HIGHWAY CONGESTIONS – CAUSE, COMPONENTS, SOLUTIONS

D.C. PUCHIANU[1], H.G. COANDĂ[2]

Congestia apare când cererea depăşeşte capacitatea, astfel încât inițiativele de construire de autostrăzi adiționale sau autostrăzi de scurtătură sunt pe deplin confirmate. Vom argumenta că nu capacitatea mică a unei autostrăzi este principalul motiv care duce la congestie, ci, cauza majoră de congestie o reprezintă funcționarea ineficientă a autostrăzilor în timpul orelor de vârf. Analizele datelor de trafic arată cum congestia reduce eficiența autostrăzii, ceea ce înseamnă că, vehiculelor le ia mai mult timp pentru a traversa porțiuni care sunt aglomerate, decât le-ar trebui dacă congestia ar fi prevenită. Compensarea acestei pierderi de eficiență printr-o extindere de capacitate este financiar imposibilă; compensarea printr-o cerere de scurtare este în mod practic imposibilă. Modul cel mai bun de a combate congestia este sporirea eficienței operaționale. Pentru a spori eficiența este necesar accesul controlului inteligent la autostrăzi prin măsurători pe rampe.

Congestion appears when the demand surpasses the capacity, so the initiative for additional highways or shortcut highways is completely confirmed. In this paper, we will therefore argue that it is not the low capacity of a highway the first cause for congestion, but the major cause of congestion represents the inefficient operation of highways during rush hours. Traffic data analyses show how the congestion reduces the highways efficiency, which means, the vehicles need more time to cross a congested portion than normal if the congestion is prevented. The compensation of this loss of efficiency through an extension of capacity is financially impossible; the compensation through a shortcut is practically impossible. The best way to prevent the congestion is to increase the operational efficiency. In order to increase efficiency, it means to have an intelligent access control of highways through ramp metering.

Keywords: congestion, efficiency, highways network, control politics, traffic ramps

1. Introduction

The development of metropolitan cities for commercial activities, services and administrative structures determines the citizens to move in cheap areas situated on town outskirts. For many years creating new highway structures solved the increasing demand of vehicles access. Because the demand increased

[1] Ph.D Student, Faculty of Control and Computers, University “Politehnica” of Bucharest, Romania
[2] Lecturer, Ph.D., Faculty of Electrical Engineering, University “Valahia” of Targoviste, Romania

year by year, the new highway structures created have started facing more and more the risk of congestion.

It was normal that the network transportation be in the focus of researches – as the data network or supply energy network, to solve the problem of congestion. Intelligent systems were created to reduce congestion. Such a system is presented in [1].

The dedicated literature has different perspectives on this subject. A recent approach has the prevention of congestion through ramp metering. Of course, different opinions can be found on: *How to measure congestion; how much congestion can be eliminated by ramp metering; how much of it is due to excess demand; what kind of management can be used to reduce it; and the relative magnitudes of recurrent vs non-recurrent congestion.* The reason for these differences is simple: *in the absence of reliable empirical knowledge of congestion and its causes, people holding different opinions won't change them.*

This work resumes a part of studies made during the elaboration of the PhD thesis titled "*Traffic optimization for intelligent highway systems*", now ready to be sustained. Some results are obtained using as measurement tool PeMS (*Performance Measurement System*) [1] in order to compare the particularity of highway congestion with communications and energy supplies network congestion.

2. Congestion components

The maximum throughput over a link - its effective capacity (empirical notion) - depends on how a link is connected to other links and the pattern of traffic, as well as its physical characteristics (features and shapes). A challenge for traffic theory is to determine the maximum throughput of a link, given the network topology and traffic pattern.

The congestion delay can be divided into (1) the portion that can be eliminated by ramp metering, and (2) the delay due to excess demand. There is a systematic procedure to calculate these two components of delay for recurrent congestion, using loop-detector data.

Recurrent congestion evolves over 3 phases (Fig. 1): increased demand is first met at free flow speeds, until the demand exceeds maximum throughput and congestion starts; both flow and speed then decrease and occupancy increases; only after demand drops well below maximum throughput does occupancy decrease and speed increase until free flow is reestablished. So the objective of ramp metering must be to maintain free flow and maximum throughput. It is a challenge to design such a ramp metering algorithm.

Congestion measures compare the actual travel time with some standards. There are two defensible standards: one is travel time under free flow conditions (at 100kmph/60mph), the other one is travel time under maximum throughput.

