Traffic Management of the Congested Urban-suburban Arterial Roads

András Szele1*, Lajos Kisgyörgy2

1 Research Centre for Transport Development, Institute for Transport Sciences Non-profit Ltd., 1119 Budapest, Than Károly utca 3–5, Hungary
2 Department of Highway and Railway Engineering, Faculty of Civil Engineering, Budapest University of Technology and Economics, 1111 Budapest, Műegyetem rakpart 3., Hungary
* Corresponding author, e-mail: szele@kti.hu

Received: 17 January 2019, Accepted: 09 September 2019, Published online: 08 November 2019

Abstract
Managing the congestion is an old and legitimate claim of the cities. However, in spite of the promising solutions, the use of passenger cars outside the city centers did not decrease. The still hot topic requires a real solution.

The goal of the traffic planning and management was to serve the traffic demands but with the increase in traffic, infrastructure development could not keep up, so this goal - theoretically - is not valid anymore. We are in the midst of a paradigm shift: the old goals are obviously not working, but new goals and the tools needed to achieve them have not yet been developed.

According to the presented research, the operation of urban-suburban rush hour congestion radically differs from the way we have assumed so far, so a new framework was developed that describes this traffic operation. This framework allows us to look for new and valid management solutions and set new strategic and operational goals. In this paper, it is proposed not to build infrastructure for traffic demands, but to adapt the traffic demands to the infrastructure, by creating the optimal saturated traffic flow. This will prevent negative traffic dynamics and impacts on the network and provide higher quality and a more reliable road network operation. The research results have led us to theoretical questions such as the right of the limitation of the free routing on the road network.

Keywords
congestion theory, traffic management, congestion management, arterial road, traffic operation

1 Introduction
The most important information on congestion management is summarized in a 2007 OECD publication [1]. According to the paper, the elimination of congestion is not possible (because of the induced and latent demands, the growing motorization and the rerouting drivers looking for free capacities) and not necessary due to their traffic control effect, but it is worth trying to prevent the creation of excessive congestion. The most problematic congestion is the recurrent congestion. The paper defines recurrent congestion as a "consequence of factors that act regularly or periodically on the transport system". The rush-hour urban-suburban congestion is the most typical and most common type of recurrent congestion. To solve the problems caused by the congestion there are the following solutions [1]:

- Integrated transport planning takes a long time for the already developed urban structure to change towards a better structure, which will result a better transport operation, so it is rather a long-term tool. However, with the annual growth of 4–5% in motorization and traffic in developed countries, this long-term solution cannot compete. The usage of this tool is needed and unavoidable but won’t help in the cities in the near future.
- Improving the reliability of the urban road system performance mostly prevents the road system from the inherent and normal effects of the road transport like accidents and roadworks. These tools are important, but won’t help in the rush hours, because those times there are no free capacities to find a reroute for the drivers in the congestion created by the accidents or roadworks.
On the other hand, the control of urban freight transport is a real challenge. Most of the vans and trucks could deliver the goods out of the busiest periods, but they don’t. The whole urban freight transport could be organized as one system, which is a preferable but rare solution.

• **Locking in the benefits of available capacity** is the fanciest way of handling the congestion. At the same time, demand management is still in its infancy, and we hardly hear about successful solutions. Parking management, on the other hand, is the oldest solution, most of the big cities have been using for a long time, but it would be necessary to intervene much more radically to reduce congestion, which is politically risky.

The congestion pricing is the most known and most successful method. It has proven its efficiency against the congestions, but it has also been proven that the congestion charge sometimes does not reduce the traffic just the drivers avoid the protected area [2]. The introduction of the congestion charge in a city is a very complex and politically sensitive process. Despite the obvious success, no new congestion charge systems have been implemented in the last years.

The road pricing is not an ordinary tool against the congestion in the urban areas, because of its unpopularity [3]. Although there are toll systems in many places, its primary purpose is not usually to reduce congestion, except Singapore, which is the best example of a deliberate, well-functioning road and congestion pricing system, although the politically risky decision-making wasn’t a problem in that situation. Access management is basically used in two places: to protect (most commonly the historical) downtowns or to provide special connections like ramp metering. It is clear, that the application area is narrow today. There is much more opportunity in this tool and this paper also discusses a similar idea.

