Cellular Data Meet Vehicular Traffic Theory: Location Area Updates and Cell Transitions for Travel Time Estimation

Andreas Janecek
Research Group Entertainment Computing
Faculty of Computer Science
University of Vienna, Austria
andreas.janecek@univie.ac.at

Danilo Valerio
Telecommunications Research Center (FTW)
Vienna, Austria
valerio@FTW.at

Karin A. