Vehicular Congestion Detection and Short-Term Forecasting: A New Model With Results

Gustavo Marfia, *Member, IEEE*, and Marco Roccetti

***Abstract*—While vehicular congestion is very often defined in terms of aggregate parameters, such as traffic volume and lane occupancies, based on recent findings, the interpretation that receives most credit is that of a state of a road where traversing vehicles experience a delay exceeding the maximum value typically incurred under light or free-flow traffic conditions. We here propose a new definition according to which a road is in a congested state (be it high or low) only when the likelihood of finding it in the same congested state is high in the near future. Based on this new definition, we devised an algorithm that, exploiting probe vehicles, for any given road 1) identifies if it is congested or not and 2) provides the estimation that a current congested state will last for at least a given time interval. Unlike any other existing approach, an important advantage of ours is that it can generally be applied to any type of road, as it neither needs any *a priori* knowledge nor requires the estimation of any road parameter (e.g., number of lanes, traffic light cycle, etc.). Further, it allows performing short-term traffic congestion forecasting for any given road. We present several field trials gathered on different urban roads whose empirical results confirm the validity of our approach.**

***Index Terms*—Intelligent transportation systems, traffic forecasting, vehicular software technology for congestion detection, vehicular traffic congestion definition.**

I. Introduction

TWO main approaches have emerged with the aim of limiting the business and societal costs of vehicular congestion. The first approach amounts to provide aggregate traffic information (such as the intensity of traffic volumes and lane occupancy rates) to transportation authorities, which, in turn, feed this information into advanced traffic management systems (ATMSs) to control traffic lights. The second approach is more pervasive and is based on the idea of gathering road traversal times from probe vehicles [1]–[7]. This information is then managed by advanced traveler's information systems (ATISs), which, in turn, supply single drivers with feedback on traffic and with suggestions on the best routes to a destination as a function of real-time traffic conditions (simplistic examples of ATIS are personal navigation devices (PNDs) [5]).

Although the importance of providing updated information on traffic conditions, on a per single vehicle basis, is widely understood, to date, the most amount of research has been devoted to devise traffic estimation and forecast algorithms to be used by ATMSs. As such, these algorithms are mostly concerned with the problem of keeping under control specific locations (e.g., highways and principal arterial roads) subject to high volumes of traffic [8], [9]. Roads where flow intensities are high, as in freeways, are easy to monitor since they usually have a small number of interconnections with other roads while representing a small subset of the whole urban map.

It is, instead, hard to obtain a comprehensive road-by-road picture of urban traffic since such roads are tightly interconnected and consequently subject to high traffic variability. Therefore, traffic information is not widely available in this specific context, with only a few of the main transportation authorities of the most densely populated cities having the tools to monitor (a small subset of) urban roads. As an example, only cities of the magnitude of Los Angeles, CA, and Milan, Italy, are currently provided with a pervasive monitoring infrastructure composed of induction loops and video cameras, which, however, do not cover the entire urban area [10].

However, things may change as soon as wireless sensor technologies enter the game and play a primary role. For example, recent market research forecasts that the 88% of PNDs mounted on vehicles in 2015 will integrate a GPS and a cellular connection [11]–[14], thus paving the way to the deployment of a mobile traffic sensing infrastructure comprised of vehicles that provide on the fly the time they spend to traverse a given portion of road. As a result, all this information sensed by a multitude of vehicles could be processed and put to good use to guide each single driver through the least congested path toward its destination.

It is hence not surprising that many researchers have recently shifted their attention to the design of mechanisms able to detect, as well as to forecast, the congestion state of any given segment of a road, even if not classifiable as a principal arterial way [15]–[18]. Most of such schemes rely on the idea of collecting and processing the traversal time data from all the vehicles that pass through a given road. Obviously, vehicles should be equipped with a GPS receiver, a wireless communication interface, and a software protocol needed to exchange data with a centralized entity. In turn, the centralized entity should process these data and subsequently distribute it to drivers, thus providing a useful aid for routing decisions.

The challenge, at this point, is to design a set of algorithms capable of detecting and forecasting traffic congestion based on a pervasive traffic sensing infrastructure [16]. Obviously, the

Manuscript received May 27, 2010; revised November 16, 2010 and April 5, 2011; accepted May 26, 2011. Date of publication June 7, 2011; date of current version September 19, 2011. This work was supported by the Italian Fondo per gli Investimenti della Ricerca di Base—Data Acquisition and MAnagement in a Sensing and COmmunicating environment Project. The review of this paper was coordinated by Dr. W. Zhuang.

The authors are with the Computer Science Department, University of Bologna, 40127 Bologna, Italy (e-mail: marfia@cs.unibo.it; roccetti@cs.unibo.it).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2011.2158866

starting point of all these research initiatives is a good definition of what vehicular congestion is for any given road segment. Indeed, such definition has been available for a long time and sounds as follows: congestion is the travel time or delay in excess of that normally incurred under light or free-flow travel conditions [19].

Although clear in theory, this definition has not found a successful algorithmic counterpart as it does not provide an unambiguous method, independent of any parameter and applicable to any road, to find the traversal time value T^*, which distinguishes a congested from a noncongested road. To overcome this problem, we provide a brand new definition that identifies the state of congestion of a given road as a state that lasts for at least S units of time and during which travel times or delays exceed the time T^* normally incurred under light or free-flow travel conditions.

Following our definition, a road is congested when the traversal times of vehicles exceed T^* and all he subsequent vehicles that enter the road within a time S keep exceeding the same value. The intuition behind this is given by observing that even when the input flow to a congested road suddenly drops, the inertia of the existing queue causes the road segment to be seen as congested also by those vehicles that enter the road within time S. From this observation, we can logically draw that if a vehicle traverses a road segment when congested, then a second vehicle will very likely experience a similar traversal time if it enters the road segment within S units of time from the first vehicle. It is exactly this kind of consideration that allows one to understand how our congestion detection definition may also be used for estimating the duration of congestion states, thus providing a simple and effective tool useful for traffic forecasting.

From the foregoing considerations, assuming S is known, an algorithm able to compute the congestion threshold T^* for any given road is straightforward. The idea at the basis of such an algorithm is as follows. Take a set of all the pairs of vehicles entering a given road, which are separated in time of at most S units, we say that a road segment is in a congested state if the number of pairs of subsequent vehicles, for both of which the traversal time exceeds T^*, is much greater than the number of pairs for which the traversal time of only the first vehicle of the pair exceeds T^* (being this ratio $N : M$, for example). Obviously, with the phrase "pairs of subsequent vehicles," we indicate, for example, a couple of vehicles i and j, with i entering that road segment earlier than j, but independent of the number of vehicles between i and j. Conversely, a road segment is in a noncongested state if the amount of pairs of subsequent vehicles, for both of which the traversal time is below T^*, is much greater than the amount of pairs for which the traversal time of only the first vehicle of the pair is lower than T^* (we can suppose this ratio is $K : H$). Assuming that $N : M$, $K : H$, and S are known, it is as easy as pie to find T^*. The rationale is that we see congestion as that state where the number of vehicles for which the traversal times are all stably high outnumber the number of vehicles for which their traversal times gracefully drop to a nonhigh value. Similarly, we deem a road segment as not congested when a much greater and steady number of vehicles with low travel times traverse that road with respect to the number of vehicles with high travel times.

The issue of quantifying the ratios $N : M$ and $H : K$ is crucial and should be left to the experience and sensibility of who is in charge of tuning our mechanism for managing traffic operations (typically the traffic operations manager). However, a good choice can be based on the following consideration. A general reasoning can be conducted observing that human beings (and drivers as well) require a high reliability on the information they process to make their decisions. Specifically, drivers perceive a road as congested when a high rate of vehicles suffers from delays, which are above a congestion threshold. The point of how high this rate should be is still open. Obviously, any value exceeding 60%/70% matches that sense of "perceived congestion" drivers have in mind. Indeed, authors of [20] identify in the 80th percentile (more precisely the 80th–50th interquartile difference) a reasonable level of reliability, which drivers require on the traffic information they receive.

Now, we have not yet provided any precise recipe for determining S. It is worthwhile to mention that a reasonable assumption of S is of great importance for our approach as it helps in determining the congestion threshold T^*. Not only, given the nature of our algorithm, S gives us the means of providing a forecast of traffic. To this aim, we can consider a study from the American National Bureau of Transportation Statistics that can be of help as it reveals that the average American driver spends 55 min a day behind the wheels [21]. Considering that the average daily traffic pattern for any driver is from home to work in the morning and back in the late afternoon, and assuming similar travel time values for both directions, we find that the average one-way travel time is equal to 27.5 min. This indicates a plausible reference value for S as an average driver will not spend more than such value stuck in traffic. As per a minimum value of S, a good choice could be that of 2 or 3 min; this being, on average, the traversal time of a typical urban road without traffic.

Once implemented, to confirm the validity of our traffic detection algorithm, we carried out real experiments amounting to 450 mi of travelled roads throughout different locations in the world. We report in this paper results from those experiments, which have validated our approach.

To conclude this section, we emphasize that our scheme differs from any other we are aware of on one important point, i.e., it computes a congestion threshold dynamically as a function of the congestion duration S. We think that this design choice brings at least three advantages over all the other alternative schemes:

1) A road's congestion threshold solely depends on the probe vehicle traversal times and does not require any other contextual information.
2) Our scheme binds the detection of congestion (or no congestion) to the prediction of its duration, while the other algorithms either ignore this relationship or separately address these two problems.
3) Our algorithm takes into account that S may change because of sudden differences in a road's capacity, traffic

light cycles, or varying weather conditions, for example, and adapts accordingly.