The amateur drivers mostly accept the first one and the professional drivers the second.

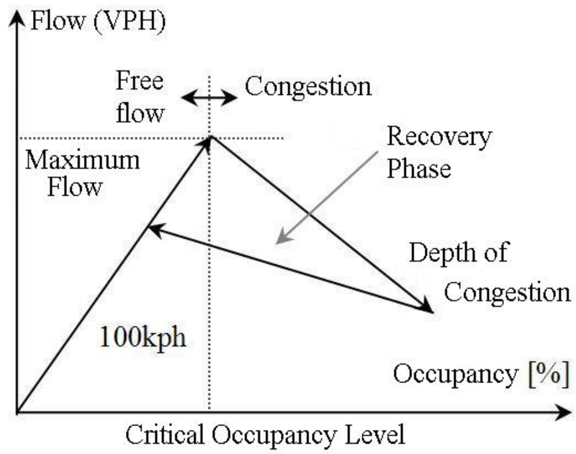

Fig. 1 Congestion model (3 phases)

The system PeMS (*Performance Measurement Systems*) was developed to determine the level of bands and sections utilization at different times, also to prevent the congestion. The system gets traffic data from detectors inside of street paving or section bands (Fig. 2), from half mile to half mile, and the reports at every 30 seconds of flow and solicitation level. The flow is the number of vehicles who pass the detector in 30 seconds, and the level of solicitation is the time spent by the vehicle in detector area. The PeMS processes the data from 4199 detectors from 1324 places in real time and gives the average values of flow and speed for 5 minutes.

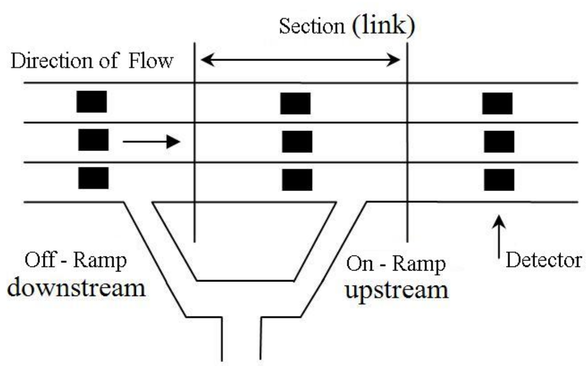

Fig. 2 Highway section

A consequence of measurements is the delay of vehicles on ramps at different times. The feedback control strategy called IMP (*Idealized ramp Metering control Policy*) allow the monitoring of solicitation level of bands through the deviation of flow on ramps every time this is over the critical predefined values. More details can be found in [6].

Consider that the vehicles running with 100 kmph when the flow is at maximum on a highway (when that area is most efficiently operated). So, the delay caused by the congestion is the additional time spent by the vehicle running with a speed under 100 kmph: a vehicle that needs 10 minutes for 10 km with 60 kmph, has a delay of 4 minutes.

The efficiency of any highway section can be defined as follows:

$$\eta = \frac{\alpha / 100}{\beta} \tag{1}$$

where α is the total number vehicle-km travel, and β is the total number vehicle-hour travel on a section at different times, for example, in the morning. The value of η overestimates the efficiency; we use this definition because it's easy to compute from input data.

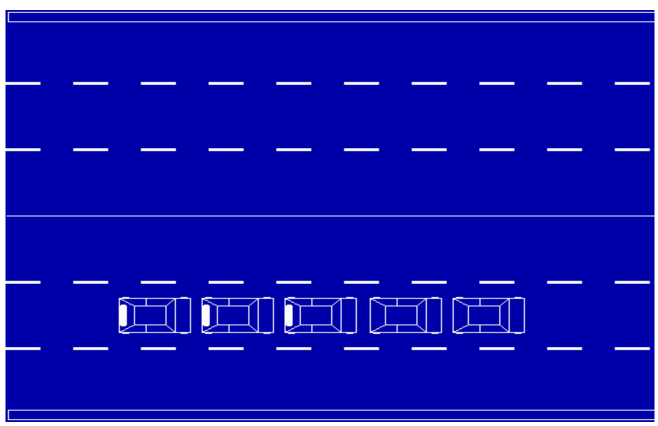

Fig.3 Platoon formation

The arguments suggest the Ideal Metering Principle (IMP): *If a metering policy keeps flow below its effective capacity in every link throughout a congestion episode, the speed will be maintained at 100 kmph, i.e. congestion will disappear. Any attempt to reduce it by increasing the metering rate will lead to congestion and an increase in the total delay.*

