Eur. Phys. J. B (2007)
DOI: 10.1140/epjb/e2007-00346-5

THE EUROPEAN PHYSICAL JOURNAL B

Anticipative control of switched queueing systems

S. Lämmer[1,a], R. Donner[1], and D. Helbing[2,3]

[1] Dresden University of Technology, A.-Schubert-Str. 23, 01062 Dresden, Germany
[2] Swiss Federal Institute of Technology, Scheuchzerstr. 70, CH-8092 Zurich, Switzerland
[3] Collegium Budapest – Institute for Advanced Study, Szentháromság utca 2, 1014 Budapest, Hungary

Received 20 August 2007 / Received in final form 8 October 2007
Published online 19 December 2007 –

Abstract. The relevant dynamics of a queueing process can be anticipated by taking future arrivals into account. If the transport from one queue to another is associated with transportation delays, as it is typical for traffic or productions networks, future arrivals to a queue are known over some time horizon and, thus, can be used for an anticipative control of the corresponding flows. A queue is controlled by switching its outflow between "on" and "off" similar to green and red traffic lights, where switching to "on" requires a non-zero setup time. Due to the presence of both continuous and discrete state variables, the queueing process is described as a hybrid dynamical system. From this formulation, we derive one observable of fundamental importance: the green time required to clear the queue. This quantity allows to detect switching time points for serving platoons without delay, i.e., in a "green wave" manner. Moreover, we quantify the cost of delaying the start of a service period or its termination in terms of additional waiting time. Our findings may serve as a basis for strategic control decisions.

PACS. 02.30.Yy Control theory – 02.30.Ks Delay and functional equations – 89.75.-k Complex systems – 89.40.-a Transportation

1 Introduction

Many real-world complex systems like vehicular traffic [1,2] or production networks [3,4] are characterized by a large number of interacting transportation processes. The aim of an efficient organization of such systems is to minimize the time required for all these processes. Typically, this optimization is difficult and demanding, as the topology of the underlying networks is composed of a potentially large number of merges and intersections at which there are conflicts between the flows on different routes [5]. To avoid physical collisions, these flows have to be controlled by devices like traffic lights [6]. The operation strategy of these devices is decisive for the optimization of the system performance.

The switching between flows from different directions leads to an accumulation of vehicles (or products) on the links which are currently not served. The corresponding effects can be mathematically described in terms of queueing theory [7–9]. In switched queuing systems, one may distinguish different states of the queue: a "no service" state and a service period, which itself is composed of a "setup" state, a "clearing" state, and an "extension" state (see Fig. 1a). In the context of vehicular traffic control, the "setup" state is essential for a safe operation making sure all vehicles have left the conflict area before the considered traffic stream enters. The time interval for which all streams need to be stopped, we will refer to as setup time τ^0. Each of the states mentioned above is associated with different dynamical regimes of the queueing process.

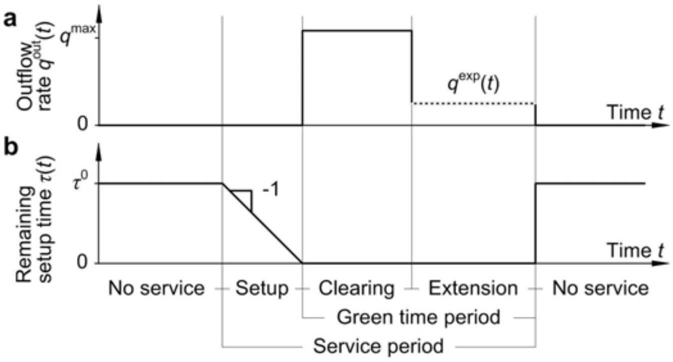

Fig. 1. State dependent values of (a) the outflow rate $q^{out}(t)$ and (b) the remaining setup time $\tau(t)$, see Section 3.2.

Traditionally, such systems are controlled by taking into account only the current state of a queue or the current length of the queue. Future arrivals to the queue are either neglected or assumed according to a certain statistics, i.e., modelled by stochastic processes [10]. Other approaches explicitly consider future arrivals to the queue.

