Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Part C 16 (2008) 535–553

TRANSPORTATION
RESEARCH
PART C

www.elsevier.com/locate/trc

Real-time merging traffic control with applications to toll plaza and work zone management

M. Papageorgiou *, I. Papamichail, A.D. Spiliopoulou, A.F. Lentzakis

Dynamic Systems and Simulation Laboratory, Technical University of Crete, 73100 Chania, Greece

Abstract

There is a number of highway transportation infrastructures where traffic flow arriving on a higher number of lanes must merge into a lower number of lanes within a limited space, e.g., merging highways, merging on-ramps, toll plazas, work zones. When the arriving traffic flow exceeds the downstream capacity, congestion is formed, and the outflow is reduced due to the congestion-induced capacity drop. A general conceptual framework for real-time merging traffic control is developed in the paper, aiming at throughput maximization or, equivalently, delay minimization. To this end, an algorithm known from local ramp metering operations is employed (ALINEA). The proposed framework is applied to new and innovative traffic control applications for real-time toll plaza and work zone management. Microscopic simulation results demonstrate the potentially achievable benefits. Further significant applications of the developed concept are outlined.

Keywords: Merging traffic control; ALINEA; Toll plaza management; Work zone management; Bottleneck control

1. Introduction

There is a number of highway transportation infrastructure facilities that call for merging of traffic flow from a higher number of lanes into a lower number of lanes within a limited space, including:

(i) Merging of two highways or freeways.
(ii) Merging of on-ramps into the freeway mainstream.
(iii) Merging within a toll plaza infrastructure.
(iv) Merging of freeway mainstream lanes due to road works.

When the total arriving merging flow is lower than the flow capacity of the downstream reduced-lane highway, the merging efficiency is usually satisfactory and there is no need for external intervention. If the total arriving flow reaches or exceeds the downstream capacity, congestion is created in the merging area and the exiting flow is reduced below the downstream capacity due to the capacity-drop phenomenon. Under these

* Corresponding author. Tel.: +30 28210 37289; fax: +30 28210 37584.
E-mail address: markos@dssl.tuc.gr (M. Papageorgiou).

0968-090X/$ - see front matter
doi:10.1016/j.trc.2007.11.002

conditions, real-time control of the arriving flow streams may be employed to guarantee a maximum exit flow from the merge area. This is indeed practiced in ramp metering or freeway-to-freeway control installations (Papageorgiou and Kotsialos, 2002) where a part of the arriving traffic flow is controlled in order to maximize the merge area throughput.

This paper proposes a general conceptual framework for real-time merging traffic control, whereby metering may be applied to all or a part of the arriving streams, via suitable modification and use of ramp-metering like algorithms, in particular ALINEA (Papageorgiou et al., 1991, 1997). The proposed framework is then applied to new and innovative traffic control applications, e.g., for real-time toll plaza management or real-time work zone management on freeways. Section 2 presents the details and options of the general real-time merging traffic control framework, while Sections 3 and 4 present specific instances of the general framework for toll plaza and work zone management, respectively, along with microscopic simulation results that demonstrate the efficiency of the proposed approach. Section 5 summarizes the conclusions and outlines a further, potentially most significant application of the general concept to combined on-ramp/mainline metering.

2. Real-time merging traffic control framework

2.1. Structure and elements

A general real-time merging traffic control system (Fig. 1) comprises a number of general elements, each of which may take a number of different forms or options:

(i) Arriving flow and queuing infrastructure (total of M lanes).
(ii) Exiting flow infrastructure (μ lanes, $\mu < M$).
(iii) Merge area.
(iv) Control devices.
(v) Real-time measurements or estimates.
(vi) Control algorithm.

Each of these elements will be discussed in the following with regard to its characteristics and possible options.

2.2. Arriving flow and queuing infrastructure

The arriving traffic flow approaches the merge area on a total number M of lanes which may belong to one single highway (as, e.g., in work zones) or to more than one highway (as, e.g., at freeway on-ramp merges or highway convergence areas).

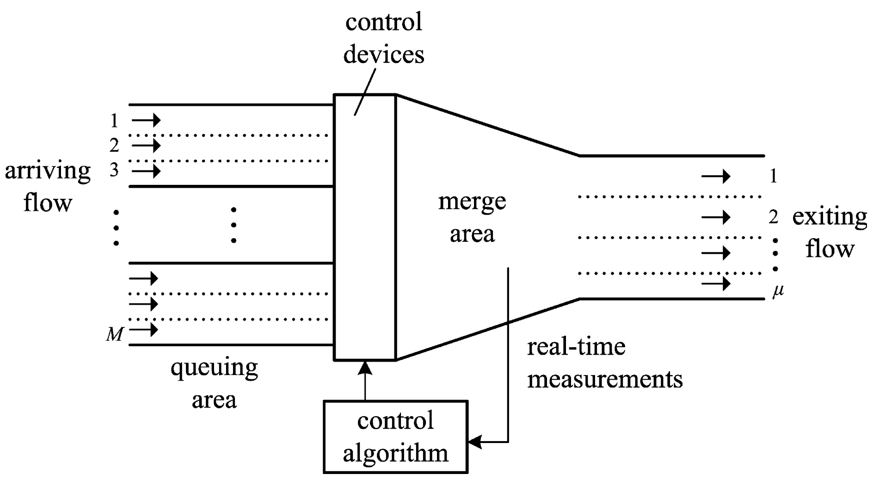

Fig. 1. A general real-time merging traffic control system framework with M entering and μ exiting lanes, $\mu < M$.

If the total arriving traffic flow occasionally exceeds the downstream infrastructure capacity (see Section 2.3), then queuing is inevitable. Within the queuing area (Fig. 1), lane changing may be:

- Fully allowed, in which case it is reasonable to assume that the individual lane queues will tend to be balanced, provided the individual lane flows (and hence the queue waiting times) are similar to each other.
- Not allowed, in which case individual lanes may have queues (and waiting times) that are different from each other.
- Partly allowed, e.g., in the case of traffic arriving from more than one highways; vehicles may be allowed to change lanes within each arriving highway but cannot switch to lanes of another highway. In this case, queues (and waiting times) may be different on the lanes of each arriving highway.

Restrictions may be present with regard to the maximum admissible queue on some lanes. For example, ramp metering is usually not allowed to create over-long ramp queues that would interfere with the adjacent street traffic. Note that generalized restrictions for all M lanes may not be feasible if the arriving traffic flow is sufficiently high.

2.3. Exiting flow infrastructure

Although a most general assumption would be to allow for the possibility of more than one highways also at the exit of the merge area, this may lead to some additional problems that would need to be addressed; since we cannot think of any real merging infrastructure of this kind, we restrict our framework to one single exit highway with μ lanes and corresponding flow capacity q_{cap}. Obviously, we must have $\mu < M$ for merging. Moreover, it is assumed that the exit flow capacity is (occasionally) lower than the maximum total arriving flow; otherwise merging traffic control is not needed.

2.4. Merge area

The vehicles arriving from a total of M lanes, change lanes appropriately within the merge area in order to fit into the μ lanes of the exit. The merging procedure may be quite complex, and drivers adapt their lane-changing behavior to the specific requirements of each kind of merging infrastructure. For example, merging conditions may be “symmetric” (if all vehicles have the same lane-changing rights and opportunities) or (more or less) “asymmetric”. “Asymmetries” may arise, e.g., because:

- Some arriving lanes have priority over others, e.g., freeway mainstream lane vehicles have priority over on-ramp vehicles (or vice versa in some infrastructures as the Paris Boulevard Périphérique); nevertheless, it is quite typical that some mainstream vehicles change from the right-most lane to other mainstream lanes in order to avoid potential conflicts with vehicles merging from the on-ramp.
- Some arriving lanes may prolong (explicitly or implicitly) into exit lanes, hence vehicles on these lanes may not need to be actively involved in lane-changing maneuvers.

Asymmetries may also characterize the infrastructure geometry; while toll plaza merging areas are typically symmetric (trapezoidal, see Fig. 3), freeway on-ramp merging or work zone merging areas are usually geometrically asymmetric. Further differences concern the size (surface) of the merge area in different infrastructure types, e.g., toll plaza merge areas are typically bigger than freeway work zone merge areas.

The variability and complexity of diverse merging infrastructure characteristics that are reflected in correspondingly diverse driver lane-changing behaviors, is a challenge for microscopic models and simulators. Indeed, the lane-changing component is a notorious weak point of microscopic simulators when applied to various merging traffic infrastructures (Kondyli et al., 2007).

