



**HAL**  
open science

## Estimation of urban traffic state with probe vehicles

Cyril Nguyen van Phu, Nadir Farhi

► **To cite this version:**

Cyril Nguyen van Phu, Nadir Farhi. Estimation of urban traffic state with probe vehicles. IEEE Transactions on Intelligent Transportation Systems, 2021, 22 (5), pp.2797-2808. 10.1109/TITS.2020.2975120 . hal-03101391

**HAL Id: hal-03101391**

**<https://hal.science/hal-03101391v1>**

Submitted on 9 Feb 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Estimation of urban traffic state with probe vehicles

Cyril Nguyen Van Phu\*, Nadir Farhi

COSYS-GRETTIA, Univ Gustave Eiffel, IFSTTAR, F-77454 Marne-la-Vallée, France

\* corresponding author

**Abstract**—We present in this paper a method to estimate urban traffic state with communicating vehicles. Vehicles moving on the links of the urban road network form queues at the traffic lights. We assume that a proportion of vehicles are equipped with localization and communication capabilities, and name them probe vehicles. First, we propose a method for the estimation of the penetration ratio of probe vehicles, as well as the vehicles arrival rate on a link. Moreover, we show that turn ratios at each junction can be estimated. Second, assuming that the turn ratios at each junction are given, we propose an estimation of the queue lengths on a 2-lanes link, by extending a 1-lane existing method. Our extension introduces vehicles assignment onto the lanes. Third, based on this approach, we propose control laws for the traffic light and for the assignment of the arriving vehicles onto the lane queues. Finally, numerical simulations are conducted with Veins framework that bi-directionally couples microscopic road traffic and communication simulators. We illustrate and discuss our propositions with the simulation results.

**Index Terms**—Intelligent transportation systems, Queuing systems.

## I. INTRODUCTION

### A. State of the art

Different techniques are traditionally used to measure road traffic parameters; for example we can cite inductive loops or video cameras. There is nowadays an infrastructure-less technique to estimate traffic flow parameters such as queue lengths : GPS localization system coupled with communicating vehicles, namely probe vehicles. This kind of equipment penetration ratio is increasing and does not need heavy set up.

Probe vehicles were historically studied for measuring travel times [1]. They also helped to estimate penetration ratio and arrival rate of vehicles (equipped and non equipped) on a link. For example, the author of [2] derived these estimations from the estimation of queue lengths at junctions, queue lengths being estimated using the information provided by the probe vehicles. Thus, we can see that in order to characterize urban road traffic state and its primary parameters such as arrival rates or penetration ratio, estimating queue lengths at junctions is an important step. Furthermore, Varaiya [3] has modeled a road network as "a controlled store-and-forward (SF) queuing network" and proposed an algorithm to control this network of queues. Indeed, minimizing delays and waiting times can be done by minimizing queue lengths at junctions controlled with traffic light signals. Hence, queue length estimation is a major measurement input data, used to control traffic light signals, and so transportation road networks.

Concerning queue lengths, in 1963, Miller [4] found an approximation of the average queue length at junctions. More recently, the authors of [5] used shockwave theory to refine queue length estimation. Some works also proposed to use

probability distribution of the queues [6]. Other works used Markov chains to model the dynamics of queue lengths [7]. The authors of [8] and [9] have addressed the queue length estimation with probe vehicles by proposing a probabilistic analytical model. In [8], the authors have estimated queue length in under-saturated traffic conditions, with the "a priori knowledge of the marginal distribution of the queue length" and using "the location information of the last probe vehicle in the queue". The authors of [9] have proposed a method to estimate the queue length, the incoming arrival rate, and the output flow, on a  $m$ -lanes link ( $m \geq 2$ ). The estimations are given for low or saturated demand with no requirement of information concerning the timings of the traffic light signal. In [9], all the lanes are assumed to be balanced (i.e. cars share the lanes of the link without any preference). Therefore, all the lanes would have the same length. In [10], the authors estimate arrival rate for low penetration ratio of equipped vehicles. The method proposed in [10] uses as input data "vehicle trajectories approaching to an intersection as well as traffic signal status". The trajectories of equipped vehicles are used to detect if a probe vehicle has stopped at the traffic light and its stopping position. With these information, the arrival rate is estimated and bounds for this arrival rate are given. In [11], the authors proposed another method. They have lower-bounded the queue length by "the location of the last stopped connected vehicle" and upper-bounded it, when the bound exists, by the location of the "closest moving connected vehicle". Once bounded, the queue length is estimated using the least-mean-square-error method and the noise is filtered using discrete wavelet transform. In 2015, the authors of [12] have addressed the two lanes case by combining discriminant models "based on time occupancy rates and impulse memories" from detectors. The proportions of total traffic volume in each lane are estimated with Kalman filter. In 2018, the authors of [13] have also addressed the two lanes case. They have measured "individual probe vehicles' shockwave speed". Then the lane each probe is moving on, is determined by discriminating the two lanes with data clustering methods. They have shown that a bivariate mixture model clustering gives the best results. Shockwave theory and LWR (Lighthill, Whitham and Richards) model [14], [15] refine the queue length estimation.

### B. Paper contribution and organization

We present here an extension of an existing method that uses probe vehicles for the estimation of urban traffic state, including penetration ratio of communicating vehicles, vehicular arrival rates, as well as the queue lengths of an urban

link. Our extension considers the general case where different destinations can be associated to the lanes, which produces different arrival rates to each lane of the urban link. We propose here to estimate the joint probability distribution of all the queue lengths of the urban link, instead of estimating only one queue length for the link, as done in [9], [10]. This distinction of the lane queues improves the estimation of the number of cars on the queues. Moreover, it gives us the possibility to control the flows of each queue separately, and then ameliorates the traffic control on the junction. We propose in addition control laws for balancing the queue lengths in a multi-lane link. We present here our method on a link of two lanes. We think that the ideas presented in this particular case could be adapted in order to address the general case ( $m$ -lanes, with  $m > 2$ ). Furthermore, we think the method proposed here could be used and extended in a decentralized manner to the network case because of the low computational effort needed for the one link case. The estimators and the control laws we propose here would permit us to perform multi-level urban traffic control, as initiated in [16] (local control) and in [17] (semi-decentralized control).

In section I we give an introduction with the related works. In section II we describe the problem statement and the notations. In section III we propose an estimation of traffic state parameters : penetration ratio of probe vehicles, vehicles arrival rate (subsection III-A), and queue lengths in the case of two incoming lanes (subsection III-B). Our method estimates queue lengths at junctions with two lanes incoming roads, under the hypothesis of under-saturated traffic (moderate/low demand without overflow queue). We also assume that the GPS localization system is not able to determine which lane a vehicle is moving on because of a typically five meters accuracy [18]. We extend the analytic model proposed in [8] to the two-lanes case by introducing a vehicle assignment model onto the lanes. In subsection III-C, we propose a control of the traffic light and an optimal assignment of the vehicles onto the lanes, in order to balance the two lanes queue lengths. In section IV, numerical simulations are conducted with Veins framework [19] which bi-directionally couples microscopic road traffic and communication simulators. Finally, we conclude in section V.

## II. PROBLEM STATEMENT

In this section we describe the main assumptions of our work and the notations used.

### A. Assumptions

*a) Road network topology:* We consider a road network composed of junctions controlled by traffic light signals, and links between junctions. We assume that all the incoming and outgoing links to/from a signalized junction have maximum two lanes. We define an entry link of the network as a link which does not have a start node. We assume that the geometry of the road network is known. A typical junction is represented on Fig. 1. We also assume that the timings of the traffic light signal are known, and specially the red times.

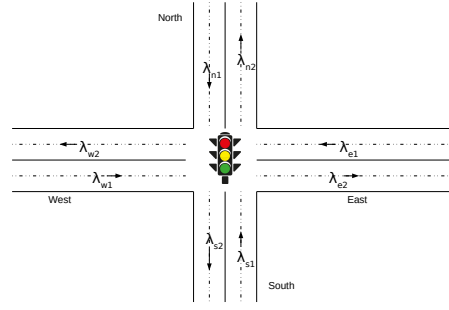


Fig. 1: A signalized junction of the road network.

*b) Traffic demand:* We assume that the travel demand is exogenous, which means that the demand is located only at the entry links of the network. We assume that the vehicles arrive onto each link  $l$  under a Poisson process of rate  $\lambda_l$ . We consider in this paper the low/moderate demand case where the Poisson arrival assumption is valid. In [2] the author discusses the Poisson arrivals assumption and recalls that this assumption is commonly used to describe arrivals at isolated intersections, specially in the case of low/moderate demand with no overflow queue. The vehicles form queues at junctions. Since we assume Poisson arrivals, we consider that the queues are empty at the beginning of each red time (no overflow queue).

*c) Probe vehicles:* We assume that a ratio  $p$  (with  $0 \leq p \leq 1$ ) of vehicles are equipped with localization and communication systems and we name them probe vehicles. The probe vehicles send their positions and speeds to a road side unit (RSU) coupled with the traffic light signal of the junction. We assume that the transmit power of the communication system embedded in every vehicle is strong enough, and that the sensitivity of the RSU is accurate enough, such that the RSU can detect every vehicle in every incoming or outgoing link of its associated junction. We consider the case where the localization system embedded in the vehicles is not accurate enough to discriminate the lane the vehicle is moving on.