[a] e-mail: traffic@stefanlaemmer.de

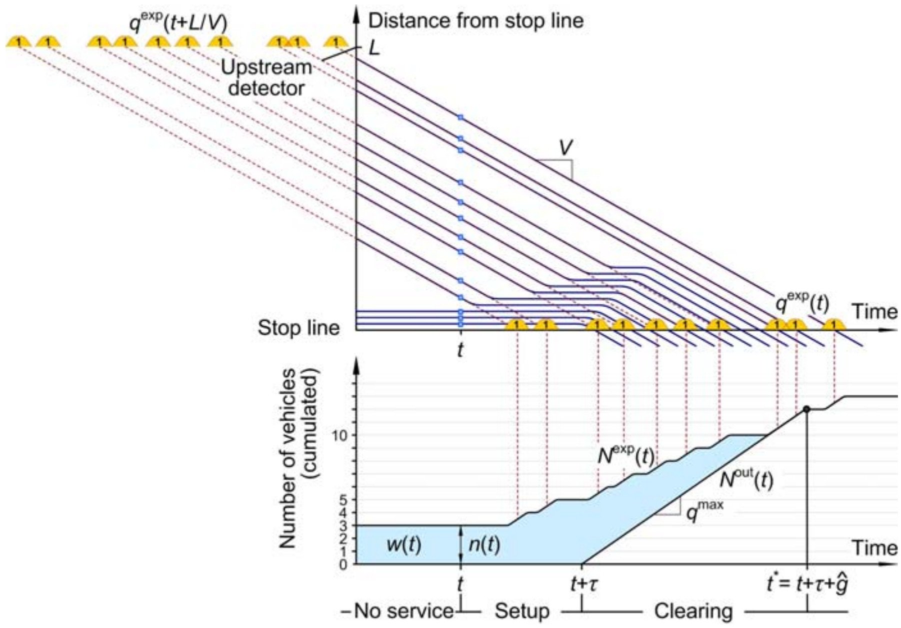

Fig. 2. Prognosis of a service process starting at time t. While the outflow vanishes for the remaining setup time τ, the number of served vehicles $N^{\text{out}}(t)$ increases at the maximum rate $q^{\max}$ during the clearing state. When this curve intersects with the time series $N^{\exp}(t)$ of the expected arrivals, the queue is fully cleared. This is the case at time $t^* = t + \tau + \hat{g}$, where $\hat{g}$ is the required green time.

De Schutter and de Moor [11,12], for example, assume the arrival rate to be constant. This allows them to formulate an extended linear complementarity problem from which they derive optimal switching schedules. Since we are considering flows in a network, however, the arrival rates are never constant but "pulsating" due to platoons, i.e., alternating between zero and large values. Therefore, an efficient prediction algorithm is urgently required to implement anticipative control strategies.

The transport of vehicles or products along a link is usually associated with transportation delays. If these delays are sufficiently well described, they allow to predict future arrivals at the subsequent nodes. In the example of road networks, this requires to evaluate upstream detector data, and, if necessary, to take the switching sequence of neighboring traffic signals into account [13–15].

Our goal is to anticipate the queueing process of one single queue over a subsequent service period. Over the corresponding time interval, which will be specified below, we assume the arrivals to the queue to be known. A natural way to detect these arrivals is to place a detector sufficiently far upstream the stop line (see Fig. 2). By assuming constant velocities in free traffic, the expected arrival flow can be estimated by a constant time-shift. A time series $N^{\exp}(t)$ represents the cumulated number of vehicles expected to reach the stop line until time t under free traffic conditions. Hence, $N^{\exp}(t)$ is defined as

$$N^{\exp}(t) = N^{\exp}(t_0) + \int_{t_0}^{t} q^{\exp}(t')\, dt', \qquad (1)$$

where $N^{\exp}(t_0)$ is some initial value at time t_0 and $q^{\exp}(t)$ the expected arrival rate. We allow $q^{\exp}(t)$ to be unsteady, but assume that it is restricted to the range $0 \le q^{\exp} \le q^{\max}$ with the saturation flow $q^{\max}$.

Besides the expected arrivals, we also need to know how many vehicles have *actually* been served until time t. This quantity is denoted by $N^{\text{out}}(t)$ and defined as

$$N^{\text{out}}(t) = N^{\text{out}}(t_0) + \int_{t_0}^{t} q^{\text{out}}(t')\, dt', \qquad (2)$$

where $q^{\text{out}}(t)$ denotes the actual outflow from the queue, e.g., measured by a detector immediately after the stop line. The two entities $N^{\exp}(t)$ and $N^{\text{out}}(t)$ are sufficient to formulate a model which describes the temporal aspects of the queueing process.

The paper is organized as follows: In Section 2, the queueing process is formulated as a hybrid dynamical system. This formulation will allow us in Section 3 to determine the green time required for clearing a queue. The fact that each additional vehicle arriving during the clearing time period is further enlarging the required green time is taken into account. Finally, in Section 4, we describe how our approach can be used to predict the waiting time while clearing the queue. The result of our paper will be a quantitative measure specifying the cost (in terms of waiting time) of delaying the start of a service period and the cost of terminating it. This measure can serve as a basis of strategic control decisions [6].

