolling drivers" often can reduce urban road congestion significantly, in some cases as effectively as technologically expensive road pricing schemes that cordon off the center city.

Keywords: Queuing, Traffic, Parking, Markovian model.

1. INTRODUCTION

We have all done it – driving around in a neighborhood seeking inexpensive on-street parking in order to avoid pricey parking lots and garages. And each of us has our own tolerance for delay vs. cost, namely the amount of time we are willing to drive seeking on-street parking vs. the amount of money we would have to pay for off-street parking. Eventually, since "time equals money," if we are unsuccessful in our patrolling for on-street parking, we reluctantly give up the effort and settle for the more costly off-street alternative.

The scenario above, which most of us have experienced, is a major cause of traffic congestion in many neighborhoods. Consider a typical recent study of street traffic congestion in urban America. This particular report is from the Park Slope section of Brooklyn, New York, a thriving commercial and residential zone (*Transportation Alternatives*, February (2007)). The purpose of the study was "…to ascertain the extent of the neighborhood's ever-worsening traffic and parking problems and to propose solutions to both." Based on data collected early in 2007, "…the study reveals an overwhelming amount of traffic is simply circling the block 'cruising' or patrolling for parking, while the curbside itself is nearly 100% filled with parked vehicles." The researchers found that

[*] Corresponding Author

ISSN: 1735-8272,

45% of total traffic and 64% of local traffic is cruising for a parking space. And the average curb occupancy rate is 94%, with "…nearly 100% occupancy at metered spaces during peak periods." This cruising behavior, creating significant traffic congestion, is the focus of our paper. The paper is a follow-on companion to a related piece by the same authors, *Urban Vehicle Congestion Pricing: A Review* (Larson and Sasanuma (2010)). In that paper we review the motivations and history of "congestion pricing'' in urban centers, where congestion pricing most often has implied charging a toll for each vehicle entering a core center area of the city during business hours. A major point of both papers, when read together, is that congestion pricing should encompass a total systems approach, including parking pricing (PP) as well as road use pricing (RP). Donald Shoup of UCLA has conducted much of the previous research in this area. In particular see his book, *The High Cost of Free Parking* (2005). Numerous other references, many quite recent, are contained in our companion paper (Larson and Sasanuma, (2010)).

2. QUEUING MODEL FOR PARKING PRICING

In the following, we develop a model that depicts patrolling drivers seeking on-street metered or free parking. As mentioned above, the model is motivated by recent data from Park Slope, Brooklyn and by extensive earlier analyses by Donald C. Shoup.

We assume that all parking spaces are occupied almost all of the time that would-be parkers are seeking parking spaces. Drivers seeking parking spaces are assumed to be driving around through the streets seeking the first available spot. As soon as one opens up, meaning a parked car is driven away, the next patrolling car virtually immediately occupies that spot. The platoon of patrolling cars is a moving queue serviced in random order. Not all would-be parkers are served in this queue, as the arrival rate of would-be parkers exceeds the departure rate of parked cars. So, we allow drivers in the patrolling queue to become discouraged, leave the queue and presumably settle for more expensive off-street parking (for instance, in a parking garage or in a parking lot).

For modeling purposes we assume an infinitely large homogeneous city with S parking spaces per square mile. We assume that the statistics of parking space availability and desirability are uniform over the city. We assume that the time any given parker occupies a parking space is a random variable W with probability density function $f_W(x)$ and mean $E[W]=1/\mu$. Prospective or would-be parkers appear in a Poisson manner at rate λA/hour, where A is defined to be the size of the area being considered (in sq. mi.). Prospective parkers will patrol looking for the first available parking space. Any unsuccessful would-be parker can become discouraged. We model this process by assuming that any would-be parker will leave the queue of patrolling would-be parkers at an individual Poisson rate of γ/hr.

There are two "large numbers" features in this system that allow us to model the queue as a Markovian system. First, regardless of the details of the probability density function (pdf) $f_W(x)$, the aggregate process of parked cars leaving parking spaces is accurately modeled as a Poisson process with rate $AS\mu$/hr. This is because the departure process from any given parking space is seen as a renewal process with inter-renewal pdf $f_W(x)$. As is well known, the merger or pooling of a large number of (sufficiently well-behaved) renewal processes converges to a Poisson process (Cox and Smith (1954)). We assume that the number of parking spaces we are considering is sufficiently large so that this approximation is very accurate. Second, the time until reneging of any would-be parker could be any well-behaved random variable having mean $1/\gamma$, not necessarily a negative exponential random variable. But, if the moving queue of patrolling would-be parkers is sufficiently large, we again have the pooling of many renewal processes --- each having the same probability density function of time until "renewal" and each starting at a random time. Such pooling will result in the aggregate process of *N* would-be parkers leaving the queue becoming a Poisson process with rate $N\gamma$, where *N* is typically large enough so that the Poisson assumption is valid.

We require one additional assumption in order to model this process efficiently. We assume that when there are zero cars patrolling in the modeled area, no parked cars leave their spaces. We know that this assumption is incorrect, but we are focusing on large queues of patrolling cars in which case the likelihood of zero patrolling cars is very small. If this assumption in an application setting is not valid, one can eliminate it by creating a larger Markovian model that includes the possibility of several or even many empty parking spaces.

In our work we will focus on a square area of the city having unit area (e.g., one square mile or one square kilometer). We will assume that this region is large enough for our saturation congestion theory to be valid. One might argue that in any actual city no would-be parker feels constrained to patrol within any arbitrary boundaries. This is true. But for every would-be parker who starts within our modeled square and then ventures out of it looking for an available parking space, there is statistically an another equivalent would-be parker who started in some near-by zone who ventures into our zone. Statistically, for everyone who leaves, there is someone who enters. We can take care of this by placing "reflecting barriers" around our zone, so that when anyone in the real system leaves, we simply reflect him or her back into the zone, creating a statistical equivalence to the real non-cordoned system.