Fortunately, the complexity and variability of microscopic merging behaviors does not reflect on the real-time merging control actions envisaged here, because the basic assumptions that justify a common real-time control approach for any merging infrastructure are of a macroscopic nature and are related to the well-known notion of a fundamental diagram. Fig. 2 displays a typical flow-density diagram for the merge area

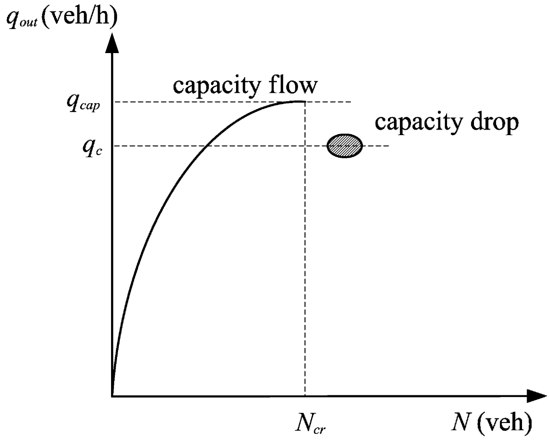

Fig. 2. Fundamental diagram of a merging area.

where the flow q_{out} is the exit flow and N is the number of vehicles included in the merge area (or in the downstream part of the merge area, see Section 2.6). When N is small, merging conflicts are scarce and swift, while the exit flow is correspondingly low. As N increases, merging conflicts may increase, but q_{out} increases as well until, for a specific value N_{cr}, the exit flow reaches the downstream capacity q_{cap}. If N increases beyond N_{cr}, merging conflicts become more serious, leading to substantial vehicle decelerations and eventual accelerations that reduce the exit flow to lower values q_{c}, where $q_{\mathrm{cap}} - q_{\mathrm{c}}$ is the capacity drop due to congestion.

The phenomenon of the capacity drop has been repeatedly observed and reported, mostly for freeway/on-ramp merge areas, see, e.g., Cassidy and Rudjanakanoknad, 2005. More specifically, it was observed that traffic flow in merge areas may increase up to some capacity value q_{cap} at relatively high mean speeds; if the total arriving flow continues to increase beyond this value, a traffic breakdown occurs with dropping speeds and increased occupancy. The measured flow q_{c} downstream of the formed congestion head (congestion discharge rate) was then found to be lower than q_{cap} by some 5–20% in different investigations. It should be noted that the precise reasons behind (and factors affecting the extent of) the capacity drop are not well understood, see, e.g., Treiber et al., 2006 for a potential explanation.

In our view, the capacity drop appears as a result of vehicle acceleration at the congestion head, i.e., from lower speeds within the congestion to higher speeds downstream of the congestion head. More specifically, individual vehicle accelerations are quite different from each other (due to different driver attitudes or different vehicle capabilities, e.g., trucks, etc.); as a consequence, "over-long" gaps are likely to be created in front of low-acceleration vehicles due to faster accelerating vehicles ahead; "over-long" in the sense that the same vehicles with the same speed would apply shorter gaps under stationary (non-accelerating) conditions. Filling of the over-long gaps via lane-changing from other lanes on multilane highways is unlikely; first, because lane-changing of vehicles is less likely in the acceleration mode; and, second, because over-long gaps are more probable on the slow lanes that are not a target for overtaking vehicles. In conclusion, longer gaps (lower density) at the same speed corresponds to lower flows during acceleration than at stationary conditions, and this may indeed be a reason for the observed capacity drop. If this view is correct, then the level of the capacity drop may depend on the speed gradient (upstream vs. downstream of the congestion head). Moreover, this view suggests that a capacity drop should be observable even at bottlenecks without lane drop, e.g., in front of tunnels, etc., and there are some indications in the literature that this is actually happening.

Real data analysis from freeway on-ramp merge areas revealed that traffic breakdown may occur at different flow values q_{cap} on different days, even under similar environmental conditions (Elefteriadou et al., 1995; Lorenz and Elefteriadou, 2001; Cassidy and Rudjanakanoknad, 2005). Naturally, flow capacity differences become even more pronounced in case of adverse environmental conditions (Keen et al., 1986). In contrast, the critical occupancy (at which capacity flow occurs) was found to be fairly stable (Cassidy and Rudjanakanoknad, 2005) even under adverse environmental conditions (Keen et al., 1986; Papageorgiou et al., 2006). These findings suggest the introduction of a control scheme that would act on the arriving traffic flow in order to maintain (whenever necessary) the number of vehicles N in the merge area close to its critical value N_{cr}, thus maximizing the exit flow q_{out}, i.e., optimizing the merge infrastructure efficiency. This is a usual ramp metering practice, e.g., by use of the local ramp metering strategy ALINEA (Papageorgiou et al., 1991, 1997).

It should be emphasized that the phenomenon of capacity drop is crucial for throughput increase via the proposed merging traffic control measures. If no capacity drop appears in a merging traffic infrastructure, the only justification for applying merging traffic control might be the potentially improved traffic safety due to more orderly vehicle merging; improved traffic safety was indeed observed in some ramp metering installations, see, e.g., Papageorgiou et al., 1997.

2.5. Control devices

There are several control devices that may be used to regulate the flows arriving in the merge area. Each control device may apply to one individual lane or to a subset of lanes:

- Traffic lights are the most popular devices to regulate arriving flows. Traffic lights are used for ramp metering, but also in some toll plaza management systems, e.g., at the San Francisco-Oakland Bay Bridge, California. Traffic lights may be applied to individual lanes or groups of lanes.
- Physical moving barriers may be used for the control of individual lanes, particularly at toll plaza booths.
- Variable speed limits may be used to regulate the mainstream traffic flow in freeway on-ramp merge areas. This may be achieved via ordinary road-side variable message signs or via emerging dual vehicle-to-infrastructure communication systems that act directly on individual vehicle speeds.

Whenever the existing infrastructure allows, the control devices (e.g., traffic lights) should be positioned sufficiently far from the actual merging area in order to allow for vehicles to accelerate sufficiently for efficient merging.

It should be emphasized that the control algorithm for merging traffic control (Section 2.7) calls for implementation of a specific flow q_i for each controlled entering lane i, and the control devices should be operated appropriately to this end. Section 3 presents a possible way of achieving this by use of traffic lights. On-going research work considers the appropriate application of variable speed limits or emerging dual vehicle-to-infrastructure systems.

In some cases, control devices may be installed only on a part of the entering lanes (e.g., in the case of ramp metering) while traffic flow on other lanes is allowed to enter the merge area freely (e.g., by-pass lanes for HOV, buses, emergency vehicles). In other cases, different control devices may be applied to different groups of entering lanes, e.g., traffic lights at the on-ramp and variable speed limits for the mainstream of a freeway on-ramp merge area.

2.6. Real-time measurements or estimates

Application of feedback control in order to maintain the number of vehicles N close to N_{cr} (Section 2.4) calls for availability of real-time measurements or estimates of N. Although this quantity may be directly measurable by use of video sensors, this possibility may be costly or difficult, particularly if the merge area is extended, as in toll plazas. An alternative possibility would be the real-time estimation of N by use of a limited number of ordinary loop detectors (Vigos et al., 2008).

Yet another possibility would be to consider the easily measurable occupancy o, rather than the number of vehicles N, as the variable under control. As a matter of fact, the diagram of Fig. 2 is valid for occupancy as well (and occupancy rather than N is typically used for ramp metering operation). This raises the question on the specific location where the occupancy should be measured. The occupancy measurement should best be placed at or just upstream from the location where serious vehicle decelerations (congestion) appear first. This is because congestion, once occurred, propagates upstream, thus the measurements feeding the control algorithm should be placed:

- not downstream of the location where congestion appears first, otherwise the congestion is not visible and the control goal ($o \approx o_{cr}$) may not be achievable;

- as close upstream from that location as possible, in order to minimize delays in the triggered control reaction.

Usually, merge congestions in freeway on-ramp merge areas appear first a few hundred meters downstream of the ramp nose, which is, in most cases, downstream of the acceleration lane drop. This is probably due to the fact that drivers are very attentive while merging from the acceleration lane onto the mainstream, thus tolerating relatively short inter-vehicle gaps at high speeds (which corresponds to accordingly high flows); after the acceleration lane drop, when drivers relax and attempt to restore their usual gaps, a traffic breakdown first occurs. In contrast, microscopic simulators create the merge congestion typically at the acceleration lane-drop location.

Placing the measurement detectors at or upstream from the typical location of first congestion appearance is crucial in order to enable the system to monitor any occurring merge congestion; related misunderstandings may lead to misapplication and control failure (Papageorgiou et al., 2007).

In conclusion, the functional merge area for pertinent merging traffic control may be different than the physical merge area. More specifically, the functional merge area for control may extend downstream to cover the location of first congestion appearance, which, in some cases, may be few hundred meters within the downstream highway, as in freeway ramp metering merges.

2.7. Control algorithm

The control algorithm receives the real-time measurements or estimates of N or o and drives the control devices in order to maintain $N \approx N_{cr}$ or $o \approx o_{cr}$ which maximizes the merge area exit flow. This is the major control goal, but there may be alternative and secondary control goals as discussed in this section. The overall control algorithm may be decomposed into three distinct parts, each with individual duties:

(a) Feedback control for exit flow regulation (ALINEA).
(b) Distribution of entering flows.
(c) Translation of control decisions.

to be discussed in the following.