2 Queueing model

In the following, the length $n(t)$ of a queue will be defined as the number of vehicles delayed at time t. Hence, the queue length $n(t)$ is given by the difference between the number of vehicles, $N^{\text{exp}}(t)$, that could already have left the link under free traffic conditions, and the number of vehicles, $N^{\text{out}}(t)$, that actually did:

$$n(t) = N^{\text{exp}}(t) - N^{\text{out}}(t). \tag{3}$$

In free traffic, when no vehicles are delayed, $N^{\text{exp}}(t)$ and $N^{\text{out}}(t)$ are equal, resulting in vanishing queue length $n(t) = 0$. Otherwise the value of $N^{\text{exp}}(t)$ is larger then $N^{\text{out}}(t)$ resulting in $n(t) > 0$. Although equation (3), similar to point queue models [9,16,17], does not care about the spatial location of the queue on the link, it describes the temporal aspects of the queueing process such as required green times and waiting times well [18].

Before we derive an equation for the temporal evolution of the queue length, we make use of the fact that the outflow rate $q^{\text{out}}(t)$ is known for a given state of the queue. As Figure 1a shows, the outflow is zero while the queue is not served, but also during the setup. While the queue is being cleared, the queued vehicles flow out at the maximum rate $q^{\max}$. Please note, that the delay due to reaction times and finite acceleration is fully covered by the setup time τ^0. If the service is extended after the queue has been fully cleared, the vehicles are served immediately when they arrive at the stop line. Thus we find $q^{\text{out}}(t) = q^{\text{exp}}(t)$ corresponding to the free traffic state.

Based on these distinct states, from equation (3) we can now derive a state dependent formulation for the temporal evolution of the queue length:

$$\frac{dn}{dt} = \begin{cases} q^{\text{exp}}(t), & \text{if "no service" or "setup"}, \\ q^{\text{exp}}(t) - q^{\max}, & \text{if "clearing the queue"}, \\ 0, & \text{if "extension of service"}. \end{cases} \tag{4}$$

For each of the states, we obtained an ordinary linear differential equation with $q^{\text{exp}}(t)$ as input variable. Since there are abrupt transitions between the different states, however, the resulting dynamics is nonlinear. The interference of both discrete states and continuous variables makes the system a hybrid dynamical system [19–21].

The transition from the state "clearing the queue" to state "extension of service" can – in contrast to all other state transitions – not directly be initiated by the traffic light itself, i.e., by switching it between red and green. The transition occurs at some time point during the service period, when the queue length $n(t)$ becomes zero. For a prognosis of the queueing process, it is therefore essential to determine this particular time point, or, in other words, to determine how much green time is required to clear the queue.

3 Green time required to dissolve a queue

To clear a queue, one must not only serve the vehicles queued up at the current time point t, but also those that will arrive during the setup time and during the clearing of the queue itself. The required green time shall be denoted by $\hat{g}$.

In the following, we present an approach for the computation of $\hat{g}$, assuming that the service period is started (or continued) at the current time point t. After that, we will study the dynamical properties of $\hat{g}$ and present a hybrid dynamical formulation of its temporal evolution.

3.1 Computation of required green time

The clearing of a queue requires a setup for a time period τ followed by a green time of duration $\hat{g}$. Since the outflow rate is maximum ($q^{\text{out}}(t) = q^{\max}$) during the clearing time period $\hat{g}$, a number of $\hat{g}q^{\max}$ vehicles will be served during this time. Therefore, at time $t^* = t + \tau + \hat{g}$ the number of served vehicles will have cumulated up to $N^{\text{out}}(t) + \hat{g}q^{\max}$. According to equation (3), the queue will be fully cleared if this value equals the number of arrived vehicles $N^{\text{exp}}(t^*)$ at the same time point. Consequently, the green time $\hat{g}$ required for clearing the queue must fulfill

$$N^{\text{out}}(t) + \hat{g}\, q^{\max} = N^{\text{exp}}(t + \tau + \hat{g}). \tag{5}$$

As Figure 2 shows, during the clearing time period, the time series of $N^{\text{out}}(t)$ is described by a linear function with the slope $q^{\max}$. Since the gradient of the expected arrivals, $N^{\text{exp}}(t)$, is always smaller than $q^{\max}$ or at most equal, both curves will intersect in only one compact region. The highest value within this region is taken as the value of the required green time $\hat{g}$. It can be easily obtained with standard bisection methods [22].