The model implies an Ideal Metering Principle, which keeps the vehicles in the back of an active ramp, so that the effective capacity of each highway section to be maintained at its critical level. The total traveling time using IMP (T_C) is the sum of traveling time (at 100% efficiency and 100 kmph) and the delay caused by ramps imposed by IMP.

$$T_C = \frac{\alpha}{100} + D_r \tag{2}$$

So the traveling time, saved by using IMP will be:

$$\beta_s = \beta - T_C = \beta - \frac{\beta}{100} - D_r \tag{3}$$

Because $\beta - \alpha / 100$ is the delay caused by congestion, results:

$$C_D = \beta_s + D_r \tag{4}$$

It can be observed that D_r can come from an excessive demand.

In contrast with the belief attributes congestion delays to the excess demand capacity, it can be noticed that the delay due to congestions consists of an important part that can be eliminated through IMP, and of a residual one which can be reduced only through the demand movement during rush hours. The demand can be shifted through other modes, or the periods, which are not rush hours [3].

The traffic observations give the command for the split in the platoon. In the first case chose the platoon in the many platoons.

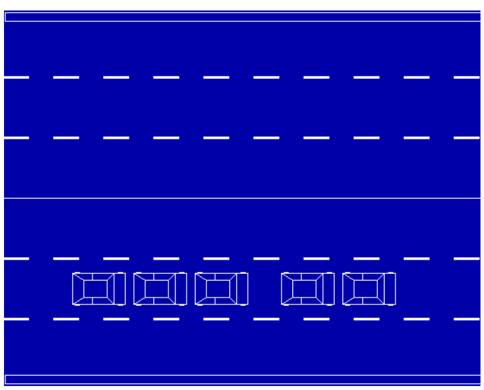

Fig..4. a. Split platoon command

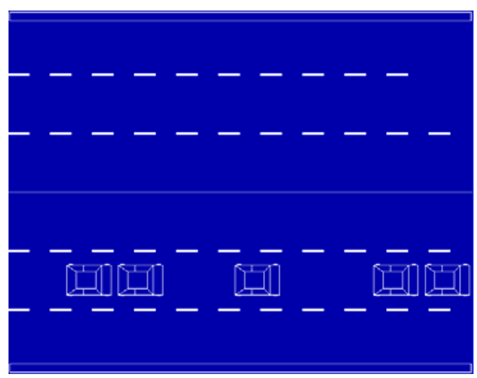

Fig. 4. b. Split platoon

A fundamental identity (5) of demand has three elements of traffic: L_v – length of vehicle, v – speed and φ - flow.

$$\rho = \frac{\varphi \cdot L_v}{v} \tag{5}$$

When the occupancy level reaches a critical level, congestion appears and the speed decreases. The critical occupancy level is specific for every section.

It can be demonstrate that:

$$\alpha = \varphi \cdot L_S \tag{6}$$

and

$$\beta = \frac{\alpha}{v} \tag{7}$$

where L_S represents the highway section length.

The ratio α/β is the section's productivity (during an hour). The maximum value of the result is:

$$\alpha_{\max} = \varphi_{\max} \cdot L_s \tag{8}$$

where φ_{max} is maximum flow noticed on the analyzed section in the study period at a speed of 100 kmph. Thus, the maximum productivity is of 100 kmph. It is defined the efficiency index of a highway section, the ratio between the actual level of productivity and the maximum level (9).

$$\eta = \frac{\alpha / \beta}{100} \tag{9}$$

The formula is also used to calculate the efficiency, not only for a section and a certain hour, but also for a whole network at any time (10):

$$\eta_{retea} = \frac{\dfrac{\sum_t \sum_k \alpha_k(t)}{\sum_t \sum_k \beta_k(t)}}{100} = \frac{\alpha_{retea} / \beta_{retea}}{100} \tag{10}$$

where *k* refers to all the network parts and *t* to all intervals of 5 minute.

The maximum flow noticed depends on the physical characteristics – the section degree and curve, the connection mode with other sections, the placement of start and stop ramps, etc. This depends on the traffic shape and the operational level of highway. In the standard interpretation, the section is shaped in insulation in order to obtain the theoretical maximum flow called *capacity* (which can be higher than the maximum flow observed). In order to prevent confusion, the empiric defined result is generally used, namely, the maximum flux [2].