*d) Turn ratios:* Fig. 2 represents the queues we consider, on a link of the road network. Probe vehicles are represented

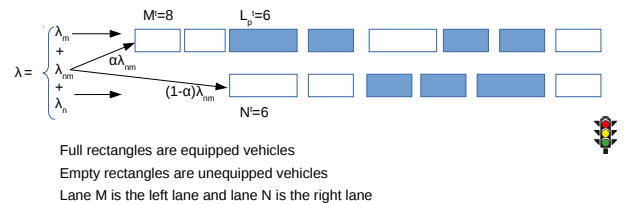


Fig. 2: Queues in 2-lanes incoming link. Vehicles that can choose both lanes are assigned onto lane  $M$  with probability  $\alpha$  and onto lane  $N$  with probability  $(1 - \alpha)$ .

by full rectangles and unequipped vehicles are represented by empty rectangles. Some vehicles are necessarily assigned to the queue on lane  $N$  (they turn right) and other vehicles are necessarily assigned to the queue on lane  $M$  (they turn left).

Vehicles going straight can choose both lanes. We will assume that a ratio  $\alpha$  (with  $0 \leq \alpha \leq 1$ ) of such vehicles going straight will choose the queue on lane  $M$ . So, on a 2-lanes link, we assume that the main flow  $\lambda$  is composed of three flows :

- 1) the flow with arrival rate  $\lambda_n$  which is necessarily assigned to lane  $N$  (vehicles turning right).
- 2) the flow with arrival rate  $\lambda_m$  which is necessarily assigned to lane  $M$  (vehicles turning left).
- 3) the flow with arrival rate  $\lambda_{nm}$  which can be assigned to both lanes  $N$  or  $M$  (vehicles going straight).

We consider that these three flows are independent and identically distributed (iid) stochastic arrivals, each one being a Poisson process. We suppose that a fraction  $\alpha$  (with  $0 \leq \alpha \leq 1$ ) of the flow  $\lambda_{nm}$  is assigned to lane  $M$  and the complement  $(1 - \alpha)$  of this same flow  $\lambda_{nm}$  is assigned to lane  $N$ . As denoted in TABLE I,  $A^t$  is the random variable representing the assignment onto the two lanes. We assume that  $A^t$  is following a Bernoulli law such that  $P(A^t = 1) = \alpha$  and  $P(A^t = 0) = 1 - \alpha$ . Thus,  $\mathbb{E}(A^t) = \alpha$ . We define :

$$\mu_N(t) := r_N(t)(\lambda_n + (1 - \alpha)\lambda_{nm}) \quad (1)$$

$$\mu_M(t) := r_M(t)(\lambda_m + \alpha\lambda_{nm}) \quad (2)$$

We will show later in Proposition 1 that  $\mu_N(t)$  and  $\mu_M(t)$  represent the average arrival rate multiplied by the red duration on respectively lane  $N$  and lane  $M$ .

Also, we assume that the turn ratios are given. Indeed, it is easy to measure the turn ratios as following : the RSU detects all the probe vehicles in its radio range area. This is because the probe vehicles embed WAVE (Wireless Access in Vehicular Environments) [20] on OBU (on board unit). In [20], the basic safety messages (BSM) broadcast periodically the location and speed of probe vehicles. This is a default feature which is also implemented as a basic function in VEINS simulator [19] that we use in this paper. So, if we know at time  $t$  the location of each probe vehicle on a given link and its unique identifier, it is enough to look at a time  $t + t_x$  ( $t_x$  being a time shift), where those vehicles are located. With this method, it is possible to estimate the turn ratios  $l_n$ ,  $l_m$  and  $l_{nm}$  which are the proportions of the main flow  $\lambda$  on the incoming link that respectively turn right, left or go straight. We will note :  $\lambda_n = l_n\lambda$ ,  $\lambda_m = l_m\lambda$ ,  $\lambda_{nm} = l_{nm}\lambda$ , with  $l_n + l_m + l_{nm} = 1$ . We assume in this paper that  $l_n$ ,  $l_m$ , and  $l_{nm}$  are given.

## B. Notations

We will use the notations of TABLE I.

## III. TRAFFIC STATE ESTIMATION

### A. Primary parameters estimation

In this section, we give a method for the estimation of the primary traffic parameters  $p$  and  $\lambda$ . We assume that every probe vehicle in the RSU radio range area is assigned to an incoming or outgoing link to/from the junction. Thus, the total number of incoming probe vehicles  $x_p(t)$  in a given link to the junction is known.

Name	Definition
$L_V$	the average vehicle length
$G_V$	the minimum distance gap between vehicles
$R$	the total red time in one cycle
$r_N(t)$	the time since the beginning of the red phase for lane N (it is 0 if we are not in red phase at time $t$ ), $0 \leq r_N(t) \leq R$ .
$r_M(t)$	the time since the beginning of the red phase for lane M (it is 0 if we are not in red phase at time $t$ ), $0 \leq r_M(t) \leq R$ .
$\lambda_n, \lambda_m, \lambda_{nm}$	the average arrival rate in vehicles/second for vehicles that respectively turn right, left or go straight.
$\lambda = \lambda_n + \lambda_m + \lambda_{nm}$	the total arrival rate for the incoming link in vehicles/second.
$l_n, l_m, l_{nm}$	the proportions (turning ratios) of the main flow $\lambda$ on the incoming link that respectively turn right, left or go straight, with $\lambda_n = l_n\lambda$ , $\lambda_m = l_m\lambda$ , $\lambda_{nm} = l_{nm}\lambda$
$x(t)$	the total number of vehicles on all the lanes of the considered link at time $t$ .
$x_p(t)$	the number of probe vehicles on all the lanes of the considered link at time $t$ .
$p, 0 \leq p \leq 1$	the penetration ratio of probe vehicles.
$N^t$	the total number of vehicles in the queue at time $t$ and lane N. In this paper, $N^t$ is assumed to be a random variable.
$M^t$	the total number of vehicles in the queue at time $t$ and lane M. In this paper, $M^t$ is assumed to be a random variable.
$A^t$	the assignment of a vehicle entering the edge at time $t$ . $A^t = 1$ if the vehicle is assigned on lane M and $A^t = 0$ if it is assigned on lane N. $A^t$ is assumed to be a random variable.
$L_p^t$	the location (in number of vehicles) of the last probe in the queue, namely the last connected vehicle, at time $t$ . $L_p^t$ is assumed to be a random variable, taking value $l_p$ .
$N_p^t$	the total number of probe vehicles in the queue at time $t$ and lane N.
$M_p^t$	the total number of probe vehicles in the queue at time $t$ and lane M.
$c_p$	the total number of probe vehicles in all the lanes and all the queues at time $t$ .

TABLE I: Notations

We consider vehicles  $i$  moving at speed  $v_i(t)$  and at a distance  $\rho_i(t)$  (depending on time  $t$ ) from the traffic light. Let us consider the following definition.

**Definition 1.** For a given threshold car-speed  $v^*$  and a given threshold car-distance  $\rho^*$  to the junction, the vehicles queue  $Q = Q(t, v^*, \rho^*)$  is defined by  $Q = \{i, v_i < v^* \text{ and } \rho_i < \rho^*\}$ .

$\rho^*$  is useful because if the queue would exceed the bound  $\rho^*$ , we could know that the assumption of low/moderate demand is not adequate. Furthermore,  $\rho^*$  is less than the edge length, so the queue keeps bounded. We then denote by  $Q_p$  the subset of  $Q$  that includes only probe vehicles,  $Q_p \subset Q$ . The total number of probes  $c_p$  in the queue is given by the cardinal (number of elements) of the set  $Q_p$ . We assume mixed vehicles (equipped and non equipped) with an average vehicle length  $L_V$  and minimum distance gap  $G_V$  between vehicles.  $\rho_0$  denotes the offset distance from the RSU to the stop line of the traffic light signal. We propose to compute  $l_p$  the last probe location in the unit of "number of vehicles" as follows. We have :

$$\max_{i \in Q_p}(\rho_i) = \rho_0 + l_p L_V + (l_p - 1)G_V \quad (3)$$

Then,

$$l_p = \lceil (\max_{i \in Q_p}(\rho_i) - \rho_0 + G_V) / (L_V + G_V) \rceil \quad (4)$$

where  $\lceil \cdot \rceil$  denotes the round operator to the nearest integer.

Given  $c_p, l_p$ , Comert [2] has derived many estimators for  $p$ , one of them being  $c_p/l_p$ , which is biased for  $p < 1$ . We follow here the same idea and propose a variation of the estimator of  $p$ . For the one lane case, we propose:

$$\hat{p} = (c_p - 1) / (l_p - 1), \text{ for } l_p > 1 \quad (5)$$

We have :  $N^t = l_p + 1/p - 1$ , where  $1/p - 1$  represents the average backlog of the queue behind the last probe. Then  $\hat{p} = c_p / N^t = c_p / (l_p + 1/p - 1)$ . Moreover, by following the same arguments of the proof in [2], it is easy to check that this estimator is unbiased for every  $p$ , i.e.  $\mathbb{E}(\hat{p}) = p, \forall p, 0 \leq p \leq 1$ .