3.2 Temporal evolution of the required green time

Although equation (5) provides the solution of $\hat{g}$ only in an implicit form, we will in the following derive an equation which describes its temporal change $d\hat{g}/dt$ as a function of the current state of the queue and the expected arrivals $q^{\text{exp}}(t)$. For that, we will from now on refer to $\hat{g}(t)$ as the solution of equation (5) determined at time t.

In a first step, equation (5) is transformed by substituting equations (1) and (2):

$$N^{\text{out}}(t_0) + \int_{t_0}^{t} q^{\text{out}}(t')\, dt' + \hat{g}(t)\, q^{\max} = N^{\text{exp}}(t_0) + \int_{t_0}^{t+\tau(t)+\hat{g}(t)} q^{\text{exp}}(t')\, dt'. \tag{6}$$

This allows one to apply the time derivative d/dt to both sides, leading to:

$$q^{\text{out}}(t) + \frac{d\hat{g}}{dt}\, q^{\max} = q^{\text{exp}}\big(t + \tau(t) + \hat{g}(t)\big) \left(1 + \frac{d\tau}{dt} + \frac{d\hat{g}}{dt}\right). \tag{7}$$

Finally, the term $d\hat{g}/dt$ can be separated (replacing $d\tau/dt = \dot{\tau}$):

$$\frac{d\hat{g}}{dt} = \frac{\big(1+\dot{\tau}\big)q^{\exp}(t^*) - q^{\text{out}}(t)}{q^{\max} - q^{\exp}(t^*)} \quad \text{with} \quad t^* = t + \tau(t) + \hat{g}(t). \tag{8}$$

The above equation provides a general expression for the temporal evolution of the predicted green time $\hat{g}(t)$ required to dissolve the queue.

In the next step, we specify equation (8) for each state of the queue. Besides the outflow rate $q^{\text{out}}(t)$, also the remaining setup time $\tau(t)$ is state dependent. As Figure 1b shows, it has the value τ^0 in the state "no service", decreases at the rate $\dot{\tau} = -1$ during the "setup", stays zero during the entire green time period (including the states "clearing the queue" and "extension of service"), and jumps back to τ^0 immediately after the service time period ends. In the following, we will obtain a hybrid dynamic formulation by inserting these values into equation (8) for each of the states.

In the state "no service" ($q^{\text{out}}(t) = 0$ and $\dot{\tau} = 0$), equation (8) reads $d\hat{g}/dt = q^{\exp}(t^*)/(q^{\max} - q^{\exp}(t^*))$. This term is non-negative, which causes the required green time $\hat{g}(t)$ to increase monotonously as long as the service has not started. More surprisingly, however, $\hat{g}(t)$ jumps to a higher value if $q^{\exp}(t^*) = q^{\max}$. This corresponds to the situation, where the arrival of a vehicle platoon is expected at time t^*. Please notice two important features: First, as it follows from equation (5), the magnitude of the jump is directly proportional to the size of the platoon. Second, between the jump at time t and the arrival of the first vehicle of the platoon at time t^*, there is exactly as much time left as required to perform the setup, to serve the currently waiting vehicles, and to serve all other vehicles arriving before the platoon. This means, if the service is initiated by a jump in $\hat{g}(t)$, the corresponding platoon will be served without any delay. This is illustrated in Figure 3.

During the setup time ($q^{\text{out}}(t) = 0$ and $\dot{\tau} = -1$), the nominator of equation (8) disappears, resulting in $d\hat{g}/dt = 0$. This means that the value of $\hat{g}$ computed before does not change during the setup. While the queue is being cleared ($q^{\text{out}}(t) = q^{\max}$ and $\dot{\tau} = 0$), we obtain $d\hat{g}/dt = -1$. Hence, with each time unit during which the queue is cleared, the remaining green time required decreases by one time unit. In both states, the time point t^* at which the queue was predicted to be cleared stays constant. This fact reflects the validity of the prognosis.

Finally, during the period of green time extension ($q^{\text{out}}(t) = q^{\exp}(t)$ and $\dot{\tau} = 0$), the nominator of equation (8) disappears again. However, when a platoon with maximum flow rate arrives ($q^{\exp}(t) = q^{\max}$), the denominator might become zero as well. The resulting jump in $\hat{g}(t)$ causes a transition to the state "clearing the queue". From this, we see that the process of serving a platoon and the process of clearing a queue, both follow the same law.