2.7.1. Feedback control

This is the central task of the control algorithm that is essentially identical for any kind of merge infrastructure. Feedback control is activated at each time interval T, whose value may be selected within the range [20, 60 s]. More specifically, at the end of each running period T, time-averaged measurements of occupancy o from the ending period (or the latest measurements or estimates of vehicle-number N) are used to calculate the entering flow to be implemented (via the control devices) in the next period in the aim of maintaining $o \approx o_{cr}$ (or $N \approx N_{cr}$). This may be achieved by use of the well-known integral-feedback (I-type) ALINEA regulator (Papageorgiou et al., 1991, 1996)

$$q(k) = q(k-1) + K_R[\hat{o} - o(k-1)], \tag{1}$$

where $k = 1, 2, \ldots$ is the discrete time index; $q(k)$ is the entering flow (veh/h) to be implemented during the new period k; $o(k-1)$ is the last measured occupancy (%) averaged over all exit lanes; $K_R > 0$ is a regulator parameter; $\hat{o}$ is the set (desired) value for the occupancy which may be set equal to o_{cr} for maximum exit flow.

The regulator (1) apparently integrates (accumulates) the regulation error $\hat{o} - o(k-1)$ from time-step to time-step. As a consequence, the flow $q(k)$ resulting from (1) could increase (or decrease) indefinitely under certain circumstances due to the well-known wind-up effect of I-type regulators that may be illustrated as follows. If $o < \hat{o}$ holds for a long period (e.g., during off-peak), then the regulation error $\hat{o} - o(k-1)$ in (1) is persistently positive and drives $q(k)$ to accordingly high values, while the actual inflow q_{in} is much lower due to lack of demand. When o eventually exceeds $\hat{o}$ (e.g., at the start of the peak hour), the regulation error becomes negative, and (1) will gradually reduce q, starting from the highest value Q that it had reached during the period of positive regulation errors; but an actual reduction of the entering flow is only achieved when q becomes smaller than the current inflow q_{in}, and this may take many time-steps, whose number is obviously larger for larger Q (and lower K_R in (1)); during this transient period, o may continue to increase and an overshooting (i.e., a persisting negative regulation error $\hat{o} - o(k-1)$) is created, whose magnitude and duration increase with increasing Q, due to the corresponding increase of the transient period.

Thus, in order to avoid this wind-up effect, the flow $q(k)$ resulting from (1) is truncated if it takes values outside a pre-specified range $[q_{min}, q_{max}]$ and the truncated value is used in (1) as $q(k-1)$ in the next time step. The range $[q_{min}, q_{max}]$ should be sufficiently wide to enable flexible control around the usual operation value $q(k) \approx q_{cap}$ without frequent bound-hitting during the peak hours. On the other hand:

- q_{max} should not be too large in order to limit the duration of the aforementioned transient period and the magnitude of the resulting overshooting.
- q_{min} should not be too small to avoid (temporarily) strong strangulations of the entering traffic flow.

The usual control goal in case of local control is outflow maximization (i.e., $q_{out} \approx q_{in} \approx q_{cap}$) which is achieved by the choice of $\hat{o} = o_{cr}$ (or $\hat{N} = N_{cr}$) in (1). It should be noted that occupancy, rather than flow, measurements are used to feed the regulator (1) because, as Fig. 2 suggests, measured flows do not uniquely characterize the prevailing traffic state, since the same flow value may reflect either free or congested traffic conditions. In addition, according to the discussion of Section 2.4, capacity values may differ from day to day; hence any selected target $\hat{q}_{cap}$ would either underload (on days where the actual capacity happens to be higher than $\hat{q}_{cap}$) or overload (on days where the actual capacity happens to be lower than $\hat{q}_{cap}$) the infrastructure.

In contrast, the critical occupancy o_{cr} (or N_{cr}), at which capacity flow occurs, appears to be more stable (Section 2.4) and may be selected as a target, i.e., $\hat{o} = o_{cr}$. The precise value of o_{cr} (or N_{cr}) may depend on the specific infrastructure and the measuring devices and can be identified experimentally. Alternatively, one may employ a real-time estimator for these quantities, e.g., as proposed in Smaragdis et al. (2004) and Kosmatopoulos et al. (2006).

The algorithm (1) leads to a stable closed-loop behavior (Papageorgiou et al., 1991; Kosmatopoulos and Papageorgiou, 2003) for a broad range of positive K_R values. In the steady-state (stationary conditions), we have in (1) $q(k) = q(k-1)$ and hence $o(k-1) = \hat{o}$, i.e., exact regulation. Due to the integral-feedback character of (1), stationary accuracy is achieved even in case of uncontrolled entering lanes or by-pass lanes for buses or HOV, etc. Stationary accuracy is also guaranteed if there are errors in the implementation of the calculated q (e.g., due to traffic light operation inaccuracies, red-light violations), see Papageorgiou et al., 1991; Kotsialos et al., 2006.

It is interesting to note that the aforementioned wind-up effect could also be circumvented if the last calculated (and truncated) flow value $q(k-1)$ in (1) would be replaced by the real measured inflow $q_{in}(k-1)$. It may be shown, however, that the stationary regulator accuracy cannot be guaranteed in this case if there are errors in the implementation of the calculated q. Moreover, some small but unnecessary metering delays may appear during light traffic if the inflow values q_{in} vary strongly over time, see Papageorgiou et al., 1991; Kotsialos et al., 2006. On the other hand, replacing $q(k-1)$ in (1) by $q_{in}(k-1)$ would utterly suppress the transient period and the related overshooting.

The control results are little sensitive to the specific value of K_R within a broad range of values. As a rule-of-thumb, K_R in (1) should be higher than 25μ veh/h/%. Extremely high values of K_R may lead to unstable closed-loop behavior.

In some cases, the merging traffic control problem may be a component of a more comprehensive network-wide control scheme that prescribes specific outflows $\hat{q}_{out}$ to the local control components. In this case, if the desired flow set-point is sufficiently smaller than capacity flow (e.g., $\hat{q}_{out} \leqslant 0.9q_{cap}$), then the problem of ambiguous characterization of the traffic state is not an issue, and hence one may use the following flow-based regulator (FL-ALINEA; see Smaragdis and Papageorgiou, 2003) instead of (1)

$$q(k) = q(k-1) + K_F[\hat{q}_{out} - q_{out}(k-1)] \tag{2}$$

with $K_F = 1$.

In some cases, a portion of the M arriving lanes may not be controllable, e.g., the mainstream lanes in ramp metering or some by-pass lanes for HOV or buses. In these cases, q in (1) or (2) reflects the flow on the controllable arriving lanes only, while the uncontrollable flows act as disturbances that do not need to be taken into account in the control algorithm thanks to its feedback character.

2.7.2. Distribution of entering flow

The feedback regulators (1) or (2) deliver, at each time step T, the total flow values q to be implemented at the exit of the controllable entering lanes. How should this total flow be distributed among individual (or among groups of) controllable entering lanes? This question cannot be answered in a general way as it may depend on various characteristics of the merging infrastructure and the pursued control goals. Here are some examples:

(i) Consider a toll plaza or work zone infrastructure where lane-changing is possible in the queuing area (Section 2.2); an equal distribution of q among the entering lanes seems reasonable, as arriving drivers are likely to join the shorter queues, thus leading to a ''spontaneous'' equalization of the individual lane queues.
(ii) Consider a toll plaza or work zone infrastructure where lane-changing is not allowed or not possible in the queuing area (e.g., due to physical separation); it may then be reasonable to actively distribute q among the entering lanes in a way that equalizes the lane queues or the corresponding waiting times. This calls for a corresponding feedback control by use of real-time measurements or estimates of individual lane queue lengths.
(iii) Consider a freeway on-ramp merge area where the on-ramp flow is controlled via traffic lights (ramp metering) while the mainstream flow is controlled via variable speed limits. The waiting queues in the ramp are not allowed to exceed an upper limit (to avoid interference with the adjacent street traffic). Under these conditions there may be a variety of different policies on how to distribute q among the mainstream and the on-ramp, taking into account also the limitation of the ramp queue length. Clearly, each potential policy must be implemented via a suitable distribution algorithm.

2.7.3. Translation of control decisions

At the end of parts (a) and (b) (Sections 2.7.1 and 2.7.2, respectively) at each time step T, each controllable entering lane (or each group of lanes with one common control device) has an assigned subflow q_i that needs to be implemented via appropriate activation of the corresponding control device. The specifics of this translation of the control decisions into control device actions cannot be specified in a general form. Even for the same control device, there may be different possible operational policies; e.g., for traffic lights we may have one-car-per green, n-cars-per-green, full traffic cycle, discrete release rates, etc., see Papageorgiou and Papamichail, 2008.

This completes the presentation of the general framework for real-time merging traffic control. Two instances of the framework, one for real-time toll plaza management and another for real-time work zone management will be presented in the following. More detailed results for both examples are reported in corresponding dedicated papers (Spiliopoulou et al., 2008; Lentzakis et al., 2008).