2.1. The Markov Birth and Death Model

Assuming one square mile of operation, we now can draw the state-rate-transition diagram for this queue, as shown in Figure 1.

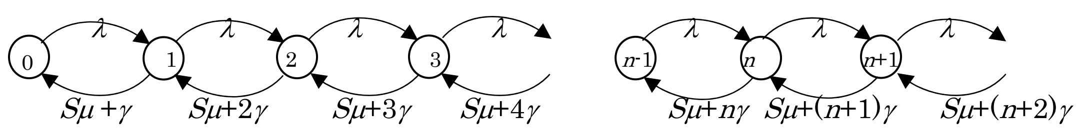

Figure 1 State-Rate-Transition Diagram for Queueing System

By the usual process of "telescoping" balance of flow equations, we can express each steady state probability P_n in terms of P_0 and a product of upward transition rates (λ's) divided by the product of downward transition rates between state *n* and state 0. The result is

$$P_n = \frac{\lambda^n}{\prod_{i=1}^{n}(S\mu + i\gamma)} P_0 \tag{1}$$

Now, invoking the requirement that the steady state probabilities sum to one, we obtain

$$(1 + \sum_{n=1}^{\infty} \frac{\lambda^n}{\prod_{i=1}^{n}(S\mu + i\gamma)}) P_0 = 1$$
,

or,

$$P_0 = 1/(1+\sum_{n=1}^{\infty}\frac{\lambda^n}{\prod_{i=1}^{n}(S\mu+i\gamma)}).$$

Hence,

$$P_n = \frac{\dfrac{\lambda^n}{\prod_{i=1}^{n}(S\mu+i\gamma)}}{(1+\sum_{m=1}^{\infty}\dfrac{\lambda^m}{\prod_{i=1}^{m}(S\mu+i\gamma)})}, \quad n=1,2,3,\ldots \tag{2}$$

For steady state to exist we require $P_0>0$, which always occurs. But we want P_0 to be very small for our approximations to be valid.

From the solutions obtained above, we can find all of the quantities of Little's Law, L, L_q, W and W_q. The basic Little's Law relationship is, of course, $L=\lambda W$. Here since "the system" is the queue only and service implies finding an empty parking space, we have the equivalences, $L = L_q$ and $W = W_q$. Assuming steady state operation, L is the time-average number of cars seeking parking spaces, or equivalently, the mean size of the patrolling queue of would-be parkers. W is the mean time that a patrolling car remains on patrol, until leaving either by finding a parking space or by frustration and reneging from the queue.

There are other performance measures of interest. The mean number of parking spaces becoming available per hour is $(1-P_0)S\mu \approx S\mu$ since $P_0 << 1$. The mean number of renegers per hour is $\lambda-(1-P_0)S\mu \approx \lambda - S\mu$, assuming $\lambda > S\mu$ (which is required for our approximations to be valid). For a random patrolling would-be parker, the probability of successfully getting a parking space is $(1-P_0)S\mu/\lambda \approx S\mu/\lambda$. This agrees with intuition. If say 100 parking spaces become available per hour and 250 would-be parkers arrive each hour, then 40% will succeed in finding a parking space and 60% will leave in frustration.

In the following we will assume that $0 < P_0 \approx 0$. This means that the queue of patrolling cars is, for all practical purposes, never empty. Under these conditions, we argue that the mean number of patrolling cars is

$$L = L_q = \frac{\lambda - S\mu}{\gamma} \tag{3}$$

This is a fundamental result for our saturated on-street parking system. When the queueing system is in saturation due to the shortage of the supply of on-street parking, the reneging rate plays a key role in determining the queue length. Kaplan (1987, 1988) introduced the same formula for the public housing applicant queue when the public housing system is saturated. He derived the formula using the law of conservation of flow (of housing applicants.) We argue the validity of Eq. (3) by changing the queue discipline from SIRO (Service In Random Order) to LCFS (Last Come, First Served). It is well known that L and L_q are invariant under the set of queue disciplines whose preferential orderings do not include customer-specific service times. The LCFS discipline is one such discipline. By LCFS here we mean the following: The next available parking space would be given instantaneously to that patrolling car that has been patrolling for the least amount of time. Usually this car would be the last to have arrived in queue. But it might be the case that the most recent car has already left the queue by reneging, in which case the next "youngest" patrolling car would be selected. The rate of successful parkings per hour is $S\mu$, and thus the fraction of would-be parkers who receive parking spaces virtually instantaneously upon arrival is $S\mu/\lambda$. The cars that do not get nearly instantaneous parking remain patrolling for an amount of time that is exponentially distributed with mean $1/\gamma$. For this revised queueing system W_q, the mean time patrolling can be written,

$$W_q \approx (0)(S\mu/\lambda) + (1/\gamma)(1 - S\mu/\lambda) = \frac{1 - S\mu/\lambda}{\gamma}$$

Since $L_q = \lambda W_q$, we can write

$$L_q \approx \frac{\lambda - S\mu}{\gamma},$$

as was to be shown.

In the above argument we use "approximately equal to" signs instead of "equals signs." This is due to the fact that there is a small but positive delay between a car's arrival in the queue of patrolling cars and its selection as a recipient of a parking space. The mean delay between the arrival of a newly patrolling car and the emergence of a newly available parking space is $1/S\mu$, assumed to be very small in contrast to $1/\gamma$.