Altogether, the state dependent evolution of the green time $\hat{g}(t)$ required for clearing a given queue may be summarized by the following hybrid dynamical equation:

$$\frac{d\hat{g}}{dt} = \begin{cases} \dfrac{q^{\exp}(t^*)}{q^{\max} - q^{\exp}(t^*)}, & \text{if "no service"}, \\ 0, & \text{if "setup"}, \\ -1, & \text{if "clearing the queue"}, \\ 0, & \text{if "extension of service"}. \end{cases} \tag{9}$$

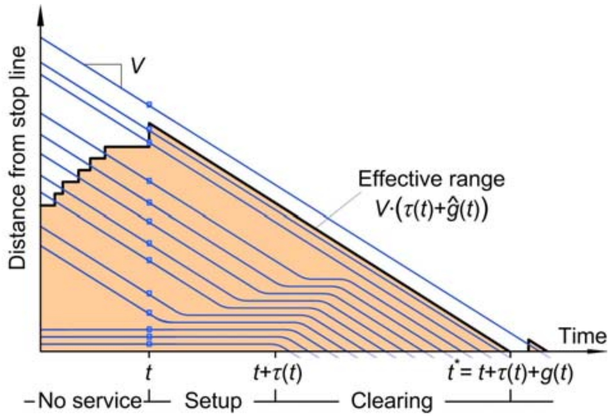

Fig. 3. Spatial interpretation of the minimum service time $\tau(t) + \hat{g}(t)$ required for clearing the queue. By multiplying its value with the free velocity V, we obtain a curve (thick line) with the following properties: (i) Whenever a vehicle passes it, the line jumps by an amount of $V/q^{\max}$ meters further upstream. (ii) The number of vehicles that have passed this line is given by $\hat{n}(t) = \hat{g}(t)q^{\max}$, see also equation (14). (iii) If vehicles arrive at the maximum flow rate, the jump magnitude is proportional to the size of the corresponding platoon. (iv) If the start of the service period is initiated by such a jump, as here at time t, the corresponding platoon is served without any delay, i.e., by a "green wave". All vehicles within this so called "effective range" will be served within a "clearing" state.

With the above analysis, we have proven the validity of the prognosis method proposed in Section 3.1. Furthermore, the entity $\hat{g}(t)$ allows us to detect platoons and to determine the right time point to start the service in order to serve them without delay, i.e., by a "green wave".

4 Waiting time

The surely most important quantity of queueing processes is the waiting time. The waiting time $w(t)$ of all vehicles up to time t is defined as the time integral of the queue length $n(t)$:

$$w(t) = w(t_0) + \int_{t_0}^{t} n(t')\,dt'. \tag{10}$$

From $dw/dt = n(t)$ and the continuity of $n(t)$, it follows directly that (as long as the queue is not yet empty) the waiting time is further increasing. This means that even if the service process has already started and if it would continue forever, some additional waiting time could not be avoided.

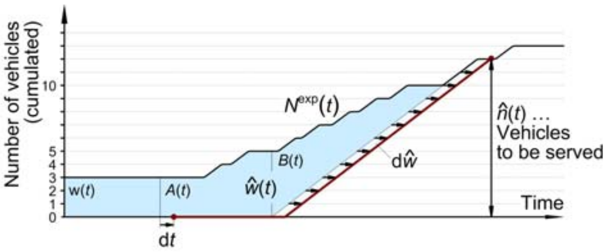

Fig. 4. The waiting time $\hat{w}(t)$ of all vehicles up to the end of the subsequent service period corresponds to the shaded area between the curves $N^{\rm exp}(t)$ and $N^{\rm out}(t)$. From equation (14), it follows that the quantity $\hat{n}(t)$, which is the number of vehicles to be served in order to dissolve the queue, corresponds to the growth rate $d\hat{w}/dt$ during the "no service" state. While serving the queue, $\hat{w}$ stays constant.

In the following, we will demonstrate how one can use the findings from the previous sections to quantify and predict the amount of waiting time that will occur during the subsequent service process. In addition, we will show that it is possible to explicitly determine with which rate this unavoidable amount of waiting time increases while the start of the service period is being delayed.

4.1 Waiting time prediction

In order to optimize the performance of the queuing processes, it is necessary to predict the waiting time $\hat{w}(t)$ of all vehicles up to the end of the following service period. For this, we first assume that the service is started (or continued) at the current time t. As Figure 4 shows, $\hat{w}$ can be defined as the sum of three terms: the waiting time $w(t)$ until time t given by equation (10), the waiting time $A(t)$ during the "setup", and the waiting time $B(t)$ while "clearing the queue". It follows that

$$\hat{w}(t) = w(t) + A(t) + B(t), \tag{11}$$

where $A(t)$ and $B(t)$ are specified in the Appendix. Please note that, after the queue has been cleared, additional waiting times cannot occur until the end of the service period. The premature termination of the service process is treated separately in Section 4.2.