The rest of this paper is organized as follows: In Section II, we review the main schemes proposed as effective means to detect traffic congestion in urban roads. In Section III, we provide both the intuition behind and the formal implementation of our algorithm. A description of the experimental scenario and with empirical results may be found in Section IV. We finally conclude in Section V.

II. Related Work

A great amount of research has been carried out over the past few years, working on the problem of evaluating the performance of urban road networks by means of new congestion metrics. In most of such works, a congested state is distinguished from a noncongested state by analyzing the traversal times of the vehicles that flow through a given road segment and comparing them with some fixed threshold [22]–[27]. In the following, we will describe how three relevant traffic detection algorithms work. The first two methods both rely on the use of probe vehicle data for the detection of congestion, whereas the third is based on the highway capacity manual delay formula for signalized intersections and will serve, as we shall see in Section IV-A, as a benchmark for the validation of our results.

The methodology described in [23] is based on the use of fixed speed thresholds to determine whether a given road segment is congested or not. In such work, vehicle probes are periodically collected from a fleet of four thousand taxis operating in Shanghai and averaged out, providing instantaneous traversal times and speeds at specific locations. In practice, traffic is classified according to the average speed experienced by a group of taxis that traverse a given road segment, as follows. If, in urban contexts, the taxis result moving at a speed that exceeds 30 km/h, hence traffic is classified as very smooth. If instead the taxis' speeds are between 25 and 30 km/h, then the traffic is smooth. Finally, if the average speed falls in one of the [16, 25), [11, 16) or [0, 11) km/h intervals, then traffic is defined as medium, congested, and very congested, respectively. Such methodology has these two important drawbacks. The first drawback amounts to the fact that an *a priori* traffic classification based on predetermined values of speed is too rigid. A given road may exist where a speed of 20 km/h cannot be considered as a symptom of congestion simply because this is the maximum speed cars can reach due, for example, to a specific traffic light cycle. The second problem is that this mechanism lacks the ability to predict how long a state of congestion will last.

The second method, which was termed surface street traffic estimation, was specifically proposed to identify congestion on signalized road segments, which are road segments whose downstream intersection is managed by a traffic light [28]. Indeed, congested traversals are distinguished from noncongested traversals by analyzing the GPS traces collected by vehicles. Two different algorithms cooperate to this aim. The first algorithm estimates the red light duration of a traffic light as the 95th percentile of the stopping duration of vehicles. The second algorithm computes two thresholds. The first threshold is an average speed, which is computed as the road segment length divided by the sum of the fifth percentile of traversal times plus the red light duration. The second threshold is a space mean speed, which is computed as the fifth percentile of the spatial mean speed values that exceed the first threshold. While the meaning of the first threshold is clear, as a vehicle that experiences an average speed below this value traversed the road with a delay that is above the free flow traversal time (FFTT) plus the red light duration, it is worth spending a few more words on the meaning of the second threshold. The space mean speed of a vehicle is the arithmetic mean of the instantaneous speed samples taken at fixed locations. Therefore, this second threshold differentiates the values given by those vehicles that traverse a road with a stop-and-go pattern from the values of those vehicles that, instead, smoothly flow through the road. Summarizing, vehicles that exceed both thresholds are classified in a free flow state, whereas vehicles that fall below both thresholds are classified as congested. While this strategy is clear, as it identifies congestion as queueing, this algorithm falls short in two main aspects. The first is that this method does not provide any forecasting information, thus being questionable as to its utility. The second is that this method focuses on segments with signalized intersections, being not clear on how it may be extended to more general cases. Finally, an approach often used to distinguish congested from noncongested states is based on the highway capacity manual (HCM) delay formula for signalized intersections [8], [29]–[31]. This formula computes the average traversal time $\bar{T}_{\mathrm{HCM}}$ of a vehicle as the sum of three values. The first d_0 is the average traversal time per vehicle in free flow conditions. The second d_1 is equal to the additional average delay per vehicle due to traffic light phases. Finally, d_2 amounts to an additional average delay experienced by a vehicle because of congestion. While d_0 can simply be obtained by dividing the length of a road segment by its speed limit, d_1 and d_2 are functions of the capacity of the road (determined by the number of lanes and the length of traffic light phases), the average amount of vehicles entering the road within a given time, and the expected duration of the given analysis conditions. In summary, the $\bar{T}_{\mathrm{HCM}}$ value may be computed on a per-vehicle basis assuming that the input traffic volume of a road matches its capacity and that the analysis period is long enough. This value is finally exploited to differentiate congested from noncongested portions of roads. While this approach exploits the duration of the analysis with the aim of capturing the future state of a road segment, it has, indeed, a limit given by its inability to adapt to the capacity fluctuations of a road segment (being the capacity of a street subject to a number of modifications caused by obstructing vehicles and maintenance works, for example). Moreover, accurate results require accurate estimates of the parameters that influence the capacity of a road, which are not easy to obtain on a large-scale basis. Summarizing, the foregoing approaches either 1) do not couple the congestion identification and forecasting problem as one, thus risking to identify as congested roads that will not last in that state any longer in time or 2) excessively rest upon statically chosen parameters, jeopardizing their adaptability to new and different road settings. We will show that our approach is successful in overcoming both of the previous problems as it does not need the *a priori* setting of any parameter. Hence, we are confident that it can represent a good candidate for traffic estimation and forecasting in pervasive urban traffic scenarios.

III. Novel Congestion Detection Model

Before proceeding with a detailed explanation of our congestion detection algorithm, we briefly account for the general context where it should be exploited (e.g., within the frame of an ATIS [32], [33]). In particular, we assume vehicles mounting an advanced PND integrated with a GPS receiver and a full-duplex communication device.

First of all, any given segment of a road must be put under observation for a duration of half a day/a day, as the idea is that, upon completion of that given road segment traversal, any vehicle sends to a centralized entity a message that includes an identifier of that road segment and its entry and exit times (i.e., the traversal time). As soon as the centralized entity has completed the observation activity and collected a sufficient number of samples from probing vehicles, it has a clear picture of the congestion states characterizing that given segment of a road. At this point, our method steps through a second phase that requires 1) choosing a value S, 2) taking all the pairs of traversal times of those vehicles where the second one entered the road no later than S units of time after the first during the given observation period, and 3) applying the methodology that we will later describe to compute the congestion threshold T^*.

If S and T^* can be found satisfying a set of requirements cleared later, then they can be put to good use as follows.

Suppose that afterward (10 min later or even 2 days after) another vehicle traverses that road experiencing a travel time equal to T and transmits this information to the centralized entity. Depending whether $T > T^*$ or $T \leq T^*$, the centralized entity is in the condition to inform all the subsequent vehicles that plan to traverse that road that they will incur, with a given likelihood (typically 80%), in a congested or not congested state of duration of at least S.

In simple words, our mechanism has not been designed to provide static information to drivers like "on Monday evening at 5 P.M. you will incur in congestion" but dynamically treats traversal times larger than the congestion threshold T^* as a symptom that is preannouncing a congestion event of duration of at least S to be advertised to the surrounding vehicles.

Obviously, the initial observation phase required to tune the system may be performed only once in a given period or repeated with a frequency to be obtained by the transportation authority on a per road basis.

Let us explain now how the algorithm needed to detect and forecast congestion works.

A. Congestion Detection and Forecasting Algorithm

We first begin by recalling the general definition given in Section I, which is as follows.

Definition 3.1: A road segment R is in a congested state if the travel times or delays of the traveling vehicles exceed the time T^* normally incurred under light or free-flow travel conditions, and this congested state lasts for at least S units of time.

Owing to this generic definition, it is possible to infer a further set of operational definitions from which a simple algorithm can be derived to identify when a road segment is in a state of high or low congestion. Let us start with a set of definitions aimed at identifying two different types of sets of vehicles entering a road segment under diverse congestion conditions.

Definition 3.2 (Platoon): Let us define a platoon $P_{t,S}$ of vehicles as a group of vehicles entering a road segment R, with the first vehicle of the platoon entering R at time t and the last vehicle entering R no later than time $t + S$.

In essence, the concept of platoon captures all those vehicles that entered a road segment within a time span S, but depending on the point in time when the first of them entered that road (beginning at time t).

To extend this concept, we can introduce the definition of a fleet F, which, for a given S, aims at addressing not only the cars of a single platoon but all those pairs of cars (i, j), separated by at most S units of time, which enter a given road segment generically during a period of observation of duration Z. This definition is as follows.

Definition 3.3 (Fleet): A fleet F_S of pairs of vehicles is defined as $F_S = \{(i, j) | (i, j) \in P_{t,S} \times P_{t,S}, \forall t \in [0, Z - S], i < j\}$, where $<$ is meant to induce a natural order between subsequent vehicles.

Definition 3.4 (High Congestion Vehicle Set): Taking a fleet F_S, $HighCongestion_{T_1^*}$ is defined as the set of all the pairs of vehicles i, j (with i entering R before j) of this fleet for which their traversal times, say T_i^* and T_j^*, both exceed the congestion threshold T_1^* (i.e., $(T_i^* > T_1^*) \wedge (T_j^* > T_1^*)$). We also define as $Noise1_{T_1^*}$ the set of all the pairs of vehicles, say h, k, for which the traversal time T_h^* of only the first vehicle h exceeds T_1^* (i.e., $(T_h^* > T_1^*) \wedge (T_k^* \leq T_1^*)$).

In essence, $HighCongestion_{T_1^*}$ represents the amount of all those vehicles suffering from a stable situation of congestion. Indeed, all their traversal times lie above the congestion threshold T_1^*. Instead, $Noise1_{T_1^*}$ represents the set of those vehicles, a part of which is leaving the congestion state. The traversal times of those vehicles lie below the congestion threshold T_1^*.

