1. Introduction

Despite the large interest given to the evaluation of queues and delays at urban signalized intersections in the last 50 years, there are still some aspects that are not yet fully grasped, and that force traffic scientists and practitioners to apply heuristic methods to approximate these performance variables. In particular, there is lack of a general approach, which is capable of explaining traffic phenomena that are closely related to the stochastic nature of traffic within these types of service systems. In this paper we focus in particular to solving the following three shortcomings, which we found in the past approaches.

First, there is no full insight into the dynamics of overflow queues at signals (especially in the transition between undersaturated and oversaturated conditions and vice versa). Overflow queues are assumed in fact positive and stationary when the arrivals are on average below the signal capacity (i.e. the degree of saturation is smaller than 1), while they grow cycle by cycle as soon as $x>1$. Therefore, there is no smooth transition between these two phases, especially if the signal is designed to operate near capacity (which is often the case of e.g. optimal pre-timed controls). Heuristics have been proposed to overcome this problem (Kimber and Hollis, 1979, Akcelik, 1980) but no scientific fundament justifies the validity of this approach. A direct consequence of this simplification is that the available formulas are hardly applicable for design and planning purposes, especially when one aims to calculate the dynamic effects of congestion on the demand and on the network performance (Viti, 2006).

A second aspect is related to the increasing interest given recently to the variability of queues and delays within these systems; interest that is increasing in line with the growing importance given to the concept of transport reliability. This aspect is closely related to -and it can actually explain- the first limitation, as it will be discussed in the following of the paper. Both shortcomings are therefore caused by a not yet clear role of the variability of arrivals and departures both within a cycle and in the cycle-to-cycle process, which are the causes of the dynamics and the uncertainty of queues, delays and waiting times at such signals (Viti, 2006a, Van Zuylen and Viti, 2007).

Finally a third aspect, which supports the research in this paper and in other works of the authors (e.g. Viti, 2006a) is the lack of a theoretic approach, which can be used generally at such systems, e.g. a methodology that enables one to adopt different arrival and departure distributions, non-stationary arrival rates, limited accumulation space, different control types etc. Referring in particular to this last aspect, some authors...
proposed to adapt widely used fixed timed control formulas like Akcelik’s (Ackelik, 1980) or the HCM delay formula (TRB, 2000) to account for the different delay that drivers experience at vehicle actuated control signals. The heuristic adaptation of formulas for fixed controls, which operate very differently from actuated control signals, denotes the weakness of these formulas, especially to explain the variability of the performance of these systems.

The above three limitations that characterize the available queue and delay formulas are, among others, the reasons for larger and larger application of higher-detailed models in practice (e.g. Cell Transmission Models (Daganzo, 1994), Cellular Automata (Nagel, 1992)) or microsimulation software packages (e.g. VISSIM (PTV, 2003)). However, these models trade off modeling accuracy with long computing times and difficult calibration processes due to the many parameters and variables. Moreover, each microscopic simulation represents only one random draw out of all possible scenarios that can occur, thus many simulation runs need to be done in order to obtain statistically valid results. This makes them unsuitable to deal with long-term planning problems and to analyze the variability of delays.

Motivated by the above reasons, we show in this paper that there is indeed a methodology, which is capable of relaxing these shortcomings: the probabilistic modeling approach. The paper is structured as follows. Section 2 explains some probabilistic aspects that determine the variability of traffic at signals. Section 3 presents the probabilistic modeling approach for fixed and vehicle actuated controls. These models, which have been developed in earlier papers by the authors are presented and explained briefly in this paper. The reader may find the complete (and elaborated) modeling formulations in e.g. (Viti, 2006). Section 4 shows that both probabilistic formulations of fixed and actuated controls are consistent with microscopic simulations running with the same assumptions. This implies that both approaches are valid options for these systems. Finally Section 5 gives the conclusions and the future perspectives of this research.

2. The variability of traffic performance at signals

The development of management strategies designed to adapt the road capacity to the actual demand (e.g. responsive controls) or vice versa (e.g. in-vehicle guidance systems) have supported an increased research over the causes and the effects produced by the variability of traffic. The knowledge of the variability of travel times is valuable for example to quantify the reliability and robustness of a transportation system and to assess the impact of traffic responsive and adaptive control systems. A deeper insight into what causes this variability can tell what part of it is systematic (or recurrent) and therefore predictable or controllable.