Let us first study the properties of $\hat{w}(t)$, using the same methodology as in the previous sections. From the general definition in equation (11), we can derive the time derivative of the predicted waiting time $\hat{w}(t)$. We obtain

$$\frac{d\hat{w}}{dt} = \hat{g}(t)\, q^{\rm max} \big(1 + \dot{\tau}(t)\big) \;-\; q^{\rm out}(t) \big(\tau(t) + \hat{g}(t)\big) \tag{12}$$

(see Appendix). The rate at which the predicted waiting time $\hat{w}$ increases during each state of the queue can be found by substituting the specific values of $q^{\rm out}(t)$, $\tau(t)$, and $\dot{\tau}$ for each of the states (see Fig. 1).

In the state "no service" ($q^{\rm out}(t) = 0$, $\tau(t) = \tau^0$, and $\dot{\tau} = 0$), equation (12) simplifies to:

$$\frac{d\hat{w}}{dt} = \hat{n}(t) \quad \text{with} \quad \hat{n}(t) = \hat{g}(t)\, q^{\rm max}. \tag{13}$$

Here, $\hat{n}(t)$ represents the number of vehicles that will have to be served with the maximum rate $q^{\rm max}$ during the "clearing" state of duration $\hat{g}(t)$. A graphical representation of $\hat{n}(t)$ is given in Figure 4. In all other states, the right hand side of equation (12) becomes zero. This indicates that the predicted value of $\hat{w}(t)$ will not change during the service period. In summary, we obtain

$$\frac{d\hat{w}}{dt} = \begin{cases} \hat{n}(t), & \text{if "no service"}, \\ 0, & \text{during the entire service period.} \end{cases} \tag{14}$$

From this, we can conclude the following: If a queue is not served at time t, the waiting time until the end of the subsequent service period increases at the rate $\hat{n}(t)$. This fact indicates an interesting analogy between $\hat{n}(t)$ and the queue length $n(t)$: while $n(t)$ is the growth rate of the current waiting time $w(t)$, the quantity $\hat{n}(t)$ stands for the growth rate of the predicted waiting time $\hat{w}(t)$. The above analysis shows that the number of vehicles to be served in order to dissolve a queue, $\hat{n}(t)$, represents a direct measure for the cost of delaying the service of a queue.

4.2 Termination of service

A transition to the state "no service", i.e., the termination of the service period, is associated with a discontinuity in $\tau(t)$, see Figure 1b. Independent of the length of the queue in this state, at the beginning of the next service period, a new setup of duration τ^0 is required. Thus, the discontinuity in $\tau(t)$ causes the predicted waiting time $\hat{w}(t)$ to jump to a higher value. The magnitude of this jump, indicating the cost of terminating the current service period, shall be denoted by $\Delta\hat{w}(t)$.

In order to develop an equation for $\Delta\hat{w}(t)$, we need to study the number of vehicles to be served, $\hat{n}(t, \tau(t))$, also as a function of the remaining setup time $\tau(t)$. Let us assume that the service period is terminated at time t. With $q^{\rm out}(t) = 0$, equation (12) reads $d\hat{w}/dt = \hat{n}(t)\,(1 + \dot{\tau}(t))$, which can be transformed into

$$\frac{d\hat{w}}{d\big(t + \tau(t)\big)} = \hat{n}\big(t, \tau(t)\big). \tag{15}$$

The above equation formulates the derivative of $\hat{w}(t)$ after the time point $t + \tau(t)$ of the earliest possible restart of the "clearing the queue" state. Since for a given time point t, the remaining setup time increases from $\tau(t)$ to τ^0, the value of $\Delta\hat{w}(t)$ can be obtained by integrating equation (15):

$$\Delta\hat{w}(t) = \int_{\tau(t)}^{\tau^0} \hat{n}(t, \tau')\, d\tau'. \tag{16}$$

From this we can conclude the following: The amount of waiting time $\Delta\hat{w}(t)$ which will occur only due to the transition to the "no service" state depends on the arrivals as well as on the width of the interval $[\tau(t), \tau^0]$. If, before time t, the queue was not served ($\tau(t) = \tau^0$), the width of the interval is zero, resulting in $\Delta\hat{w}(t) = 0$ as expected. If the queue has already performed a setup ($\tau(t) < \tau^0$), the width of the interval corresponds to the setup time already elapsed. It becomes largest when the green time period has already started. From this, one may see that the cost of terminating a service period grows with the setup time already elapsed.

