New Insights from the Analysis of Free Flow Vehicular Traffic in Highways

Marco Gramaglia*†, Pablo Serrano†, José A. Hernández†, Maria Calderon†, Carlos J. Bernardos†
* Institute IMDEA Networks, Spain
Email: marco.gramaglia@imdea.org

†Universidad Carlos III de Madrid, Spain
Email:{pablo, jahgutie, maria, cjbc}@it.uc3m.es

***Abstract*—Building vehicular networks in roads and highways is a challenging research topic with a large number of applications ranging from the prevention of traffic jams and car collisions to efficient route planning. Analyzing the distance between vehicles in roads is a key factor in, e.g., designing vehicular networks protocols or planning a supporting infrastructure to improve connectivity.**

This work proposes a Gaussian-exponential mixture model to characterize the time distance between vehicles in a highway lane, based on measurements collected at different locations in several highways of the city of Madrid. The model arises from the observed behavior that some vehicles travel very close together, like in a burst mode, showing Gaussian inter-arrival times, while other vehicles are isolated, showing exponentially distributed inter-arrival times. The experiments show that such a Gaussian-exponential mixture model accurately characterizes inter-vehicle times as observed from real traces.

I. Introduction

Very likely in the future, vehicles will be provided with communications equipment that not only allows them to share traffic-related information but also provide a vehicular network infrastructure with services like in a ubiquitous computing fashion. A number of suitable applications for such a vehicular network have already been proposed in the literature, ranging from exchanging information between vehicles to prevent traffic jams and car collisions [1]. Other applications already proposed in the literature include local hazard warning, efficient route planning and coordination of traffic flows [2].

Indeed, both inter-vehicle (or vehicle-to-vehicle) [3] and vehicle-to-infrastructure communication [4] have been an active research topic in the past few years, and a number of working groups and standardization bodies have already put the first stones towards defining a common framework in this communications arena, see for instance IEEE 1609[1], ISO TC204[2], ETSI TC ITS[3] and the Car to Car Communications Consortium[4] (C2C-CC).

However, before all such applications become true, it is necessary to identify the requirements to build a vehicular network efficiently and effectively. Thus, characterizing the distribution of vehicles and their speed in highways is a required step in order to be able to design mechanisms that can operate efficiently in the vehicle environment and that can cope with its particular nature. Analyzing the connectivity level of a vehicular network requires to build a model of vehicular traffic from real measurements. Depending on which technology the vehicles use to communicate among themselves, it is necessary one model or another. When vehicles communicate using wireless short-range technologies (i.e., IEEE 802.11p [5]), the traffic model must consider the aggregation of vehicles from different lanes in a highway. However, for the case of infrared communications [6], or the recent cases of visible light communications [7], [8], a model for single-lane traffic is necessary.

A number of previous modeling studies have attempted to characterize traffic inter-arrival times in the past [9], [10]. For instance, the authors in [10] claim that vehicle inter-arrival times follow an exponential distribution, while the authors in [9] consider the log-normal distribution as the most suitable one to characterize inter-vehicular times. Furthermore, these two studies also assume that inter-arrival times are independent and identically distributed random variables. However, as we will show in the paper, this is not the case as actually there are some dependences between consecutive vehicles, especially at highways with dense traffic.

Indeed, we show from our traces that many vehicles travel close together like in a burst fashion, while others are isolated. Such bursty vehicles exhibit Gaussian inter-arrival times, while the others show exponentially distributed times. Based on these observations, this work proposes an exponential-Gaussian mixture model to characterize the inter-arrival times of vehicles, as observed from real traffic measurements collected at several highways in the surroundings of the city of Madrid.

The rest of this article is organized as follows: Section II briefly describes the measurement set under study. Section III performs a deep analysis of the traces, showing the basic facts that motivates the Gaussian-exponential mixture model. Section IV introduces the model and confirms its improved accuracy to model the behavior observed in the traces. Finally, Section V summarizes the main findings of this work and describes our new research directions.

[1]http://www.standards.its.dot.gov/fact_sheet.asp?f=80
[2]http://www.isotc204wg16.org/
[3]http://portal.etsi.org/its/its_tor.asp
[4]http://www.car-to-car.org/

Fig. 1. Location of the highways and the measurement points.

II. Data set description

Table I summarizes the data set used in this work. This comprises $M = 15$ traces collected at different locations on four important highways in the surroundings of the city of Madrid, in Spain. Note that the traces are presented in decreasing order of traffic, in terms of vehicles per hour. The considered highways are:

- M-40 and M-50 which are two orbital highways that encircle the City of Madrid. Their speed limits are 100 and 120 km/h respectively.
- A-6 is a highway linking the north-west of Spain and Madrid. Its speed limit is 120 km/h.
- A-1 is a highway linking the north-east of Spain and Madrid. Its speed limit is also 120 km/h.

Fig. 1 shows the four highways along with the location of the measuring points (A to E in the figure).

The traces were collected in the mornings of four consecutive working days, at two different time intervals: from 8.30 to 9.00 AM and from 11.30 to 12.00 AM. For trace j ($j = 1, \ldots, 15$), the i-th vehicle measured ($i = 1, \ldots, N_j$, where N_j refers to the number of samples in the j-th trace) generates the following values: arrival time T_i, speed v_i and lane in the highway. In the following, t_i shall denote the inter-arrival time between the i-th and the $i-1$-th vehicles (i.e., $t_i = T_i - T_{i-1}$). Hence, the i-th measurement of the j-th trace comprises the tuple $\{v_i, t_i\}_j$ (we discard the sample corresponding to the first arriving vehicle of the trace as its inter-arrival time cannot be defined). Note that, according to the classification in [9], these traffic traces can be considered as free-flow non-rush hour traffic with moderate traffic volume and high speed.