The stochastic nature of the demand is widely acknowledged to be one cause of travel time variability (see e.g. (Clark, 2005)). Day-to-day as well as within-day demand variations are in fact observed in all transportation networks. These large variations of the traffic states have a degree of predictability, which can be improved by selecting an appropriate model (e.g. modeling travel times by specifying the time of the day, the day of the week and even the season as determinants of this variability). There are also variations in the traffic demand, which cannot be explained by specifying external factors. Part of travel time variability remains unexplained as it stems from the random nature of human behavior, i.e. their driving behavior and their travel choices. Therefore, a considerable degree of randomness in the traffic flows is observed at the same time of the day and for days in which the activities are repeated with regularity.
Although the variability of traffic flows can be measured systematically using e.g. loop detectors, there is still no clear insight into the effect of this variability on the performance of a transportation network, especially in an urban context, where flows interact and influence each other at road junctions and cause delays one to another.

The largest part of the delay experienced in an urban network by the drivers is caused by traffic controlled signals. Traffic control on intersections is done by means of traffic lights that are most of the times operated automatically using a cyclic sequence of green, amber and red lights. The timing of the green and red duration can be fixed and predetermined (fixed time or pre-timed control) or arrival-based (actuated controls).

Actuated controlled intersections work very differently from fixed or pre-phased intersections, since signal settings are not input of traffic managers but they are determined by the arrival distribution in time. Since the stochastic nature of the arrivals, different headway distributions and different flow rates can be observed from cycle to cycle. The assigned green times are thus variable according to the variability of queues forming during the red and green phases and to the variability of vehicle headways. If one considers that the queue formed during the red phase depends on the number of vehicles arriving during that phase and on the length of the red phase itself, which depends on the green time extensions given to all conflicting streams, the relationship of these variables to the expected delay experienced by one traveler is quite complex.

Very little consideration has been given to the estimation of the queue variability within these control systems, and studies refer solely to fixed time controls. Haight (Haight, 1959) firstly derived a probability distribution of the overflow queue length assuming Poisson arrivals and constant headway in the service process. This approach was extended by Mung (Mung, 1996) for a general arrival distribution. Both Haight’s and Mung’s models are characterized by a high complexity. Newell (Newell, 1971) formulates mathematically this problem using renewal theory. This approach inspired the work of Olszewski (Olszewski, 1990), who investigated the queue length distribution in time using a Markov Chain process. Inspired by the modeling approach of Olszewski, we analyzed the queuing behavior at fixed controls and the behavior of the standard deviation in relation with the average value and to different arrival and departure processes. Later, this approach was developed for vehicle actuated controls, where green and red times are stochastic as well as the overflow queue (Viti, 2006a). In this paper we represent this formulation and we check the consistency with respect to higher order models, i.e. microsimulation.

### 3. Probabilistic formulation of queues and delays

**Fixed and pre-timed controls**

Fixed or pre-timed signalized intersections are governed by a cyclic mechanism, which allows one to calculate the overflow queuing process independently on the queuing process which occurs within a cycle (i.e. how vehicles queue up and are served within the three signal phases. Van Zuylen (Van Zuylen, 1985) described a Markov model for queues at isolated intersections assuming Poisson arrivals and normally distributed saturation flows. Olszewski (Olszewski, 1990) independently developed the idea of applying the Markov chain technique to signal control problems.

The stochastic queuing process is defined in its discrete time case to be a sequence of stochastic variables in time, where a state $Q_{t+1}$ is described by the previous state $Q_t$ and the number of arrivals $a_t$ and departures $d_t$ during the interval $[t, t+1]$, according to the simple relationship $Q_{t+1} = \max\{Q_t + a_t - d_t, 0\}$. Assumed the probability distribution of the arrivals known and deterministic service rate, one can compute the probability of the
queue length in time by first computing the transition probability from one cycle to another:

\[ q_{ij}(t) = \Pr(i = j + a_t - d_t), \quad a_t \in [0, a_{\text{max}}], \quad d_t \in [0, d_{\text{max}}] \]  

which represents the probability that the queue length moves from a state \( j \) at time \( t-1 \) to state \( i \) at time \( t \). The probability of a zero overflow queue comprises all cases where \( j + a_t - d_t \leq 0 \), while if a maximum value for the queue \( Q_{\text{max}} \) is assumed (which can represent the maximum number of vehicles that can buffer at the road section without creating spillback effects), this value will comprise all values for which \( j + a_t - d_t \geq Q_{\text{max}} \).

The queue length probability at time \( t \) is therefore given by the following formula (2):

\[ \Pr(j,t) = \sum_{i=0}^{Q_{\text{max}}} \Pr(i) \cdot q_{ij}(t) \]  

This formulation allows one to compute numerically the full probability distribution of the overflow queue length for any specified probability for the arrivals and the departures. This formulation enables also one to evaluate non-stationary demand conditions; as a consequence the queuing process is described as a smooth process between these two phases.

Figure 1 shows how Formula (2) is consistent with Akcelik’s formula if one assumes Poisson arrivals and deterministic departure rates, zero initial queue and fixed evaluation period of 15 minutes. To be noted that the latter cannot, on the other hand, be applied with different assumptions than the ones considered in the figure, limiting its applicability.

