Delays at Toll Booths — Why Wait in Line?

DENOS C. GAZIS and RALPH E. GOMORY

Ralph E. Gomory has been president of the Alfred P. Sloan Foundation since 1989. Before that he was senior vice president for science and technology at IBM. He has served in many capacities in academic, industrial and governmental organizations, and is a member of the National Academy of Science and the National Academy of Engineering. He has received the Lanchester Prize in 1963, the John von Neumann Theory Prize in 1984 and the National Medal of Science, awarded by the President in 1988.

Denos C. Gazis is assistant director of the Department of System Technology and Science at IBM Research in Yorktown Heights, New York. He previously served in various research, review and management positions covering a wide range of research and development activities in computer technology. He is currently leading IBM activities on Intelligent Vehicle-Highway Systems. He received the Lanchester Prize of Operations Research for his pioneering work on traffic theory.

ALMOST EVERY driver has had the frustrating experience of waiting at a congested system of toll gates. Some drivers suffer this experience daily, some twice daily, in their regular commute. At times, toll gates cause queuing for several miles upstream of the toll plaza. These delays to drivers, this waste of fuel, and the increased pollution are generally accepted as inevitable fruits of civilization. But, in fact, the problem is tractable. Good solutions are at hand now and even better ones are not far off. The difficulties that prevent a solution today are more institutional than technological.

There are activities for which an optimization, or at least great improvement, is in principle possible, but is prevented by a mismatch of jurisdictional and operational boundaries. Traffic

TRANSPORTATION QUARTERLY, Vol. 48, No. 2, Spring 1994 (107-114)

control problems frequently belong in this category. Typically, a managing authority of a traffic facility does not derive any direct benefit from optimizing toll booth arrangements to reduce or eliminate traffic waiting times. Instead, benefits accrue to the users of the system, in the absence of any mechanism for assessing them for the cost of the improvements. Thus, improvement of toll collection systems has relied on the good will of managing authorities, and sometimes public uproar. In some instances, improvements in operational management (installation of automated toll gates, for example), made in order to reduce labor costs and the possible misappropriation of toll receipts, have also produced some reduction of delays.

Management of a traffic system should be carried out with an eye toward reducing the "social cost" of the system. The social cost should include the cost of construction and operation of the system, and also the cost to society of any deficiencies in the system; for example, the cost of traffic delays, fuel, pollution, etc. Taking this point of view in the case of delays caused by toll gates, it is easy to conclude that all is not well in the way toll collection is managed today. It also is apparent that relatively painless solutions exist that can be implemented now.

HOW TOLL BOOTH DELAYS CAN BE ELIMINATED

The key to eliminating toll booth delays is understanding that they are primarily caused by lack of *throughput* at toll plazas, not by the fact that vehicles must slow down to pay a toll.

Traffic throughput at a toll plaza is the maximum number of vehicles per hour that can pass through. If this number is less than the capacity of the highway to bring cars up to the toll plaza, the toll plaza will represent a potential bottleneck in the system. If during rush hour, the highway is actually used to capacity, cars will be delivered to the toll plaza faster than they can go through it, a backup will begin and this backup will continue and get longer and longer. This backup, with its attendant delays, will continue until, say at the end of the rush hour, the number of arriving vehicles falls to less than the highway capacity and finally to less than the throughput of the toll plaza, and the backup starts to decrease. It is this backup that causes the 10, 20, 30 or 40-minute delays commonly found at many toll plazas.

Any plan that provides throughput at the toll plaza equal to or, better yet, a little greater than the capacity of the highway leading to it will *almost completely eliminate* toll booth delays. We say almost to allow both for statistical fluctuations in the arrival of cars and the fact that vehicles may have to slow down to pay tolls.[1] This last source of delay is very small compared to that caused by backup. To see this, imagine a trip down a completely empty highway. Slowing down to pay tolls on a completely empty road adds perhaps 30 seconds at each toll plaza. So it is the lack of *throughput* with its attendant backup that causes almost all substantial toll plaza delays. However, there is no reason today, why toll plazas can not have throughput equal to the capacity of their highways.