TABLE I
Data set collection.

Trace #	Highway [location in Fig. 1]	Veh/h	Avg. Speed [Km/h]	Lane
1	M-40 [D]	2038	102.64	L
2	A-6 [B]	1800	65.62	R
3	A-6 [B]	1744	68.20	R
4	A-6 [B]	1688	68.33	R
5	A-6 [B]	1522	85.88	L
6	A-6 [B]	1506	85.82	L
7	A-6 [A]	1466	89.13	C
8	A-6 [B]	1090	77.38	C
9	M-50 [E]	914	81.47	C
10	M-50 [E]	888	108.97	L
11	M-50 [E]	684	97.33	R
12	A-6 [B]	593	87.61	R
13	A-1 [C]	482	97.71	R
14	A-1 [C]	462	96.08	R
15	A-1 [C]	462	95.42	R

III. Analysis of the arrival process

In this section we start by analyzing the vehicle arrival process at a given lane in a highway for some representative cases of Table I. We first analyze the vehicle inter-arrival times t_i, and then we take advantage of the travel speeds v_i to build a model that is able to both mimic the figures resulting from the traces and provide some insight on drivers' behavior.

A. Is the arrival process a Poisson process?

We first focus on the inter-arrival times between vehicles in a highway lane. This is, we are interested in analyzing whether the vehicle arrival process can be modeled by a Poisson process or not. Our motivation is that the classical "Poissonian assumption" is very common in the literature, partially because of its inherent analytical tractability (see, e.g., [9], [11]). For the arrival process to be Poissonian, the inter-arrival times should follow an exponential random variable [12]. In order to characterize the time between vehicles t_i's, then, we first analyze if their empirical distribution function matches the one from an exponential.

The cumulative distribution function (CDF) F of an exponential random variable of mean λ^{-1} is given by

$$F(t) = 1 - e^{-\lambda t}.$$

Therefore, for ease of presentation it is often better to used the Complementary CDF (CCDF), as its logarithm follows a straight line with slope $-\lambda$, i.e.,

$$\log(1 - F(t)) = -\lambda t.$$

In order to analyze the time between cars, we first normalize each of the traces from Table I by their respective means λ_i^{-1}, and then compute the empirical CCDF. We plot the resulting figures in logarithmic scale in Fig. 2 using dashed lines for Traces 1–5, along with the theoretical CCDF of an exponential random variable with $\lambda = 1$ using a solid line. It can be seen that the experimental data largely deviates from the exponential random variable, and therefore *we must reject the hypothesis that vehicle arrival times follow a Poisson*

distribution. In order to have more statistical confidence of this result, we perform a Kolmogorov-Smirnoff (K-S) goodness of fit test [13] on the complete set of data, which rejects the hypothesis of exponential inter-arrival times with 95% confidence *in all the cases* from Table I.

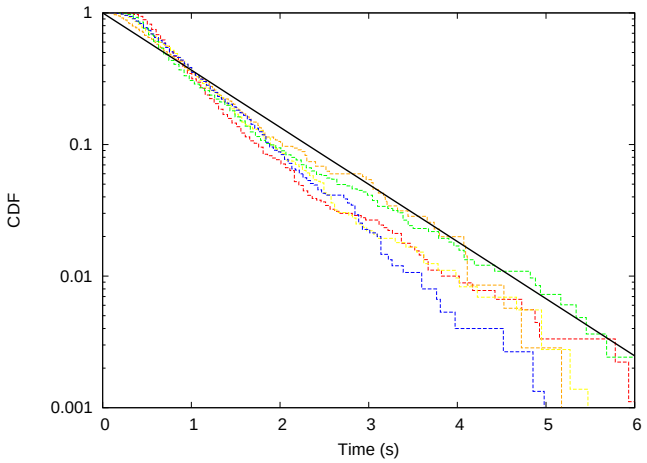


Fig. 2. Empirical CCDFs of the normalized time between arrivals (dashed lines) and theoretical CCDF of an exponential distribution (solid line).

In addition to the above finding, there are two other observations that can be derived from Fig. 2. First, inter-arrival times seems to be *shifted*, i.e., for each trace there seems to be a minimum value of t before the CCDF starts to decrease. Given that we are analyzing the arrival process in a highway lane, this constitutes a quite expected result, as vehicles do not overlap. Second, the experiments also show that, after a time threshold around 2 and 3 s, the experimental CCDFs show an exponential decay (i.e., a straight line in the log CCDF plot), which suggests that the exponential behavior is still somehow present in the traces.

B. On the dependences between consecutive vehicles

Note that, had the CDF of the inter-arrival times resembled that of an exponential random variable, this would *not* have implied that the times between vehicles t_i are independent and identically distributed (i.i.d.) random variables. Indeed, it is often observed that cars travel together following some sort of *bursts*, even in non-congested highways. Actually, it is quite common to observe this phenomenon when a number of cars are following a slow moving vehicle (e.g., a truck) that cannot be overtaken easily. In these situations the distances between vehicles are typically short, and *their traveling speeds are similar*, a behavior that has to introduce some dependences in the arrival process.

In order to analyze if this behavior can be observed in our traces, we compute the correlation plot of the speed sequences v_i, this resulting in Fig. 3. From the figure, we can see that i) for some traces, there exists a noticeable correlation between speeds, and ii) such correlation in some cases is particularly high for consecutive cars (Lag=1).