5 Summary

The methodology presented above allows us to anticipate the relevant dynamics of a queueing process with arbitrary arrivals. The arrival times of individual vehicles are assumed to be known over the anticipation time horizon, which makes the method particularly well suited for systems with transportation delays, e.g., vehicular traffic or production. By describing the dynamics of the queuing process with both discrete and continuous state variables, i.e., using a hybrid dynamical systems approach, we could give a compact formulation of how the state variables will develop in time.

Given the time series $N^{\exp}(t)$ of the expected arrivals and the number $N^{\text{out}}(t)$ of served vehicles we have derived an expression for the green time $\hat{g}(t)$ required to clear the queue. Its value can be obtained by solving equation (5) numerically in each time step (see Fig. 2). After we have studied its analytical properties in Section 3.2, the quantity $\hat{g}(t)$ turned out to be an observable of fundamental importance for the following reasons:

1. If a platoon is expected to meet the queue, $\hat{g}(t)$ jumps to a higher value. The magnitude of the jump is proportional to the size of the platoon, where the constant of proportionality is the maximum flow rate $q^{\max}$.
2. Between the jump and the arrival of the corresponding platoon, there is exactly as much time as required to perform a setup, to clear the existing queue, and to serve the other vehicles arriving before the platoon.
3. If a controller starts the service period immediately after a jump in $\hat{g}(t)$ was detected, the corresponding platoon will be served in a "green wave" manner. This principle is illustrated in Figure 3.

Since we succeeded to specify the evolution of the state variables for a subsequent service period, we have also been able to anticipate the waiting time $\hat{w}(t)$ until the queue is cleared. If the service period has not yet started, the anticipated waiting time $\hat{w}(t)$ increases at the rate $\hat{n}(t) = \hat{g}(t)q^{\max}$. Surprisingly, it is again the observable $\hat{g}(t)$, that can be used to quantify the cost of delaying the service period (in terms of waiting time).

The results outlined above open new perspectives for the control of switched queueing systems. In contrast to existing controllers, which basically use the current queue length $n(t)$ as an input variable, our framework allows to explicitly consider future arrivals to the queue. In order to implement an anticipation horizon into existing control policies, the similarity between relation equations (10) and (14) seems to be of highest relevance. In the same way, as the queue length $n(t)$ is the time derivative of the waiting time $w(t)$ up to the current time point t, we present another observable $\hat{n}(t)$ being the time derivative of the waiting time $\hat{w}(t)$ up to the end of the next service period.

The potential of the presented findings for developing real-time optimization techniques for the decentralized control of material flows in networks will be discussed in a forthcoming paper.

The authors are grateful to the German Research Foundation (DFG research projects He 2789/5-1,8-1) for partial financial support of this project.

Mathematical supplement

The predicted waiting time $\hat{w}(t)$ is given by the sum in equation (11). The first term, $w(t)$, is specified in equation (10). The second term, $A(t)$, is the waiting time during the "setup" and can be written as integral of the queue length $n(t')$ over the time interval $t' = [t, t+\tau(t)]$:

$$A(t) = \int_t^{t+\tau(t)} N^{\exp}(t') - N^{\text{out}}(t)\, dt'. \tag{17}$$

Please note that the outflow is zero, which causes $N^{\text{out}}(t)$ to stay constant in t'. The third term, $B(t)$, is the waiting time during the state "clearing the queue", in which the cumulated number of served vehicles follows a linear function with slope $q^{\max}$.

$$B(t) = \int\limits_{t+\tau(t)}^{t+\tau(t)+\hat{g}(t)} N^{\exp}(t') - \Big[N^{\text{out}}(t) + \Big(t' - \big(t+\tau(t)\big)\Big)\, q^{\max}\Big]\, dt' \tag{18}$$

The time derivative of $\hat{w}(t)$ is the sum of the time derivatives of its terms,

$$\frac{d\hat{w}}{dt} = \frac{dw}{dt} + \frac{dA}{dt} + \frac{dB}{dt}. \tag{19}$$