3. Real-time toll plaza management

3.1. Infrastructure description

Toll plaza management aims at smooth, safe and efficient servicing of vehicles in the related facilities. Proposed management measures for toll plazas address lane configuration, signing, markings, geometric design, toll collection technologies (FHWA, 2006). In some rare cases, traffic lights may be also installed (e.g., in the toll plaza of the San Francisco-Oakland Bay Bridge, CA). In this section, a simulated case study will be presented in order to demonstrate the potential benefits of real-time merging traffic control for maximum throughput.

Toll plazas contain usually a higher number of toll-paying booths than of exiting lanes. Thus, vehicles exiting the booths need to merge before reaching the exiting highway lanes. During peak hours, the rate of booth-exiting vehicles may be higher than the downstream infrastructure capacity q_{cap}, in which case a congestion may form that reduces the exit flow below q_{cap} (capacity drop). Based on the reasoning presented in Section 2, it may be worthwhile to apply merging traffic control in order to maximize the toll plaza throughput, or, equivalently, minimize the average vehicle delays, or, equivalently, minimize the upstream extent of the formed queue.

The particular infrastructure considered in order to demonstrate, via microscopic simulation, the potential of real-time toll plaza management, mimics the Bay Bridge (California) toll plaza with $M = 15$ entering (booth) lanes and is sketched in Fig. 3. Traffic arrives on 15 lanes and reaches, after 275 m, the booth area (30 m) which is not explicitly modeled (no booth delays) because it is of little relevance in the present context. After the booth area there is a trapezoidal merging area followed by the exit highway with 5 lanes and capacity $q_{cap} \approx 10{,}500$ veh/h. Lane changing is possible anywhere except in the booth area.

The specific arrival demand scenario considered is stochastic; the average demand is increased in small steps, starting with 3000 veh/h (in an empty system) and reaching 12,000 veh/h (which exceeds the downstream capacity q_{cap}) within the first 10 min; during the next 10 min, the average demand is maintained constant at this high value (peak hour); during time $t \in [20,\ 30\ \text{min}]$, the demand is reduced stepwise back to 3000 veh/h and then drops to zero until the end of the investigation (40 min) in order to have an empty system again. This scenario includes a total of about 4400 veh served.

The feedback algorithm (1) is used for merging traffic control, albeit by use of real-time measurements or estimates of the number N of vehicles included in the trapezoidal merge area (Fig. 3) rather than occupancy o. The regulator is activated every $T = 30$ s, and the required bounds are set $q_{min} = 4500$ veh/h and $q_{max} = 13{,}000$ veh/h. The flow value q delivered by (1) is eventually distributed evenly among the 15 entering lanes by use of 15 traffic lights (one per lane) that are placed at the exit of the booth area, just upstream from the merge area (Fig. 3). Each traffic light has a constant green phase of 4 s that allows in average some two vehicles to pass. A signal cycle c (in s) consists of a (constant) green phase and a (variable) red phase and allows some two vehicles to pass. Thus, the implementation of a flow q (veh/h) calls for a traffic cycle (equal for all lanes) that satisfies

$$q = (2M/c)3600 \Rightarrow c = 7200M/q. \tag{3}$$

The red phase is not allowed to be lower than 2 s, hence there is a minimum cycle length $c_{min} = 6$ s and c resulting from (3) is truncated if it is lower than c_{min}. Note that c_{min} is never reached here since $q_{max} = 13{,}000$ veh/h delivers $c = 8.3$ s in (3).

To avoid bursty entering flows that may appear if all 15 lanes assume their green and red phases simultaneously, the 15 lanes are subdivided in three groups of five lanes each, namely $\{1,4,7,10,13\}$, $\{2,5,8,11,14\}$, $\{3,6,9,12,15\}$, respectively, and a shift (offset) is introduced for the cycle start of each group relative to the cycles of the other groups.

Note that the regulator time step T is not an integer-multiple of the cycle c in general, which creates a synchronization problem that is solved as follows. As soon as $q(k)$ is delivered by the feedback algorithm, the corresponding cycle $c(k)$ is calculated via (3). This cycle is held for application at the end of the last (running)

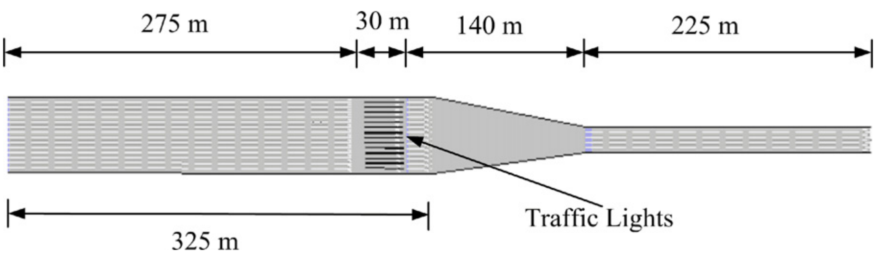

Fig. 3. A toll plaza infrastructure mimicking the Bay Bridge (California) toll plaza.

cycle $c(k-1)$ of the previous period $k-1$. The resulting implementation error is deemed minor, particularly in view of the automatic disturbance-rejection property of (1) (see Section 2.7.1).

The described infrastructure was simulated by use of the microscopic simulator AIMSUN NG Professional Edition 5.0.1 using the simulator's default parameters and a simulation time step of 0.75 s. The trapezoidal merging area (Fig. 3) is modeled as a "junction", and this is actually a first reason for using measurements of N rather than of occupancy, as AIMSUN does not allow the placement of loop detectors inside "junctions". On the other hand, this allows us to test the feedback algorithm (1) with this measurement option as well.

The simulated merging control software is interfaced with the simulator via the AIMSUN API (Application Programming Interface) that allows us to emulate a real-time closed-loop operation. More specifically, the simulator delivers every T the current number N of vehicles in the junction, based on which the control software calculates the corresponding traffic light settings and returns them to the simulator for application.

The merging behavior of vehicles is a well-known weak point of microscopic simulators (Kondyli et al., 2007), particularly in the case of somewhat unusual merging infrastructures. Therefore, the produced simulation results should be understood as a demonstration of potential benefits achievable via merging traffic control, rather than as a quantitatively reliable assessment of achievable benefits. It should be emphasized that the macroscopic fundamental diagram "emerges" from the microscopic simulation of individual vehicle movements and is not impose in any way in our simulations.

Since AIMSUN is a stochastic simulator, different runs with different seeds may produce quite different results. To address this issue, the usual practice is to run several simulations, each with different seed, for each simulated scenario. This practice is followed in the reported investigations with 10 replications per scenario. The main evaluation criterion used is the average vehicle delay (AVD) (in s/veh/km) that is delivered by AIMSUN. Note that vehicle delay expresses the actual travel time minus the (minimum possible) travel time that would be needed for a vehicle to travel a distance at free speed. The AVD expresses the average delay-per-kilometer of a vehicle, taking into account all vehicles that have completed their trip during the simulation.

3.2. Simulation results

In the no-control case, the arriving vehicles continue their trip in the merge area and exit without noteworthy problems as long as the arriving demand is lower than the exit capacity q_{cap}. When the arriving demand approaches q_{cap}, serious vehicle merging conflicts with strong vehicle decelerations are observed that lead to the formation of a congestion close to the exit of the merge area; the congestion spills back onto the 15 entering lanes without ever reaching the simulated system entrance. The resulting mean AVD (in s/veh/km) for 10 replications is 164 with a standard deviation of 15.5, and highest/lowest AVD values of 190/137, respectively.

Fig. 4a and b display the trajectories of the outflow q_{out} and of the number of vehicles N, respectively, for one particular simulation run with an individual AVD = 168 s/veh/km close to the mean AVD of the corresponding 10 replications. The outflow q_{out} is roughly following the arriving demand increase until about $t = 10$ min where it shortly approaches the capacity q_{cap}. The vehicle number N is increasing slowly (due to

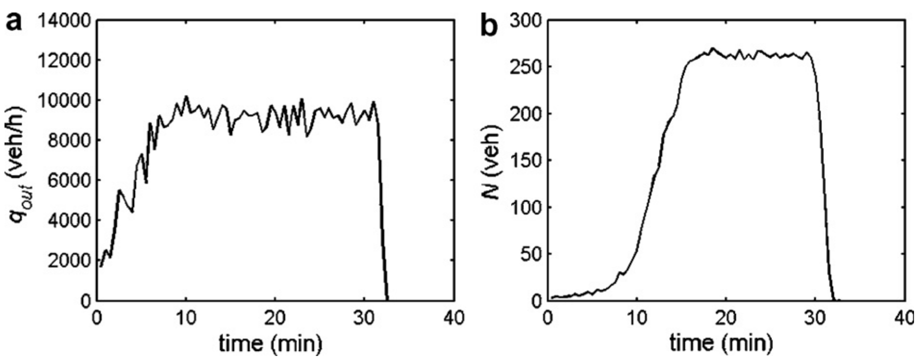

Fig. 4. Outflow q_{out} (a) and vehicle-number N (b) in the no-control case for the toll plaza example.

increasing flow) during the same period. After $t = 10$ min, the vehicle accumulation N in the merge area has a steep increase (due to the formed congestion) and stabilizes eventually at a value around 260 veh; at that state, the merge area is full of vehicles due to congestion. Soon after the first appearance of the congestion (e.g., after $t = 10$ min) the outflow reduces (capacity drop) to values around 9500 veh/h. When the demand is decreased and the formed queue is served, the system is emptied (at around $t = 32$ min).