If the calculation of the overflow queue is sufficient for example to explain the average delay experienced in a cycle, no information is given on the way this overflow queue is created within a cycle. Recently the authors have developed a probabilistic formulation of the queuing process also within a cycle, which helps at justifying the existence of the overflow queue from cycle-to-cycle and to give an exact formulation of the control delay ([Van Zuylen, 2006 #171], {Van Zuylen, 2007 #185}). We leave these formulations out of this paper for sake of simplicity.
**Vehicle actuated controls**

An actuated controller operates signals according to actual arrival of vehicles at the intersection. Green times and cycle times are determined by the number of vehicles arriving at the intersection during the red phase and their headway distribution during the green phase. This green time is usually constrained to be smaller than a maximum value, which is mainly determined by the intersection geometry. Therefore, the assigned green times and the delay incurred are stochastic variables too. The computation of green times is then subdivided into three terms: the green time given to serve the vehicles queuing up during the red phase, the one given to the vehicles queuing while the green phase is started and the green time extension given to vehicles arriving in sequence with short headways.

If $a_i(\tau)$ is the arrival rate (in vehicles per second), and $r_i(\tau)$ is the red time at the previous cycle for stream $i$, one can compute the probability of a certain number of vehicles $k$ queuing up during the red phase (of length $\rho$) as:

$$P[Q_i'(\tau) = k] = \int_{\rho=\rho_{\text{min}}}^{\rho_{\text{max}}} (P(a_i(\tau) \cdot \rho = k) \cdot P(r_i(\tau) = \rho))d\rho$$  \hspace{1cm} (3)

The probability of a green time $g_i'(t)$ needed to clear the queue at the end of the red phase to be a value $l$ is therefore given by:

$$P(g_i'(\tau) = l) = \sum_{k \leq l} P(Q_i'(\tau) = k)$$  \hspace{1cm} (4)

While clearing the queue formed during the red phase, other vehicles may join the queue. These vehicles are computed by replacing the probability of red time in formula (3) with the green time of formula (4). The probability of green time due to all vehicles in queue $g_i^O$ is given by computing the joint probability of green due to vehicles arriving during the red phase and the ones during the green phase (see Viti 2006). The probability of green time extension is computed by computing all sequences of vehicles with headway shorter than the unit extension $\tau$. If one computes the probability distribution of a sequence of $n$ vehicles at times $0 < t_1 < t_2 < ... < t_n = t$ the probability of observing this sequence with $t_2 - t_1 < \tau$, $t_3 - t_2 < \tau$, etc. is given by the following formula:

$$P(t_{\text{ext}} = t) = \sum_{a=0}^{n_{\text{max}}} P(t_1 < t_2 < ... < t_n = t) \cdot P(n, t) \hspace{1cm} \text{s.t.} \hspace{0.5cm} t_1 < \tau, t_2 - t_1 < \tau, ..., t_n - t_{n-1} < \tau$$  \hspace{1cm} (5)

The probability of having an extension of exactly $t$ seconds is then given by:

$$P(t_{\text{ext}} = t) = \sum_{a=0}^{n_{\text{max}}} P(t_{\text{ext}} = t) \cdot P(g^{\text{max}}_i - g^O_{i} \geq t)$$  \hspace{1cm} (6)

The probability of a total green time $g_{\text{tot}}$ is finally given by computing the joint probability of green given to clear the queue and the green time extension. Overflow queues are likely to occur only when the intersection is oversaturated and the maximum green extension is met. The corresponding probability is computed by the following formula:

$$P(Q^O_i(t) = q) = \sum_{k = g_{\text{max}}^q} P(Q_i(t) = k)$$  \hspace{1cm} (7)

Since an eventual overflow queue should be cleared in the next green phase, formula (3) should also consider that, apart from the arrivals, also the eventual overflow queue should be served (see Viti, 2006) for details). Last step is to derive the probability
distribution of the red times at the previous cycle. The corresponding probability of a red time to be a certain value \( r(t) = s \) is thus computed with the following formula:

\[
P(r(t) = \rho) = P(\sum_{j=1}^{n} g_{i,j}^{\text{tot}}(t-1) + TL = \rho)
\]

(8)

The probabilistic formulation described helps at getting more insight into the way green and red times are variable depending on the variability of the arrivals. The extension of the model to stochastic departures as well is possible, but it is left for future research.

**Figure 2: Average green times for different combinations of flows**

Figure 2 shows numerical examples of expected green times assigned to one stream in a signal with only two flow streams converging to the intersection if minimum and maximum green times are fixed to respectively 10 and 60 seconds.