Definition 3.5 (Low Congestion Vehicle Set): Taking the same fleet F_S, $NoCongestion_{T_2^*}$ is defined to be the set of all the pairs of vehicles of this fleet, say i, j (with i entering R before j), for which their traversal times, say T_i^* and T_j^*, are both below the congestion threshold T_2^* (i.e., $(T_i^* < T_2^*) \wedge (T_j^* < T_2^*)$). We also define as $Noise2_{T_2^*}$ the set of all the pairs of vehicles, say h, k, for which the traversal time T_h^* of only the first vehicle h is below T_2^* (i.e., $(T_h^* < T_2^*) \wedge (T_k^* \geq T_2^*)$).

Similarly, as before, $NoCongestion_{T_2^*}$ groups all the cars incurring in a stable situation of no congestion (i.e., all their traversal times are below T_2^*). In this case, instead, $Noise2_{T_2^*}$ represents a situation where the second car of most pairs is entering in a new congestion state as their traversal times exceed T_2^*.

Given the preceding sets, we now introduce the indicator functions of those sets with the aim of providing a tool with which the number of vehicles of a fleet F_S entering a road segment R can be counted under different congestion conditions.

Definition 3.6 (High Congestion State): Let $\mathbf{1}_{HighCongestion_{T_1^*}} : F_S \rightarrow \{0,1\}$ be defined as

$$\mathbf{1}_{HighCongestion_{T_1^*}}((i,j)) = \begin{cases} 1, & (i,j) \in HighCongestion_{T_1^*} \\ 0, & (i,j) \notin HighCongestion_{T_1^*}. \end{cases}$$

Let $\mathbf{1}_{Noise1_{T_1^*}} : F_S \rightarrow \{0,1\}$ be defined as

$$\mathbf{1}_{Noise1_{T_1^*}}((h,k)) = \begin{cases} 1, & (h,k) \in Noise1_{T_1^*} \\ 0, & (h,k) \notin Noise1_{T_1^*}. \end{cases}$$

Definition 3.7 (No Congestion State): Similarly, for noncongested states, let $\mathbf{1}_{NoCongestion_{T_2^*}} : F_S \rightarrow \{0,1\}$ be defined as

$$\mathbf{1}_{NoCongestion_{T_2^*}}((i,j)) = \begin{cases} 1, & (i,j) \in NoCongestion_{T_2^*} \\ 0, & (i,j) \notin NoCongestion_{T_2^*}. \end{cases}$$

Let $\mathbf{1}_{Noise2_{T_2^*}} : F_S \rightarrow \{0,1\}$ be defined as

$$\mathbf{1}_{Noise2_{T_2^*}}((h,k)) = \begin{cases} 1, & (h,k) \in Noise2_{T_2^*} \\ 0, & (h,k) \notin Noise2_{T_2^*}. \end{cases}$$

At this point, we are able to count 1) the number of pairs of vehicles that suffers high congestion versus the number of pairs of vehicles that is leaving a congested situation, and 2) the number of pairs of vehicles that does not suffer congestion versus the number of pairs of vehicles that is entering in a congestion state. This allows us to provide the following propositions on which our congestion detection algorithm is based.

Proposition 3.1 (Congestion): A given road segment R is congested for a period of length at least S if the following holds:

$$\frac{\sum_{(i,j)\in F_S} \mathbf{1}_{HighCongestion_{T_1^*}}(i,j)}{\sum_{(i,j)\in F_S} \mathbf{1}_{Noise1_{T_1^*}}(i,j)} \geq \frac{N}{M} \tag{1}$$

with $N/M = 80\%/20\%$, as discussed in Section I [20].

The same can be drawn for a noncongested state as follows.

Proposition 3.2 (No Congestion): A given road segment R is not congested for a period of length at least S if the following holds:

$$\frac{\sum_{(i,j)\in F_S} \mathbf{1}_{NoCongestion_{T_2^*}}(i,j)}{\sum_{(i,j)\in F_S} \mathbf{1}_{Noise2_{T_2^*}}(i,j)} \geq \frac{H}{K} \tag{2}$$

again with $H/K = 80\%/20\%$ [20].

The rationale of Proposition 3.1 is simply that if the number of cars in a stable situation of congestion largely exceeds the number of cars that are leaving a congestion state, then that state can be confirmed as a congested state. Conversely, Proposition 3.2 states that if the number of cars suffering no congestion outnumbers the set of those entering in a congested situation, then that state can be confirmed as a noncongested state.

B. Efficient Implementation

Up to this point, we have devised a mathematical model that can be used to distinguish a congested state of a road segment R from a noncongested state based on the computation of two congestion thresholds, namely, T_1^* and T_2^*. Before explaining how to compute the values of T_1^* and T_2^*, we have to anticipate here that it is not only a matter of the choice of an adequate mathematical model but a matter of its application to only those real cases where it can return an effective solution that meets the drivers' needs in terms of knowledge about the fact if a given road is congested or not. We here appeal to a principle of reality, which says that it makes sense to apply our algorithm to all those roads where congested and noncongested states alternate, whereas it does not make sense to apply it when only one of the two mentioned states can be found on a road at any given time because in such a case there is no reason for using any existing model. In this latter case, we would find paradoxical T_1^* and T_2^* values. Take, for example, this case showing when our algorithm should not be used. Consider a road segment R for which we know for sure it is congested, with a certain set of vehicles running on it. Take now Proposition 3.1 and apply it to the traversal times returned by those vehicles. Obviously, this proposition will be able to find a T_1^* value that is exceeded by the traversal times of all the vehicles. Take then Proposition 3.2 and apply it to the set of the same traversal times as before. Obviously, again, a new value T_2^* can be obtained exceeding all those traversal times. In this way, we would have obtained the paradox of having $T_2^* > T_1^*$ without any possibility of understanding if our road is in a congested state or not.

Indeed, this is not a problem of our algorithm, but its application was wrong. Instead, the right way to apply our procedure is that of working on a set of data that, for a given road R, can be sampled both from states of congestion and from states of no congestion. From an operational viewpoint, this means that, for our algorithm to work correctly, a given road must be put under observation for a period whose duration is long enough to capture both congested and noncongested traffic situations. Hence, if the traversal times are sampled correctly, then applying both Propositions 3.1 and 3.2 returns two threshold values T_1^* and T_2^* ordered according to their natural way, that is, $T_2^* \leq T_1^*$.

Based on the preceding considerations, an efficient way to implement the statements of both Propositions 3.1 and 3.2 goes through two different steps. The first step amounts to searching the pair $(\bar{T}_1^*, \bar{T}_2^*)$ that maximizes 1) the number of pairs of vehicles whose traversal times are both larger than $\bar{T}_1^*$ (congestion) and 2) the number of pairs of vehicles whose traversal times are both below $\bar{T}_2^*$ (no congestion). Simultaneously, of minimal size should be 3) the amount of pairs of vehicles for which the traversal times of only the latter vehicle is smaller than $\bar{T}_1^*$ and 4) the amount of pairs of vehicles for which the traversal time of only the latter vehicle is larger than $\bar{T}_2^*$ (noisy conditions). All this can be obtained with (3), shown at the bottom of the next page. where $\Delta = \sum_{(l,m)\in F_S} \mathbf{1}_{HighCongestion_{T_1^*}}(l,m) - \sum_{(a,b)\in F_S} \mathbf{1}_{NoCongestion_{T_2^*}}(a,b)$.

In essence, Δ is a term accounting for a few noisy traversal time values that could have a negative effect on the computation of the congestion threshold. This is the problem of a situation where two well-defined and different clusters of traversal time values coexist with a few isolated samples (either very high or very low), which, in turn, may have the effect of shifting the value of the congestion threshold. Δ, therefore, ensures that the two clusters of samples contain the maximum number of points each by minimizing the difference between their sizes. In other words, subtracting Δ guarantees that if, for example, an isolated point lies along the x–y bisector right below (or also right above) all the other points plotted on a congestion graph, our mechanism returns a solution where the $\bar{T}_1^*$ and $\bar{T}_2^*$ values simply separate the two clusters, whereas it is excluded the possibility that a solution exists where $\bar{T}_1^*$ and $\bar{T}_2^*$ separate the isolated points from the union set that contains the two clusters.

The second step amounts to taking the just computed $(\bar{T}_1^*, \bar{T}_2^*)$ values, respectively replacing them in (1) and (2), and finally checking if the inequalities are satisfied. The motivation behind the execution of this second step is that Step 1 could complete, giving us a percentage of pairs of vehicles in the state of congestion equal to N, with $N < 80\%$, thus resulting in a percentage of pairs of vehicles in a noisy situation above the level of 20% (we name this kind of check $\texttt{Check1}(\bar{T}_1^*)$). A similar situation could occur also with the percentage of pairs of vehicles in a noncongested state, which could be less than 80% (we name this kind of check $\texttt{Check2}(\bar{T}_1^*)$).

Unfortunately, a reason for the checks to fail could be that of having chosen a too large duration S for the state of congestion of interest. This would mean that for many pairs of subsequent cars the following holds: The congested (or noncongested) state a first vehicle incurs in does not last in time as a second vehicle no longer finds the same state. However, this could be a problem simply concerned with the duration of the S we have chosen, whereas a smaller value for S could exist, in principle, for which both the subsequent cars incur the same state of congestion. The idea is hence that of looking for such a value by reducing S until a situation is captured where both the subsequent vehicles of the pair experience a similar state of congestion (or no congestion).