The first term, $dw/dt = n(t)$, directly follows from equation (10). For the second term, we obtain

$$\frac{dA}{dt} = \Big[N^{\exp}\big(t+\tau(t)\big) - N^{\text{out}}(t)\Big] \underbrace{\frac{d\big(t+\tau(t)\big)}{dt}}_{=1+\dot{\tau}} - \Big[\underbrace{N^{\exp}(t) - N^{\text{out}}(t)}_{=n(t),\text{ see equation (3)}}\Big] + \underbrace{\int_t^{t+\tau(t)} -q^{\text{out}}(t')\, dt'}_{=-\tau(t)\, q^{\text{out}}(t)} \tag{20}$$

and for the third one we get

$$\frac{dB}{dt} = \Big[\underbrace{N^{\text{exp}}\big(t+\tau(t)+\hat{g}(t)\big) - N^{\text{out}}(t)}_{=\hat{g}(t)\,q^{\text{max}},\ \text{see equation (5)}} - \hat{g}(t)\,q^{\text{max}}\Big] \cdot \frac{d\big(t+\tau(t)+\hat{g}(t)\big)}{dt} - \Big[N^{\text{exp}}\big(t+\tau(t)\big) - N^{\text{out}}(t)\Big] \underbrace{\frac{d\big(t+\tau(t)\big)}{dt}}_{=1+\dot{\tau}} + \underbrace{\int_{t+\tau(t)}^{t+\tau(t)+\hat{g}(t)} \left[-q^{\text{out}}(t') + q^{\text{max}} \frac{d\big(t+\tau(t)\big)}{dt}\right] dt'}_{=\hat{g}(t)\,q^{\text{max}}\,(1+\dot{\tau}) - \hat{g}(t)\,q^{\text{out}}(t)} . \quad (21)$$

When computing the sum in equation (19), several terms cancel out each other: The second term in equation (20) is the same as $dw/dt = n(t)$ and the first term in equation (20) is the same as the second in equation (21). Finally, there only remain the third terms in equations (20) and (21). The result is given in equation (12).

References

1. T. Nagatani, Rep. Progr. Phys. **65**, 1331 (2002)
2. D. Helbing, Rev. Mod. Phys. **73**, 1067 (2001)
3. J. Perkins, P.R. Kumar, IEEE Trans. Automat. Control **34**, 139 (1989)
4. D. Helbing, S. Lämmer, in *Networks of Interacting Machines: Production Organization in Complex Industrial Systems and Biological Cells*, edited by D. Armbruster, A. Mikhailov, K. Kaneko (World Scientific, Singapore, 2005), pp. 33–66
5. D. Helbing, S. Lämmer, J.P. Lebacque, in *Optimal Control and Dynamic Games*, edited by C. Deissenberg, R.F. Hartl (Springer, Dortrecht, 2005), pp. 239–274
6. S. Lämmer, Ph.D. thesis, University of Technology, Dresden, 2007
7. T. van Woensel, N. Vandaele, Asia-Pacific J. Oper. Res. (accepted) (2006)
8. G.F. Newell, *Applications of queueing theory* (Chapman and Hall, 1971)
9. S.K. Bose, *An Introduction to Queueing Systems* (Springer, 2001)
10. R.J. Troutbeck, Transportation Sci. **20**, 272 (1986)
11. B. de Schutter, B. de Moor, in *Hybrid Systems V*, edited by P. Antsaklis, W. Kohn, M. Lemmon, A. Nerode, S. Sastry (Springer-Verlag, Berlin, 1999), Vol. 1567 of *Lecture Notes in Computer Science*, pp. 70–85
12. B. de Schutter, Eur. J. Oper. Res. **139**, 400 (2002)
13. G.F. Newell, Oper. Res. **7**, 589 (1959)
14. P.G. Michalopoulos, V. Pisharody, Transportation Sci. **14**, 365 (1980)
15. M. McDonald, N.B. Hounsell, in *Concise Encyclopedia of Traffic & Transportation Systems*, edited by M. Papageorgiou (Pergamon, Exeter, 1991), Advances in Systems Control and Information Engineering, pp. 400–408
16. G.F. Newell, *The effect of queues on the traffic assignment to freeways*, in *Proc. 7th Int. Symposium on Transportation and Traffic Theory*, edited by T. Sasaki, T. Yamaoka (Institute of Systems Science Research, Kyoto, Japan, 1977), pp. 311–340
17. C.F. Daganzo, Transportation Sci. **32**, 3 (1998)
18. D. Helbing, J. Siegmeier, S. Lämmer, Networks and Heterogeneous Media **2**, 193 (2007)
19. A.V. Savkin, *Controllability of complex switched server queueing networks modelled as hybrid dynamical systems*, in *Proc. 37th IEEE Conf. Decis. & Control*, Vol. 4, 4289 (1998)
20. A.V. Savkin, R.J. Evans, *Hybrid Dynamical Systems* (Birkhäuser, Boston, 2002)
21. B. de Schutter, SIAM J. Control Optim. **39**, 835 (2000)
22. G.B. Arfken, H.J. Weber, *Mathematical Methods for Physicists*, 4th edn. (Academic Press, 1995)