• **Congestion management measures release or provide new capacity.** Operational traffic management means improving operation through ITS [4, 5], providing road traffic information, traffic guidance, signal coordination, dynamic speed control, and using traffic management centers. These tools result a better traffic operation with higher efficiency, although the existence of the latent demand questions even these mostly low-cost capacity extensions. Promoting public transport is a really good idea, one of the most commonly used tools. Unfortunately, even this tool is just releasing some capacity, which will be used by the induced traffic. The mobility management is another fancy tool, involves ride sharing, using slow modes such walking and biking, trying to achieve a better modal shift by mobility management in companies, dematerializing transport by online and remote activities, and lastly changing travel behavior by education and campaigns. But again, using these tools itself is just results a temporarily releasing road network. Modifying the existing infrastructure and building new infrastructure are obviously a dead end. Most congestion could be temporarily eliminated or at least significantly reduced by some capacity release or extension if cities had the right amount of resources. However, this has not been a goal since latent demand would fill up the new capacities. The capacity extension, as a solution is quite common even today.

The congestion management tools aim to remove bottlenecks, to maintain the operation of the network, use of the existing capacities and manage demand [6–10]. The positive effects of the traffic management solutions tried to prove many times [4, 11]. In evaluating the results, it is worth taking into account that behind the results in congestion alleviation was almost always the better utilization of the existing capacities, which is ultimately capacity increasing. Longer time series about the operation of traffic management systems have not been investigated so far. It is strongly believed that the induced, growing, and rerouting traffic eventually found and used these new capacities. For the congested urban-suburban road network the inbound flow control [12] could be an effective solution. The essence of the tool is to keep the queues out of the city and let in only the amount of vehicles, which won’t cause congestion in the prevented zone. Experience is unknown, but such limitations can significantly reduce the extent of urban congestion by freeing capacity, hopefully not just temporarily. Currently, the most advanced solution to the congested signal-controlled road network is the SCOOT [13], which coordinates the traffic demands on the network. The system is based on the continuous traffic monitoring and synchronizes the traffic lights in a way that minimizes the loss of vehicles.

The most important comment is for all the listed and existing solutions that without serious limitation there is no way to have a better traffic operation. From this point of
view, these tools are just not radical enough, because they are trying to be politically safe. Unfortunately, with this act they lose all the profound impacts on the congestion.

If we are talking about the traffic management, the ultimate question is that there is any possible method to improve the quality of traffic operation in a widely congested state, which is more and more frequent, without straight or latent capacity increase?

The future of the traffic management tools is not the direct or indirect increase in capacity, but rather to tackle the traffic on the existing infrastructure by admitting its limitations, and mainly, the improvement of the quality of the congested traffic operation.

It is fundamentally to consider that the conventional traffic planning logic is still alive, so the goal of the planning process is still to serve the traffic demand, despite this goal has long been unreachable. Many cities have a dedicated goal of reducing road traffic, especially because the necessary capacities in urban areas cannot be provided. The old-fashioned design and management methods and the growing demands together paralyze the designers.

New methods are needed, and the new methods require new goals. Essentially, there is a need for a paradigm shift that aims achievable management goals and can reach them. Understanding and accepting the dynamics presented in the next section would greatly help to understand the impact of the interventions on a network level, so can base the desired paradigm shift.

2 The theory of the traffic operation of congested urban-suburban arterial roads

The results presented in this paper are based on a comprehensive research [14, 15] investigating the congested urban-suburban road networks consist an arterial road and the joining local roads complemented by the alternative routes used. On those networks, the more important intersections are typically signalized, but there are non-signalized nodes for the local flows, which can have significant traffic, especially during the congestion. The operation of metropolitan grid street networks should be further explored, because other phenomena may appear. The results led us to a new concept, which nomenclature and one of the most important phenomena can be seen in Fig. 1.

One of the most important parts of the theory is the key cross-section. It is generally one branch of an intersection, but it could be a signalized pedestrian crossing or a railway crossing, so it is better to call it a cross-section.

The low capacity of this cross-section causes the congestion. Many times, it may seem that there are more key cross-sections on the congested road network, but except the densest downtown street networks, there is usually only one key cross-section, which capacity is firstly exhausted. The first queues created here, and it is the starting point for the congestion of the arterial road, called primary congestion. This congestion then spreads to the crossing, weaving, merging flows, and the alternative routes, where the secondary congestions [14] developed. Finally, the condition of the traffic is transformed into a widely congested state involved the whole network. An important feature of the urban-suburban congested road network is that in the rush hour every arterial road is congested due to the lack of capacity of the key cross-sections [15].

The key cross-section is usually in a signalized intersection on the arterial road. Since the capacity is the lowest here, there is no congestion developing in the next intersection, so the exit is not limited. This cross-section cannot be avoided on the network, only by using another congested arterial road with another key cross-section. The exit flow of the key cross-section in the congestion is equal to the capacity of the key cross-section. The support of the entering flow, representing by the long queue, could and will fulfill all the free capacities in the key cross-section.

The second important point of the theory is the main flow (Fig. 2). This flow is the strongest, biggest one, usually going towards the city center in the morning, outbounds in the afternoon, or towards the tourist attraction on the weekend and during the vacation period using the arterial roads. Its prevalent presence defines all the traffic...
operation. It is an important detail that surprisingly big part of this flow tends to avoid the congested arterial road by using alternative routes [14] (see Figs. 1 and 2). These drivers, somewhere in the network, have to rejoin to the traffic of the arterial road to cross the key cross-section, creating high traffic volumes on local roads.

The drivers using alternative routes gain a significant time compared to the drivers remain on the arterial road. This traffic operation creates unfavorable traffic dynamics, which promotes the usage of the alternative routes, because the arterial road’s drivers are getting slower because of the cars (re)merging in the front of the queue.

Massively used alternative routes often lead the traffic into silent residential streets (causing environmental, safety, and social issues) and involves intersections built for much less traffic could become dangerous.

In earlier papers on the topic, the concept of the vitality of the traffic directions [15] was introduced, which describes the phenomenon when the traffic of the local roads significantly grows, at the expense of the traffic on the arterial road in the congested period. On a road network with recurrent congestion all traffic directions have an ability, which are determined by the location, design and regulations. These ability shows to what extent can that certain traffic direction get its traffic into the congestion-causing cross section. This ability named the vitality of a direction. To quantify the vitality proposed is the quotient of the traffic volume of the examined direction and the egressing traffic volume of the congestion-causing cross section (Fig. 3). The unit of measurement is the percent (%). The capacity of the congestion-causing cross section is finite. As a result, if the vitality of a direction increases than the vitality of another direction will decreases. If the vitality of the minor directions increases than the vitality of the main direction will decreases. The aggregated vitality of all minor direction is the reciprocal of the vitality of the main direction. It’s an assumption that the concept of vitality will play an important role in the description, understanding and designing of road networks with recurrent congestion.

High vitality means that the analyzed traffic flow grows significantly, when the arterial road is in a congested condition (Fig. 4).

The traffic operation of recurrent congestion can be shortly characterized as the low speed of the arterial road traffic is generated by the alternative road user drivers (re)merging close to the key cross-section. This phenomenon influences hardly the traffic operation, because usually the flow of the congested arterial road has the lowest vitality.

3 New goals for the traffic management of congested urban-suburban arterial roads

Congestion is an inevitable part of the urban-suburban areas, its elimination is not possible and not necessary due to their traffic control effect [1], since these congestions are keeping away the latent traffic demands. So the main goal is not to eliminate congestion, but to control, regulate, and limit it.
The self-destructive traffic dynamics of the congested road network described above have to be eliminated. Today it is quite common that the traffic flow on the arterial road is much slower than on the alternative routes. This traffic operation is proposed to be transformed. The aimed operation form is a relatively fast arterial road and definitely slower alternative routes. This also means that a large part of the waiting time during the congestion is moved from the arterial road to the local roads. However, it is necessary to prevent the creation of alternative routes, which undermine the favorable traffic operation dynamics. For the extra waiting times, drivers on the local roads can be compensated with better travel times on the arterial road.

The second strategic goal is to find new goals instead of to serve the traffic volume during the design process. The design goals were mostly focusing on the determination of traffic volumes. Sometimes certain features have been checked, like the LOS. Unfortunately, the traffic volume-based design method no longer valid in the congested network, so it is worth concentrating on new goals.