While no rule prevents us from starting with a value of S equal to the length of the observation period Z (e.g., half a day), such a choice would very probably result in a waste of time, as for values of S close to Z no correlation can be observed between the traversal times of subsequent vehicles (hence, our methodology would fail). Hence, reminding the reference value of 27.5 min as discussed in Section I, our algorithm starts its search from $S = 27.5 \times 3 = 82.5$ min and completes when it reaches the final value of 2 min. Obviously, if the search is completed without the possibility of identifying a congestion or noncongestion state, whatever the value of the chosen S, this means that for that given road segment, it is not possible to distinguish any congestion state of interest. In such a particular case, our algorithm completes by returning an adequate alert message.

To conclude, we sketch in Table I the main phases of the algorithm we have so far discussed. It shows an algorithm that, after some iterations on the S values, finds the congestion thresholds $\bar{T}_1^*$ and $\bar{T}_2^*$ of a given road segment (provided that both these thresholds exist).

TABLE I
CONGESTION THRESHOLD DETECTION ALGORITHM

	input: A list of traversal times collected during an observation period of duration Z. **output:** $S, \bar{T}_1^*$ and $\bar{T}_2^*$.
1.	$S \leftarrow 82.5$ minutes;
2.	$(\bar{T}_1^*, \bar{T}_2^*) \leftarrow T^*$ s.t. $\texttt{Max}(T_1^*, T_2^*)$;
3.	**while** $\neg\texttt{Check1}(T_1^*) \wedge \neg\texttt{Check2}(T_2^*) \wedge S > 2$ **do**
4.	$S \leftarrow S - 1$ minutes;
5.	$(\bar{T}_1^*, \bar{T}_2^*) \leftarrow \texttt{Max}\ (T_1^*, T_2^*)$;
6.	**end**

C. Representing Congestion With Graphs

We now proceed showing how the results of our algorithm can be graphically represented. To do so, we define what a congestion graph is.

Simply, a congestion graph of a road segment R is a scatter graph of points $(x = T_i, y = T_j)$, where the x- and y-axis values of each point represent the traversal time of a pair of subsequent vehicles. In particular, for each pair of vehicles, the x value of a point on the graph is equal to the traversal time of the first vehicle of the pair that entered R, whereas the y value is equal to the traversal time of the second vehicle that entered R within S seconds later. Each congestion graph can be indexed with a value of S, with S being the maximum difference in time between the moment when the first vehicle of a pair entered R and the moment when the second vehicle of the same pair entered R for any given pair of vehicles on that graph.

As significant examples, consider the leftmost, central, and rightmost graphs in Fig. 1 representing three different congestion graphs for three different S values. The leftmost graph has been filled with data coming from vehicles running on a road segment that either suffers from congestion or not, depending on the specific moment of the day (in particular that road was under observation for 7 h). Within any given S, there was a traversal of only two cars. As expected, running our algorithm on the data shown on this graph returns the values of S, $\bar{T}_1^*$, $\bar{T}_2^*$, N, M, H, and K, respectively, as follows: $\bar{T}_1^* = 93$ s, $\bar{T}_2^* = 89$ s, $N = 92\%$, $M = 8\%$, $H = 84\%$, and $K = 16\%$.

$$(\bar{T}_1^*, \bar{T}_2^*) = (T_1^*, T_2^*)$$
$$s.t. \left\{ \max_{T_1^*, T_2^*} \sum_{(i,j) \in F_S} \left(\mathbf{1}_{HighCongestion_{T_1^*}}(i,j) + \mathbf{1}_{NoCongestion_{T_2^*}}(i,j) - \mathbf{1}_{Noise1_{T_1^*}}(i,j) - \mathbf{1}_{Noise2_{T_2^*}}(i,j) \right) - |\Delta| \right\} \quad (3)$$

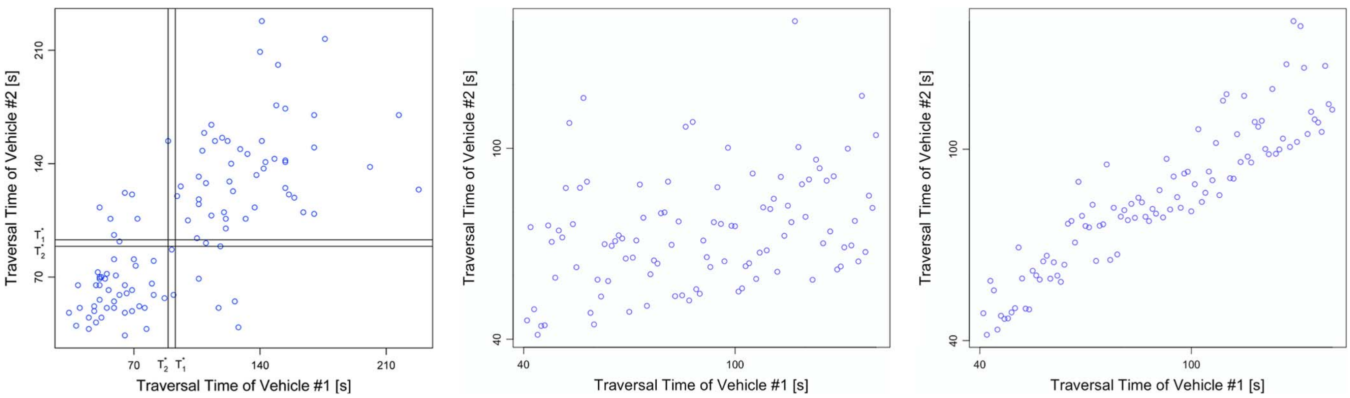

Fig. 1. (Left) Congestion graph with (1) and (2) satisfied and $S = 6$ min. (Center) Congestion graph with $S = 3$ h, with (1) and (2) not satisfied. (Right) Congestion graph with $S = 5$ s, no interesting situation.

Indeed, the $\bar{T}_1^*$ and $\bar{T}_2^*$ thresholds are plotted on the graph by means of two different couples of intersecting lines.

An interesting, even if expected, phenomenon, shown by this example, is that if $\bar{T}_1^*$ and $\bar{T}_2^*$ exist, as returned by our algorithm, then they take very close values (often almost coincident). This is a natural consequence of the physical reality where in practice, only one congestion threshold exists, above which, we have congestion, and below which, we do not. Further, in all the experiments we have carried out (and discussed in Section IV), the value $\delta = [(\bar{T}_1^* - \bar{T}_2^*)/\bar{T}_1^*] \times 100\%$ was always smaller than 3%. For this reason, with the aim of simplifying this matter, in the remainder of this paper, we will exploit the value $\bar{T}_1^*$ as a unique representative of the congestion threshold of a given road.

The congestion graph at the center of Fig. 1 represents instead a clear situation where no reasonable values for $\bar{T}_1^*$ and $\bar{T}_2^*$ can be found. This is not surprising as the data for the traversal times plotted in this graph were sampled with a value of $S = 3$ h for a road where congestion states did last always less than 30 min. Clearly, if we take such a huge value for S in this situation, we are making the mistake of establishing a correlation between two cars that traversed the same road segment in very different moments subject to very different states of congestion.

As a final and interesting example, consider the case of the rightmost graph in Fig. 1. There were sampled data with an extremely small value for S ($S = 5$ s). If we tried to apply the conditions expressed in (1) and (2) to this graph, we would find values for $\bar{T}_1^*$ and $\bar{T}_2^*$ that could apparently look like reasonable. Again, this would be a mistake as the problem is that there is no interest in identifying congestion when its duration is 5 s. Indeed, there is no reasonable congestion state that lasts so shortly. This is why our algorithm has a lower limit on $S = 2$ min.

Obviously, with the cases of the central and the rightmost graphs in Fig. 1, we wanted to represent the abnormal situations we would get if we used wrong values for S. Specifically, if we chose a value of S that is too large, we would have obtained sparse points in the graph, accounting for a situation where no relationship between vehicle pairs exist. Instead, if S was chosen too low, we would have obtained points concentrated along the x–y bisector, accounting for a situation where vehicles are too close to each other to be useful for any kind of decision.

To better explain these specific experiments, the rightmost graph reports the traversal time pairs recorded in a situation where the road was observed for only half an hour, since with $S = 5$ s, after half an hour, we obtained a sufficient number of sample points. Obviously, the unhealthy choice of $S = 5$ was deliberately taken to demonstrate that our mechanism works for values of S that are at least as long as a couple of minutes.

Vice versa, with the central plot, we took into consideration a different road section where our objective was that of demonstrating that if S is too large (3 h), then there is little or no correlation between the traversal times of subsequent vehicles. To emphasize this, we plotted a sufficient number of pairs of traversal times (comparable to those of the other examples) collected between 2 and 3 h one from the other. Clearly, in this case, the observation period was a few days long (we could only collect a few samples during a day).

D. Running the Formula

To test its validity, we have input to (3) a fleet of pairs of vehicles where the balance between congested and noncongested vehicles varied in all possible ways.

In particular, given 100 simulated pairs of traversal times, we tested all the following combinations (where NC stands for the set of pairs of traversal times both below the threshold $T^* = 100$ s, and CON stands for the set of pairs of traversal times both above the threshold T^*): $\{|NC| = 10, |CON| = 90\}$, $\{|NC| = 20, |CON| = 80\}$, $\{|NC| = 30, |CON| = 70\}$, $\{|NC| = 40, |CON| = 60\}$, $\{|NC| = 50, |CON| = 50\}$, $\{|NC| = 60, |CON| = 40\}$, $\{|NC| = 70, |CON| = 30\}$, $\{|NC| = 80, |CON| = 20\}$, and $\{|NC| = 90, |CON| = 10\}$.

We ran 100 experiments for each specific combination, where each traversal time (above or below the threshold) was taken randomly from a uniform distribution. In Fig. 2, we plotted the average computed value of T^* with superimposed shape of the objective function as a function of the number of pairs that fell in each of the two given sets. As can be observed, T^* was exactly equal to 100 when the average value of the objective function reached its maximum, and the numbers of pairs of vehicles that fell in the two sets were equal (50–50). However, even when the sizes of the two sets became unbalanced (e.g., 90 congested versus 10 noncongested, and

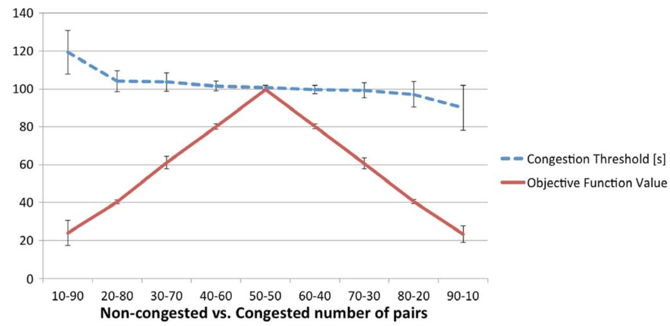

Fig. 2. T^* and objective function as a function of the number of pairs of traversal times corresponding to vehicles in the noncongested and congested sets (average values + standard deviation).

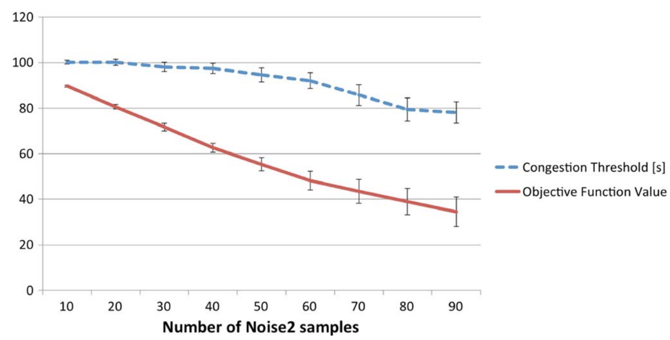

Fig. 3. T^* and objective function as a function of the number of pairs of traversal times corresponding to vehicles in a noisy set (average values + standard deviation).

vice versa), our formula found a T^* value that was remarkably close to the correct value.

We also developed a further experiment where, keeping fixed the number of pairs of traversal times corresponding to congested and noncongested vehicles (50–50), we gradually increased the number of pairs of traversal times corresponding to vehicles that fell in the $\mathbf{1}_{Noise2_{T_2^*}}(i,j)$ set. Our expectation was that our algorithm would compute a threshold of $T^* = 100$ s until the point that the amount of introduced noise became too large. This is what exactly happened, as reported in Fig. 3, where, again, on the y-axis to the T^* values (seconds) the shape of the objective function is superimposed. This result is another confirmation that our algorithm works.

IV. Experimental Assessment

We carried out a set of nine different experiments to verify the validity of our congestion detection and forecasting algorithm discussed in the previous section. Each of these experiments was conducted by managing the vehicular data sampled and transmitted by a set of cars driven over a real section of the road. With the term section, we both indicate a single road segment and two or more adjacent segments separated by one or more intersections. Eight out of nine sections of roads taken into consideration were in Los Angeles, whereas one was in Pisa, Italy. All the information concerning these roads is listed in Table II, where the road identifier and the name of the road, the section of the road under analysis, its length, its traversal time under free flow conditions, its entire traffic light cycle time (CT), and, finally, the green time (GT) duration of the cycle are provided. For the sake of conciseness, we will identify a given road, and its section under analysis, by the sole use of its corresponding road identifier in all the tables that follow, displaying the result of our experiments.

Of a certain importance is also the consideration that all the examined roads in Los Angeles are wide and provided with an advanced ATMS infrastructure, whereas the road in Pisa is narrow and crowded with both pedestrian and vehicular traffic. The motivation behind the choice of these two cities is that they represent two very different traffic situations, both from a traffic management and driving style standpoint. To collect data, each vehicle was equipped with an onboard system comprised of a laptop with a GPS and an evolution-data-optimized interface used to store a digital map of the area under analysis. As discussed in the previous section, upon traversal of a given road section R, a car transmitted its traversal time to a centralized entity. As soon as a sufficient amount of data required to distinguish congestion from noncongestion on R was available (typically after a dozen hours during the daytime), our system computed an estimate of $\bar{T}^*$. The road section traversal times were sampled by performing loops on tracks over those roads. Tracks were, in general, comprised of a first part of the road section chosen as it presented a high varying traffic pattern plus a second part with little or no traffic. The rationale underlying this choice was to be able to perform subsequent observations of a given road section as close in time so as to exploit the same set of cars. In Pisa, for example, we chose a track that included Via Bonaini, Via B. Croce, P.zza Guerrazzi, and Via Gian Battista Queirolo, as Via B. Croce could become very crowded, whereas the other road sections in the track rarely experienced intense vehicular flows (the track is highlighted in Fig. 4).

In the following section, we are going to present the results we have obtained from our experiments.

A. Results

Our results are shown in Table III, where for each road is respectively listed the number of loops on tracks, the congestion threshold $\bar{T}^*$ computed using our mechanism, the duration of the congestion S, our measure N of how many pairs of subsequent vehicles suffer a stable congestion situation, and the measure H of how many pairs of subsequent vehicles experience no congestion. Further, to verify the validity of the results we have obtained, we have computed for each examined road section the value $\hat{T} = T_{\mathrm{FFTT}} + (CT - GT)$ based on the methodology proposed in [28]. $\hat{T}$ amounts to the time a car would spend traversing that road, in case it incurs not a traffic congestion situation but that of waiting for the traffic light situated at the end of the intersection to become green. Cars are assumed to stop in queue for an entire red light time (this resembles a typical situation where drivers enjoy an experience of a very moderate congestion).

Values for $\hat{T}$ have been inserted in Table IV as they represent a figure of merit against which our obtained $\bar{T}^*$ results must be contrasted.

Examining the results in Table IV, we can draw the following considerations. First, roads from 1 to 5 were all the roads where

TABLE II
EXPERIMENT INFORMATION: LOCATION, ROAD SECTION, ROAD LENGTH, FFTT, TRAFFIC LIGHT CT, AND GT

	Road	Section, Direction	Road section length [m]	T_{FFTT} [s]	CT [s]	GT [s]
1	Via B. Croce	P.zza Guerrazzi-Via Queirolo, left	380	34	85	55
2	S. Monica Blvd	Veteran-Sepulveda, left	380	61	120	15
3	S. Monica Blvd	Wilshire-Roxbury, straight	280	17	90	54
4	S. Monica Blvd	Wilshire-Bedford, right	390	30	90	54
5	Lincoln Blvd	Fiji-Venice, back	2300	205	120	60
6	Wilshire Blvd	Midvale-Westwood, right	130	7	150	80
7	S. Monica Blvd	Roxbury-Bedford, right	100	7	90	54
8	Wilshire Blvd	Veteran-Westwood, right	340	33	150	80
9	S. Monica Blvd	Westwood-Sepulveda, right	680	75	120	50

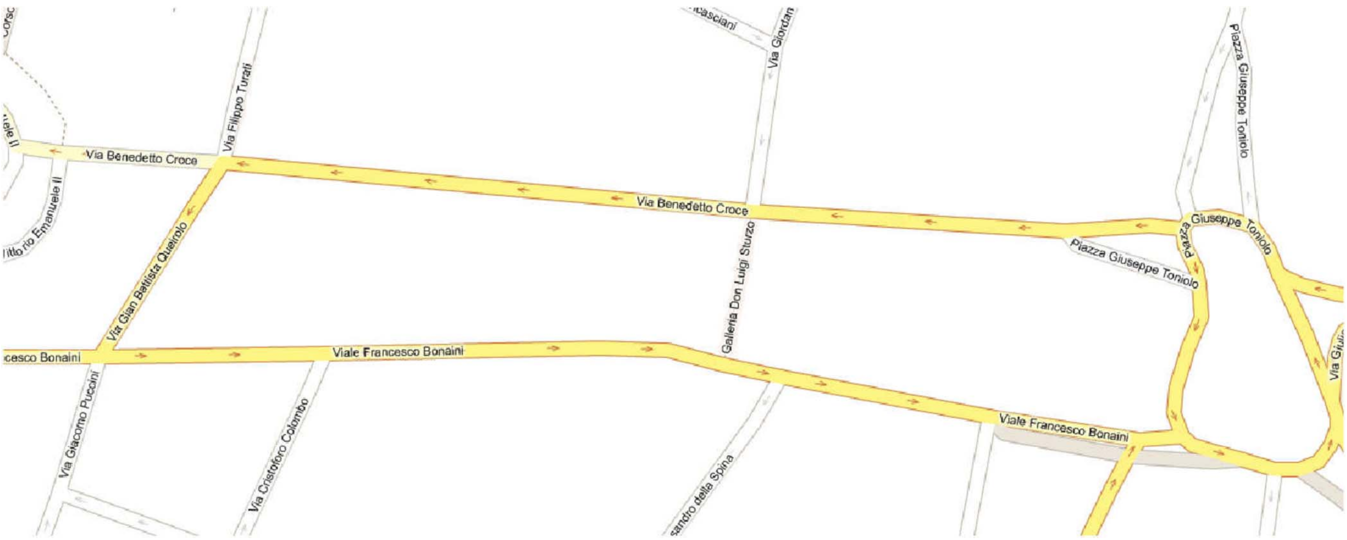

Fig. 4. Experiment site map in Pisa.