When merging traffic control is applied, the traffic lights are operated with a short red phase (corresponding to $q_{\max} = 13{,}000$ veh/h) for as long as N in the merge area is lower than the set value $\hat{N}$, according to (1). Note that some slight delays may occur to some vehicles that encounter the red phase, but there are no real queues in front of the traffic lights and the system outflow is mainly determined by the arriving demand (no real metering). The slight delays could be avoided if the traffic lights are switched on only when needed (i.e., when N approaches $\hat{N}$). As the arriving demand increases, N increases, and, when $N(k) > \hat{N}$, the regulator (1) starts its actual entering flow control operation aiming at maintaining $N(k)$ near $\hat{N}$. At this time, a queue is formed just upstream from the traffic lights which increases, without ever reaching the simulated system entrance.

As mentioned earlier, the ultimate goal of merging traffic control is outflow maximization which leads to minimization of the average delay and minimization of the upstream queue extent. Outflow maximization is achieved via an appropriate specification of $\hat{N}$ in (1). To investigate this issue, a series of simulation runs were carried out with different $\hat{N}$-values and 10 replications for each such value; the utilized K_R value in (1) is set equal to 500 h^{-1}. Fig. 5 displays some of the obtained results in terms of the AVD; more specifically Fig. 5 displays, for each investigated $\hat{N}$-value in the range [16, 50], the AVDs of the 10 simulation replications and the corresponding mean AVD value. For comparison, Fig. 5 also displays (horizontal lines) the mean, highest and lowest AVD values of the 10 replications of the no-control case.

Fig. 5 gives rise to the following observations:

- For $\hat{N}$ values around 20 veh, the mean AVD is minimized, apparently due to outflow maximization. This corresponds to the N_{cr}-value mentioned earlier.
- For $\hat{N}$ lower than $N_{cr} = 20$ veh, the mean AVD increases because the downstream highway “starves for flow”, i.e., the resulting q_{out} is (unnecessarily) restricted to values lower than q_{cap}, and hence, the downstream highway capacity is underutilized.

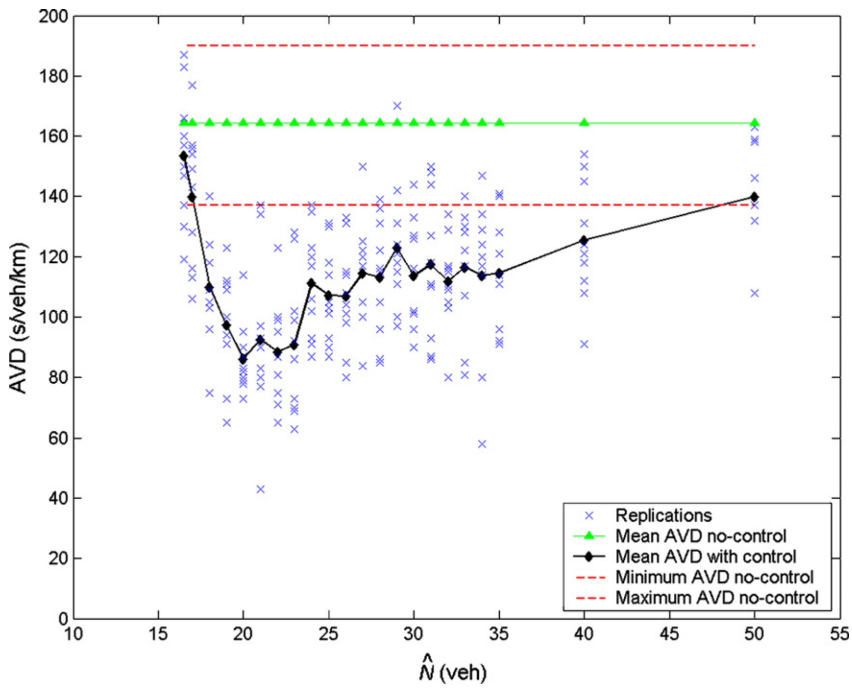

Fig. 5. Average vehicle delay versus $\hat{N}$ with and without merging traffic control for the toll plaza example.

- For $\hat{N}$ higher than $N_{cr} = 20$ veh, the mean AVD increases due to lower q_{out}, but the latter has now a different background than for $\hat{N} < N_{cr}$: the higher $\hat{N}$ (beyond N_{cr}), the more frequent the appearance of vehicle merging conflicts that lead to strong decelerations and reduced outflows.
- Although the mean AVD in the control case is clearly lower than the mean AVD of the no-control case for all $\hat{N} \in [16, 50]$, individual control replications may be worse than individual no-control replications. Merely in the range $\hat{N} \in [18, 26]$, there is virtually no control replication providing worse results than the best no-control replication.
- The mean AVD for $\hat{N} = N_{cr}$ is 90 s/veh/km, which is 45% lower than the mean AVD of the no-control case.

Fig. 6a and b display the trajectories of the outflow q_{out} and of the vehicle-number N, respectively, for one particular simulation run with $\hat{N} = 20$ veh and an individual AVD = 90 s/veh/km close to the mean AVD of the corresponding 10 replications. The outflow q_{out} is roughly following the arriving demand increase until about $t = 8$ min; at this time, N is seen to exceed $\hat{N}$ (Fig. 6b), and the regulator (1) is activated; for reasons mentioned in Section 2.7.1, N exhibits an overshooting (up to $N \approx 35$ veh), followed by a short undershooting, before being stabilized by the regulator action to values around 20 veh. The relatively strong N-variations during the transient period $t \in [8, 12$ min$]$ reduce temporarily the outflow q_{out}; in contrast, q_{out} is seen during the stationary control period $t \in [13, 26$ min$]$ to attain values around 10,500 veh/h, i.e., around the downstream highway capacity q_{cap}. Comparing Figs. 4 and 6, merging traffic control increases the toll plaza outflow by some 1000 veh/h or 10.5% compared to the no-control case during the stationary period. As a result, the formed queue for the system under control is emptied at around $t = 27$ min, i.e., some 5 min earlier than in the no-control case.

The reported merging traffic control simulation results were produced with a value $K_R = 500$ h^{-1} in (1). It is important to investigate the sensitivity of the control results with respect to the value of K_R, since high sensitivity may imply a high effort of fine-tuning in field applications. Fig. 7 displays the mean AVD values resulting from merging traffic control with several $\hat{N}$ values, against K_R within the range [100, 800]. Note that each point in Fig. 7 corresponds to a specific couple of utilized K_R, $\hat{N}$ values, and that 10 simulation replications were carried out for each such couple.

Fig. 7 shows that the control results (in terms of the mean AVD) for all considered $\hat{N}$-values are virtually identical for $K_R \geqslant 300$. For $K_R < 300$, the initial overshooting/undershooting observed in Fig. 6b become more pronounced (for reasons detailed in Section 2.7.1), leading to temporarily lower outflows q_{out} during the transient period and hence increased mean AVD. In contrast, the control behavior for $K_R < 300$ is similar to the one displayed in Fig. 6 after the transient period (i.e., in the stationary phase). Thus, if the transient period duration (some 5 min) is negligible compared to the whole peak period duration, then K_R values of 100 or 200 are virtually equally efficient as higher values. On the other hand, it may be shown that the closed-loop system becomes unstable beyond some (very high) value of K_R.

In conclusion, there is a broad range of K_R values, for which the merging traffic control behavior is virtually equally efficient, and hence the need for fine-tuning K_R in field applications is expected to be minor.

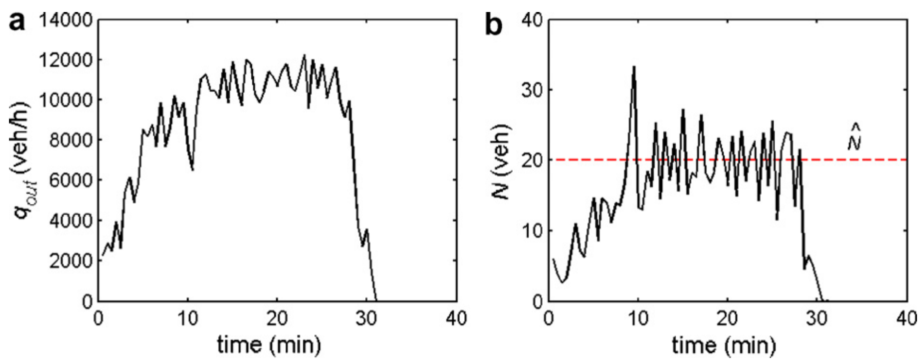

Fig. 6. Outflow q_{out} (a) and vehicle-number N (b) in the merging control case with $\hat{N} = 20$ veh, $K_R = 500$ h^{-1} for the toll plaza example.