### 4. Comparison with microsimulations

The traffic system is a complex combination of physical and behavioral mechanisms, while the probabilistic models presented are simply a combination of mathematical relationships between probability distributions. Validation of the analytical and the Markov models should be made using field data. It is rather difficult acquire such dataset, since several days of observations should be made in order to make valid estimates of mean and standard deviation of queues. Even if one can collect the queue length dynamics for several days it is rather unlikely that one can observe periods of stationary demand conditions, which are long enough to observe equilibrium conditions for the average overflow queue length. A microscopic simulation program may represent in this case alternative to field data. We used this approach and compared the results of the probabilistic model with the microscopic program VISSIM (PTV, 2003 #49).

To obtain comparable results we run the program with hundreds of different random generations for the demand. The fixed time model has been compared both with stationary and with non-stationary (stepwise) demand conditions, showing very good consistency, as shown by the non-stationary demand case in Figure 3.
The difference between the two models is more visible when comparing the standard deviations, especially when the demand is highest. This difference can be appointed at the arrival profile of VISSIM as compared with the assumption of Poisson arrivals in the Markov model. One possible explanation of the above inconsistencies between models can be in the car following characteristics of VISSIM. The arrivals are generated in the microscopic program at 2.5 km from the signal and only the loading of the vehicles is assumed as Poisson. In reality, while vehicles drive along the section they tend to disperse or form platoons according to the car-following behavior assumed in VISSIM. Therefore, the arrival at the signals should not be Poisson.

To test the consistency of the vehicle actuated control model we consider again a simple intersection which serves only two flow streams. The program has been run 100 times for a given combination for the demand \((a_{AB}, a_{CD})\) in order to obtain a statistically valid representation of the variability of traffic. The simulation program has the option to load the input demand using a Poisson distribution as well as the model presented. It has to be pointed out that while the input demand in the system can be controlled, the arrivals at the intersections should not be distributed anymore as Poisson, since car-following criteria modify this distribution in the long sections upstream the signal.

Service rates are in the microscopic program variable and, a priori, unknown (due to e.g. acceleration/deceleration effects and reaction times). In the probabilistic model they are assumed fixed and known. Therefore, an average service rate has been calculated from a first simulation with a total demand condition below the saturation flow of the intersection. Assumed initially that the saturation flow of the intersection would be at least 1800 veh/h, a demand of 1600 veh/h was loaded for one hour of simulation time. The demand was split respectively to 1000 veh/h and 600 veh/h to the two streams. A vehicle actuated control program was developed at the Delft University of Technology and implemented in VISSIM, TRAFCOD (Furth, 1999).

Comparison of the model calibrated on this scenario using only the above average service rate and the same parameters set in TRAFCOD is satisfactorily using low
Queue delays and signal controls: a general approach. Viti, Van Zuylen

demand conditions. Both expected green times and standard deviation for the two intersections were found very close to the microsimulation results. This result has been used to test the prediction capability of the probabilistic approach under a different demand condition. A large demand condition, close to the total capacity of the intersection, was simulated. A total demand of 1800 veh/h was loaded split respectively into 1000 veh/h and 800 veh/h among the two flow streams. To give a visual example of the similarity of the results between the microscopic program and the probabilistic model the cumulative distribution of green times was compared in Figure 4.

![Figure 4: Comparison between the Markov model and multiple runs of the microscopic program VISSIM for the actuated control case](image)

5. Conclusions and further research

Despite the large interest given in the last 50 years, there is not yet full insight into the causes and the effects of delays at signalized intersections. This paper discusses three shortcomings still characterizing the queue and delay models available. The first is a non-smooth transition between expected queues in undersaturated and oversaturated conditions in most of delay models. Heuristics have been applied to resolve this issue but the method has no theoretic fundament. Moreover there is lack of a comprehensive methodology, which enables one to quantify the variability of such queues and delays, and thus to measure the reliability of such systems. Finally there is lack of a methodology, which is flexible enough to enclose all different characteristics of signalized intersections, e.g. multiple lanes, different signal types etc.

We showed in this paper that a probabilistic modeling approach may enable one to overcome all these limitations by simply assuming a probability distribution for the number of arrivals and departures within a cycle. This assumption allows one to calculate the distribution of queues and delays at fixed time controls. We showed further that the same approach can be used to calculate the expected green and red times at vehicle actuated controls as well as the expectation value of the overflow queue in oversaturated conditions. The numerical results of this methodology have been compared with microsimulations, showing very good consistency between each other. However, the probabilistic
modeling approach requires much less computation times and parameters to be calibrated, making it more suitable for e.g. planning and design problems.

6. Acknowledgments
The authors would like to thank the Dutch Ministry of Transport (AVV), who financially supported the PhD thesis of the corresponding author.

7. References
Akcelik R., 1980. Time dependent expression for delay, stop rate and queue length at traffic signals. Australian Road Research Board, Internal Report, AIR 367-1