The above formula for L_q can be re-written in the form: $L_q \approx \frac{S\mu W_q}{1-\gamma W_q}$ or $\frac{L_q}{S} \approx \frac{\mu W_q}{1-\gamma W_q}$. This representation is practically useful if we are unable to observe λ. Here, $\frac{L_q}{S}$ is the expected number of cruising vehicles per on-street parking space. For example, if $\frac{L_q}{S} \approx 0.2$ and there are 20 on-street parking spaces along a certain street, it can be considered that these 20 parking spaces are generating 4 cruising vehicles on average when parking spaces are fully occupied.

Similarly, $\frac{S\mu}{\lambda}$, the probability that a random arrival obtains an on-street parking space, can be represented as $\frac{S\mu}{\lambda} \approx 1 - \gamma W_q$. Both $\frac{L_q}{S}$ and $\frac{S\mu}{\lambda}$ are important performance indices for on-street parking. Larson and Sasanuma (2010) examined these performance indices for Newbury Street on-street parking in Boston and obtained $\frac{L_q}{S} \approx 0.14$ and $\frac{S\mu}{\lambda} \approx 0.35$.

In the following two subsections we model explicitly two alternative ways of implementing the LCFS queue discipline, as discussed above. These analyses are to show the operational feasibility of the revised but highly fictional LCFS queue discipline. The “real system” at all times is still assumed to follow the SIRO queue discipline.

2.2. Random Walk

Assuming the postulated LCFS queue discipline, one can model the arrival of a newly patrolling car as an entry into “state 1” an infinite random walk on the non-negative integers, where state 0 implies that the car transitions to a trap state -- signifying successful assignment to a parking space. Transitioning to any higher state j+1, $j \geq 1$, indicates that the position in queue has been changed upward from j to j+1. Due to the LCFS discipline, higher states imply less likelihood of eventually receiving a parking space. If we define

$\beta_0 \equiv$P{car enters the trap state}=

P{car transitions down one state in the random walk} =

P{car obtains a parking space},

then we can write

$$\beta_0 = P\{\text{first transition is to the trap state}\} + (1 - P\{\text{first transition is to the trap state}\})\beta_0{}^2$$

The reason for the term $\beta_0{}^2$ is the fact that if the car has transitioned into state 2, then to be awarded a parking space it must first transition down to state 1 and then eventually to state 0. Each transition down one state occurs with probability β_0, and the transition processes in each case are independent. Under the LCFS queue discipline, the probability that the first transition is to the trap state is equal to the probability that a parking spot becomes available before the next arrival, and that is equal to $S\mu/(S\mu+\lambda)$. Thus we can write,

$$\beta_0 = \frac{S\mu}{S\mu+\lambda} + \frac{\lambda}{S\mu+\lambda}\beta_0{}^2$$

The solution to this quadratic equation is $\beta_0 = S\mu/\lambda$, and that agrees with our intuition and previous results.

There is a subtlety in the derivation, as the argument appears to ignore reneging. Since reneging can occur, the “cars” in the argument are in fact ordered slots: youngest slot in queue, 2nd youngest slot in queue, etc. The car occupant of any slot may change due to reneging. Once that is seen, the results are seen to be valid, even in the presence of reneging.

2.3. Queuing Newly Available Parking Spaces

If one does not wish to consider the LCFS policy analyzed above, perhaps due to unrealistic demands on tracking newly arriving cars, one can accomplish the same objective by using a queue discipline that we will call **NCNS**, *Next Come, Next Served*. In this scheme each newly available parking slot enters a queue of other newly available parking slots, and this queue is depleted by newly arriving cars seeking parking slots. Any driver in a car lucky enough to arrive when this queue of available parking slots is nonempty is immediately given a parking slot. All others are denied slots forever, and they join the other patrollers who eventually renege after patrolling a random time having mean $1/\gamma$. This process can be modeled as an M/M/1 queue, with state i indicating the number of available parking slots (i = 0, 1, 2,…), and with upward transition rates $S\mu$ and downward transition rates λ. Since $\lambda > S\mu$, we know that the queue is stable and possesses a steady state solution. Using well-known results from the M/M/1 queue, we immediately have,

P{an empty parking space is available at a random time}$=1-P_0 = S\mu/\lambda < 1$.

Since Poisson Arrival See Time Averages (Wolff, 1982), we have

P{a random arrival obtains a parking space}$=1-P_0 = S\mu/\lambda < 1$,

as expected.

In steady state, the mean number of free parking spaces is,

$$N_p = \sum_{n=1}^{\infty} nP_n = P_0 \sum_{n=1}^{\infty} n(S\mu/\lambda)^n = \frac{\lambda - S\mu}{\lambda} \sum_{n=1}^{\infty} n(S\mu/\lambda)^n = \frac{S\mu}{\lambda - S\mu}$$

For example, if $\lambda = 2S\mu$, then $N_p = 1$ free parking space. One free parking space would remain free for an amount of time equal to the time of the next driver seeking a parking space, having mean $1/\lambda$. Usually this time is quite small in contrast other times in the system. More generally, in this instance Little's Law states that $N_p = S\mu W_p$, so we have the mean time that a newly available parking space remains available is

$$W_p = \frac{1}{\lambda - S\mu}$$

As an example, if $\lambda = 100$ cars per hour and $S\mu = 40$ cars per hour, then W_p = (1/60) hour = 1 minute. Again, this time is small in contrast to other times in the system, and all of our results are correct within acceptable "engineering approximations."

In conclusion, we can feasibly implement a car-to-parking-space queue discipline that supports Eq. (3), using either LCFS or NCNS. But we remember that the actual or "real" discipline is still assumed to be SIRO.

2.4 The Distribution of Patrolling Cars

Using the above logic, we see that the entire system, conceptually augmented with either LCFS or NCNS queue discipline; can be viewed as a Poisson arrival queue with infinite number of servers, i.e., an $M/G/\infty$ queue. "Service" occurs for any car the instant the car obtains a parking space or reneges from patrolling. The distribution of numbers of patrolling cars in the system is not affected by our augmented queueing discipline. Mean service time M can be written,

$$M = (0)\frac{S\mu}{\lambda} + \left(\frac{\lambda - S\mu}{\lambda}\right)\frac{1}{\gamma} = \frac{\lambda - S\mu}{\lambda\gamma}$$

The Poisson process arrival rate is λ. For the $\mathrm{M/G/\infty}$ queue having arrival rate λ and mean service time M, the steady state probability distribution of the number N of customers in the system is well-known to be Poisson with mean λM, i.e.,

$$P(N=n)=\frac{(\lambda M)^n}{n!}e^{-\lambda M}, \quad n=0,1,2,\ldots$$

In this case, we can write the probability that there are N cars patrolling for parking spaces is equal to

$$P(N=n)=\frac{\left(\frac{\lambda-S\mu}{\gamma}\right)^n}{n!}e^{-\frac{\lambda-S\mu}{\gamma}}, \quad n=0,1,2,\ldots$$

Here again we see that the mean number of patrolling cars is equal to $\frac{\lambda-S\mu}{\gamma}$, the result of Eq. (3). But now we know that the entire distribution – assuming our saturation conditions hold – is Poisson. Finally, as saturation grows worse, that is as λ increases towards ever-greater congestion, the Poisson distribution becomes a Gaussian or Normal distribution. Also, as λ increases, the coefficient of variation for the number of patrolling cars $\sqrt{\frac{\gamma}{\lambda-S\mu}}$ approaches zero, and therefore, the number of patrolling cars becomes more stable around its mean $L_q \approx \frac{\lambda-S\mu}{\gamma}$. This suggests that it is plausible to use the mean L_q as a deterministic quantity when we approximate the properties of a heavily congested system (See Kaplan (1987)). For example, we can estimate an approximate (Poisson) exit rate of any patrolling car from the SIRO system as

$$\gamma+\frac{S\mu}{L_q}\approx\gamma+\frac{1-\gamma W_q}{W_q}=\frac{1}{W_q}$$

The next step to take with this model is to place hourly prices on on-street parking and off-street parking. Then one makes certain model parameters dependent on these prices, especially the ***price difference*** between on-street and off-street parking. These ideas build on the suggestions of Shoup (2005). As the price difference between on-street and off-street parking becomes less, one should have the rate γ at which one leaves the queue of patrolling cars increase. That is, the desire to find an on-street parking space and the patience it requires in the patrolling queue will decrease as the price advantage of on-street parking decreases. Eventually as one gets closer to price parity, our approximate assumption of an endless queue of patrolling cars becomes invalid and we must modify the model accordingly. Shoup's stated objective is to raise on-street prices so that one has roughly 15% of the on-street parking spaces available in steady state. For the model, this would require extending the state-rate-transition diagram down significantly into unsaturated states but still allowing the artifice of stopping at some left-most nonzero state that has very small steady state probability. We do not see the need to model the system all the way down to zero parking spaces being occupied.

3. CONGESTION PRICING AND QUEUEING THEORY[1]

Andreatta and Odoni (2003) showed how we can set congestion charge (CC) by following the economic principle: *The congestion cost caused by the entrance of a driver to a queueing system consists of the cost of delay to this driver (internal cost) plus the cost of additional delay to all other users caused by this driver (external cost).* For example, if a driver enters into a congested road and experiences 5 minutes delay, the internal cost to him is the cost of 5 minutes. However, when the road is very congested, the entrance of this driver may delay each of 7 other drivers one additional minute. Then the external cost generated by him is the cost of 7 minutes to the other drivers. Economists argue that in order to achieve the most efficient use of the road facility, this external cost should be burdened by each driver. In economic terms, the external cost should be internalized (Vickrey (1969), and Carlin and Park (1970)): They claimed that, "*Optimal use of a transportation facility cannot be achieved unless each additional (marginal) user pays for all the additional costs that this user imposes on all other users and on the facility itself. A congestion toll not only contributes to maximizing social economic welfare, but is also necessary to reach such a result.*" Vickrey (1959) (republished in 1994) also pointed out the importance of a variable pricing system for on-street parking spaces in order to ensure some vacancy to accommodate the demand and avoid unnecessary traffic congestion caused by on-street parking shortages.

We follow economic principles to obtain the "optimal" congestion pricing. Consider a queueing facility with a single type of user in steady state and let

$\lambda =$ demand rate per unit of time by road users.

c = cost of delay per unit time per user.

C = total cost of delay per unit time incurred in the system.

L_q = expected number of users in queue.

W_q = expected delay time in queue for a random user.

We can also assume that $L=L_q$ and $W=W_q$, as in our parking model.

Then the time-average total delay cost per unit time can be written,

$$C = cL_q = c\lambda W_q$$

Where Little's Law is used. The *marginal delay cost* (*MC*) imposed by an additional road user can be obtained as, $MC = \frac{dC}{d\lambda} = cW_q + c\lambda \frac{dW_q}{d\lambda}$.

The first term on the right is the internal cost experienced by the additional road user, and the second is the external cost due to the increase in the expected delay, $\frac{dW_q}{d\lambda}$, resulting from the

[1] This discussion is based on Andreatta, G., and Odoni, A.R. (2003) 'Analysis of Market-Based Demand Management Strategies for Airports and En Route Airspace', in *Operations Research in Space and Air*, Ciriani, T. (ed.) Kluwer Academic Publishers, Boston; also Professors Odoni, A.R., and Larson, R.C. (2006) 'Lecture Notes' in *Logistical & Transportation Planning Methods*, MIT, Cambridge, MA.

increased traffic created by this user. Hence, we can write two components of the marginal delay cost *MC* as follows:

(1) Marginal internal cost: $MC_i = cW_q$

(2) Marginal external cost: $MC_e = c\lambda \dfrac{dW_q}{d\lambda}$.