TABLE III
ROAD DATA: NUMBER OF LOOPS, CONGESTION THRESHOLD $\bar{T}^*$, S, N, AND H

	# of loops	$\bar{T}^*$ [s]	S [s]	N	H
1	111	93	362	92%	84%
2	134	175	608	80%	87%
3	77	62	987	94%	99%
4	77	82	987	92%	99%
5	30	354	900	100%	97%
6	71	36	454	39%	98%
7	77	42	987	46%	83%
8	71	74	454	37%	100%
9	67	121	493	90%	54%

TABLE IV
COMPARISON: $\hat{T}$ AND $\bar{T}^*$

	$\hat{T}$ [s]	$\bar{T}^*$ [s]
1	64	93
2	166	175
3	53	62
4	63	82
5	265	354
6	77	36
7	43	42
8	103	74
9	145	121

situations of high traffic congestion alternated with situations of no congestion. This is revealed by the values of N and H that were both above the threshold of 80%. The validity of these results is also witnessed by the value of the congestion threshold $\bar{T}^*$ we derived, which was always larger than the value of $\hat{T}$, thus confirming that our algorithm was able to find a congestion threshold value above which cars really incurred in congestion. Indeed, the alternating between states of high congestion and no congestion revealed by our algorithm matches with the empirical knowledge of the traffic situation over those roads.

Of particular interest are also roads from 6 to 8. We deliberately chose these road sections as they are well known to be roads that almost never experience states of congestion. Our results confirm this fact in two different ways. First, for each of these roads, the value $\hat{T}$ is always larger than that provided by our algorithm, thus confirming that cars over these roads enjoy a smooth drive, rarely incurring in a red light, due to the existence of a green wave. Second, the very small values we obtained for N (corresponding to failures of the `Check1` procedure in our algorithm) further corroborate the fact that almost no congestion is visible over those roads.

Finally, a very specific case is given by road # 9. This road section would seem to reveal a stable high congestion state as resulting from a high value of N and a small value of H. However, the obtained $\bar{T}^*$ value is smaller than the value of $\hat{T}$, thus providing a contrasting argument against our initial statement. The motivation for this paradox is as follows: Indeed, this road section often experiences severe congestion on S. Monica Blvd. Nonetheless, the drivers involved in our experimentation were instructed to take a right turn on Sepulved Blvd. to maintain

TABLE V
ROAD DATA: NUMBER OF LOOPS, CONGESTION THRESHOLD $\breve{T}$, $C_{\breve{T}}$, AND $NC_{\breve{T}}$

	$\breve{T}$ [s]	$C_{\breve{T}}$	$NC_{\breve{T}}$
1	55	76%	24%
2	49	100%	0%
3	36	70%	30%
4	50	75%	25%
5	296	47%	53%
6	17	61%	39%
7	13	71%	29%
8	44	77%	26%
9	87	84%	16%

TABLE VI
COMPARISON: $\bar{T}_{\mathrm{HCM}}$ AND PERCENTAGE DIFFERENCE

	$\bar{T}_{HCM}$ [s]	$\frac{\bar{T}_{HCM}-\bar{T}^*}{\bar{T}^*} \times 100$ (%)
1	78	-16%
2	147.5	-16%
3	54	-13%
4	67	-18%
5	341.4	-4%
6	35	-3%
7	33	-21%
8	75	+1%
9	122	+1%

the route on the predetermined tracks. Hence, as this traffic light permits to turn right on red, only very seldom our cars incurred in the delay given by a red light time. This motivates the low value of $\bar{T}^*$, particularly in comparison with that of $\hat{T}$. To overcome this, we carried out a couple of additional experiments with cars going straight at that intersection. As expected in this case, the value of $\bar{T}^*$ always surpassed that of $\hat{T}$.

Moreover, in addition to what was just discussed, we regarded as important to further extend our experimentation to include a comparison of our results with those that can be obtained exploiting the scheme proposed in [23], where samples coming from probe vehicles are contrasted with a set of static thresholds to determine the existence of congestion. While the authors of [23] defined four different thresholds to distinguish among five different levels of congestion, we are simply interested in differentiating between only two situations: the presence of congestion (including medium, congested, and very congested states in [23]) or the absence of congestion (including smooth and very smooth states in [23]). Hence, as the authors distinguish between these two situations with a 25 km/h threshold on roads where the enforced speed limit is 40 km/h, we adapted such a value to the roads we drove on, maintaining the same threshold value for the experiment carried out on the narrow road in Pisa, whereas we increased to 28 km/h the congestion threshold for the experiments conducted on the wide Los Angeles streets. At this point, we converted these speed thresholds into traversal time thresholds $\breve{T}$ and inserted those values in Table V. Table V also reports the percentages of traversals classified as congested $(C_{\breve{T}})$ and those classified as noncongested $(NC_{\breve{T}})$, computed based on the scheme in [23].

Examining $C_{\breve{T}}$ and $NC_{\breve{T}}$, it is evident how the method described in [23] has the tendency to overestimate the situations of congestion, whereas the states of no congestion are underestimated. This bad attitude is confirmed not only by contrasting the values reported in Table III with those obtained with our method in Table V but by resorting to the empirical knowledge of the traffic situation over those roads as well.

As a comparison with [23] cannot be considered sufficient to validate our method because of the tendency of [23] to overestimate congestion, we carried out an additional and final comparison. In particular, we contrasted the values of $\bar{T}^*$ with those provided by the HCM method for each of the road sections under analysis. As reported in Section II, the HCM method exploits the value $\bar{T}_{\mathrm{HCM}}$ that should be interpreted as the average traversal time cars experience when driving on a given road as a function of the intensity of traffic and of the peak capacity of that road [30]. We provide in Table VI a comparison between $\bar{T}^*$ and $\bar{T}_{\mathrm{HCM}}$. As can be seen from the rightmost column of Table VI, the two parameters give almost converging results. In particular, for roads # 5, 6, 8, and 9, the matching between $\bar{T}^*$ and HCM is nearly perfect. This is due to the fact that the Los Angeles Transportation Authority has provided very accurate and updated estimates for the peak capacities of those road sections to be used in the HCM formula [10]. Instead, the values of the peak capacities for roads # 1, 2, 3, 4, and 7 to be used in the HCM method were less precise as simply drawn from the HCM general manual. This motivates why the difference between $\bar{T}^*$ and $\bar{T}_{\mathrm{HCM}}$ in percentage may exceed 20% in these particular cases.

B. *Study of Two Specific Cases*

To provide further evidence of the validity of our approach, we here discuss in more detail a set of specific cases drawn from the two tracks shown in Fig. 2 (traversing Via B. Croce, Pisa, in a counterclockwise direction) and Fig. 3 (traversing N. S. Monica Blvd., Los Angeles, in a clockwise direction), respectively. In particular, we studied the cases and contrasted the results coming from a comparison between roads # 1 and 4, as well as those coming from a comparison between roads # 3 and 7. The motivation for choosing these cases is as follows.

Roads # 1 and 4 were chosen as they exhibited similarities (e.g., length, traffic light phase and speed limit), but differed in terms of the number of traffic lights (two on the S. Monica Blvd. and one at the end of the via B. Croce), lanes reserved for through traffic (two for S. Monica Blvd. and one for via B. Croce), and driving discipline (although we cannot provide any supporting data, our drivers reported that traffic discipline was more strictly adhered to in Los Angeles). The purpose of this first comparison is to argue how these characteristics influenced the value of S on both roads.

The case of roads # 3 and 7 was taken into consideration as these two sections are different but consecutive parts of the same street. Nonetheless, as discussed previously, our algorithm was able to distinguish congestion from noncongestion on road # 3, whereas it was unable to perform as well on road # 7. Our aim, hence, is to better clarify such a situation, showing how the results given by our algorithm should be interpreted, avoiding a misuse that may lead to a contradictory and inefficient selection of travel routes (see Fig. 5).

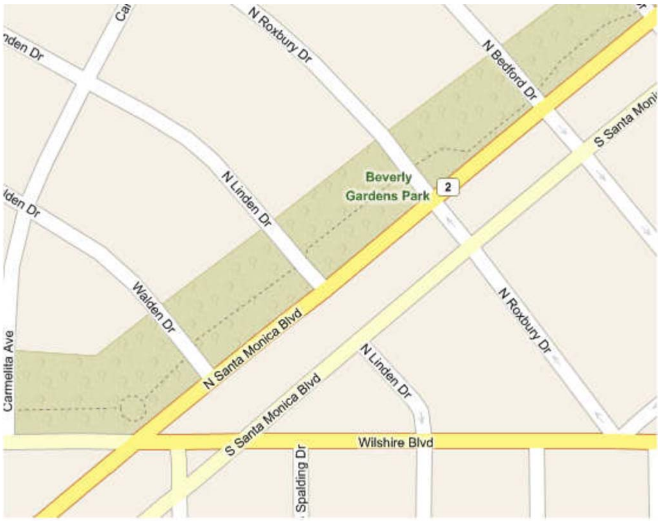

Fig. 5. Experiment site map on S. Monica Blvd., between Wilshire Blvd. and Bedford Dr.