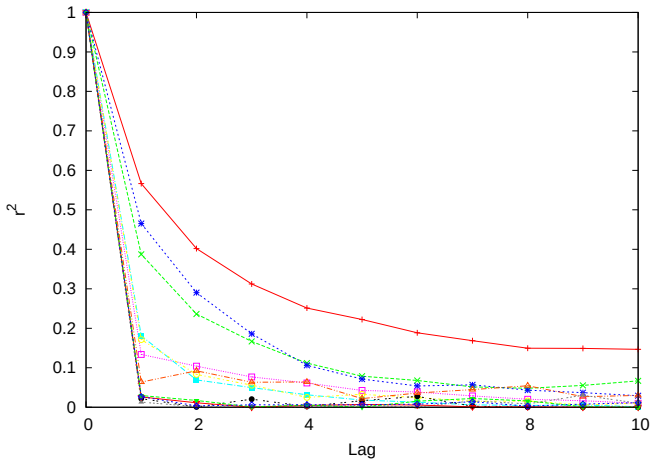


Fig. 3. Correlation plot of the sequences of speeds v_i.

Based on these results, we conjecture that, for some traces, some vehicles travel like bursts (i.e., relatively close together with similar speeds), while others are somehow isolated from the rest (i.e., with large distances between them and no dependences with the previous or next car). We will analyze with more detail this behavior in the next section.

C. Identifying bursts of vehicles

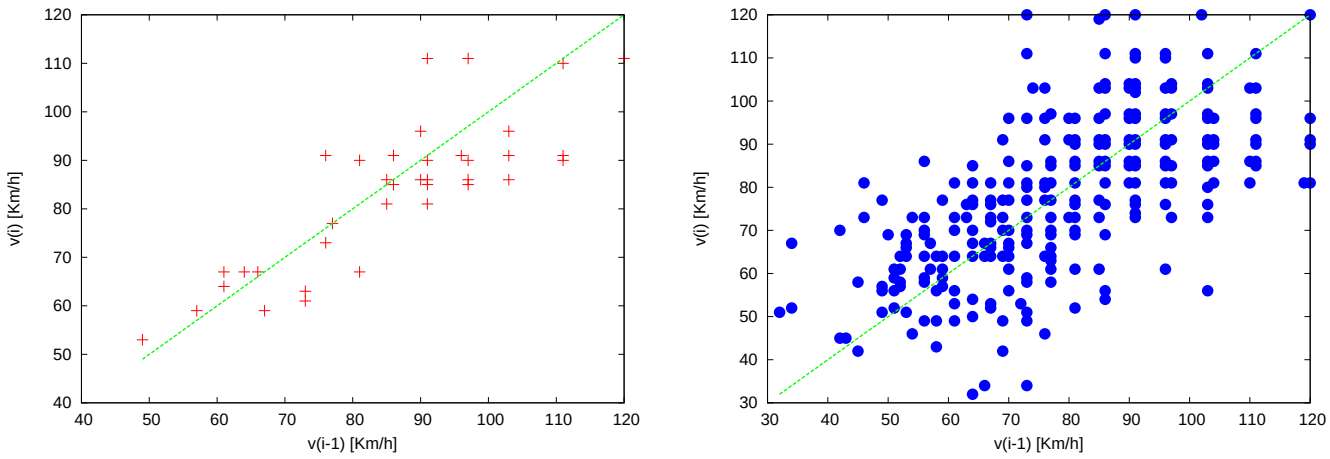
We claim that vehicles belonging to the same burst can be identified because they travel at a similar speed to that of the head of the burst, and the distances between them are typically short. To prove this assumption, we analyze the relationship between consecutive speeds, i.e., v_i and v_{i+1}, depending on the time t_{i+1} between them. We proceed as follows. Let δ denote a time threshold. Based on this δ, for each trace we construct two sets of points: in one set (denoted as $S_{\leq\delta}$) we put those speed pairs from cars that were relatively close, while in the other set (denoted as $S_{>\delta}$) we put the other pairs, i.e.,

$$(v_i, v_{i+1}) \in S_{\leq\delta} \text{ if } t_{i+1} \leq \delta,$$
$$(v_i, v_{i+1}) \in S_{>\delta} \text{ if } t_{i+1} > \delta.$$

Fig. 4 shows the corresponding scatter plots of speeds for the case of Trace 9 with $\delta = 1\ s$. That is, for all pairs of vehicles that are separated by less than one second we plot in Fig. 4(a) their speeds vs. the speed of its predecessor, while we do the same for vehicles traveling more than one second apart in Fig. 4(b). It can be seen that, for the case of $S_{\leq\delta}$, consecutive vehicles have similar speeds (the scatter plot is clearly placed around the $y = x$ line), while for the case of $S_{>\delta}$ this behavior cannot be observed.

In order to quantify the speed similarity, we compute the Pearson product-moment correlation coefficient r for each data set. For the case of $S_{\leq\delta}$ the resulting value is $r = 0.84$, which can be considered as a clear indicator of a "strong" correlation[5], while for the case of $S_{>\delta}$ the resulting value is

[5]At least, according to the somehow arbitrary thresholds proposed in Cohen, J. (1988), Statistical power analysis for the behavioral sciences (2nd ed.) Hillsdale, NJ. : Lawrence Erlbaum Associates.