Vickrey suggested that the marginal external cost MC_e should be imposed on each road user in order to realize socially "optimal" utilization of road resources. Hence, the congestion charge (CC) should be set equal to MC_e.

4. THE PARKING PRICING MODEL

From our previous work, for the parking process in saturation, the total delay cost per unit time and associated marginal delay cost are

$$C = cL_q \approx (\lambda - S\mu)\frac{c}{\gamma} \tag{4}$$

and

$$MC = \frac{\partial C}{\partial \lambda} \approx \frac{c}{\gamma} \tag{5}$$

We can also obtain the marginal internal cost and marginal external cost,

$$MC_i = cW_q \approx \left\lfloor 1 - \left(\frac{S\mu}{\lambda}\right)\right\rfloor \frac{c}{\gamma} \tag{6}$$

$$MC_e = c\lambda \frac{\partial W_q}{\partial \lambda} \approx \frac{S\mu}{\lambda} \cdot \frac{c}{\gamma} \tag{7}$$

The ratio of MC_e/MC_i is

$$r = \frac{MC_e}{MC_i} \approx \frac{\frac{S\mu}{\lambda} \cdot \frac{c}{\gamma}}{\left(1 - \frac{S\mu}{\lambda}\right) \cdot \frac{c}{\gamma}} = \frac{\frac{S\mu}{\lambda}}{1 - \frac{S\mu}{\lambda}} \tag{8}$$

Here, we observe an interesting result. For a given *c*, the marginal delay cost in the system is dependent only on γ and does not depend on $S\mu$ or λ. In a sense, in saturation each additional would-be parker "brings with him or her" an average of $1/\gamma$ of delay, to be incurred by somebody or some combination of people. However, the marginal internal cost, the marginal external cost and their ratio *r* are dependent on $\dfrac{S\mu}{\lambda}$, which is the success probability for would-be parkers to find on-street parking spaces. Eqs. (6) and (7) show that the marginal external cost MC_e is proportional to the parking success rate whereas the marginal internal cost MC_i is proportional to the parking failure rate. If half of would-be parkers will find a parking space, then marginal internal and external costs are equal. If 90% of all would-be parkers are denied parking, then the external cost MC_e associated with one new would-be parker is only $0.1c/\gamma$, whereas the internal cost MC_i is $0.9c/\gamma$. This is due to the fact that 90% of the time our new would-be parker arrives, he will be denied parking and will have to incur the mean patrolling time (cost) $1/\gamma$ almost all by himself; he denies others only 10% of the time.

Because of the reneging effect, the internal cost has a cap of c/γ, which prevents the congestion from becoming infinitely long. The reneging behavioral change that cruising drivers make while they are in the system is often ignored in economics-based congestion pricing theory. However, this behavioral change plays a critical role in determining properties of the saturated system. Our result suggests that if short supply is expected, it is important to design a system that gives cruising drivers incentives to renege.

5. TRADING OFF COST SAVINGS AND CONVENIENCE

Economists like to speak of "optimal" charges for those imposing external costs, this problem being no exception. But it is difficult to operationalize this concept. What precisely is meant by optimal? Optimal is an absolute word requiring a precise and unambiguous objective function and set of constraints. We do not have those conditions in the context of on-street vs. off-street parking. And how do the fees collected get distributed to aggrieved parties? As operations researchers and not as economists, we tend to think of drivers as decision makers who weigh their options and act accordingly.

Without significant empirical research, it is not possible to know precisely how would-be parkers would behave in our "patrolling queue" situation. But we can make some plausible first-order assumptions, presented in a transparent manner for review and critique. First, it seems clear that some drivers would value their time more than others, and those would tend to leave the queue of patrolling drivers more quickly than others. Second, a driver's willingness to spend time in the patrolling queue would rise or fall with the price differential between on-street and off-street parking, with higher price differentials meaning more willingness to spend time looking for less expensive on-street parking. Third, any unsuccessful patrolling driver will eventually become discouraged, "cut his losses," and leave the queue for more expensive off-street parking.

We can develop a simple model reflecting these assumptions. Suppose there are D categories of drivers, where category d, $1 \le d \le D$, has a self-assessed value for time of W_d dollars per hour. We assume the categories are rank-ordered such that $W_1 \ge W_2 \ge W_3 \ge ... \ge W_D$. Let p_d be the fraction of all would-be parkers belonging to category d, $1 \le d \le D$. Clearly, $\sum_{d=1}^{D} p_d = 1$. Let Δ be the hourly parking price differential (in dollars) between off-street and on-street parking, with the on-street parking being less expensive. We now need a decision criterion for a patrolling driver to leave the queue and accept the more expensive off-street parking. One plausible criterion is this: When the value of the time already invested in patrolling for a less expensive on-street parking space equals the price differential between off-street and on-street parking, then the expected values of the respective options – when including sunk costs – become equal. But the variance of costs for continued patrolling is large, whereas the variance of cost associated with the off-street option is zero (a known, published parking fee). Thus, the decision rule is to leave the queue and switch to off-street parking when the sunk cost of time invested becomes equal to the parking price differential. This set of assumptions provides a basis for evaluating the resultant reneging parameter γ as a function of the price differential Δ. The mean time that a category *d* patrolling driver would remain patrolling is $\frac{1}{\gamma_d} = \frac{\Delta}{W_d}$.