In some situation, it can be used better the term (11):

$$\hat{\eta} = \frac{\varphi \cdot v}{\varphi_{\max} \cdot v_{\max}} \tag{11}$$

which gives $\hat{\eta} = 0{,}45$ instead 0,5.

The highway section can be modeled like a queue system, which will offer the clients (vehicles) customary service. Therefore, the serving time of the client will be (12):

$$T_S = \frac{L_S}{v} \tag{12}$$

The system serves vehicles in parallel that the demand will be ε:

$$\varepsilon = \frac{v}{L_S} \cdot \varphi \tag{13}$$

Its maximum level is:

$$\varepsilon_{\max} = \frac{100}{L_S} \cdot \varphi_{\max} \tag{14}$$

and it can be thus defined as the proportion between the current level and the maximum one.

3. Congestion on other types of network

Similar to the highway networks, the data transfer system or electrical energy transportation system is organized in network links with high capacity. Because these networks receive demands both from authorized and uncontrolled users, there is always the risk of congestion.

Exploring the similarities and the differences considering congestion for the three types of network, watching the experience regarding the size of demands, rutting and checking, we were able to notice:

- the patterns regard demands in transportation and communication systems present similarities: the user wants to move entities from a node to another (vehicles or data packets). In the electrical energy powered systems, the user enforces the load of some nodes powered by others nodes.
- in the transportation networks, IMP can control direct the demand. In data networks is an *admission control*: at the input of network, some flows or connections could not be admitted according with some criteria. The admission control is used also in the electrical energy powered systems, by accessing the resources reserved for emergency.
- the direction of data packets is controlled inside of routers placed in network's nodes. The energy flows governed by physical lows are generated by various shape of sources, respective receives through various demands. Therefore the routes of these flows cannot be directly controlled. Similar to the previous system, the routes chosen by drivers cannot be controlled. They prefer the routes with small transit time, but such a time in congestion depends on traffic flow, and this one depends on transit time. So, it appears a system of equation similar to the one in the case of energy powering systems governed by energy flows and phase angles. The differences in demand, routing and control could be in all three networks.

- a transmission link within the energy powering systems is in congestion if the energy surpasses the thermal capacity of the link. To prevent the fall of line, deviations towards other links are necessary; the only way to reduce the overcharge is to change the shape of energy generated or consumed. In the data networks, the nodes can be in congestion and in some situations, the data packets can be lost when the buffers capacity is surpassed.

4. Conclusions

The paper argues that a big part of congestion can exist due to an inefficient operation. One of the reasons for which the strategies to avoid congestion are not applied is the belief due to which the congestion is governed by demand and the ramp measurements can only shift the delay on ramps. The main contribution of our analysis is that the intelligent measurements shift only a portion of delay through ramps, the rest of delay is avoided. The implementation and simulation of the IMP feedback control strategy has proved this assertion. Other contributions proposed in this paper are: the modeling of the highway as a queue system, the emphasis of the interdependence between the traffic events expressed by messages for variety signalling and the proposed formula to determine the critical occupancy level specific for every highway section. Of course, the problem remains open, the final conclusions are away.

The transport systems, the supply energy systems and the data networks dare congestion. For data networks, with the increasing of demand increased the line capacity or the power of equipments. Until now, this was the strategy for transport system or the operators of supply networks. Today the option is to increase the efficiency throughout *intelligent control strategies.*

R E F E R E N C E S

[1] * Freeway Performance Measurement Project [online:http://transacct.eecs.berkeley.edu]

[2] *C. Chen, Z. Jia, P. Varaiya*, Causes and cures of Highway Congestion, IEEE Control Systems Magazine, pp. 26-32, December 2001

[3] *D.R. Drew, C.J. Keese*, Freeway level of service as influenced by volume and capacity, Highway Research Record, no.99, pp. 1-47, 1965

[4] *Z. Jia, P. Varaiya, C. Chen, K. Petty, A. Skabardonis*, Congestion, excess demand, and effective capacity in California freeway, 2000

[5] *S. H. Low, D.E. Lapsley*, Optimization flow control, I: basic algorithm and convergence, Transaction on Networking, vol. 7, no.6, pp. 861-875, 1999

[6] *M. Papageorgiou, A. Kotsialos*, Freeway ramp metering: on overview, Proc. IEEE Intelligent Transportation Systems Conference, Dearborn, MI, October, 2000