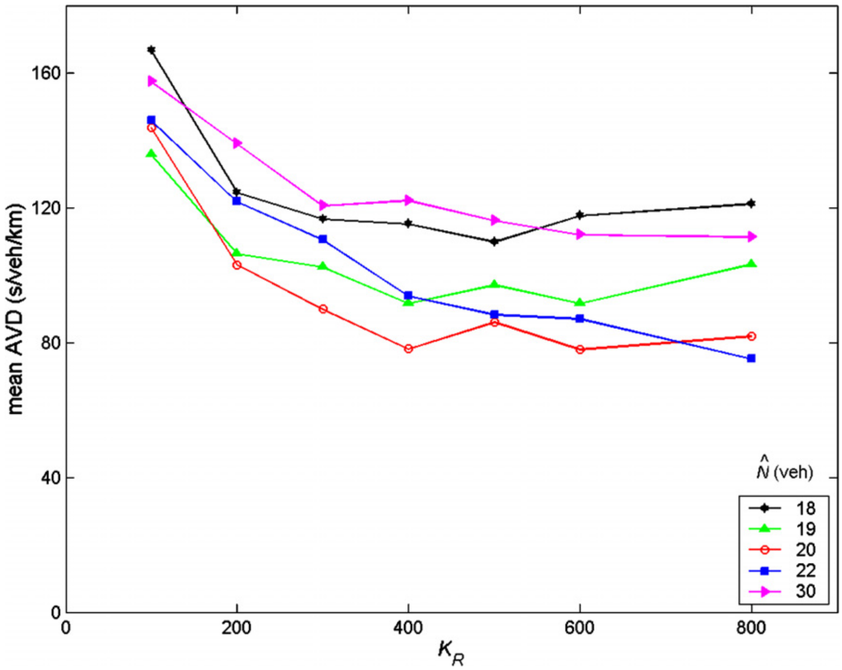

Fig. 7. Mean AVD versus K_R for different set-values $\hat{N}$ for the toll plaza example.

More details on toll plaza merging traffic control and related simulation results may be found in Spiliopoulou (2007) and Spiliopoulou et al. (2008).

4. Real-time work zone management

4.1. Infrastructure description

Work zones on freeways necessitate the drop of one or more lanes at the work zone entrance which may create traffic flow disruptions and congestion upstream from the work zone due to reduced capacity. Work zone management aims at smooth, safe and efficient passage of vehicles, with particular focus on the work zone entrance where vehicles need to merge into fewer lanes. Proposed management measures for work zones address speed limitations as well as signing, markings, geometric design (FHWA, 2005). In this section, a simulated case study of a hypothetical work zone will be presented in order to demonstrate the potential benefits of real-time merging traffic control for maximum throughput.

Vehicles approaching the work zone need to merge into a lower number of lanes. If the rate of arriving vehicles exceeds the capacity q_{cap} of the remaining lanes in the work zone, a congestion will form and the entering flow will be reduced below q_{cap} (capacity drop). Based on the reasoning of Section 2, it may be worthwhile to apply merging traffic control in order to maximize the work zone throughput. Clearly, a higher throughput through the work zone is associated with lower vehicle delays and reduced congestion extent upstream from the work zone entrance.

The particular hypothetical infrastructure considered in order to demonstrate, via microscopic simulation, the potential of real-time work zone management, has $M = 3$ entering and $\mu = 1$ leaving lanes as sketched in Fig. 8. The total length of simulated freeway is 750 m while the trapezoidal merging area is 20 m long, followed by the work-zone restricted one lane freeway portion with a capacity $q_{cap} \approx 2300$ veh/h.

The specific arrival demand scenario is stochastic; the average demand is increased in small steps, starting at level zero (in an empty freeway) and reaching 2500 veh/h (which exceeds q_{cap}) within the first 10 min; during the next 10 min, the demand is maintained constant at this high value; while, in the third 10-min period, the

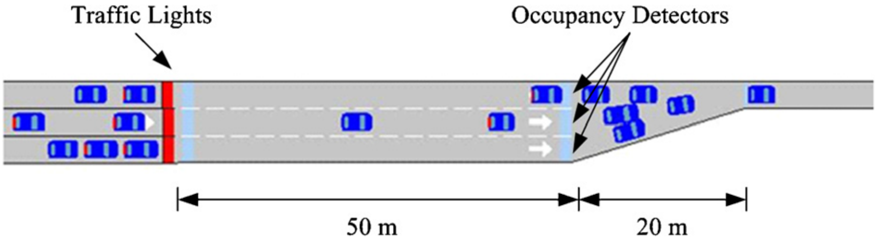

Fig. 8. A hypothetical work zone entrance.

demand decreases stepwise back to zero and is maintained there until the end of the investigation (40 min) in order to have an empty system again. This scenario includes a total of about 830 veh passing.

The feedback algorithm (1) is used for merging traffic control with $T = 30$ s, $q_{\min} = 1000$ veh/h, $q_{\max} = 3000$ veh/h. The feedback algorithm is fed with occupancy measurements collected just upstream from the trapezoidal merge area (Fig. 8); note that possible congestion due to merging conflicts starts within the merging area in this simulated environment; thus, any appearing congestion moves upstream and is captured soon by the emulated loop detectors. In a real environment, it would be probably better to place the loop detectors within the merging area (upstream from the typical location of first congestion appearance) for faster congestion capture and corresponding reaction; however, the merging area was modeled here as a "junction", and AIMSUN does not allow for the placement of loop detectors within "junctions".

The flow value q delivered by (1) is eventually distributed evenly among the three entering lanes by use of three respective traffic lights that are placed 50 m upstream from the merge area (Fig. 8). The traffic lights are operated as described in Section 3.1 with a constant green phase of 4 s and a variable red phase. A shift (offset) is applied to the signal cycle start of each traffic light relative to the cycles of the other traffic lights in order to minimize simultaneous vehicle departures. The use of traffic lights in work zones was recently proposed by Wei and Pavithran (2006), albeit with fixed signal settings rather than the real-time control considered here.

The real-time work zone management example considered assumes that all upstream lanes are signal-controlled for reasons of equity and safety. In case of partial metering (e.g., metering of the merging lanes only), traffic safety could be affected due to unbalanced upstream queues motivating drivers to frequent lane changes towards the free lanes; while drivers preferring to stay in their (metered) lanes would suffer higher delays than drivers on free lanes. Anyhow, as mentioned earlier, the presence of uncontrolled lanes (e.g., for buses, HOV, etc.) does not pose any problem to the presented method.

Simulations are carried out with AIMSUN NG Professional Edition 5.1.1 as described in Section 3.1, with 10 simulation replications for each investigated scenario, and the average vehicle delay (AVD, in s/veh/km) is used as the main evaluation criterion.

4.2. Simulation results

In the no-control case, arriving vehicles continue their trip in the merge area and the work zone area without noteworthy problems as long as the arriving demand is lower than q_{cap}. When the arriving demand approaches q_{cap}, serious vehicle merging conflicts lead to the formation of a congestion in the merge area; the congestion spills back onto the three entering lanes without ever reaching the simulated system entrance. The congestion is captured by the occupancy detectors when it reaches the corresponding location (Fig. 8). The resulting mean AVD (in s/veh/km) for 10 replications is 217 with a standard deviation of 38.8 and highest/lowest AVD values of 269/158, respectively.

Fig. 9a and b display the trajectories of the merge area outflow q_{out} and occupancy o, respectively, for one particular simulation run with AVD = 217 s/veh/km (i.e., equal to the mean AVD of 10 replications). The outflow q_{out} is roughly following the arriving demand increase until about $t = 10$ min where it shortly approaches the capacity q_{cap}. The occupancy is increasing slowly (due to increasing flow) during the same period. After $t = 10$ min, the occupancy is seen to increase sharply (due to the formed congestion) and stabilizes eventually at a value around 85%. The sharp occupancy increase is accompanied by an outflow decrease

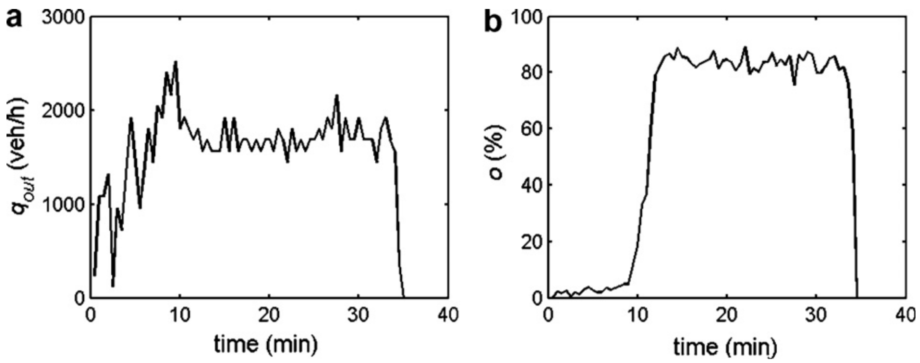

Fig. 9. Merge area outflow q_{out} (a) and occupancy o (b) in the no-control case for the work zone example.