For two lanes, we introduce :

$\kappa := \min(\mu_n, \mu_m) / \max(\mu_n, \mu_m)$ . We can see that  $\kappa$  does depend only on turn ratios but not on the arrival rate for the link, because  $\kappa$  is a ratio.

$$\kappa = \frac{\min(r_N(t)(l_n + (1 - \alpha)l_{nm}), r_M(t)(l_m + \alpha l_{nm}))}{\max(r_N(t)(l_n + (1 - \alpha)l_{nm}), r_M(t)(l_m + \alpha l_{nm}))} \quad (6)$$

We consider queue lengths on lanes  $N$  and  $M$  respectively equal to  $n$  and  $m$ . We propose :

$$\hat{p} = \frac{c_p}{n + m} \quad (7)$$

By the way, in our case, the length  $n$  of queue  $N$  can be estimated with the number of arrivals on lane  $N$  during  $r_N(t)$  which is  $\mu_n$ . As  $\mu_n + \mu_m = \max(\mu_n, \mu_m) + \min(\mu_n, \mu_m)$  and by estimating  $\max(\mu_n, \mu_m) = l_p + 1/p - 1$ , where  $1/p - 1$  represents the backlog of the queue behind the last probe, we can write :

$$\mu_n + \mu_m = \max(\mu_n, \mu_m) \left( 1 + \frac{\min(\mu_n, \mu_m)}{\max(\mu_n, \mu_m)} \right) \quad (8)$$

$$\mu_n + \mu_m = (l_p + 1/p - 1)(1 + \kappa) \quad (9)$$

We introduce  $c_\kappa = c_p / (1 + \kappa)$  and replace (9) in (7). Finally, we get the following equation :

$$\hat{p} = \frac{c_\kappa}{l_p + (1 - p)/p} \quad (10)$$

Hence, solving in  $\hat{p}$ , by putting  $p = \hat{p}$  :

$$\hat{p} = (c_\kappa - 1) / (l_p - 1), \text{ for } l_p > 1 \text{ and } c_p > 1 \quad (11)$$

which extends (5) for the case of two lanes. Similarly, we can check that this estimator for two lanes is unbiased, by following again the same arguments as in [2]. In Appendix A, we propose another method to compute  $\hat{p}$  which is based on the calculus of an unbiased estimator and gives the same result of formula (11).

We propose to compute  $\lambda$  with formula (12) by simply accumulating probe vehicles on the entire radio range area of the RSU during red time, and using  $\hat{x} = x_p/p$ .  $\lambda$  should be computed when all the lanes of the considered link have a red light at the traffic light. We formulate this as following, where  $t_0$  is the starting time for the red light on both lanes and  $t_1$  is the time after which the set of the two lanes are not at red light.

$$\hat{\lambda} = \frac{x_p(t_1) - x_p(t_0)}{p(t_1 - t_0)} \quad (12)$$

## B. Queue length estimation

Once  $p$  and  $\lambda$  are estimated with probe vehicles, we can refine our traffic state estimation (queue lengths). We propose in this section to estimate all the queue lengths associated to all the lanes on a link of the road network. We will propose a model that uses vehicular assignment onto the lanes, for links composed of two incoming lanes. In a first step, an analytical probability distribution formulation of the queue lengths, without using the information from the probe vehicles, will be presented. Then, we will use the information provided by the probe vehicles : while generalizing the work for 1-lane road done in [8] to the 2-lanes case, we will refine our analytical formulation. We recall here that we can not directly detect the lanes on which the probe vehicles are moving because of insufficient accuracy of GPS localization system [18], which makes the problem not obvious.

a) *Distribution probability law of the 2-lanes without having the information provided by the probes:* We first propose an estimation of the probability distribution  $P(N^t = n, M^t = m)$  without having any information from the probe vehicles.

### Proposition 1.

$$P(N^t = n, M^t = m) = \frac{\mu_N(t)^n e^{-\mu_N(t)}}{n!} \frac{\mu_M(t)^m e^{-\mu_M(t)}}{m!} \quad (13)$$

*Proof.* We subdivide the Poisson process of rate  $\lambda_{nm}$  common to the two lanes. The common arrival of rate  $\lambda_{nm}$  is splitted with probability  $\alpha$  to lane  $M$  and probability  $(1 - \alpha)$  to lane  $N$ . The two produced flows are independent random flows each one following Poisson process of parameters respectively  $\alpha\lambda_{nm}$  for the flow assigned to lane  $M$  and  $(1 - \alpha)\lambda_{nm}$  for the flow assigned to lane  $N$ . Furthermore, the splitted Poisson processes are independent; see subdividing Poisson process in reference [21].

By combination, arrivals on lane  $N$  is the sum of two independent Poisson processes. Using the stationary property of Poisson processes, we can show that the number of arrivals in  $[0, r_N(t)]$  on lane  $N$  is a Poisson process of parameter  $\mu_N$ . Similarly, the number of arrivals on lane  $M$  in  $[0, r_M(t)]$  is a Poisson process of parameter  $\mu_M$ . As these two arrival flows on lanes  $N$  and  $M$  are independent, then the bivariate distribution probability law of the two queue lengths is the product of two Poisson Law of parameters  $\mu_N$  and  $\mu_M$ .  $\square$

b) *Distribution probability law of the 2-lanes queue lengths with the information provided by the probe vehicles:*

We present here the conditional probability law of the two queue lengths, taking into consideration the information provided by the probe vehicles, specially the location of the last probe  $l_p$  and the total number of probes in the two lanes queues  $c_p$ . We recall here  $N_p^t$  and  $M_p^t$  are the number of probe vehicles respectively on the lane  $N$  and on the lane  $M$ , at time  $t$ .

**Proposition 2.**

- If  $lp \leq \max(n, m)$  and  $c_p \leq n + m$ , then

$$P(N^t = n, M^t = m | L_p^t = l_p, N_p^t + M_p^t = c_p) = \frac{\binom{l_p - 1 + \min(l_p, n, m)}{c_p - 1} (1 - p)^{n+m} P(N^t = n, M^t = m)}{\sum_{\substack{j, k \geq 0 \\ \text{subject to} \\ \max(j, k) \geq l_p \\ j+k \geq c_p}} \binom{l_p - 1 + \min(l_p, j, k)}{c_p - 1} (1 - p)^{j+k} P(N^t = j, M^t = k)}.$$

- Otherwise,

$$P(N^t = n, M^t = m | L_p^t = l_p, N_p^t + M_p^t = c_p) = 0.$$

*Proof.* By Bayes' rule we have

$$P(N^t = n, M^t = m | L_p^t = l_p, N_p^t + M_p^t = c_p) = \frac{P(N^t = n, M^t = m, L_p^t = l_p, N_p^t + M_p^t = c_p)}{P(L_p^t = l_p, N_p^t + M_p^t = c_p)} \quad (14)$$

Then the numerator in (14) is written

$$P(N^t = n, M^t = m, L_p^t = l_p, N_p^t + M_p^t = c_p) = P(L_p^t = l_p | N_p^t + M_p^t = c_p, N^t = n, M^t = m) P(N_p^t + M_p^t = c_p | N^t = n, M^t = m) P(N^t = n, M^t = m) \quad (15)$$

We have

- $P(L_p^t = l_p | N_p^t + M_p^t = c_p, N^t = n, M^t = m) =$

$$\binom{l_p - 1 + \min(l_p, n, m)}{c_p - 1} / \binom{n+m}{c_p}.$$

- $P(N_p^t + M_p^t = c_p | N^t = n, M^t = m) =$

$$\binom{n+m}{c_p} p^{c_p} (1 - p)^{n+m-c_p}.$$

For the calculus of  $P(L_p^t = l_p | N_p^t + M_p^t = c_p, N^t = n, M^t = m)$ , we followed the same ideas as those of section 3 in [8]. Indeed, we will use the example of Fig. 2 where  $l_p = 6$ ,  $M^t = 8$ ,  $N^t = 6$  and  $c_p = 7$ . The probability is then computed by selecting the total number of events where  $L_p = l_p = 6$  divided by the sample space. The sample space, which is composed of all the last probe possible locations is given by  $\binom{n+m}{c_p} = \binom{14}{7}$ . For  $L_p = l_p = 6$  we must have all the probes in the preceding locations. The event space has a number of events corresponding to choosing  $(c_p - 1) = 6$  probes among  $l_p - 1 + \min(l_p, n, m) = 6 - 1 + 6 = 11$  positions available. Here, the event space has a total number of elements given by  $\binom{l_p - 1 + \min(l_p, n, m)}{c_p - 1} = \binom{11}{6}$ . This is why :

$$P(L_p^t = l_p | N_p^t + M_p^t = c_p, N^t = n, M^t = m) = \frac{\binom{l_p - 1 + \min(l_p, n, m)}{c_p - 1}}{\binom{n+m}{c_p}} \quad (16)$$

For the calculus of  $P(N_p^t + M_p^t = c_p | N^t = n, M^t = m)$  we have  $c_p$  probe vehicles among  $n + m$  total vehicles. The probability for a vehicle to be a probe vehicle is  $p$  and the probability to be unequipped is  $(1 - p)$ . The configurations considered in this case are  $c_p$  vehicles equipped and  $(n + m -$

$c_p)$  vehicles unequipped. The number of combinations of such configurations is  $\binom{n+m}{c_p}$ . This is why :

$$P(N_p^t + M_p^t = c_p | N^t = n, M^t = m) = \binom{n+m}{c_p} p^{c_p} (1 - p)^{n+m-c_p} \quad (17)$$

So the numerator in (14) is given by :

$$P(N^t = n, M^t = m, L_p^t = l_p, N_p^t + M_p^t = c_p) = \binom{l_p - 1 + \min(l_p, n, m)}{c_p - 1} p^{c_p} (1 - p)^{n+m-c_p} P(N^t = n, M^t = m)$$

The denominator in (14) is the marginal distribution probability of  $P(N^t = j, M^t = k, L_p = l_p, N_p^t + M_p^t = c_p)$  on  $(j, k)$ . Therefore, the ideas to compute this probability are the same as the ideas used to compute the numerator of (14). We notice here that the last probe position (in the unit number of vehicles) is necessarily less than or equal to the maximum of the queue lengths, since the last probe is necessarily in one of the two lanes queues. Similarly, the total number of probes  $c_p$  is less than or equal to the total number of vehicles in the queues, since the probes are in the queues. Therefore, we can write :

$$P(L_p^t = l_p, N_p^t + M_p^t = c_p) = \sum_{\substack{j, k \geq 0 \\ \max(j, k) \geq l_p \\ j+k \geq c_p}} P(N^t = j, M^t = k, L_p = l_p, N_p^t + M_p^t = c_p).$$