A straightforward improvement to toll plaza operations is the addition of gates. However, if this addition is made abreast of the existing gates, it is deficient in two ways. First, it requires more roadway width at the line of gates, and this additional right-of-way may not be feasibly available. Second, it suffers from the "Edie effect," named after Leslie Edie who studied the operation of toll gates operated by the Port of New York Authority in the early 1950s. Edie observed[2] that addition of toll gates abreast reaches a point of diminishing returns because of the penalties of weaving maneuvers and the concomitant confusion to drivers. He found that

1. Cars arrive at the toll gate at some measurable average rate, but the gaps between arrivals are not constant. Instead, they vary randomly according to some statistical distribution function. At the toll gate, the cars are "served" at some average service rate, but again the service times are not constant, but vary according to some other statistical distribution function. Whatever the choice of distribution functions for arrival times and service times, they involve a variance, or spread, of these times around a mean. And this spread causes delays to increase rapidly with demand, even when this demand is substantially below the capacity of the service system reflected by the mean service time. When a customer takes a little less than the mean time to be serviced, it is not always possible for the next car in line to use the service time that has been gained, simply because the next car is not yet at the gate because of the arrival time variance. On the other hand, when a car takes a little longer than average to be served, more often than not there will be pile-up cars behind which take some time to untangle. As demand increases toward the theoretical maximum capacity of the queue server, (i.e. the toll gates), the delay suddenly bends sharply upward and climbs precipitously. This behavior was demonstrated by a computer simulation study carried out by Professor Mitsuru Saito of the City University of New York and one of his graduate students, Mr. Wei Tao Chen. Saito and Chen simulated traffic delays at toll gates along the Garden State Parkway in New Jersey. They found that traffic delays would about triple, from less than a half minute to as much as 1.5 minutes, as the demand increased from 85% of service capacity to 95% of service capacity.

2. His study earned him the first Lanchester Prize of the Operations Research Society of America and Johns Hopkins University, in 1954. It was also popularized in a book entitled, *The Scientist Speculates*. In a chapter of that book entitled "Edie's Number," the number 3 is defined as Edie's number.

when the fan-out from highway lanes to toll gates exceeded three lanes, additional gates would yield very small improvement of throughput. Careful design of a toll plaza can raise this point of diminishing returns to about four lanes, but this fan-out of four lanes per approach roadway lane is generally the practical maximum limit in the design of toll gates today, assuming there is enough real estate available.

Lane fan-outs of 4:1 may be enough to match highway capacity if some of the gates can accept tokens or exact change, and if everything works according to plan. Experience suggests a number closer to 5-5.5 toll booths per approach roadway lane being required to match toll plaza throughput to the capacity of the highway. Consequently, what is needed is to augment today's toll plazas by some means to increase their throughput by about 25% to 35%. Fortunately, this can be done. In fact, there is a great deal of interest currently in mobilizing "high tech" in order to solve the toll delay problem. Before discussing these promising high tech approaches, it is worth pointing out that some "low-tech" approaches are readily available now.

One low-tech approach involves the addition of gates slightly downstream or upstream of the existing toll plaza. The basic idea of this plan, which is used at some locations today, is to use roadway length as a substitute for additional right-of-way width, which may not be available. This toll plaza layout technique avoids the Edie effect, by staggering additional gates along the length of roadway entering or leaving an existing toll plaza, rather than extending the existing bank of toll gates perpendicular to the highway. Some existing gates are used to pass traffic through to or from added gates downstream or upstream from the existing row of gates. This technique was used to expand toll plaza traffic capacity at the Henry Hudson and Bronx Whitestone Bridges in New York City, where it has appreciably reduced backups.

While this low-tech approach is especially promising, there are others. A second solution of the same general type is to use an existing cashier-manned gate to collect from alternative "platoons" of n cars. The first platoon is waved through to pay at a downstream gate placed at a distance sufficient for accommodating storage of a platoon. The second platoon pays at the first toll gate. This technique has been tried in several places and appears to work quite well. Studies have shown that the service capacity may be increased by as much as 30%. Thus, there are low-tech techniques that can work to substantially reduce or eliminate traffic delay at toll plazas.

A solution that promises to eliminate toll delays is the use of Automatic Vehicle Identification (AVI) in conjunction with fully automated toll gates. In such a system, a vehicle carries an identification tag (e.g., an RF tag) that is read by an instrumented toll gate, while the cars go through without need to slow down. The driver is billed periodically, or alternatively "reloads" the vehicle/subscriber tag periodically with electronic tokens. Ultimately, automated toll gates may be expected to have a throughput equal to a highway lane, and operate at speeds equal to the speed limit. However, current tests are generally carried out at low approach speeds and they produce throughput somewhat lower than the capacity of a highway lane; say around 80% of lane capacity. A cost/benefit analysis may be carried out in order to determine advantages of dedicating one or more gates to such an operation. A sufficient number of cars must be tagged to utilize the instrumented gate capacity on any given day, and the tagged cars must be provided unimpeded access to the instrumented gate through proper design of the approach lanes. Otherwise this solution will fail to match throughput of the toll plaza with the capacity of the highway.

In view of the fact that only a 25% to 35% increase in capacity is needed to avoid back-up at toll plazas, even a partial deployment of AVI, provided there are enough enrolled users, would be enough to solve the problem. While the AVI solution may be the ultimate answer for the future, we must face the fact that we are destined to live without it for a considerable transition period. During this period, we can eliminate toll gate delays through addition of downstream and/or upstream gates. We can then add a gradual introduction of AVI gates.