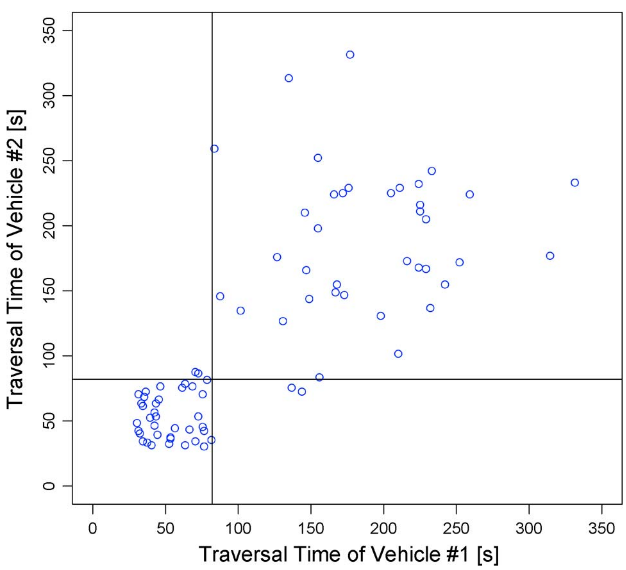

Fig. 6. Congestion graph for S. Monica Blvd., between Wilshire and Bedford.

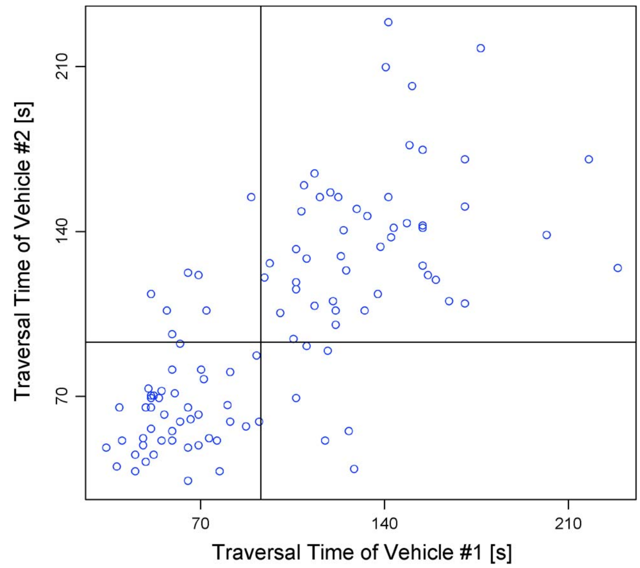

Fig. 7. Congestion graph for Via B. Croce, between Piazza Guerrazzi and via Queirolo.

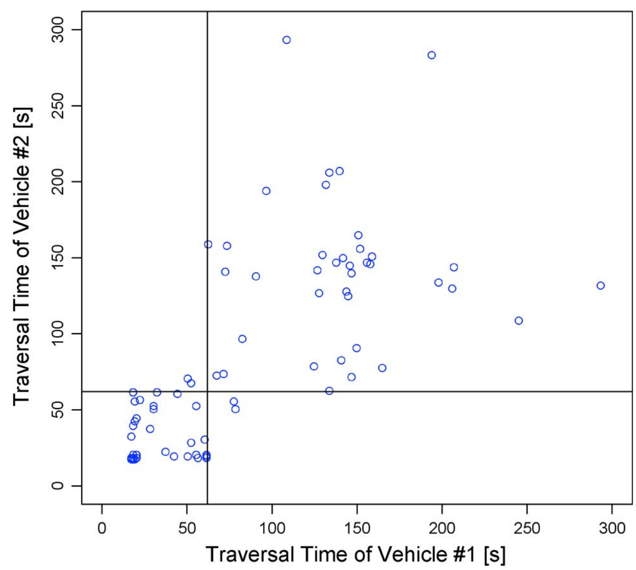

Fig. 8. Congestion graph for S. Monica Blvd., between Wilshire and Roxbury.

Road # 1 Versus Road # 4: What is interesting that emerges from this comparison is that these two roads, namely, # 1 and 4, have very different values of S in spite of a series of similarities in terms of both road characteristics (see Table II) and the results provided by our algorithm (see Table III).

A clear explanation of this phenomenon can be given by observing the congestion graph for these two road sections (Figs. 6 and 7, respectively). What emerges is the following. The points in the graph of Fig. 6 are more clustered and concentrated almost exclusively in the two regions of congestion (top-right area) and of no congestion (bottom-left area). Hence, as in this case, it is easier to intercept both congested and noncongested states; also, the horizon of predictability (the S value) grows larger. This does not apply, instead, to the graph in Fig. 7, where the points are in some sense more scattered on the left semiplane. Here, the 16% of points lies in the top-left area (noisy situation). As a result, the value of S concerning the length of forecasting drops lower.

Obviously, there are real facts behind the explanation we just provided. The fact essentially is related to the number of lanes per considered road. Normally, roads with higher capacities (two lanes or more) or even smaller capacity fluctuations (no presence of obstructing vehicles) experience a faster transition from a congested to a noncongested state. This is exactly what happened to the two-lane road # 4, as confirmed in Fig. 6, whereas the relatively high percentage of scattered points in the left semiplane of Fig. 7 reveals that road # 1 is a one-lane street that easily transitions from a noncongested to a congested state.

Road # 3 Versus Road # 7: What could seem a paradox here is that although road # 3 and 7 are sections of roads belonging to the same street, the former alternates states of no

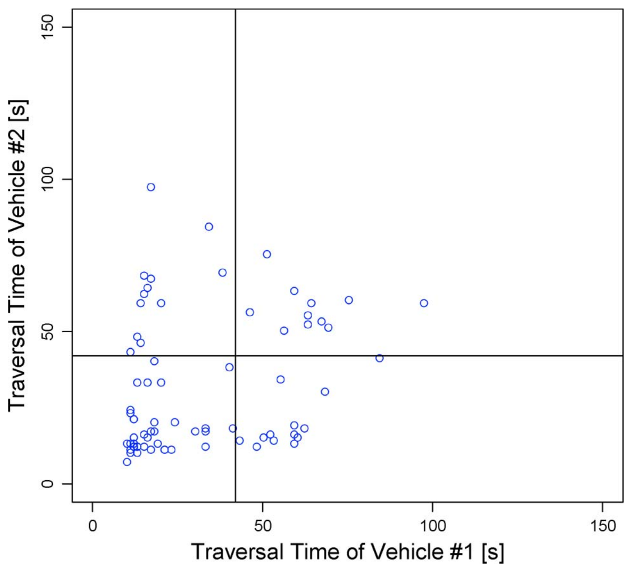

Fig. 9. Congestion graph for S. Monica Blvd., between Roxbury and Bedford.

congestion to states of congestion, whereas the latter almost always experiences a noncongested situation. These situations are highlighted by the corresponding graphs in Fig. 8 (road # 3) and Fig. 9 (road # 7). Indeed, the graph in Fig. 8 has almost no scattering, whereas the graph in Fig. 9 presents a high percentage of scattered points, particularly in the right semiplane devoted to represent congestion. Again, this graphical pattern is easy to explain based on what happens in reality. Indeed, road # 7 has the road section between Roxbury Dr. and Bedford Dr. of very short length. Further, the traffic light at Bedford Dr. is coordinated with all the downstream traffic lights, thus easing the outflow of vehicles that are stuck in queue. For these two reasons, this street experiences a fortunate situation where the shift from a congested to a noncongested state is made easier.

To conclude, the cases we here discussed confirm that our algorithm is really precise in identifying even those situations where abnormal facts may occur.

C. *Dealing With Nonrecurrent Traffic Congestion Causes*

Our mechanism has been designed with the aim of detecting congestion in the situation when both it is determined by recurrent traffic patterns and its cause is due to nonrecurrent (or abnormal) events. The suitability of our mechanism in recognizing congestion in all types of circumstances derives from its ability to follow the evolution in time of the congestion threshold $\bar{T}^*$ of a given road segment as any of its physical properties change (e.g., diminished capacity due to maintenance work) or as it experiences nonrecurrent traffic patterns (e.g., accidents).

To better explain how our mechanism can deal with nonrecurrent congestion events, take the following example where a two-lane road, due to scheduled maintenance work, is reduced to a single-lane road starting from 1 P.M. In such a scenario, clearly, vehicles running through that road between 8 A.M. and 1 P.M. enjoy a noncongested situation. After 1 P.M., since the capacity of the road is halved, vehicles will take longer to traverse it, thus experiencing congestion. If a sufficient number of vehicles traversed that road during the morning, as well as after 1 P.M., then our mechanism would have been able to determine an adequate congestion threshold $\bar{T}^*$ (of value, say, 100 s) that allows one to distinguish states of no congestion from states of congestion. We can also assume that the maintenance work lasts for a few days. Suppose now that the day after, at a given time, an accident occurs, blocking the flow of cars for a very long time (i.e., a large number of cars incur in a state of severe congestion). Under these circumstances, our scheme would have identified a much larger congestion threshold, say 200 s (provided that a sufficient number of cars has incurred in this abnormal event). Up to this point, we have described how our algorithm works in its present form. Obviously, it is not difficult to devise an extension of our model that is able to distinguish different causes of congestion (recurrent and nonrecurrent) simply by observing how the congestion threshold fluctuates provided that a sufficient number of cars experience that specific traffic situation.

As a real example, take into consideration what we observed during one of our experiments in Via Benedetto Croce in Pisa. During a situation where a moderate level of congestion was experienced (the traversal times fluctuated around 100 s), a pedestrian suddenly fell on the sidewalk and, very rapidly, a few minutes later an ambulance arrived, at first stopping and then slowing the flow of vehicles. The problem was fixed, taking no more than 15 min; thus, only a limited amount of vehicles experienced this abnormal event, with traversal times reaching approximately 220 s. As the duration of this event was too short, it did not significantly influence the magnitude of the value of the congestion threshold $\bar{T}^*$. Needless to say, a longer duration of this abnormal event would have had a more serious impact on the value of the congestion threshold, thus allowing one to distinguish it as a new and more severe cause of congestion, with respect to the previous situation (moderate congestion).