□

c) *Estimators:* The distribution probability law of the couple  $(N^t, M^t)$  is known; see Proposition 2 . As,  $\mathbb{E}(N^t, M^t) = (\mathbb{E}(N^t), \mathbb{E}(M^t))$ , one way to estimate the two queue lengths is to derive each queue length separately from the couple, by computing the expectation of  $N^t$  and  $M^t$  separately. We propose the following estimator for queue length on lane  $N$  :

$$\mathbb{E}(N^t | L_p^t = l_p, N_p^t + M_p^t = c_p) = \sum_{n \geq 0} n \sum_{k \geq 0} P(N^t = n, M^t = k | L_p^t = l_p, N_p^t + M_p^t = c_p) \quad (18)$$

Similarly for the queue length on lane  $M$ , we will choose :

$$\mathbb{E}(M^t | L_p^t = l_p, N_p^t + M_p^t = c_p) = \sum_{m \geq 0} m \sum_{j \geq 0} P(N^t = j, M^t = m | L_p^t = l_p, N_p^t + M_p^t = c_p) \quad (19)$$

C. *Traffic light control and optimal assignment of vehicles onto the lanes*

We are interested here in the equilibration of the two queue lengths with respect to the two parameters  $\alpha$  and  $\bar{r} := r_N / r_M$ . We exclude here the case  $r_N = r_M = 0$  where both lanes have

green light, and where no queue is formed by assumption; and therefore, the assignment onto the queues is meaningless. We use notations  $\mathbb{E}_{(\alpha, r_N)}(N^t) := \mathbb{E}(N^t)$  and  $\mathbb{E}_{(\alpha, r_M)}(M^t) := \mathbb{E}(M^t)$  in order to emphasize the dependence of these two expectations on the parameters  $\alpha, r_N$  and  $r_M$ . Let us now define  $f(\alpha, \bar{r})$  as follows.

$$\begin{aligned} f(\alpha, \bar{r}) &:= |\mathbb{E}_{(\alpha, r_N)}(N^t) - \mathbb{E}_{(\alpha, r_M)}(M^t)| / \mathbb{E}_{(\alpha, r_N)}(N^t) \\ &= |\mu_N - \mu_M| / \mu_N \\ &= \frac{\lambda |r_N(l_n + (1-\alpha)l_{nm}) - r_M(l_m + \alpha l_{nm})|}{\lambda r_N(l_n + (1-\alpha)l_{nm})} \\ &= \frac{|\bar{r}(l_n + (1-\alpha)l_{nm}) - (l_m + \alpha l_{nm})|}{\bar{r}(l_n + (1-\alpha)l_{nm})}. \end{aligned}$$

We are interested here in the minimization of  $f(\alpha, \bar{r})$  with respect to the two parameters  $\alpha$  and  $\bar{r}$ , which permits the equilibration of the two queue lengths. Let us use the notations.

$$r^*(\alpha) := \arg \min_{\bar{r}} f(\alpha, \bar{r}). \quad (20)$$

$$\alpha^*(\bar{r}) := \arg \min_{\alpha} f(\alpha, \bar{r}). \quad (21)$$

Proposition 3 and Proposition 4 below determines  $r^*(\alpha)$  and  $\alpha^*(\bar{r})$  respectively.

**Proposition 3.**  $\forall \alpha \in [0, 1], r^*(\alpha) = \frac{l_m + \alpha l_{nm}}{l_n + (1-\alpha)l_{nm}}$ , and  $f(\alpha, r^*(\alpha)) = 0$ .

*Proof.*  $\forall \alpha \in [0, 1], \bar{r} = \bar{r}_0 = \frac{l_m + \alpha l_{nm}}{l_n + (1-\alpha)l_{nm}}$  implies

$$\begin{aligned} \mathbb{E}_{(\alpha, r_N)}(N^t) &= r_N(\lambda_n + (1-\alpha)\lambda_{nm}) \\ &= (r_M \bar{r})(\lambda_n + (1-\alpha)\lambda_{nm}) \\ &= r_M(\bar{r}_0(\lambda_n + (1-\alpha)\lambda_{nm})) \\ &= r_M(\lambda_m + \alpha \lambda_{nm}) = \mathbb{E}_{(\alpha, r_M)}(M^t). \end{aligned}$$

Therefore,  $f(\alpha, \bar{r}_0) = 0$ . Thus,  $r^*(\alpha) = \bar{r}_0 = \frac{l_m + \alpha l_{nm}}{l_n + (1-\alpha)l_{nm}}$ .  $\square$

**Proposition 4.**  $\forall \bar{r} \geq 0$ ,

$$\alpha^*(\bar{r}) = \max \left( 0, \min \left( 1, \frac{\bar{r}l_n + \bar{r}l_{nm} - l_m}{l_{nm}(\bar{r} + 1)} \right) \right).$$

Moreover, if  $\bar{r} \in I := [\frac{l_m}{l_n + l_{nm}}, \frac{l_m + l_{nm}}{l_n}]$ , then

$$\alpha^*(\bar{r}) = \frac{\bar{r}l_n + \bar{r}l_{nm} - l_m}{l_{nm}(\bar{r} + 1)}, \text{ and } f(\alpha^*(\bar{r}), \bar{r}) = 0.$$

*Proof.* For any  $\bar{r} \geq 0$ ,  $\alpha^*(\bar{r})$  is simply the argument of the minimization of  $f(\alpha, \bar{r})$  with respect to  $\alpha$ , projected into the interval  $[0, 1]$ . In the case where  $\bar{r} \in I := [\frac{l_m}{l_n + l_{nm}}, \frac{l_m + l_{nm}}{l_n}]$ , we can easily check that the constraint  $\alpha^*(\bar{r}) \in [0, 1]$  is not activated, and then we do not need to project into the interval  $[0, 1]$ . Moreover, in this case,  $\alpha^*(\bar{r})$  cancels  $f(\alpha, \bar{r})$ .  $\square$

We notice here that the calculus of the optimal assignment proportion  $\alpha^*(\bar{r})$  of the vehicles going straight onto the lanes is done in deterministic at the macroscopic level (proportion of vehicular flow). The realization of the optimal assignment proportion  $\alpha^*(\bar{r})$  is done randomly at the microscopic level, as it has already been mentioned in section II.A.d: every vehicle going straight is randomly assigned to the the lanes M and N

with probabilities  $\alpha$  and  $(1-\alpha)$  respectively, assuming that the vehicles going straight that can choose both lanes will choose the shortest queue. By equilibrating the two queues with  $r^*(\alpha)$  or  $\alpha^*(\bar{r})$ , we can avoid spill-back onto the links of the network and by that reduce the risk of congestion. For example, the optimal  $r^*(\alpha)$  given in Proposition 3 can be taken into account as an additional constraint in the optimization problem of the traffic light split on every intersection, in such a way that the optimal traffic light setting will automatically balance the queue lengths on the incoming links of the intersections, which should help to avoid spill-back at the network level.

#### IV. SIMULATION RESULTS, EXAMPLES AND DISCUSSION

We present in this section the results of numerical simulations we conducted with Veins framework [19] which combines the microscopic road traffic simulator SUMO [22] with the communication simulator OMNET++ [23]. The road network is one simple junction with links composed of two incoming lanes described in Fig. 1. The junction is controlled by a traffic light with a cycle duration of 90 s. The traffic demand is coming from West towards East, North and South. We vary the arrival rates and turn ratios depending on the scenarios, as we mentioned in TABLE II. The messages we use to detect the location of the vehicles are the Basic Safety Messages (BSM) that are sent in broadcast by every probe vehicle. Given the road network topology and the data provided by the probe vehicles, we can know on which incoming/outgoing link each probe vehicle is located.

##### A. Primary parameters estimation

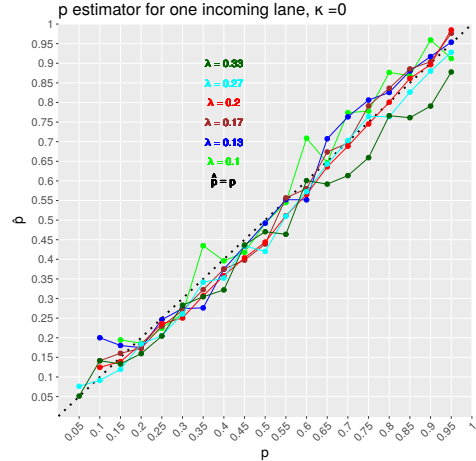


Fig. 3: Estimated penetration ratio  $p$  for a one lane incoming link, depending on penetration ratio for various demand scenarios. The arrival demand levels ( $\lambda$ ) are given in (vehicles/s). Simulated time = 40 min.

In this part, we illustrate estimation of primary parameters we proposed in section III-A. Fig. 3 represents the estimated penetration ratio  $\hat{p}$ , given by formula (5), associated in this figure to the real penetration ratio  $p$ , in the case of an incoming link of one lane. Fig. 4 represents the estimated penetration



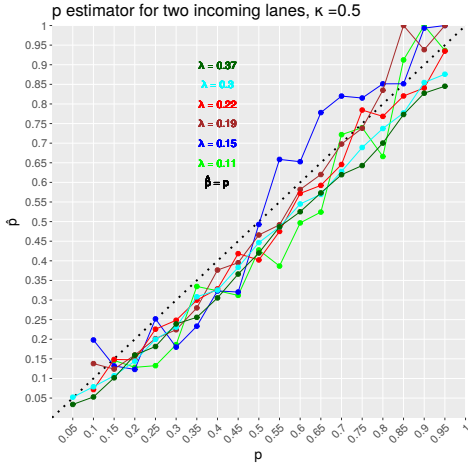


Fig. 4: Estimated penetration ratio  $p$  for a two lanes incoming link, depending on penetration ratio for various demand scenarios. The arrival demand levels ( $\lambda$ ) are given in (vehicles/s). Simulated time = 40 min.  $\kappa = 0.5$ .

ratio  $\hat{p}$ , given by formula (11), associated in this figure to the real penetration ratio  $p$ , in the case of an incoming link of two lanes. Ideally,  $\hat{p} = p$ , forming a line of slope 1 drawn in discontinuous black on the figure.