(a) Scatter plot of $S_{\leq\delta}$

(b) Scatter plot of $S_{>\delta}$

Fig. 4. Scatter plot of the speed of consecutive vehicles (v_{i-1}, v_i) depending on the time distance t_i.

noticeable smaller, $r = 0.6$, although still non negligible[6]. Hence, *the time between vehicles and their speeds are not independent random variables* since, as shown, vehicles that are close together (at least, for the case of $t_i \leq 1\ s$) present high speed correlation.

So far, we have only studied the case of $\delta = 1\ s$ for a given trace. In order to analyze the impact of this parameter, and whether we can observe a similar behavior in other traces or not, we perform a sweep on δ between $0.1\ s$ and $6\ s$, and for each delta value we compute the correlation parameter r^2 for the data set $S_{\leq\delta}$. We plot the resulting values in Fig. 5, where the 6 traces with the highest density (in terms of vehicles per distance) are plot in the left figure, and the other traces are plot in the right figure. The results can be summarized as follows:

- For the case of Fig. 5(a) (dense traces), when distances between vehicles are small, i.e., $\delta < 1.5\ s$, there is a relatively large correlation between speeds.
- In the same figure, when $\delta \in (1.5, 2.5)$ there is a steep descent in r^2.
- Finally, for the case of dense traces, when $\delta > 2.5\ s$ the behavior of r is relatively flat and independent to changes on δ.
- On the other hand, for the case of Fig. 5(b) (small densities), there is no apparent correlation between consecutive speeds regardless of the value of the parameter δ, as the behavior of r^2 is practically flat.

These results, derived from the use of formal statistical tools, have actually a very intuitive explanation: for the case of sparse lanes, drivers are quite unlikely to travel in bursts, and therefore there is no apparent correlation between the speeds of consecutive vehicles regardless of the time distances between them. On the other hand, for the case of dense lanes vehicles are more likely to travel in bursts, and therefore the correlation values are large if the relative distances are small. However, still in this case there is a threshold value for this distance ($\delta \approx 2.5\ s$), that once crossed the relation between one vehicle and its predecessor vanishes.

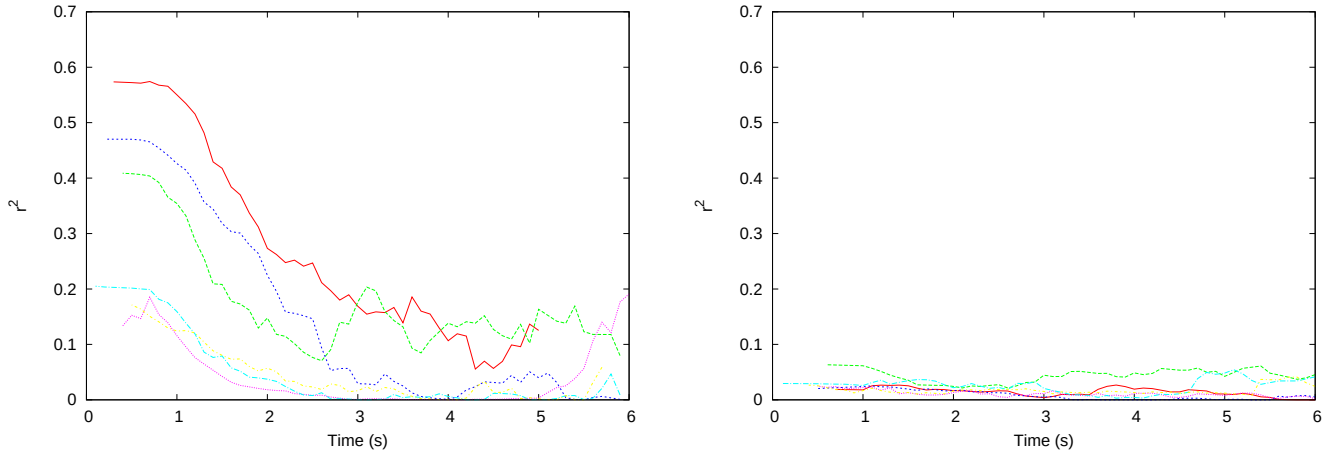
Based on the above, we could label vehicles as *bursty* or *isolated*, depending on whether they are traveling together (with similar speeds) or not. In order to look insight the underlying distribution of the arrival process, we analyze the distribution of the inter-arrival time of Fig. 4, but dividing the t_i's into two sets according to the identified threshold $\delta_t = 2.5\ s$: in set τ_B we put those values below this δ_t threshold (i.e., the ones that correspond to bursty arrivals), while in set τ_I we put the values above the threshold (the ones corresponding to isolated arrivals),

$$t_i \in \tau_B \text{ if } t_i \leq 2.5\ s,$$

$$t_i \in \tau_I \text{ if } t_i > 2.5\ s.$$

The resulting histograms for each set of data are plot in Fig. 6. According to these histograms, we further confirm the behavior that we have identified, i.e., there is a δ threshold that separates two very different types of arrivals. On one side, we have that when vehicles are relatively separated (τ_I) the empirical distribution seems to match the probability distribution function (PDF) of an exponential random variable, which we plot in a continuous line in Fig. 6(b). This way, we have that when vehicles are not traveling in bursts the arrival process could match a Poisson process. On the other side, however, we observe that when the time distances are relatively short (τ_B) the empirical distribution does not resemble that of an exponential random variable, but instead it is similar to the PDF of a normal variable, which we plot in a continuous line in Fig. 6(a). Note that, because of the way we have performed the division into τ_I and τ_B, both random variables are truncated (more specifically, the normal variable is truncated between 0 and 2.5 s, while the exponential variable is shifted 2.5 s).