Including all *D* categories, weighed by their respective relative frequencies, the resulting relationship can be written,

$$\frac{1}{\gamma} = \Delta \sum_{d=1}^{D} \frac{p_d}{W_d} \tag{9}$$

As a numerical example, consider one-hour parking with $D = 3$; $p_d = 1/3$ for $d = 1, 2, 3$; $W_1 = 100,\ W_2 = 25,\ W_3 = 10.$ Then

$$\frac{1}{\gamma} = \frac{\Delta}{3}\left(\frac{1}{100} + \frac{1}{25} + \frac{1}{10}\right) = \frac{\Delta}{3}(0.01 + 0.04 + 0.10) = 0.05\Delta$$

If Δ = US$10/hr. then $(1/\gamma)$= 0.5 hr. = 30 minutes. If Δ = US$20/hr., then $(1/\gamma)$ is doubled to 60 minutes. One socially appealing aspect of the driver behavior assumed in this model is that the successful on-street parkers are differentially more likely to be poorer people who value their time less than others. Those who value their time highly will tend to leave the queue more quickly and pay the higher off-street parking rates.

Using the result (9), the marginal delay cost *MC* (5) becomes

$$MC \approx \frac{c}{\gamma} = c\Delta \sum_{d=1}^{D} \frac{p_d}{W_d}$$

Hence, the marginal delay cost by cruising drivers is expected to be large in a city where a large price differential Δ is observed.

6. EXTENDING THE MODEL TO INCLUDE HETEROGENEOUS DRIVERS

In this section, we confirm the intuition that "poorer people are more likely to be successful on-street parkers than richer people". Assume there are two types of drivers, or "would-be-parkers," Type 1 and Type 2, whose corresponding arrival rates and reneging rates are λ_i and γ_i (i=1, 2), respectively. We construct a 2-dimensional state-rate-transition diagram for the Markovian queue created by two types of drivers. Assume that each state is represented by the ordered pair n_1 and n_2, which correspond to the respective numbers of Type 1 and Type 2 drivers in the system. The state-rate-transition diagram is shown in Figure 2. As before, we continue to assume that $0 < P_{00} \approx 0$, but now for this 2-dimensional system. Again as before, we assume that the road is congested, with either type of driver able to fill all available parking spaces: $\lambda_1 \geq S\mu$ and $\lambda_2 \geq S\mu$.

We can write a set of balance-of-flow equations, where the balanced flows occur across complete horizontal cuts of the network of Figure 2,

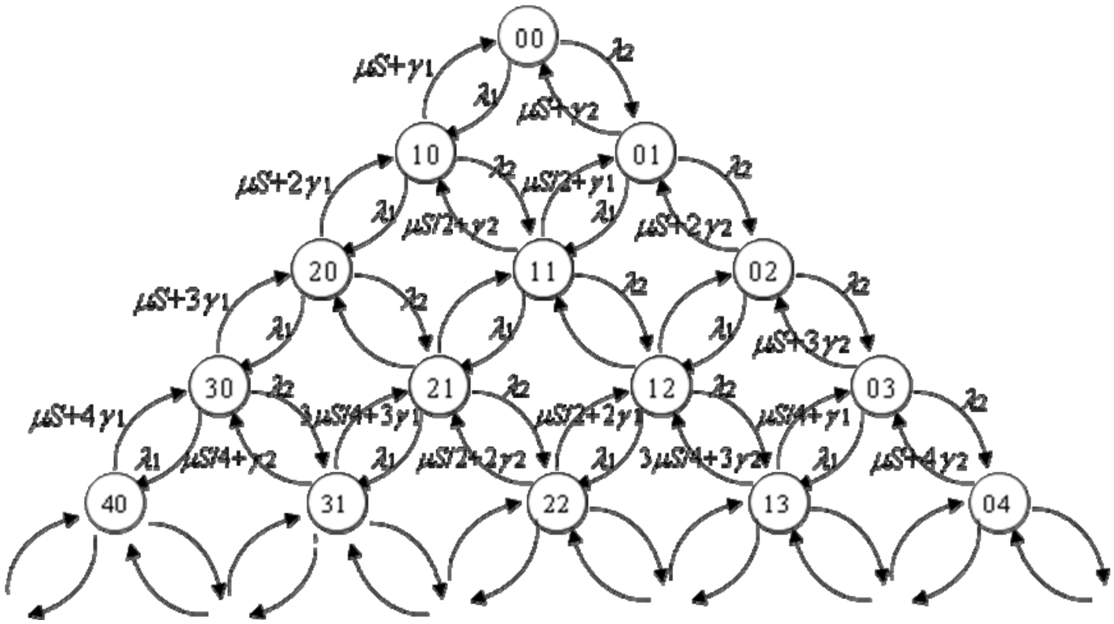

Figure 2 State-Rate-Transition Diagram for Queueing System with Two Types of Drivers

$$(\lambda_1+\lambda_2)P_{00}=(S\mu+\gamma_1)P_{10}+(S\mu+\gamma_2)P_{01}=S\mu(P_{10}+P_{01})+\gamma_1\left(P_{10}+0P_{01}\right)+\gamma_2(P_{01}+0P_{10})$$

$$(\lambda_1+\lambda_2)(P_{10}+P_{01})=(S\mu+2\gamma_1)P_{20}+(\frac{1}{2}S\mu+\gamma_1)P_{11}+(\frac{1}{2}S\mu+\gamma_2)P_{11}+(S\mu+2\gamma_2)P_{02}$$
$$=S\mu(P_{20}+P_{11}+P_{02})+\gamma_1\left(2P_{20}+P_{11}+0P_{02}\right)+\gamma_2\left(2P_{02}+P_{11}+0P_{20}\right)$$