(capacity drop) to values around 1800 veh/h. When the demand is decreased and the congestion is dissolved, the system is emptied (at around $t = 35$ min).

When merging traffic control is applied, the traffic lights are operated with a short red phase (corresponding to $q_{\max} = 3000$ veh/h) for as long as o is lower than the set value $\hat{o}$, according to (1). As the arriving demand increases, o increases, and, when $o > \hat{o}$, the regulator (1) starts its actual operation aiming at maintaining $o(k)$ near $\hat{o}$; this leads to a queue upstream from the traffic lights which increases, without ever reaching the simulated system entrance.

To determine the optimal set-point $\hat{o}$ for throughput maximization, a series of simulation runs were carried out with different $\hat{o}$-values and 10 replications for each value; the utilized K_{R} value in (1) is set equal to 100 veh/h/%. Fig. 10 displays the AVDs of the 10 replications and the mean AVD for each investigated $\hat{o} \in [5\%, 10\%]$; along with the mean, highest and lowest AVD values of the 10 replications for the no-control case (horizontal lines).

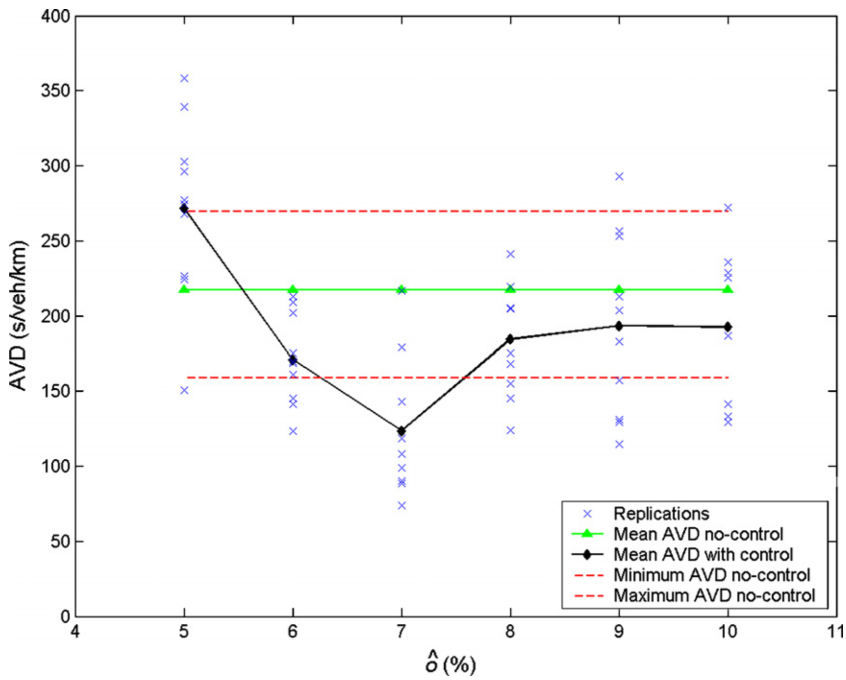

Fig. 10. Average vehicle delay versus $\hat{o}$ with and without merging traffic control for the work zone example.

It may be seen in Fig. 10 that the mean AVD under control is minimized for $\hat{o} = 7\%$. At first view, this value may appear very low for a critical occupancy; recall, however, that o is measured on three lanes (Fig. 8) to maximize the flow of one single lane (downstream). The mean AVD is lower than its no-control counterpart for all $\hat{o}$ except $\hat{o} = 5\%$; the mean AVD for $\hat{o} = 7\%$ is 123 s/veh/km which is 43% lower than the mean AVD of the no-control case. However, in contrast to Section 3.2, there are some control replications with a higher AVD than the best no-control AVD, even for the optimal set-point $\hat{o} = 7\%$.

Fig. 11a and b display the trajectories of q_{out} and o, respectively, for one particular simulation run with $\hat{o} = 7\%$ and AVD = 118 s/veh/km which is close to the mean AVD of the corresponding 10 replications. The outflow q_{out} is roughly following the arriving demand increase until about $t = 9$ min, at which time o exceeds $\hat{o}$ and the regulator (1) is activated; for reasons mentioned in Section 2.7.1, o exhibits a strong overshooting (reaching 20%), followed by an undershooting.

In contrast to Section 3.2 (Fig. 6b), the period after the initial overshooting is characterized by occasional, randomly appearing, significant departures of the occupancy under control to values that may reach 30% or more. The observed occupancy spikes are due to occasionally appearing (simulated) vehicle merging conflicts that lead to provisionally strong vehicle decelerations in the merge area, i.e., small congestions that are captured by the occupancy detectors; this triggers a corresponding regulator reaction that drives the occupancy back to values close to $\hat{o}$, albeit with a corresponding undershooting. Thus, the regular merging control period (after the transient period) comprises a number of short-lived occupancy spikes and eventual undershootings, during which the outflow is seriously reduced. Between the spikes, there are subperiods of proper regulation, with $o(k) \approx \hat{o}$ and $q_{\text{out}} \approx q_{\text{cap}} \approx 2300$ veh/h.

It is interesting to note that the manifest probability of occurrence of the spikes increases in our simulations with increasing $\hat{o}$ beyond 7%. In contrast, for $\hat{o} = 6\%$ there are virtually no spikes (other than the initial overshooting) observed in the corresponding 10 simulation replications, i.e., the occupancy is maintained around 6% while $q_{\text{out}} \approx 2100$ veh/h, i.e., lower than capacity. However, the average outflow for $\hat{o} = 7\%$ is apparently higher than $q_{\text{out}} \approx 2100$ veh/h of the case $\hat{o} = 6\%$, hence the higher mean AVD of the latter in Fig. 10.

It is difficult to judge whether the observed occupancy spikes are due to the geometry of the particular merge area or due to the typically limited accuracy of the vehicle-merging behavior by the microscopic simulator (this is what we suspect) or both. Related field data are probably the most reliable means to answer this question. In any case, even in the presence of occupancy spikes, the merge area outflow is increased in average with merging control and $\hat{o} = 7\%$, leading to the mentioned mean AVD decrease of 43% against the no-control case, while the formed queue in the control case ($\hat{o} = 7\%$) is emptied at around $t = 31$ min, i.e., 4 min earlier than in the no-control case.

Fig. 12 displays the mean AVDs for merging traffic control with $\hat{o}$ equal to 6% and 7%, against K_R within the range $[50, 300]$. It may be seen that AVD is improved for $K_R > 100$ in the case $\hat{o} = 6\%$, due to better handling of the initial overshooting as explained in Section 3.2 (Fig. 7). For $\hat{o} = 7\%$, best results are obtained in the range $K_R \in [100, 200]$; for $K_R < 100$, the control performance deteriorates due to slower handling of the occurring spikes by the regulator; for $K_R > 200$, the control performance deteriorates due to the regulator's overreaction to the occurring occupancy spikes.

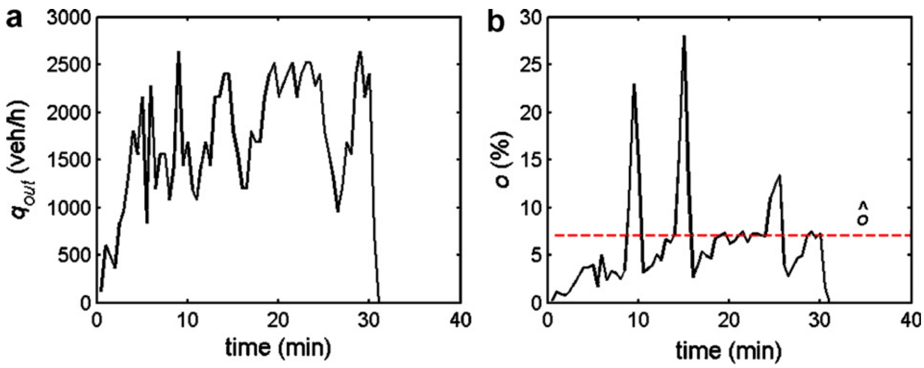

Fig. 11. Outflow q_{out} (a) and occupancy o (b) in the merging control case with $\hat{o} = 7\%$, $K_R = 100$ veh/h/% for the work zone example.