If we are looking for new goals for traffic operation, then it is worth reversing the today's design logic. By optimizing the network elements by not the traffic volume, but the quality of the traffic operation, it is possible to achieve goals that can effectively improve the traffic operation of the arterial road. What are these features? The key features of traffic are the traffic volume, the speed, and the vehicle density. The traditional design methodology takes into account only the traffic volumes, because the speed is given, or sometimes it is checked in the form of delays or waiting times. In the design process of the congested network it would be quite self-explanatory to design the speed of the traffic, as the low speed is the most characteristic feature of congestion. Doing this, the aim of the design is to provide a certain travel speed, and the results are the traffic volume and the vehicle density.

With one step forward, it could be better to achieve two goals at once. In addition to the speed of the traffic flow, the traffic volume could also be defined as a goal, for example by taking into account the outbound traffic volume in the key cross-section, which is determining the capacity of the network. These two factors together determine the desired speed of the traffic flow and the traffic volume, which can be flow at this speed. Since congestion is inevitable, it is advisable to design a traffic operation that is saturated but has relatively high speed. The main goal is not to maximize the capacity (although it is the most naturally goal), but to maintain a better traffic operation, close to the available capacity. In other words: not to have shorter queues but have relatively fast queues on the arterial roads.

Since the main goal is the optimal operation of the arterial road, therefore, traffic flow with such characteristics is called optimal saturated traffic flow.

Therefore, the design of the optimal saturated traffic flow on the arterial road can be the goal of the design, within the limits of the existing infrastructure, with special attention to the characteristics of congestion of urban-suburban arterial roads. The result is a controlled traffic flow: planned traffic volume on the arterial road and the local roads, and planned traffic speed on the arterial road, which could be higher than today, because the self-destructive traffic dynamics is eliminated.

4 Operational goals of creating optimal saturated traffic flow

The arterial roads have an important role in traffic operation: these capacitive roads are designed for high traffic volumes and environmental conflicts are less typical. These are the most important channels of public transport and urban freight transport. Based on these, the logical and only acceptable location of the main flow is here. So, the first and most important goal of the optimal saturated traffic flow is that the prevalent part of the main flow has to remain on the arterial road and should not look for alternative routes.

The second goal is to make the optimal saturated traffic flow on the arterial road as fast as possible but still in a congested condition. Keeping this condition has two goals: on the one hand, when the arterial road traffic is relatively fast, drivers are less likely to look for alternative routes and, on the other hand, faster traffic could defense themselves from the merging local flows which are tend to ruin the traffic operation. As a result, alternative road use decreases considerably, which means a better traffic operation and a more favorable traffic balance.

Protecting residential streets and handling the dangerous intersections due to increasing traffic is the next operational goal.

Support the crossing and weaving flows. In addition to the prevalent presence of the main flow on the congested urban-suburban road network, there are a number of traffic flows which are complicated by the congestion. These are typically the crossing and weaving flows, which suffer serious delays because of the congested arterial road. These traffic flows suffer from congestion and often unnecessarily waiting. So the last goal could be to
minimize the delays of these flows. The relatively high speed of the arterial road considerably reduces the delays of these traffic flows, by speeding up the weaving flows.

5 Creating the optimal saturated traffic flow

Creating the optimal saturated traffic flow is a network-level design task, typically a special design of a mostly signal-controlled road network. The process consists of five phases (Fig. 5).

In the first phase, it is needed to learn the network to be designed. To find the right answers, it is essential to have a detailed and thorough knowledge of traffic. The following information is required:

- Identification of the main flow in terms of traffic volume, period, vehicle composition, etc.,
- Identification of the key cross-section and its main features (traffic volumes and their change in time, signal phase, etc.),
- Features of the congestion on the arterial road: travel speeds and their changes in time, traffic disturbances occurring in the flow,
- Traffic volumes and waiting times of the local flows and their change in time,
- Alternative routes used, their traffic volumes, the joining intersections, and local flows with high vitality,
- The traffic volumes of the crossing and weaving flows on the examined network and their change in time.

Determining the data above is not easy. Identifying the main flow, in particular separating it from local and weaving flows or determining travel speeds of a queue requires new data collection techniques. Neither to define the limit of the examined road network is a simple task. But without this data, the planning process of the optimal saturated traffic flow cannot start.