For those who have suffered through long toll both delays, it may seem almost superfluous to explain why improvement is worthwhile. Nevertheless, we will show next that from almost any standpoint there is a social cost to living with toll booth delays that far outweighs the cost of getting rid of them.

THE SOCIAL COST OF TOLL DELAYS

There are many instances when the cost of wasted fuel alone exceeds the value of the tolls collected. For example, an average passenger car crawling through a congested queue in front of a toll gate consumes about two gallons of gasoline per hour. At $1.25 per gallon, this amounts to $2.50 per hour in gasoline consumption. A 10-minute delay costs over 40 cents in gasoline alone, but it has been known to occur before toll gates collecting 25 cents.

If we take into account not only the cost of gasoline but also the cost of air pollution and the cost of wasted time, it is reasonable to put a price tag of at least $15 per hour on delays. This translates to 25 cents per minute, leading to the conclusion that the cost of delays generally far exceeds the monetary value of the toll.

To be more quantitative, assume, for example, that a 3-lane highway reaches its peak capacity of about 1800 cars per lane per hour, for a total of 5400 cars per hour. Assume that the capacity of the toll gates is 25% below the peak capacity of the highway, or 4050 cars per hour. This means that in one hour of peak traffic, some 1350 cars will be queued in front of the gates. Assuming that after that the traffic input falls to 3050 cars/hr, (about 25% less than the gate capacity), it will take 1350/1000 = 1.35 hrs for the accumulated queue to dissipate. The cumulative delay of all cars during the rush hour will be

$$D = (1/2) \times 1350 \times (1+1.35) = 1586 \text{ car hours}$$

The social cost of this delay is $24,000 in round numbers. If it is repeated day after day for just 300 days out of the year, it will amount to $7.2 million yearly, *at just one toll plaza, in one direction only*. The maximum individual delay is suffered by a driver arriving an hour into the rush period, and is equal to 0.33 hours (20 minutes). The delay to other drivers rises from 0 to 20 minutes within an hour, and then falls gradually back to zero after 1.35 more hours[3].

The above computations do not even describe the total impact of the toll gate delay. A 3-lane highway, together with the up-stream toll plaza, requires over 1.5 miles to store 1350 cars. But traffic does not change abruptly from free-flowing to stationary. The effect of a queue propagates further upstream, slowing down traffic, interfering with traffic at entrance ramps, and extending

3. Of course, traffic demand does not change abruptly from the maximum rate to some other. Rather, it begins to taper off near the end of the rush period, and tends to some lower rate after the end of the rush period. However, the resulting delays to drivers, and their social cost, are comparable to those computed above on the basis of two stages of demand.

the deleterious impact of the toll gate delays to the contiguous road network, a fact very familiar to all drivers.

Even neglecting these additional effects of queuing at toll gates, computations such as the above can be used to derive a curve showing the social cost of toll delays versus toll gate capacity. These costs are remarkably low compared to the findings of Saito and Chen, where they cite estimates of $0.15 million for the construction of additional toll booths.

With such an apparently strong case for building additional lanes, perhaps we should comment at this point about counterarguments. It may be argued that at the entrance to a crowded area, such as Manhattan, toll gates do not cause any incremental delay to drivers, but simply distribute the delay between delay at the toll gate and delay during the rest of the trip. This argument is only plausible if we are confident that the congestion in the area is largely caused by the traffic from the toll booths and is therefore eased by preventing passage through those booths. Another argument in favor of disregarding toll delays is that they may be a form of "congestion pricing." This is the practice of imposing a toll for entering a congested area. Even in this case, it is unreasonable and wasteful to impose a congestion price in the form of delay, with its concomitant environmental and energy budget.

It is difficult in any case to see why one should build expensive 3-lane highways and then provide them with a toll capacity that matches a 2-lane road. If this outcome were desirable, as the various arguments against improving passage through the tolls would claim, the result could be obtained far more cheaply by building a 2-lane road. It is difficult to make a plausible case for highways whose toll booths do not match roadway capacity.

Drivers waiting in line for congested toll booths should know that their wait is the intended, or more unlikely unintended, consequence of some governmental decisions. It is not the inexorable workings of some law of nature — it is already avoidable.

CONCLUSION

This discussion points strongly to the conclusion that the social cost of toll booth delays far outweighs the cost of eliminating them. It also shows that this social loss and delay is avoidable, and that it is avoidable *now*. We hope that drivers will keep this in mind as they wait at congested toll plazas across the country. Perhaps this consciousness will create pressure to implement solutions that are in fact at hand.