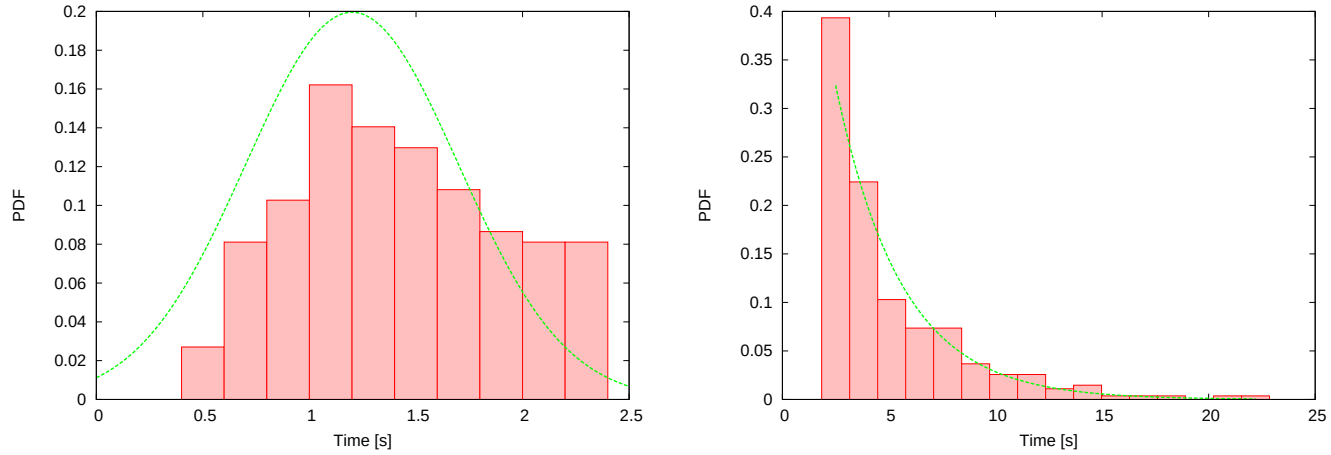
[6]Note that the traffic regulations necessarily introduce some correlation, as drivers tend to stay nearer to the speed limits. As we will see next, only in very sparse roads the correlation between speeds vanishes.



(a) Dense traces

(b) Sparse traces

Fig. 5. Correlation of the speed between consecutive vehicles when their relative distances is $t_i \leq \delta$.



(a) Set τ_B ($t_i \leq 2.5$ s)

(b) Set τ_I ($t_i > 2.5$ s)

Fig. 6. Histogram of the time between arrivals.

We try to provide an explanation for the observed results in the following. When vehicles are in a burst, i.e., the case of τ_B, drivers are "traveling together", and therefore the time between vehicles is very tied to the drivers' *reaction time*. We argue that this reaction time is around a few seconds, and that is reasonable to assume that it follows a Gaussian distribution (some human-related random variables that follow a normal distribution are, e.g., the distribution of height or the I.Q.). When vehicles are not bursty but isolated, the case of τ_I, the arrival process does indeed match a Poisson process, like other arrival process quite common in the literature (e.g., customers arrival in a queue, number of phone calls at a call center).

We claim, based on the above, that *the arrival process in a highway can be modeled using two different random variables: one exponential and one Gaussian*. However, we note that throughout the above analysis we have not derived a rigorous methodology to estimate the parameters of these variables, nor have assessed whether the resulting function will be able to mimic the observed experimental values or not. Furthermore, the value $\delta_t = 2.5$ s used to make the *hard* decision between exponential and normal can be seen as "upper bound" for all traces (as seen in Fig. 5), but not necessarily the value that fits best. In addition, the hard decision itself of classifying samples into τ_B and τ_I results in the truncation of the used variables, which could prevent a proper fitting of the model.

In the next section we first formally define the random variable that we claim can model the observed behavior; then, we will describe a methodology to obtain the parameters for this model; finally, we will assess the relative ability of the proposed variable to model the behavior observed in the real traces, as compared against previous proposals.

IV. An Exponential-Gaussian Mixture Model of Inter-Arrival Times

Based on the results from the previous section, we propose a Gaussian-exponential mixture model to characterize the vehicle inter-arrival times in a highway lane. More specifically, we propose a weighted mixture of:

- A Gaussian random variable, to mimic the behavior caused by vehicles traveling together in bursts (as seen in τ_B).
- A shifted exponential variable, to model the Poisson arrival process of isolated cars (the one observed in τ_I).

This way, our proposed random variable $f_A(t)$ to model the time between vehicles arriving at a given point in a highway lane is formally defined as

$$f_A(t) = w_G \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} + w_E \lambda e^{-\lambda(t-m)},$$

with the following set Θ of parameters:

- μ and σ, which are the mean and standard deviation of the Gaussian random variable.
- λ and m, which are the rate and the shift of the exponential random variable.
- w_G and w_E, which are the the weights of each distribution (and therefore $w_N + w_E = 1$).