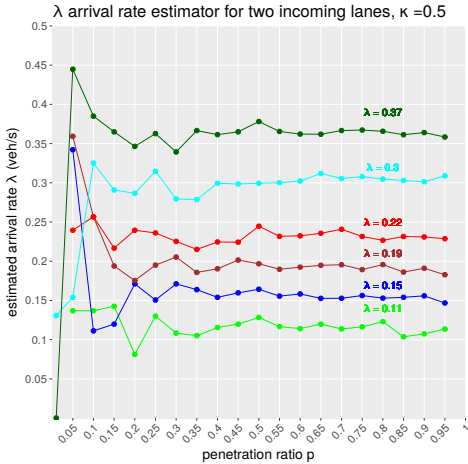


Fig. 5:  $(\lambda_n + \lambda_m + \lambda_{nm})$  in vehicles/second for a two lanes incoming link, depending on penetration ratio for various demand scenarios. Simulated time = 40 min.  $\kappa = 0.5$ .

Fig. 5 represents arrival rate estimated as given by formula (12) of section III-A in the case of an incoming link of two lanes for  $\kappa = 0.5$ . We can see here that the estimation is better when  $p$  gets higher. As there are more data and as the arrival rate is higher, the estimation of arrival rate is more accurate.

### B. Probability distributions

In this part, we illustrate the probability distribution queue lengths as proposed in section III-B. We assume the demand is coming from the West of the junction as described in TABLE II with  $r_N = r_M$  i.e.  $\bar{r} = 1$ .

These different scenarios include different possibilities concerning the demand such as : symmetric (Scenario S3) or asymmetric arrivals (scenarios S1, S2, S4, S5). We vary the arrival demand  $\lambda_n, \lambda_m$  and  $\lambda_{nm}$  and derive the optimal  $\alpha^*(1)$  by Proposition 4. We obtain the values  $\alpha^*(1) = [0.1, 0.25, 0.5, 0.75, 0.9]$ . We do not consider the cases where  $\bar{r} \neq 1$ .

Scenario	S1	S2	S3	S4	S5
$\alpha^*$	0.1	0.25	0.5	0.75	0.9
$\lambda_{nm}$	125	100	50	100	125
$\lambda_m$	200	125	200	75	100
$\lambda_n$	100	75	200	125	200
Arrival rates	Amount of vehicles for 1200 s				

TABLE II: Demand for different scenarios (simulated time=1200 s) and for  $r_N = r_M$

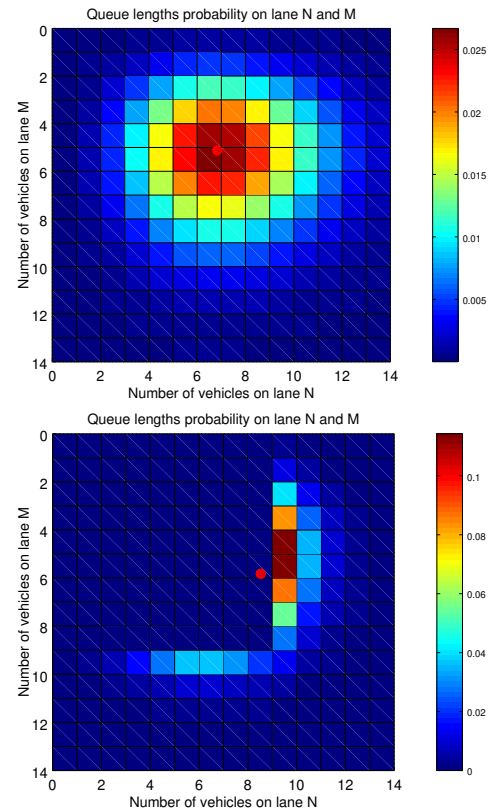


Fig. 6: On the top : probability law of Proposition 1. On the bottom : probability law of Proposition 2. Scenario with  $\lambda_n = 1/6, \lambda_m = 1/12, \lambda_{nm} = 1/24$  (veh/s)(thus  $\alpha^*(1) = 1$ ),  $r(t) = 41$  s,  $p = 0.55$ ,  $c_p = 8$ ,  $l_p = 9$ , at time  $t = 760$ s. Red dot is the expectation of the probability distribution. Simulation time=1200s.

a) Example: probability distribution of the two queues lanes without and with the information provided by the probe vehicles: We draw on top of Fig. 6 the probability distribution law  $P(N^t, M^t)$  (Proposition 1), and on bottom of Fig. 6 the conditional probability distribution law  $P(N^t, M^t | L_p, N_p + M_p)$  (Proposition 2) for scenario of Fig. 6. On top of Fig. 6, we can see that the total number of vehicles in the queue is estimated to  $(N^t = 6, M^t = 4)$ , for a total of 10 vehicles



in the queue. There is an asymmetry in the distribution probability law because of the asymmetric demand and because the common flow is not strong enough to equilibrate the two queue lanes. We can see on bottom of Fig. 6 that the conditional distribution clearly discriminates the two queue lengths and keeps track of the asymmetry. In this case, the parameters are  $p = 0.55$ ,  $c_p = 8$ ,  $l_p = 9$ . We compute  $\kappa = 0.75$ . Following the same ideas as above, we have  $\hat{p} = (8/1.75 - 1)/(9 - 1) = 0.45$ . It is probable that there are not many cars behind  $l_p$ , maybe 1 vehicle. Therefore, as  $l_p = 9$ , the biggest lane should contain around 10 vehicles. Given the asymmetry of the distribution law  $P(N^t, M^t)$ , the conditional probability  $P(N^t, M^t | L_p, N_p + M_p)$  will favour the lane with the highest arrival rate (lane  $N$ ). Then the queue on lane  $M$  should contain very few vehicles and will be around the same lane length estimation as in the top figure. In this example, the conditional distribution probability calculus emphasizes the asymmetry of the two lanes.

*b) Results for the scenarios of TABLE II:* For each scenario we measure the maximum and average queue lengths as estimated by SUMO microscopic road traffic simulator. We have used different seeds for the random number generator of the simulators, taking the average value for TABLE III and IV. We notice that SUMO queue length is measured in such a way that any vehicle with a speed greater than 0.1 m/s is not considered in the queue. We vary  $p \in [0, 1]$  for each scenario and compute :

- MAE(P2):=the mean absolute error between the estimated queue lengths as given by the estimator based on Proposition 2 and SUMO queue lengths on a subset of the data.
- MAE(P1):=the mean absolute error between the estimated queue lengths as given by  $\hat{N} = \mu_n$ ,  $\hat{M} = \mu_m$  (Proposition 1) and SUMO queue lengths on a subset of the data.
- MAE( $l_p$ ):=the mean absolute error between the estimated queue lengths as given by  $\max(\hat{N}, \hat{M}) = l_p$  and  $\min(\hat{N}, \hat{M}) = \kappa l_p$ , for the queue lengths N and M, and SUMO queue lengths on a subset of the data.

We give here the results for the scenarios of TABLE II in TABLE III and TABLE IV. We comment and emphasize some tendencies on the results. We notice on these tables that the error is decreasing as  $p$  tends to 1 : Proposition 2 is getting more accurate as more data is given in input. Even if the estimator based on  $l_p$  (MAE( $l_p$ )) is getting more accurate as  $p$  increases, we notice that the performances of our estimator based on Proposition 2 are generally better. Furthermore, if we compare MAE(P1) and MAE(P2) we notice that the performances of Proposition 2 based estimator are in general more accurate than the ones of the other estimators, especially when  $p$  increases. We notice finally that MAE( $l_p$ ) gives in

Results \ Scenarios	S1	S2	S3	S4	S5
Average SUMO Queue Length	3.54	3.21	4.33	3.21	4.77
Max SUMO Queue Length	10.07	10.07	14.12	9.71	15.95
$p = 0.05$ MAE(P2)	1.59	1.11	1.38	1.19	1.84
MAE(P1)	1.40	1.14	1.45	1.20	1.97
MAE( $l_p$ )	3.91	3.71	4.38	3.71	5.61
MAPE(P2,R)(%)	34.00	29.65	25.89	29.72	29.28
$p = 0.10$ MAE(P2)	1.51	1.02	1.29	1.12	1.75
MAE(P1)	1.40	1.14	1.45	1.20	1.97
MAE( $l_p$ )	2.01	2.28	1.96	2.13	2.60
MAPE(P2,R)(%)	31.64	26.50	23.35	26.05	26.45
$p = 0.15$ MAE(P2)	1.52	1.04	1.28	1.06	1.58
MAE(P1)	1.40	1.14	1.45	1.20	1.97
MAE( $l_p$ )	1.77	1.52	1.64	1.53	1.95
MAPE(P2,R)(%)	29.33	27.35	20.35	23.34	20.71
$p = 0.20$ MAE(P2)	1.47	0.96	1.20	0.98	1.43
MAE(P1)	1.40	1.14	1.45	1.20	1.97
MAE( $l_p$ )	1.85	1.28	1.40	1.16	1.44
MAPE(P2,R)(%)	24.71	25.40	19.52	19.78	20.63
$p = 0.50$ MAE(P2)	1.47	0.80	0.98	0.79	1.10
MAE(P1)	1.40	1.14	1.45	1.20	1.97
MAE( $l_p$ )	2.33	1.14	1.24	0.99	0.89
MAPE(P2,R)(%)	28.48	21.30	15.53	14.09	15.68
$p = 0.70$ MAE(P2)	1.33	0.73	0.93	0.70	1.13
MAE(P1)	1.40	1.14	1.45	1.20	1.97
MAE( $l_p$ )	2.48	1.25	1.30	0.97	0.90
MAPE(P2,R)(%)	24.24	17.19	15.68	12.36	14.74
$p = 0.90$ MAE(P2)	1.10	0.74	0.97	0.70	1.21
MAE(P1)	1.40	1.14	1.45	1.20	1.97
MAE( $l_p$ )	2.62	1.39	1.39	1.04	0.95
MAPE(P2,R)(%)	19.51	16.53	15.00	12.93	17.30

TABLE III: Results in the unit "number of vehicles" for queue on lane N, estimated vs SUMO queue length.