$$(\lambda_1+\lambda_2)(P_{20}+P_{11}+P_{02})=(S\mu+3\gamma_1)P_{30}+(\frac{2}{3}S\mu+2\gamma_1)P_{21}+(\frac{1}{3}S\mu+\gamma_1)P_{12}$$
$$+(\frac{1}{3}S\mu+\gamma_2)P_{21}+(\frac{2}{3}S\mu+2\gamma_2)P_{12}+(S\mu+3\gamma_2)P_{03}$$
$$=S\mu(P_{30}+P_{21}+P_{12}+P_{03})+\gamma_1\left(3P_{30}+2P_{21}+P_{12}+0P_{03}\right)+\gamma_2\left(3P_{03}+2P_{12}+P_{21}+0P_{30}\right)$$

$$(\lambda_1+\lambda_2)(P_{(n-1)0}+P_{(n-2)1}+...+P_{0(n-1)})=S\mu(P_{n0}+P_{(n-1)1}+...+P_{0n})$$
$$+\gamma_1\left(nP_{n0}+(n-1)P_{(n-1)1}+...+1P_{1(n-1)}+0P_{0n}\right)+\gamma_2\left(nP_{0n}+(n-1)P_{1(n-1)}+...+1P_{(n-1)1}+0P_{n0}\right)$$

…

Adding up the countably infinite set of balance equations, we obtain

$$(\lambda_1+\lambda_2)(\sum_{n,m=0}^{\infty}P_{nm})=S\mu(\sum_{n,m=0}^{\infty}P_{nm}-P_{00})+\gamma_1\left(\sum_{n,m=0}^{\infty}nP_{nm}\right)+\gamma_2\left(\sum_{n,m=0}^{\infty}mP_{nm}\right)$$

Using the assumption $P_{00}\approx 0$, invoking the normalizing condition $\sum_{n,m=0}^{\infty}P_{nm}=1$, and using the definitions $L_1=\sum_{n,m=0}^{\infty}nP_{nm}$ and $L_2=\sum_{n,m=0}^{\infty}mP_{nm}$, we obtain

$$\lambda_1 + \lambda_2 = S\mu + \gamma_1 L_1 + \gamma_2 L_2 \tag{10}$$

Eq. (10) above is an extension of the homogeneous case result Eq. (3). We need to derive one more equation to solve for L_1 and L_2. In order to do this, consider the mean number of Type 1 and Type 2 renegers per hour, which are

$\sum_{n,m=0}^{\infty} n\gamma_1 P_{nm} = \gamma_1\left[\sum_{n,m=0}^{\infty} nP_{nm}\right] = \gamma_1 L_1$ and $\gamma_2 L_2$, respectively. Using these, the steady state mean number of parking spaces available and taken by Type 1 and Type 2 parkers per hour are $\lambda_1 - \gamma_1 L_1$ and $\lambda_2 - \gamma_2 L_2$, respectively. Note that the sum of the mean number of parking spaces available and taken by Type 1 and Type 2 parkers per hour is $(\lambda_1 - \gamma_1 L_1) + (\lambda_2 - \gamma_2 L_2) = S\mu$, using Eq. (10). Note also that both $\lambda_1 - \gamma_1 L_1$ and $\lambda_2 - \gamma_2 L_2$ are positive because the mean number of renegers $\gamma_1 L_1$ and $\gamma_2 L_2$ must be less than the arrival rate λ_1 and λ_2, respectively, in steady state.

We now argue that the proportion of parking spaces taken hourly by Type 1 (Type 2) drivers is equal to the proportion of cruising drivers who are Type 1 (Type 2). For if not, then Type 1 (Type 2) drivers would be more or less skilled than Type 2 (Type 1) drivers at finding parking spaces. Due to the SIRO queue discipline that rewards that driver, Type 1 or Type 2, who just happens to be closest to the newly available parking space, each type of driver is by definition equally skilled. And clearly the proportion of parking spaces taken per hour by Type 1 (Type 2) drivers is equal to the fraction of parking spaces occupied by Type 1 (Type 2) drivers. For if not, then the parking time statistics of the two types of drivers would differ, and this is not allowed in our model.

Invoking these results, we can write

$$\frac{\lambda_1 - \gamma_1 L_1}{L_1} = \frac{\lambda_2 - \gamma_2 L_2}{L_2}, \quad \text{or, simplifying,} \quad \gamma_2 - \gamma_1 = \frac{\lambda_2}{L_2} - \frac{\lambda_1}{L_1} \tag{11}$$

Combining Eqs. (10) and (11), we have

$$\gamma_2 - \gamma_1 = \frac{\lambda_2}{L_2} - \frac{\lambda_1}{\dfrac{\lambda_1 + \lambda_2 - S\mu}{\gamma_1} - \dfrac{\gamma_2}{\gamma_1} L_2} \tag{12}$$

and

$$\gamma_2 - \gamma_1 = \frac{\lambda_2}{\dfrac{\lambda_1 + \lambda_2 - S\mu}{\gamma_2} - \dfrac{\gamma_1}{\gamma_2} L_1} - \frac{\lambda_1}{L_1} \tag{13}$$

Since both $\lambda_1 - \gamma_1 L_1$ and $\lambda_2 - \gamma_2 L_2$ are positive, the denominators in Eqs. (12) and (13) are all positive. Therefore, unique positive solutions for both L_1 and L_2 are guaranteed in the above equations. Analytical solutions can be obtained for both L_1 and L_2 using the quadratic formula.