The traffic design begins in the second phase. The most important step is to determine the travel speed of the planned congestion. According to the Fig. 4, the average travel speed on the arterial road is typically 12–15km/h. At this speed range, intensive alternative road use is observed, hence it is recommended to increase the travel speed range to 15–20 km/h from where a more favorable traffic operation can be observed. From the measured values in the phase one a number of vehicles at a period on the arterial road could be derived. To maintain the planned travel speed a certain number of vehicles could be on the arterial road and this number could change only in a very narrow limit. In addition, this number of vehicles may be different on different sections of the arterial road due to a number of factors (number of lanes, design speed, crossing or weaving flows, etc.). In this step, essentially, the traffic of the planned congestion is determined and the traffic condition in which the arterial road will work.

In the third phase, the capacity of the key cross-section plays the most important role. With the help of this, it could be calculated how many vehicles should enter the congested arterial road in a given period in order to maintain the planned vehicle number and travel speed but not to overload and slow down the system. It is worthwhile to give a rank in the form of minimum and maximum traffic flows because too many vehicles could create greater congestion and alternative route usage begins. On the other hand, in the case of too few vehicles, congestion can disappear, which could attract new traffic from elsewhere.

In the fourth phase, the traffic volume calculated in the previous step is distributed between the entry points feeding the congested arterial road. With this act, it can be provided more space for the main flow on the arterial road. Based on the Fig. 4 and other professional considerations, the traffic volume of the key cross-section during congestion should be predominantly, at least 50 % consist of the main flow versus the experienced 10–20 % ratio. For the local traffic demands should provide a shorter green time. It is important, that the traffic condition of the local roads can't be better than the condition on the arterial road. By doing so, it could be prevented the usage of alternative routes and the forming of traffic flows with high vitality. As a result of the process, the vehicles from the local roads will have less green time in the new traffic operation, so they will form longer queues in the local roads.
If the green times for the local roads could be more than the minimum for the local flows, it is not necessary to allocate the full green time between the traffic directions. More preferably, the gained green time could be dedicated to pedestrians or cyclists, to make a better transport quality. It is advisable to plan the optimal saturated traffic flow in the unit of the cycle time of the key cross-section, in order to adapt to the changes. This also gives the opportunity to take advantage of the synchronization possibilities inherent in the system.

The design starts in the key cross-section by distributing the entering traffic flow between the main flow and the local flows. The targeted minimum 50% participation rate of the arterial road flow in the traffic has to be given here, in the form of traffic volume per cycle time. The remained capacity of the intersections has to be distributed between the local roads and other transport modes. Planning process means a distribution of the capacity intersection by intersection.

In a more complex network, it may be possible to provide dynamically changing capacities in time and space within the framework of the optimum traffic flow of the arterial road, for example using a system like SCOOT. It can be the topic of further research.

In the fifth design phase, the design results are checked and fine-tuned. In particular, checking the adequacy of the following factors is important:

• The travel speed of the arterial road on the whole regulated section has to be around 15–20 km/h,
• The travel speeds of the alternative routes have to be lower than the speed of the arterial road,
• The waiting times on the local roads have to be longer than before the implementation of the new traffic control.

Fine-tuning could focus on the handling of the crossing and weaving flows avoid the key cross-section. High-level serving these needs is important because it could significantly reduce loss done by congestion.

The key to the system's ability to function is limiting the access to the congested arterial road, so it is important not to allow for any flows to have high vitality, thus creating an alternative route. The issue of non-signalized intersections must be handled; the system cannot operate if there are unlimited access points to the arterial road.

It is also important to enforce the rules. It is desirable to set up automatic control methods for sanctioning run the red traffic light, due to the longer waiting times on the local roads, which, for the first time, could be annoying for the drivers. Communication is a key point here. Local drivers have to know that the longer red lights are for reducing traffic in their area and after joining to the arterial road, they could gain a lot from the new operation form.

Further measures may also be required to ensure optimal saturated traffic flow. Thus, for example, the capacity of the key cross-section should be maintained during congestion, bulk lane changes or other disturbances should not reduce the volume of the outbound traffic. As a last step, it is recommended to synchronize the signalization of the intersections along the arterial road, according to the demand of the arterial road in the optimal saturated condition.

Particular attention should be paid for the period of the creation of the congestion and the way how the traffic of the arterial road changes to the optimal saturated traffic flow. One of the most important tasks is to define the change of the traffic distribution during the transition.