Our model is, then, an *unknown* mixture of a normal and an exponential random variables, whose parameters are also *unknown*. This is a well-known problem in statistics, namely computing the maximum likelihood of a set of parameters (the unobserved latent variables specified above) using a set of observed data (the inter-arrival times), than can be solved using an expectation-maximization (EM) algorithm [14], [15]. To estimate the required set of parameters, we implement the following EM-based algorithm using Matlab:

1) We initially set $m = 0$, i.e., we consider an unshifted exponential random variable.
2) We run the EM algorithm for either 200 iterations or until the algorithm converged to a set of $\{\mu, \sigma, \lambda, w_G\}$ parameters.
3) We compute the log-likelihood (denoted as LL) of the sequence of time between arrivals t_i with the obtained Θ, i.e.,
$$LL = \log(\mathcal{L}(t_i|\Theta)).$$
4) The above LL can be seen, after a run of the EM algorithm, as a function f_{EM} of the set of observations t_i and the shift parameter of the exponential m, i.e., $LL = f_{\text{EM}}(t_i, m)$ that, given m, assess the likelihood of the data according to the obtained model. Therefore, f_{EM} quantifies the goodness of a given value of m.
5) We finally perform a sweep on the m parameter from 0 to 3 s in steps of 0.05 s, to obtain the value that maximizes the LL (i.e., the function f_{EM} defined above).

In order to illustrate the operation of the EM algorithm, we plot in Fig. 7 (above) the weighted components of the exponential and the normal random variables for the case of trace 10 (a sparse trace) and trace 3 (a dense trace). We also plot in Fig. 7 (below) the histograms of the empirical traces, along with the resulting theoretical PDFs from the operation of our algorithm. It can be seen in Fig. 7(a) that, for the case of a sparse trace, the normal component is relatively small and most of the observed behavior can be captured using a shifted exponential variable. On the other hand, for the case of a dense trace, Fig. 7(b) shows that the observed values are most likely to be modeled with a Gaussian variable, while the relative component of the exponential variable (in this case shifted) is quite small.

TABLE II
RESULTING VALUES OF Θ FOR THE TRACES OF TABLE I.

Trace #	w_N	w_E	μ	σ	λ	m
1	0.58	0.42	1.44	0.46	0.36	0.50
2	0.65	0.45	1.45	0.35	0.31	0.80
3	0.63	0.37	1.50	0.50	0.31	0.80
4	0.50	0.50	1.06	0.34	0.49	0.65
5	0.37	0.63	1.77	0.66	0.37	0.50
6	0.30	0.70	1.87	0.64	0.37	0.40
7	0.02	0.98	0.85	0.09	0.12	0.50
8	0.01	0.99	1.11	0.48	0.13	0.55
9	0.05	0.95	2.44	0.87	0.13	1.00
10	0.04	0.96	0.73	0.10	0.16	0.40
11	0.11	0.89	1.52	0.41	0.11	0.65
12	0.09	0.91	1.08	0.41	0.17	0.90
13	0.14	1.04	0.30	0.86	0.22	1.00
14	0.21	1.57	0.58	0.79	0.26	1.10
15	0.30	1.84	0.69	0.70	0.38	1.00

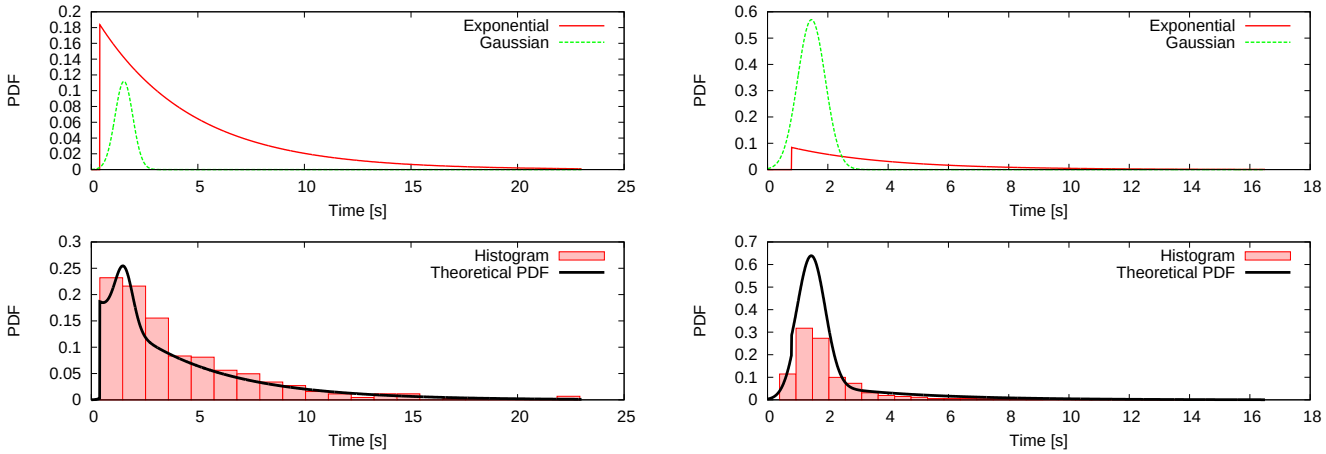
In order to obtain a better representation of the matching between the empirical and the obtained distributions, we plot in Fig. 8 the resulting Q-Q plot for the same two traces used in the previous figure. The results visually confirm that for both for the case of dense traffic and the case of sparse traffic, our proposed model as described by $f_A(t)$ is able to mimic the observed values of the distribution of inter-arrival times.

We run our EM algorithm on the 15 traces considered in the paper, with the resulting values shown in Table II. It is interesting to observe that these figures, derived from a numerical search as defined by the algorithm, are indeed quite intuitive and further confirm the results from the analysis of Section III and the proposed methodology. Indeed, we can emphasize the following three results: i) the relative weight of the exponential variable w_E increases as the density of vehicles decrease (note that the results are presented in decreasing order of densities, see Table I); ii) by observing the values of $\mu+\sigma$, it is clear that distances between vehicles when they travel in bursts are typically well below the identified threshold of 2.5 s; iii) in all cases, the shift parameter m of the exponential random variable is placed between the two extreme values used in our numerical search.