The method extends to three or any number of different types of drivers: if there are n types of drivers in the system, we can obtain the proportion of each in the system by solving n equations

$$\lambda_1 + \lambda_2 + \ldots + \lambda_n = S\mu + \gamma_1 L_1 + \gamma_2 L_2 + \ldots + \gamma_n L_n$$

and

$$\frac{\lambda_1 - \gamma_1 L_1}{L_1} = \frac{\lambda_2 - \gamma_2 L_2}{L_2} = \ldots = \frac{\lambda_n - \gamma_n L_n}{L_n}.$$

For simple illustrative purposes, consider a numerical example. Assume there are two types of drivers: 200 less affluent people per hour arrive to the system and their per-person reneging rate is 1/hr., and 200 affluent people per hour arrive to the same system and their per-person reneging rate is 3/hr. Both types of drivers are trying to find on-street parking spaces which capacity is $S\mu = 50$/hr. In this case, one could argue that less affluent people value their time at a rate of 1/3 that of affluent people. By placing numbers in Eqs. (12) and (13), we obtain

$$3-1 = \frac{200}{L_2} - \frac{200}{\frac{200+200-50}{1} - \frac{3}{1}L_2} \quad \text{and} \quad 3-1 = \frac{200}{\frac{200+200-50}{3} - \frac{1}{3}L_1} - \frac{200}{L_1}$$

Solving, we have $L_1 = 164$ and $L_2 = 62$. Hence, the ratio of less affluent and affluent in parking spaces are

Less Affluent : Affluent $= L_1 : L_2 = 164 : 62 = 73\% : 27\%$.

The interpretation is as follows: Even though less affluents' arrival rate is half of the total arrival rate, less affluent people occupy nearly three quarters of the on-street parking spaces because of their lower reneging rate, their greater "patience" while cruising for an available parking space. Furthermore, the success rate of finding available parking spaces for less affluent and affluent are $\frac{\lambda_1 - \gamma_1 L_1}{\lambda_1} = \frac{200 - 1 \cdot 164}{200} = 18\%$ and $\frac{\lambda_2 - \gamma_2 L_2}{\lambda_2} = \frac{200 - 3 \cdot 62}{200} = 7\%$, respectively.

Therefore, in terms of distributional equity, the provision of on-street parking spaces can be seen as "good" because poorer people tend to utilize inexpensive parking more often than richer people. However, the result also suggests that poorer people are more apt to patrol than richer people, thereby maintaining levels of street congestion that may be found unacceptable. The way to fix that problem is to raise the price of on-street parking, and that would increase the reneging rate of both poorer and richer people since the price advantage of patrolling for on-street parking diminishes.

7. CONCLUSION

We developed a new queueing model of the problem automobile drivers cruising streets looking for on-street parking spaces. We found that in near saturation conditions (1) the queueing delay is inversely proportional to reneging rate; (2) the distribution of number of patrolling drivers follows a Poisson distribution, (3) the marginal delay cost imposed by an additional road user becomes constant as a result of reneging when on-street parking spaces are full, and (4) the congestion charge (CC) is calculated as the marginal external cost. We extended the homogeneous model to heterogeneous model with two types of drivers. We found that the successful on-street parkers are differentially more likely to be less affluent people who value their time less than others.

ACKNOWLEDGEMENTS

We are grateful for helpful discussions and valuable information from Amedeo R. Odoni, Akira Hibiki, and Robert Irwin. We also thank Cordell Hull for awarding a research grant to MIT to help support this research. In addition, we thank the MIT-Portugal Program, which also supported this research.

REFERENCES

[1] Andreatta G., Odoni A.R. (2003), Analysis of Market-Based Demand Management Strategies for Airports and En Route Airspace; Operations Research in Space and Air, Ciriani, T.(ed.), Kluwer Academic Publishers; Boston.

[2] Carlin A., Park R.E. (1970), Marginal Cost Pricing of Airport Runway Capacity; *American Economic Review* 60; 310-318.

[3] Cox D.R., Smith W.L. (1954), On the Superposition of Renewal Processes; *Biometrika* **41;** 91-99.

[4] Downs A. (2004), Still Stuck in Traffic: Coping with Peak Hour Traffic Congestion; Brookings Institution Press, Washington, D.C.

[5] Kaplan E.H. (1987), Analyzing Tenant Assignment Policies; *Management Science* **33;** 395-408.

[6] Kaplan E.H. (1988), A Public Housing Queue with Reneging and Task-Specific Servers; *Decision Sciences* **19;** 383-391.

[7] Larson R.C., Sasanuma K. (2010), Urban Vehicle Congestion Pricing: A Review; *Journal of Industrial and Systems Engineering* **3;** 227-242.

[8] Lyons G., Dudley G., Slater E., Parkhurst G., Slater L. (2004), Evidence-Base Review – Attitudes to Road Pricing; Final Report to the Department for Transport, UK. Bristol: Centre for Transport and Society.

[9] Shoup D.C. (2005), The High Cost of Free Parking; APA (American Planning Association) Planners Press, Chicago.

[10] Vickrey W. (1969), Congestion Theory and Transport Investment; *American Economic Review Proceedings* **59;** 251-260.

[11] Vickrey W. (1994), Statement to the Joint Committee on Washington, D.C., Metropolitan Problems, The Economizing of Curb Parking Space – A Suggestion for a New Approach to Parking Meters; *Journal of Urban Economics* **36;** 42-65 (republished).

[12] Wolff R.W. (1982), Poisson Arrivals See Time Averages; *Operations Research* **30;** 223-231.

[13] Odoni A.R., Larson R.C. (2006), Lecture Notes in Logistical & Transportation Planning Methods, Massachusetts Institute of Technology, Cambridge, MA. http://web.mit.edu/urban_or_book/www/

[14] Shoup D.C. (2006), Cruising for parking; *Transport Policy* **13;** 479-486. http://shoup.bol.ucla.edu/Cruising.pdf

[15] Transportation Alternatives (February 2007), No Vacancy: Park Slope's Parking Problem and How to Fix It. http://www.transalt.org/campaigns/reclaiming/novacancy.pdf

[16] Victoria Transport Policy Institute, *Road Pricing: Congestion Pricing, Value Pricing, Toll Roads and HOT Lanes*. http://www.vtpi.org/tdm/tdm35.htm