Results \ Scenarios	S1	S2	S3	S4	S5
Average SUMO Queue Length	4.79	3.39	4.20	2.89	3.84
Max SUMO Queue Length	15.22	9.71	14.11	8.61	12.28
$p = 0.05$ MAE(P2)	1.98	1.24	1.39	1.26	1.36
MAE(P1)	2.02	1.25	1.35	1.22	1.30
MAE( $l_p$ )	5.71	3.94	4.27	3.23	4.22
MAPE(P2,R)(%)	21.08	28.26	22.48	32.15	29.79
$p = 0.10$ MAE(P2)	1.89	1.17	1.30	1.25	1.37
MAE(P1)	2.02	1.25	1.37	1.22	1.30
MAE( $l_p$ )	2.81	2.53	1.93	1.88	1.93
MAPE(P2,R)(%)	17.38	24.20	19.45	30.06	27.66
$p = 0.15$ MAE(P2)	1.68	1.09	1.22	1.20	1.29
MAE(P1)	2.02	1.25	1.37	1.22	1.30
MAE( $l_p$ )	1.94	1.54	1.45	1.44	1.68
MAPE(P2,R)(%)	15.85	19.08	18.25	27.27	26.71
$p = 0.20$ MAE(P2)	1.55	1.05	1.19	1.12	1.28
MAE(P1)	2.02	1.25	1.37	1.22	1.30
MAE( $l_p$ )	1.54	1.25	1.33	1.20	1.58
MAPE(P2,R)(%)	17.17	21.62	17.82	24.47	31.12
$p = 0.50$ MAE(P2)	1.14	0.90	1.03	0.89	1.23
MAE(P1)	2.02	1.25	1.37	1.22	1.30
MAE( $l_p$ )	0.89	1.15	1.41	1.30	2.00
MAPE(P2,R)(%)	11.89	16.53	16.11	18.88	27.11
$p = 0.70$ MAE(P2)	1.23	0.82	0.96	0.77	1.14
MAE(P1)	2.02	1.25	1.37	1.22	1.30
MAE( $l_p$ )	0.96	1.13	1.44	1.40	2.16
MAPE(P2,R)(%)	13.90	15.71	14.25	13.47	25.14
$p = 0.90$ MAE(P2)	1.28	0.80	0.91	0.72	1.00
MAE(P1)	2.02	1.25	1.37	1.22	1.30
MAE( $l_p$ )	0.96	1.19	1.56	1.50	2.30
MAPE(P2,R)(%)	16.89	15.19	13.81	12.96	21.44

TABLE IV: Results in the unit "number of vehicles" for queue on lane M, estimated vs SUMO queue length.

general less accurate results than the estimators based on Propositions 1 and 2.<sup>1</sup>

We think that the main source of difference between our estimations and SUMO queue lengths is the assignment model. Indeed, drivers do not always choose the shortest queue for their assignment in SUMO. In fact, in SUMO, there is a kind of thresholds on the difference between the queue lengths, beyond which drivers choose the shortest queue. Moreover, the drivers do not have the same behavior, i.e. the choice is stochastic, i.e. the probability of taking the shortest queue increases with the threshold on the difference on the queue lengths. Another source of uncertainty and error is the variance of the arrival flows. Indeed, estimation of the queue length takes into account the average arrival flow  $\lambda$ , but in the simulation there is a variance of the time arrivals, which is then retrieved as an error of measurement.

We also give for information the order of magnitude of the following indicator in TABLE III and TABLE IV:  $MAPE(P2,R)$ :=the mean percentage absolute error between the estimated queue lengths as given by the estimator based on Proposition 2 and SUMO queue lengths on a subset of the data, at the end of the red time. Indeed, taking into account the MAPE at the beginning of the red time is meaningless (since the queues are not formed yet). This indicator rather emphasizes the difference between our assignment model  $\alpha^*$  and SUMO assignment model. This indicators varies between around 10% to 30% depending on the penetration ratio and the simulation scenarios.

In Fig. 7 and Fig. 8 we give the two lanes queue lengths in the scenario S4 where  $r_N = r_M$ . We estimate queue lengths for  $r_N > 0$  and  $r_M > 0$ . We notice that the estimation is more accurate as  $p$  gets higher.

### C. Traffic light control and vehicles assignment onto the lanes

In this part, we assume symmetric demand,  $\lambda_n = \lambda_m = \lambda_{nm} = 0.17$  vehicles/second. We assume the traffic light cycle includes a phase of 8 seconds, where green light is given to lane  $M$ , while red light is given to lane  $N$ . Hence, the red duration on lanes  $N$  and  $M$  are different :  $r_N \neq r_M$ ,  $\bar{r} \neq 1$ . We also assume  $\alpha = \alpha^*(\bar{r})$  given by Proposition 4.

On Fig. 9 we draw  $\bar{r}$  in blue and  $\alpha^*(\bar{r})$  in green depending on time. At the beginning of the cycle, the red is only for the lane  $N$ . The lane  $M$  is at this time with green light. Then, the two red lights are simply increasing as a line of slope 1. Starting from 20 seconds,  $\bar{r} \in I$  and  $\alpha^*(\bar{r}) = 1$ . This is because all the vehicles are assigned to lane  $M$  which is the shortest queue (we recall it was at green light until then). Then,  $\alpha^*(\bar{r})$  decreases slowly to reach approximately 0.6 which means that the two queue lengths are more equilibrated as the red durations on lanes  $N$  and  $M$  are getting less different. We notice that  $\alpha^*(\bar{r})$  would tend to  $1/2$  if the red time goes to infinity. Concerning  $\bar{r} = r_n/r_m$ , it is representing how the

<sup>1</sup>A significant source of error for this model is the fact that the estimation is done in real numbers, while the measured number of vehicles is done in integer ones. In the general case, the error due to discretization is about 0.5 vehicles. This is very big in the cases where the queue length is small, in particular at the beginning of the red time. If for example, the queue length is 0.5 vehicles in average, then we have 100% error, due only to the discretization.

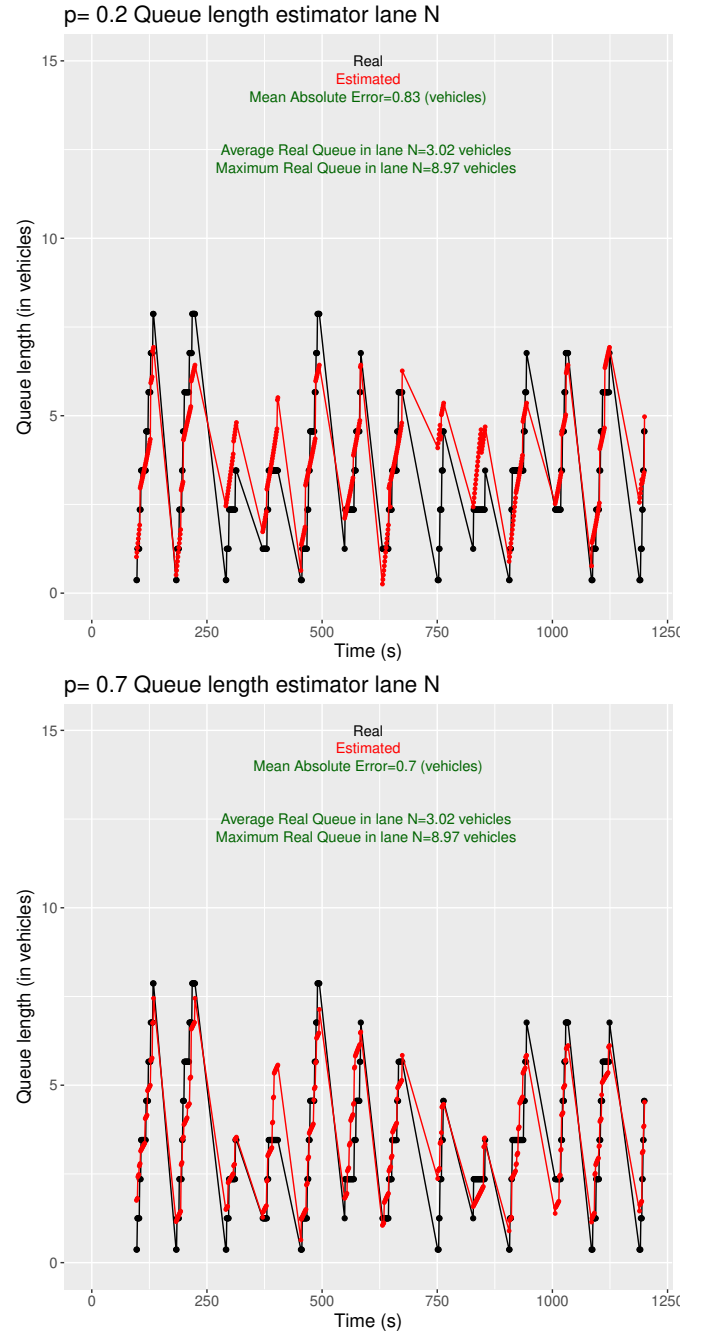


Fig. 7: Queue lengths estimator as given by Proposition 2, for varying  $p = 0.2$ ,  $p = 0.7$  and lane  $N$ ,  $\bar{r} = 1$

difference in the red lights durations is decreasing as time is going on.  $\bar{r}$  decrease is due to a constant offset (corresponding to the duration where lane  $M$  is at green light while lane  $N$  is at red light) which becomes less significant as the red light durations are increasing. We notice that  $\bar{r}$  would tend to 1 if the red time goes to infinity.