In summary, while other mechanisms exist, which are designed to detect nonrecurrent congestion states [34], [35], these typically address the problem of identifying the cause of the events that are at the basis of congestion (e.g., accidents). Instead, our mechanism, as the aforementioned examples confirm, is concerned with the issue of revealing the severity and duration of a congestion state. Nonetheless, our mechanism can easily be adapted to distinguish recurrent congestion from nonrecurrent congestion, thus also resulting in a valuable tool for detecting accident events, for example.

V. Conclusion

We have presented a simple and efficient general-purpose vehicular congestion detection and short-term forecasting algorithm. Our algorithm has been checked on a real testbed driving over 450 mi throughout Pisa and Los Angeles. We have proposed a new definition of congestion, where a road is congested only when the likelihood of finding it in the same congested state is high in the near future. This makes our algorithm easy to implement and effective in providing significant results. Given its characteristics, we believe that this algorithm is well suited for guiding drivers around critical traffic states.

REFERENCES

[1] J. Rybicki, B. Scheuermann, W. Kiess, C. Lochert, P. Fallahi, and M. Mauve, "Challenge: Peers on wheels—A road to new traffic information systems," in *Proc. 13th Annu. ACM Int. Conf. Mobile Comput. Netw.*, Montreal, QC, Canada, 2007, pp. 215–221.

[2] K. Sanwal and J. Walrand, "Vehicles as probes," Univ. California, Berkeley, Rep. UCB-ITS-PWP-95-11, 1995.

[3] G. Marfia, G. Pau, and M. Roccetti, "On developing smart applications for VANETs: Where are we now? Some insights on technical issues and open problems," in *Proc. IEEE Int. Conf. Ultra Modern Telecommun. Workshops*, St. Petersburg, Russia, 2009, pp. 1–6.

[4] M. Roccetti, M. Gerla, C. E. Palazzi, S. Ferretti, and G. Pau, "First responders' crystal ball: How to scry the emergency from a remote vehicle," in *Proc. IEEE Int. Conf. Perform., Comput., Commun.*, New Orleans, LA, 2007, pp. 556–561.

[5] Garmin, PND-Based Mobile Resource Management Solutions, May 27, 2010. [Online]. Available: http://www.garmin.com/

[6] M. Roccetti, G. Marfia, and A. Amoroso, "An optimal 1D vehicular accident warning algorithm for realistic scenarios," in *Proc. IEEE Symp. Comput. Commun.*, Riccione, Italy, 2010, pp. 145–150.

[7] C. E. Palazzi, S. Ferretti, and M. Roccetti, "An inter-vehicular communication architecture for safety and entertainment," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 1, pp. 90–99, Sep. 2009.

[8] L. Yunteng, Y. Xiaoguang, and W. Zhen, "Quantification of congestion in signalized intersection based on loop detector data," in *Proc. 10th IEEE Intell. Transp. Syst. Conf.*, Seattle, WA, 2007, pp. 904–909.

[9] C. H. Wei and Y. Lee, "Development of freeway travel time forecasting models by integrating different sources of traffic data," *IEEE Trans. Veh. Technol.*, vol. 56, no. 6, pp. 3682–3694, Nov. 2007.

[10] Los Angeles Dept. Trans., Department of Transportation Studies and Reports, May 27, 2010. [Online]. Available: http://ladot.lacity.org/studies_reports.htm

[11] Berg Insight, Berg Insight Says 88 Percent of PNDs Sold in 2015 Will Have Integrated Cellular Connectivity, May 27, 2010. [Online]. Available: http://www.berginsight.com

[12] Google Mobile, Google Maps Navigation, May 27, 2010. [Online]. Available: http://www.google.com/mobile/navigation

[13] Android, Introducing Android 2.2, May 27, 2010. [Online]. Available: http://www.android.com

[14] S. Savasta, M. Pini, and G. Marfia, "Performance assessment of a commercial GPS receiver for networking applications," in *Proc. 5th IEEE Consum. Commun. Netw. Conf.*, Las Vegas, NV, 2008, pp. 613–617.

[15] P. Jungme, C. Zhihang, L. Kiliaris, M. L. Kuang, M. A. Masrur, A. M. Phillips, and Y. L. Murphey, "Intelligent vehicle power control based on machine learning of optimal control parameters and prediction of road type and traffic congestion," *IEEE Trans. Veh. Technol.*, vol. 58, no. 9, pp. 4741–4756, Jul. 2009.

[16] R. L. Bertini, "You are the traffic jam: An examination of congestion measures," in *Proc. 85th Annu. Meet. Transp. Res. Board*, 2006, pp. 1–17.

[17] E. Vlahogianni, J. Golias, and M. Karlaftis, "Short-term traffic forecasting: Overview of objectives and methods," *Transp. Rev.*, vol. 24, no. 5, pp. 533–557, Sep. 2004.

[18] "Arterial speed study," Southern California Assoc. Gov., Los Angeles, CA, Rep. 04-006, 2005.

[19] T. Lomax, S. Turner, and G. Shunk, "Quantifying congestion," Transp. Res. Board, Washington, DC, Rep. 398, 1997.

[20] H. X. Liu, W. Recker, and A. Chen, "Uncovering the contribution of travel time reliability to dynamic route choice using real-time loop data," *Transp. Res. A—Pol. Pract.*, vol. 38, no. 6, pp. 435–453, Jul. 2004.

[21] Research and Innovative Technology Administration, National Household Travel Survey Daily Travel Quick Facts, May 27, 2010. [Online]. Available: http://www.bts.gov/programs/national_household_travel_survey/daily_travel.html

[22] H. Wen, Z. Hu, J. Guo, L. Zhu, and J. Sun, "Operational analysis on Beijing road network during the Olympic games," *J. Transp. Syst. Eng. Inf. Technol.*, vol. 8, no. 6, pp. 32–37, Dec. 2008.

[23] Y. Chen, L. Gao, Z. Li, and Y. Liu, "A new method for urban traffic state estimation based on vehicle tracking algorithm," in *Proc. 10th IEEE Intell. Transp. Syst. Conf.*, Seattle, WA, 2007, pp. 1097–1101.

[24] B. S. Kerner, C. Demir, R. G. Herrtwich, S. L. Klenov, H. Rehborn, M. Aleksi, and A. Haug, "Traffic state detection with floating car data in road networks," in *Proc. 8th Int. IEEE Conf. Intell. Transp. Syst.*, Vienna, Austria, 2005, pp. 44–49.

[25] C. de Fabritiis, R. Ragona, and G. Valenti, "Traffic estimation and prediction based on real time floating car data," in *Proc. 11th Int. IEEE Conf. Intell. Transp. Syst.*, Beijing, China, 2008, pp. 197–203.

[26] C. A. Quiroga, "Performance measures and data requirements for congestion management systems," *Transp. Res. C—Emer. Technol.*, vol. 8, no. 1–6, pp. 287–306, Feb.–Dec. 2000.

[27] B. Coifman, "Identifying the onset of congestion rapidly with existing traffic detectors," *Transp. Res. A—Pol. Pract.*, vol. 37, no. 3, pp. 277–291, Mar. 2003.

[28] J. Yoon, B. Noble, and M. Liu, "Surface street traffic estimation," in *Proc. 5th ACM Int. Conf. Mobile Syst., Appl. Services*, San Juan, PR, 2007, pp. 220–232.

[29] F. V. Webster, "Traffic signal settings," Great Britain Road Res. Lab., London, U.K., Rep. 39, 1958.

[30] Transportation Research Board Technical Staff, Highway Capacity Manual, Washington, DC, 2000.

[31] R. Akcelick, "New approximate expressions for delay, stop rate and queue length at isolated signals," in *Proc. Int. Conf. Road Traffic Signalling*, 1982.

[32] L. D'Acierno, A. Carteni, and B. Montella, "Estimation of urban traffic conditions using an Automatic Vehicle Location (AVL) system," *Eur. J. Oper. Res.*, vol. 196, no. 2, pp. 719–736, Jul. 2009.

[33] D. Levinson, "The value of advanced traveler information systems for route choice," *Transp. Res. C—Emer. Technol.*, vol. 11, no. 1, pp. 75–87, Feb. 2003.

[34] H. Dia and K. Thomas, "Development and evaluation of arterial incident detection models using fusion of simulated probe vehicle and loop detector data," *Inf. Fusion*, vol. 12, no. 1, pp. 20–27, Jan. 2011.

[35] A. Skabardonis, P. Varaiya, and K. F. Petty, "Measuring recurrent and nonrecurrent traffic congestion," *Transp. Res. Rec.*, vol. 1856, pp. 118–124, 2003.

Gustavo Marfia (M'11) received the Laurea degree in telecommunications engineering from the University of Pisa, Pisa, Italy, in 2003 and the Ph.D. degree in computer science from the University of California, Los Angeles, in 2009.

He is currently a Postdoctoral Researcher with the Department of Computer Science, University of Bologna, Bologna, Italy. His research interests include biomedical, transportation, and multimedia networking systems, fields in which he has published over 40 technical papers.

Marco Roccetti received the Italian Laurea degree in electronic engineering from the University of Bologna, Bologna, Italy.

He has been a Full Professor with the Computer Science Department, University of Bologna, since 2000. His research interests include computer entertainment, intelligent transportation systems, and web-based applications, fields in which he has authored more than 200 technical papers.

Prof. Roccetti serves as an Associate Editor for many international journals and is active in several Italian and international projects.