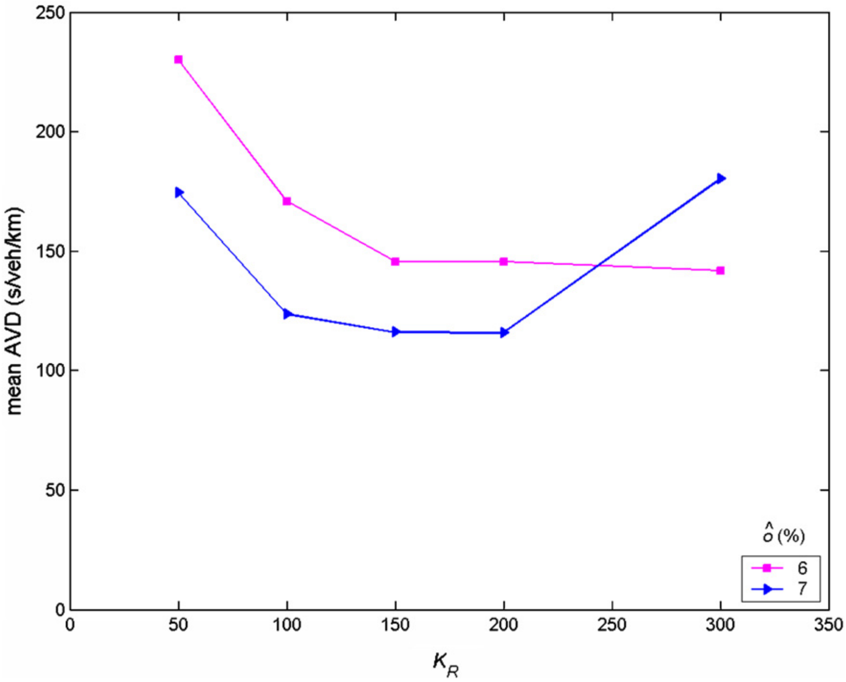

Fig. 12. Mean AVD versus K_R for different set values $\hat{o}$ for the work zone example.

In conclusion, the selection of optimal values for $\hat{o}$ and K_R appears somewhat more cumbersome in this application compared to Section 3.2, but the range of values leading to significant improvements compared to the no-control case is still reasonably broad. Field trials would provide more evidence about the actual occupancy behavior under control and, most importantly, about the achievable level of benefits via merging traffic control.

More details on work zone merging traffic control and related simulation results may be found in Lentzakis et al. (2008).

5. Conclusions and future work

A general framework was developed for merging traffic control in a variety of highway transportation infrastructures where vehicles are merging from a higher to a lower number of lanes. The control framework employs the well-known local ramp metering strategy ALINEA and aims at throughput maximization (or, equivalently, delay minimization) in the corresponding merging traffic facilities. The concept is demonstrated via microscopic simulation for the particular innovative applications of real-time toll plaza management and real-time work zone management with considerable achievable benefits.

A further innovative (and potentially more significant) application of the proposed concept may be the combined on-ramp and mainline metering at freeway junctions. Ramp metering may be applied at freeway on-ramp merges in order to maximize the freeway throughput downstream of the on-ramp. However, the available storage space of on-ramps is usually limited, e.g., because the ramp queue should not intervene with the adjacent street traffic; hence, ramp metering is usually released if the ramp storage space is fully covered with vehicles, and, as a consequence, a merge area congestion is created and the freeway throughput is reduced due to capacity drop. The proposed framework could be used for sustained throughput maximization via combined on-ramp and mainline metering (upstream from the ramp), whereby mainline metering may be effectuated by use of any of the control devices mentioned in Section 2.5. This approach bears some similarities with the one proposed independently by Chang et al. (2007).

Extending this idea further, the local on-ramp/mainline metering system could be understood as a component of a more comprehensive network-wide freeway traffic control system that decides about the (optimal) outflows to be realized by each local component by use of (2). The ultimate vision would be a comprehensive freeway traffic control system that exploits not only the (usually limited) ramp storage space but also the mainline towards an orderly, efficient and safe operation that would replace the chaotic traffic conditions observed daily on most metropolitan freeways around the world. Corresponding investigations are in progress and will be reported in due time.

As a final remark, the presented general concept for merging traffic control may be also applied for throughput maximization in transportation infrastructures where no actual merging takes place because $M \leqslant \mu$, i.e., the number M of arriving lanes is equal or smaller than the number μ of exiting lanes, provided the downstream capacity q_{cap} is lower than the arriving traffic flow. This may be the case at the entrance of tunnels or other facilities or highway parts where the downstream the capacity q_{cap} is lower than the capacity of the upstream infrastructure parts. Traffic control at the entrance of tunnels has in fact been practiced occasionally, albeit sometimes with somewhat vague goals and control algorithms.

References

Cassidy, M.J., Rudjanakanoknad, J., 2005. Increasing the capacity of an isolated merge by metering its on-ramp. Transportation Research Part B 39, 896–913.

Chang, H., Wang, Y., Zhang, J., Ioannou, P., 2007. An integrated roadway controller and its evaluation by microscopic simulator VISSIM. In: European Control Conference 2007, Kos, Greece, pp. 2436–2441, (CD-ROM).

Elefteriadou, L., Roess, R.P., Mc Shane, W.R., 1995. Probabilistic nature of breakdown at freeway merge junctions. Transportation Research Record 1484, 80–89.

FHWA, 2005. Developing and implementing transportation management plans for work zones. Report of the Federal Highway Administration, U.S. Department of Transportation.

FHWA, 2006. State of the practice: traffic control strategies for toll plazas. Report of the Federal Highway Administration, U.S. Department of Transportation.

Keen, K.G., Schoffield, M.J., Hay, G.C., 1986. Ramp metering access control on M6 freeway. In: Proc. 2nd IEE International Conference on Road Traffic Control, London, UK.

Kondyli, A., Duret, A., Elefteriadou, L., 2007. Evaluation of CORSIM and AIMSUN for freeway merging segments under breakdown conditions. In: 86th Annual Meeting of the Transportation Research Board, Washington, DC, USA, (CD-ROM).

Kosmatopoulos, E.B., Papageorgiou, M., 2003. Stability analysis of the freeway ramp metering control strategy ALINEA. In: 11th IEEE Mediterranean Conference on Control & Automation, Rhodes, Greece, (CD-ROM).

Kosmatopoulos, E., Papageorgiou, M., Manolis, D., Hayden, J., Higginson, R., McCabe, K., Rayman, N., 2006. Real-time estimation of the critical occupancy for maximum motorway throughput. Transportation Research Record 1959, 65–76.

Kotsialos, A., Papageorgiou, M., Hayden, J., Higginson, R., McCabe, K., Rayman, N., 2006. Discrete release rate impact on ramp metering performance. IEE Proceedings of the Intelligent Transportation Systems 153, 85–96.

Lentzakis, A.F., Spiliopoulou, A.D., Papamichail, I., Papageorgiou, M., Wang, Y., 2008. Real-time work zone management for throughput maximization. In: 87th Annual Meeting of the Transportation Research Board, Washington, DC, USA, paper 08-0772.

Lorenz, M., Elefteriadou, L., 2001. Defining highway capacity as a function of the breakdown probability. Transportation Research Record 1776, 43–51.

Papageorgiou, M., Kotsialos, A., 2002. Freeway ramp metering: an overview. IEEE Transactions on Intelligent Transportation Systems 3, 271–281.

Papageorgiou, M., Papamichail, I., 2008. Overview of traffic signal operation policies for ramp metering. In: 87th Annual Meeting of the Transportation Research Board, Washington, DC, USA, paper 08-0777.

Papageorgiou, M., Hadj-Salem, H., Blosseville, J.M., 1991. ALINEA: a local feedback control law for on-ramp metering. Transportation Research Record 1320, 58–64.

Papageorgiou, M., Haj-Salem, H., Middelham, F., 1997. ALINEA local ramp metering: summary of field results. Transportation Research Record 1603, 90–98.

Papageorgiou, M., Kosmatopoulos, E., Papamichail, I., Wang, Y., 2007. ALINEA maximises freeway throughput – an answer to flawed criticism. Traffic Engineering and Control 48, 271–276.

Papageorgiou, M., Kosmatopoulos, E., Protopapas, M., Papamichail, I., 2006. Evaluation of the effects of variable speed limits on freeway traffic using M42 traffic data. Internal Report 2006–25, Dynamic Systems and Simulation Laboratory, Technical University of Crete, Chania, Greece.

Smaragdis, E., Papageorgiou, M., 2003. A series of new local ramp metering strategies. Transportation Research Record 1856, 74–86.

Smaragdis, E., Papageorgiou, M., Kosmatopoulos, E., 2004. A flow-maximizing adaptive local ramp metering strategy. Transportation Research Part B 38, 251–270.

Spiliopoulou, A.D., 2007. Vehicle flow control in a freeway toll plaza (in Greek). Diploma Thesis, Dynamic Systems and Simulation Laboratory, Technical University of Crete, Chania, Greece.

Spiliopoulou, A.D., Papamichail, I., Papageorgiou, M., 2008. Real-time toll plaza management for throughput maximization. In: 87th Annual Meeting of the Transportation Research Board, Washington, DC, USA, paper 08-0771.

Treiber, M., Kesting, A., Helbing, D., 2006. Understanding widely scattered traffic flows, the capacity drop, and platoons as effects of variance-driven time gaps. Physical Review E 74, 016123.

Vigos, G., Papageorgiou, M., Wang, Y., 2008. Real-time estimation of vehicle-count within signalized links. Transportation Research Part C 16 (1), 18–35.

Wei, H., Pavithran, M., 2006. Concept of dynamic merge metering approach for work zone traffic control. In: Preprints 11th IFAC Symposium on Control in Transportation Systems, Delft, The Netherlands, pp. 374–379.