The interval  $I = \left[ \frac{\lambda_m}{\lambda_n + \lambda_{nm}}, \frac{\lambda_m + \lambda_{nm}}{\lambda_n} \right]$  as a function of  $\lambda_{nm}$  is represented on Fig. 10. We notice on Fig. 10 that as the common flow  $\lambda_{nm}$  gets higher, the interval  $I$  gets larger. Therefore, as the common flow  $\lambda_{nm}$  gets higher, there is more freedom to assign the vehicles onto a lane or another.

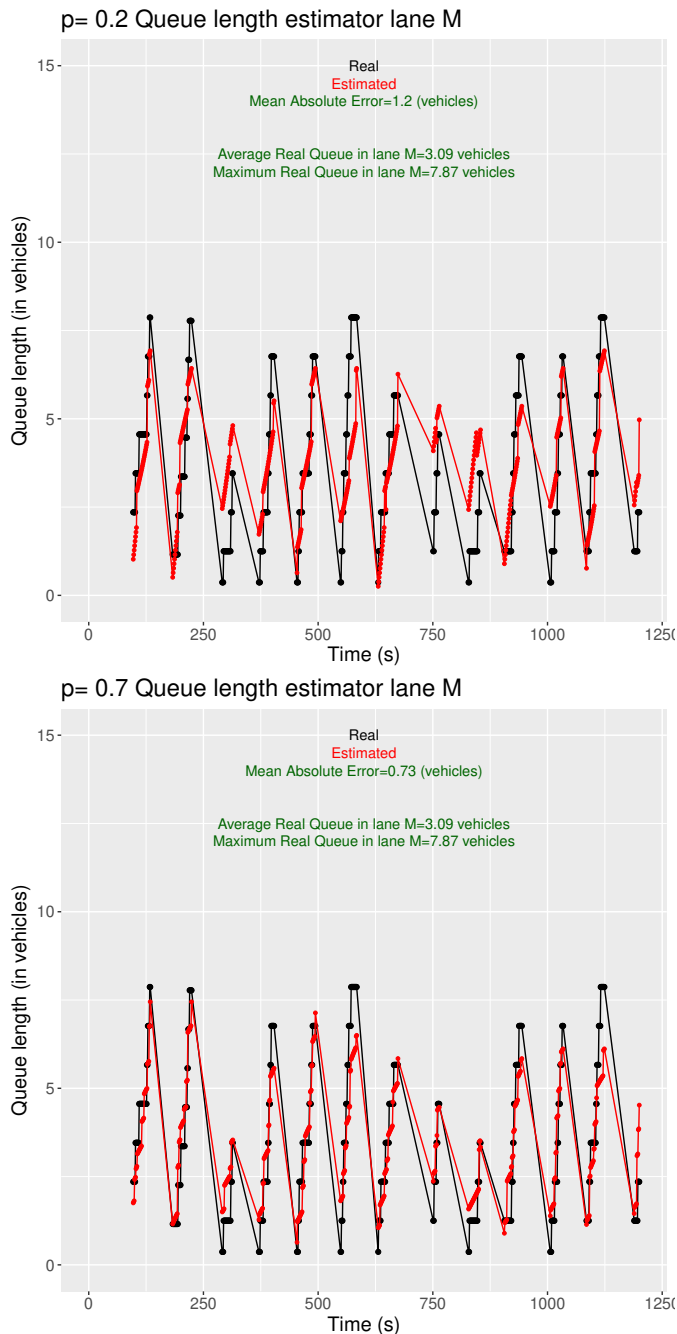


Fig. 8: Queue lengths estimator as given by Proposition 2, for varying  $p = 0.2$ ,  $p = 0.7$  and lane M,  $\bar{r} = 1$

#### D. Communication network performances

We give here some information on the communication network performances that we have measured in simulation. We have checked that the communication performances are not disturbing the estimation of the queue lengths. As the order of magnitude of the end-to-end-delay<sup>2</sup> is very low (around 0.2 ms), we don't expect significant consequence on

<sup>2</sup>The end-to-end delay is a communication indicator of performance that measures the delay from the time a message is sent from a communicating vehicle until the time it is received by the receiver (in our case the receiver is the RSU).

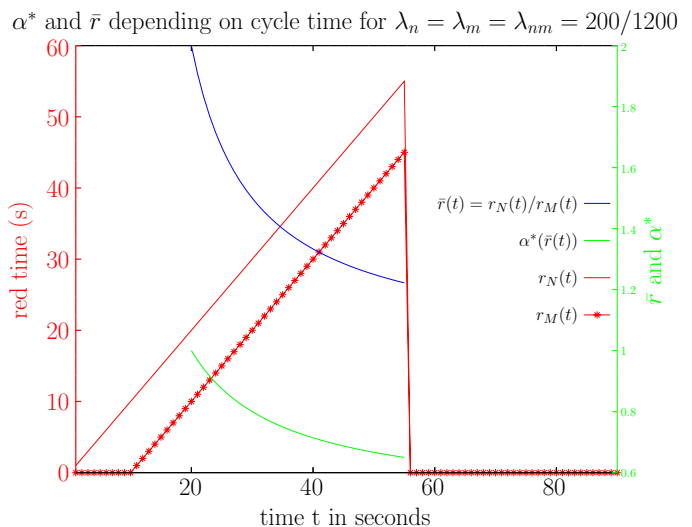


Fig. 9:  $\alpha^*$  and  $\bar{r}$  depending on cycle time of 90s for  $\lambda_n = \lambda_m = \lambda_{nm} = \frac{200}{1200} = 0.17$  vehicles/second.

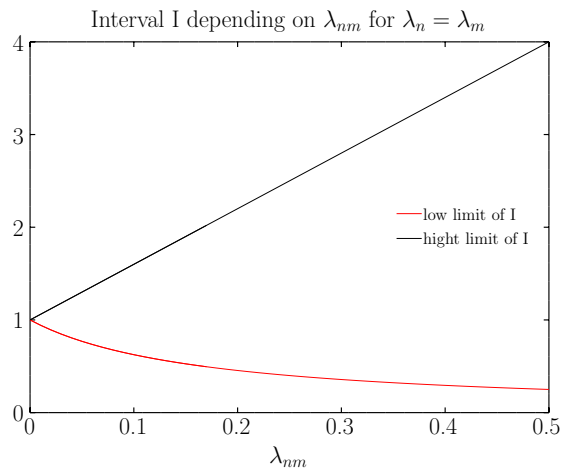


Fig. 10:  $I = \left[ \frac{\lambda_m}{\lambda_n + \lambda_{nm}}, \frac{\lambda_m + \lambda_{nm}}{\lambda_n} \right]$  depending on  $\lambda_{nm}$  for  $\lambda_n = \lambda_m = 1/6$

the queue length estimation application as it could happen in scenarios where more vehicles would communicate, and cause significant delays such as described in [16].

#### V. CONCLUSION AND PERSPECTIVES

In this paper, we have proposed a method for the estimation of urban traffic state. We give estimations for the penetration ratio of probe vehicles and for the vehicles arrival rate, on any link of the road network. Knowing the arrival rate of the incoming flow and its composition, we have computed the joint probability distribution of the queue lengths in the case of two lanes link. For this purpose, we have proposed a simple assignment model of vehicles onto the lanes. In addition, we have refined the probability distribution of the queue lengths with the information provided by the probe vehicles. A control of the traffic light has been proposed in order to balance

the queues of the two lanes. Moreover, we have proposed a formula for computing the optimal assignment of the vehicles onto the lanes. Numerical simulations have been conducted with Veins framework, and the work presented here has been evaluated. Road traffic control could benefit from the queue length estimations we presented in the present paper, in order to improve travel conditions. We think the ideas we have given in this paper could be extended to a link of any number of lanes.