Finally, in order to assess the ability of the proposed model to capture the observed behavior, comparing it against previous work, we proceed as follows[7]. For each of the traces in Table I we compute the parameters for three different models:

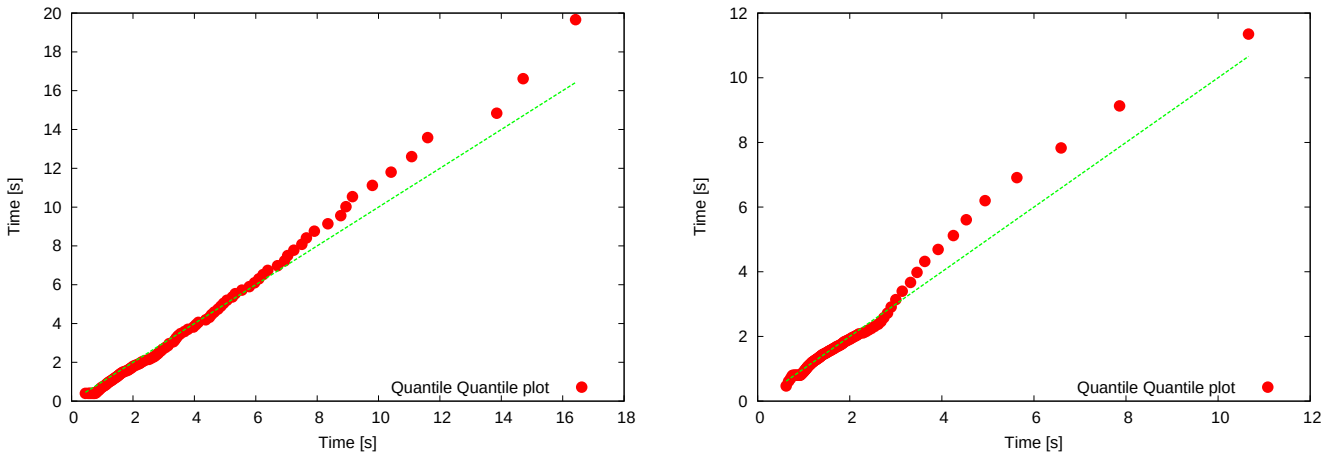
- The proposed mixture, with the set of parameters Θ as given by our EM algorithm.
- An exponential random variable (as proposed in [10]), computing its maximum likelihood rate λ as given by Matlab.

[7]Note that in Section III we performed a K-S test to reject the hypothesis that the inter-arrival times followed an exponential random variable with mean one. However, for the case of distributions with estimated parameters, the values used for the standard K-S test are invalid [16].



(a) Sparse trace (b) Dense trace

Fig. 7. Components of the Gaussian-exponential mixture and resulting PDF for two traces.



(a) Sparse trace (b) Dense trace

Fig. 8. Q-Q plot for the resulting theoretical and empirical distributions.

- A log-normal distribution (as proposed in [9]), with its maximum likelihood parameters μ and σ as computed by Matlab for a given set t_i.

For each of the traces, we compute the log-likelihood of the observed values for these tree model, denoted as LL_{EM}, LL_E and LL_L, respectively, with the obtained results presented in Table III. We also show in the table the estimated parameters (λ^{-1}, μ and σ) for the other exponential and log-normal models. For each trace j we mark with a bold font the largest value of LL out of the three obtained likelihoods (in case of a numerical tie, we mark both numbers). The results can be summarized as follows:

- Our model achieves the highest performance, as it provides the largest likelihood in 86.6 % of the cases (13 out of the 15 traces).
- The log-normal distribution is the second best model, being able to outperform our model in 2 out of the 15 traces.
- Finally, the exponential random variable is never able to match the likelihood of the other two alternatives.

Based on these results, we conclude that our model is the best suited to mimic the behavior observed in a highway lane. In addition, note that the model not only provides a better "numerical" performance that those of [9], [10], but also it provides valuable insight on drivers' behavior, some of which have been neglected so far, e.g., the dependence between consecutive arrivals. As we describe in the next section, part of our future work consists on revisiting this assumption on the i.i.d. of the random variable modeling the vehicles' arrivals.

V. Summary and Future Work

This work has analyzed the inter-arrival times of vehicles from a set measured traces collected at different locations in the city of Madrid. Next, we summarize the major findings of this work:

- Vehicle arrivals, in general, do not follow a Poisson process.

TABLE III
RESULTING LOG-LIKELIHOOD OF THE THREE MODELS CONSIDERED FOR THE INTER-ARRIVAL TIME BETWEEN VEHICLES.

Trace #	LL_{EM}	LL_E	λ^{-1}	LL_L	μ	σ
1	**-1.29**	-1.59	1.76	-1.32	0.32	0.64
2	**-1.18**	-1.51	1.99	**-1.18**	0.52	0.53
3	**-1.15**	-1.50	2.06	-1.19	1.05	0.85
4	**-1.18**	-1.47	2.12	**-1.18**	0.55	0.57
5	-1.24	-1.41	2.36	**-1.23**	0.66	0.62
6	-1.22	-1.39	2.45	**-1.20**	0.71	0.61
7	**-1.08**	-1.15	3.49	**-1.08**	0.96	0.75
8	**-1.12**	-1.19	3.29	**-1.12**	0.90	0.76
9	**-1.02**	-1.08	3.92	**-1.02**	1.07	0.76
10	**-1.01**	-1.06	4.05	-1.02	1.05	0.85
11	**-0.87**	-0.90	5.24	-0.89	1.28	0.89
12	**-0.80**	-0.84	5.85	-0.81	1.49	0.78
13	**-0.66**	-0.69	7.31	**-0.66**	1.71	0.76
14	**-0.69**	-0.72	7.37	-0.70	1.73	0.77
15	**-0.68**	-0.70	7.78	-0.69	1.72	0.84