6 The threshold of creating the optimal saturated traffic flow

The concept of excessive congestion [1] determines the intervention threshold of the congested road section or network. The congestion is excessive when "the marginal costs to society of congestion exceed the marginal costs of efforts to reduce congestion".

By calculating the cost and the benefits of a congestion management system like congestion charge in Stockholm and London, [17–19] it is clearly can be seen the effectiveness of the implementations. These calculations based on the concept of the excessive congestion and take into account all the externalities. However, in other cases, this theoretical approach is in practice hardly could be used. The main reason behind this, that the benefits of tackle congestion are generated on a social level, doesn't result real money. But the money spent on congestion management (the costs) is the part of the city budget. Consequently, in practice, the money spent on managing congestion is compared by decision-makers to the direct, easy-to-measure, visible, and generally much lower benefit. In this approach, most of the congestion is no longer perceived excessive and only minimal interventions are cost-effective. The surprisingly low level of efforts and resources for managing congestion in reality justified this approach. The bigger the city, the bigger the chance is that all the costs and benefits will be taken into account during the decision making. However, if the concept of excessive congestion cannot be used in practice, then there is no theoretical basis for the intervention threshold.
In our view, it is worth changing every urban-suburban, recurrently congested arterial road by using the concept of the optimal saturated traffic flow. In this way, the self-destructive, negative dynamics of the rush hour congestion could be eliminated. Most road networks are unlimitedly accessed by the vehicle. If there is free capacity, drivers will look for and will find it. Consequently, it is only reasonable to handle the congestion if it is possible to keep away the rerouted traffic flows. That means, to maintain a high-quality traffic operation on the arterial road, it is needed to limit the access to it. Therefore, there is a good chance that the access to the arterial road itself should also be limited at the first suitable signalized intersection or to solve this problem it is needed to plan the whole network together. The optimal saturated traffic flow could work on its own only when it is ensured that no new traffic from other arterial roads comes. However, in this condition, it is very worthwhile to use this tool.

If we could create optimal saturated traffic flow on all arterial roads in a city, probably the road transport would be much better; cities are likely to become interoperable and easily accessible. Looking further, planning the whole city road network together could create a territorial optimal saturated traffic flow. This question should be further researched. These questions involve some philosophical problem of the network design, such as the right of the free routing on the road networks. The basic principle of today's network planning is that (almost) all traffic flow is justified. The introduced topic deeply affects the operation of the autonomous vehicles in the near future [20].

Looking ahead a bit far, in the era of autonomous vehicles, the right of the limitation of the routing may become an essential problem.

7 Conclusions
Currently, there are no good solutions to the problems of the congested road networks. Taking advantage of the potential inherent in existing tools needs much more political courage. The lack of popularity of the necessary limitations is a serious obstacle to the introduction of any working solution. This is not expected to change in the long run, so any new approach and opportunity may be important. The objectives outlined in the paper and the presented tools are essentially for facilitating the introduction of the necessary limitations. This fact has a particular importance because the motorization boom that is currently in progress is likely to predict a more serious traffic situation in the future.

The results in this article are based on a wide-ranging congestion research that investigated the traffic operation of the congested road networks.

It is necessary to introduce a working methodology, but there are a lot of unclear theoretically questions and a lot of ambitious decisions. Traffic planning is just in a paradigm shift. To accelerate and ground this shift, the introduction of new strategic goals is recommended as:

- The self-destructive traffic dynamics of the congested road network have to be eliminated,
- Should set new goals instead of serving the traffic volume during the design process,
- The aimed operation form is a relatively fast arterial road and definitely slower alternative routes,
- The aim of the design should be to provide a certain speed.

These strategic goals together target the creation of the optimal saturated traffic flow.

New operational goals are needed for the new strategic goals. The followings are recommended:

- The main flow has to remain on the arterial road and should not look for alternative routes,
- The optimal saturated traffic flow on the arterial road has to be as fast as possible but still in a congested condition,
- The delays of the crossing and weaving flows, which suffer serious delays in the intersections because of the congested arterial road, should be minimized.

After the goals, the design process is presented. The design objectives and the methodology presented ultimately describe a politically acceptable form of the limitation of the free access to public roads. As further research, it is recommended to develop the methodology applied to an arterial road and to investigate the territorial optimal saturated traffic flow.

References