## REFERENCES

- [1] I. Lin, R. He, and A. L. Kornhauser, "Estimating Nationwide Link Speed Distribution Using Probe Position Data," *Journal of Intelligent Transportation Systems*, vol. 12, no. 1, pp. 29–37, 2008. [Online]. Available: <https://www.tandfonline.com/doi/abs/10.1080/15472450701849667>
- [2] G. Comert, "Queue length estimation from probe vehicles at isolated intersections: Estimators for primary parameters," *European Journal of Operational Research*, vol. 252, no. 2, pp. 502 – 521, 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0377221716000850>
- [3] P. Varaiya, "Max pressure control of a network of signalized intersections," *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 177–195, 2013. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0968090X13001782>
- [4] A. J. Miller, "Settings for fixed-cycle traffic signals," *Journal of the Operational Research Society*, vol. 14, no. 4, pp. 373–386, 1963.
- [5] M. Ramezani and N. Geroliminis, "Exploiting probe data to estimate the queue profile in urban networks," in *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, Oct 2013, pp. 1817–1822.
- [6] D. Heidemann, "Queue length and delay distributions at traffic signals," *Transportation Research Part B: Methodological*, vol. 28, no. 5, pp. 377 – 389, 1994. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/0191261594900361>
- [7] F. Viti and H. van Zuylen, "Modeling queues at signalized intersections," *Transportation Research Record: Journal of the Transportation Research Board*, no. 1883, pp. 68–77, 2004.
- [8] G. Comert and M. Cetin, "Queue length estimation from probe vehicle location and the impacts of sample size," *European Journal of Operational Research*, vol. 197, no. 1, pp. 196 – 202, 2009. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0377221708004918>
- [9] M. R. Shahrabaki, A. A. Safavi, M. Papageorgiou, and I. Papamichail, "A data fusion approach for real-time traffic state estimation in urban signalized links," *Transportation Research Part C: Emerging Technologies*, vol. 92, pp. 525 – 548, 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0968090X18305321>
- [10] J. Zheng and H. X. Liu, "Estimating traffic volumes for signalized intersections using connected vehicle data," *Transportation Research Part C: Emerging Technologies*, vol. 79, pp. 347 – 362, 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0968090X17300827>
- [11] K. Tiaprasert, Y. Zhang, X. B. Wang, and X. Zeng, "Queue Length Estimation Using Connected Vehicle Technology for Adaptive Signal Control," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 2129–2140, Aug 2015.
- [12] S. Lee, S. Wong, and Y. Li, "Real-time estimation of lane-based queue lengths at isolated signalized junctions," *Transportation Research Part C: Emerging Technologies*, vol. 56, pp. 1–17, 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0968090X15000984>
- [13] S. Y. R. Rompis, M. Cetin, and F. Habtemichael, "Probe vehicle lane identification for queue length estimation at intersections," *Journal of Intelligent Transportation Systems*, vol. 22, no. 1, pp. 10–25, 2018. [Online]. Available: <https://doi.org/10.1080/15472450.2017.1300887>
- [14] M. J. Lighthill and J. B. Whitham, "On kinematic waves ii. a theory of traffic flow on long crowded roads," *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 229, no. 1178, pp. 317–345, 1955. [Online]. Available: <http://rspa.royalsocietypublishing.org/content/229/1178/317>
- [15] P. I. Richards, "Shock waves on the highway," *Operations Research*, vol. 4, no. 1, pp. 42–51, 1956. [Online]. Available: <https://doi.org/10.1287/opre.4.1.42>
- [16] C. N. V. Phu, N. Farhi, H. Haj-Salem, and J. P. Lebacque, "A vehicle-to-infrastructure communication based algorithm for urban traffic control," in *2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, June 2017, pp. 651–656.
- [17] N. Farhi, C. N. V. Phu, M. Amir, H. Haj-Salem, and J. P. Lebacque, "A semi-decentralized control strategy for urban traffic," *Transportation Research Procedia*, vol. 10, pp. 41 – 50, 2015, 18th Euro Working Group on Transportation, EWGT 2015, 14-16 July 2015, Delft, The Netherlands. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S2352146515002410>
- [18] F. van Diggelen and P. Enge, "The world's first gps mooc and worldwide laboratory using smartphones," in *Proceedings of the 28th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2015)*, 2015, pp. 361–369.
- [19] C. Sommer, R. German, and F. Dressler, "Bidirectionally Coupled Network and Road Traffic Simulation for Improved IVC Analysis," *IEEE Transactions on Mobile Computing*, vol. 10, no. 1, p. 3–15, January 2011.
- [20] "Ieee guide for wireless access in vehicular environments (wave) - architecture," *IEEE Std 1609.0-2013*, pp. 1–78, March 2014.
- [21] R. Gallager, *Stochastic Processes: Theory for Applications*, ser. Stochastic Processes: Theory for Applications. Cambridge University Press, 2013.
- [22] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker, "Recent development and applications of SUMO - Simulation of Urban MObility," *International Journal On Advances in Systems and Measurements*, vol. 5, no. 3&4, pp. 128–138, December 2012.
- [23] A. Varga, "The omnet++ discrete event simulation system," in *In ESM'01*, 2001.

## APPENDIX A

### CALCULUS OF THE BIAS OF $p$ ESTIMATOR

We consider queue lengths on lanes  $N$  and  $M$  respectively equal to  $n$  and  $m$ . We propose :

$$\hat{p} = \frac{c_p}{n + m} \quad (22)$$

By the way, in our case, the length  $n$  of queue  $N$  can be estimated with the number of arrivals on lane  $N$  during  $r_N(t)$  which is  $\mu_n$ . As  $\mu_n + \mu_m = \max(\mu_n, \mu_m) + \min(\mu_n, \mu_m)$  and by estimating  $\max(\mu_n, \mu_m) = l_p + i_c/p$ , where  $0 \leq i_c \leq 1$  and  $i_c/p$  represents the backlog of the queue (unequipped vehicles following  $l_p$ ). We want to determine  $0 \leq i_c \leq 1$  and compute  $\hat{p}$  such that the estimation of  $p$  is without bias. We can write :

$$\mu_n + \mu_m = \max(\mu_n, \mu_m) \left( 1 + \frac{\min(\mu_n, \mu_m)}{\max(\mu_n, \mu_m)} \right) \quad (23)$$

$$\mu_n + \mu_m = (l_p + i_c/p)(1 + \kappa) \quad (24)$$

We introduce  $c_\kappa = c_p/(1 + \kappa)$  and replace (24) in (22). Finally, we get the following equation :

$$\hat{p} = \frac{c_\kappa}{l_p + i_c/p} \quad (25)$$

$$\hat{p} = (c_\kappa - i_c)/l_p \quad (26)$$

We know from [2] how to compute the expectation of  $\hat{p}$  and we follow the same ideas below :

$$\begin{aligned} P(L_p^t = l_p, N_p^t + M_p^t = c_p) &= P(N_p^t + M_p^t = c_p | L_p^t = l_p) \\ &P(L_p^t = l_p) \end{aligned} \quad (27)$$

$$P(L_p^t = l_p, N_p^t + M_p^t = c_p) = \binom{l_p - 1 + \min(l_p, n, m)}{c_p - 1} p^{c_p - 1} (1 - p)^{l_p - 1 + \min(l_p, n, m)} P(L_p^t = l_p) \quad (28)$$

Solving this equation gives :

$$i_c = 1 - p \quad (36)$$

Finally, we replace  $i_c$  in (26) :

$$\hat{p} = \frac{c_\kappa - (1 - p)}{l_p} \quad (37)$$

$$\hat{p} = \frac{c_\kappa - 1}{l_p - 1} \quad (38)$$

$$\mathbb{E}\left(\frac{c_p/(1 + \kappa) - i_c}{l_p}\right) = \sum_{l_p \geq 1} \sum_{c_p=1}^{l_p - 1 + \min(l_p, n, m)} \frac{c_p/(1 + \kappa) - i_c}{l_p} \binom{l_p - 1 + \min(l_p, n, m)}{c_p - 1} p^{c_p - 1} (1 - p)^{l_p - 1 + \min(l_p, n, m)} P(L_p^t = l_p) \quad (29)$$

$$= \sum_{l_p \geq 1} \frac{1}{l_p(1 + \kappa)} \sum_{c_p=1}^{l_p - 1 + \min(l_p, n, m)} (c_p - 1 + 1 - i_c(1 + \kappa)) \binom{l_p - 1 + \min(l_p, n, m)}{c_p - 1} p^{c_p - 1} (1 - p)^{l_p - 1 + \min(l_p, n, m)} P(L_p^t = l_p) \quad (30)$$

To derive the next equation we use two arguments :

- the expectation of a binomial probability distribution law  $\mathbb{E}(\mathbb{B}(n_x, p)) = n_x p$  with  $n_x = l_p - 1 + \min(l_p, n, m)$  in our case.
- and the formula of Newton  $(a + b)^{m_x} = \sum_{k=0}^{m_x} \binom{m_x}{k} a^k b^{m_x - k}$ , with  $a = p$  and  $b = 1 - p$  in our case.

$$\mathbb{E}\left(\frac{c_p/(1 + \kappa) - i_c}{l_p}\right) = \sum_{l_p \geq 1} \frac{1}{l_p(1 + \kappa)} (p(l_p - 1 + \min(l_p, n, m)) + 1 - i_c(1 + \kappa)) P(L_p^t = l_p) \quad (31)$$

We replace  $\min(l_p, n, m) = \kappa \max(l_p, n, m) = \kappa(l_p + i_c/p)$  :

$$\mathbb{E}\left(\frac{c_p/(1 + \kappa) - i_c}{l_p}\right) = \sum_{l_p \geq 1} \frac{1}{l_p(1 + \kappa)} (p(l_p - 1 + \kappa(l_p + i_c/p)) + 1 - i_c(1 + \kappa)) P(L_p^t = l_p) \quad (32)$$

$$\mathbb{E}\left(\frac{c_p/(1 + \kappa) - i_c}{l_p}\right) = \frac{p}{1 + \kappa} \mathbb{E}\left(\frac{l_p - 1}{l_p}\right) + \frac{p\kappa}{1 + \kappa} + \frac{i_c\kappa}{1 + \kappa} \mathbb{E}\left(\frac{1}{l_p}\right) + \frac{1 - i_c(1 + \kappa)}{1 + \kappa} \mathbb{E}\left(\frac{1}{l_p}\right) \quad (33)$$

$$\mathbb{E}\left(\frac{c_p/(1 + \kappa) - i_c}{l_p}\right) = p + \frac{\mathbb{E}\left(\frac{1}{l_p}\right)}{1 + \kappa} (1 - p - i_c) \quad (34)$$

To get an estimator without bias, we write :

$$\mathbb{E}\left(\frac{c_p/(1 + \kappa) - i_c}{l_p}\right) = p \quad (35)$$