- There is some speed-correlation between vehicles such that we can identify two types of cars: those traveling together at short distances with similar speeds (what we called bursty behavior), and those traveling at large distances with uncorrelated speeds (what we referred to as isolated).
- For the case of bursty vehicles, the inter-arrival times can be modeled with a Gaussian distribution. Isolated arrivals, on the other hand, exhibit an exponential behavior.
- Based on this findings, we propose a Gaussian-exponential mixture model that characterizes the two groups. Its model parameters are estimated using the Expectation-Maximization algorithm, which outputs the most likely parameter values from a given set of measurements.
- We further assessed the improved accuracy of the proposed model as compared against previous proposals.

As future work, we are currently considering the following topics. First, we are revisiting the assumption on the i.i.d. of the random variables that characterize the arrival process of vehicles. We are also working on the analysis of the distribution of vehicles' speed. Furthermore, our future work also includes coming up with a model that is able to generate arrival patterns that are able to capture the behavior that we observe on the real traces. Last, but not least, we will also validate our current and future findings with additional traces, representing additional types of roads and different traffic situations (e.g., hour of the day, number of lanes, etc.).

ACKNOWLEDGEMENTS

The authors would like to acknowledge the Spanish Directorate General of Traffic for kindly proving us with the empirical traces used in this work.

The authors would like to acknowledge the support of the MEDIANET project (S-2009/TIC-1468) to the development of this work. The research leading to these results has also received funding from the European Community's Seventh Framework Programme (FP7-ICT-2009-5) under grant agreement n. 258053 (MEDIEVAL project).

REFERENCES

[1] R. S. Y. Fallah, C.L. Huang and H. Krishnan, "Analysis of Information Dissemination in Vehicular Ad-Hoc Networks with Application to Cooperative Vehicle Safety Systems," *IEEE Transactions on Vehicular Technology*, no. 99, October 2010.

[2] M. Li, W. Wang, and Y. Zhang, "Research on driver experience based route planning method," in *Intelligent Vehicles Symposium (IV), 2010 IEEE*, 2010, pp. 78 –82.

[3] J. Watson, M. Pellerito, C. Gladden, and H. Fu, "Simulation and Analysis of Extended Brake Lights for Inter-Vehicle Communication Networks," in *Distributed Computing Systems Workshop (ICDCSW'07)*, Ontorio (Canada), September 2007, pp. 87–92.

[4] R. Baldessari, C. J. Bernardos, and M. Calderon, "GeoSAC - Scalable Address Autoconfiguration for VANET Using Geographic Networking Concepts," in *PIMRC 2008*, Cannes (France), September 2008.

[5] "IEEE 802.11p/D2.01, Draft Amendement to Part 11: Wireless Medium Access Control (MAC) and Physical Layer (PHY) specifications: Wireless Access in Vehicular Environments," March 1997.

[6] S. Yamada, "The strategy and deployment plan for vics," *IEEE Communications Magazine*, vol. 34, no. 10, pp. 94–97, October 1996.

[7] C. Liu, "The Case for Vehicular Visible Light Communication (V2LC): Architecture, Services and Experiments," M.S. Thesis, Rice University, Houston, TX,, June 2010.

[8] S. Okada, T. Yendo, T. Yamazato, T. Fujii, M. Tanimoto, and Y. Kimura, "On-vehicle Receiver for Distant Visible Light Road-to-Vehicle Communication," in *IEEE Intelligent Vehicles Symposium*, Shaanxi (China), July 2009.

[9] N. Wisitpongphan, F. Bai, P. Mudalige, V. Sadekar, and O. Tonguz, "Routing in Sparse Vehicular Ad Hoc Wireless Networks," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 8, pp. 1538–1555, October 2007.

[10] H. Reijmers and R. Prasad, "A Model Based Connectivity Improvement Strategy for Vehicular Ad hoc Networks," in *IEEE Vehicular Technology Conference - VTC 1996)*, Ottawa (Canada), September 1996, pp. 1785–1789.

[11] Y. Yang, Z. Mi, J. Yang, and G. Liu, "A model based connectivity improvement strategy for vehicular ad hoc networks," in *Vehicular Technology Conference Fall (VTC 2010-Fall), 2010 IEEE 72nd*, 2010, pp. 1 –5.

[12] S. M. Ross, *Introduction to Probability Models, Ninth Edition*. Orlando, FL, USA: Academic Press, Inc., 2006.

[13] J. H. Pollard, *A Handbook of Numerical and Statistical Techniques*. CUP Archive, 1979.

[14] J. Bilmes, "A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models," ICSI TR-97-021, Tech. Rep., 1998.

[15] J. A. Hernández and I. W. Phillips, "Weibull mixture model to characterise end-to-end internet delay at coarse time-scales," *IEE Proc. Communications*, vol. 153, no. 2, pp. 295–304, April 2006.

[16] J. Durbin, "Kolmogorov-smirnov tests when parameters are estimated," in *Empirical Distributions and Processes*, ser. Lecture Notes in Mathematics, P. Gaenssler and P. Rvsz, Eds. Springer Berlin / Heidelberg, 1976, vol. 566, pp. 33–44, 10.1007/BFb0096877.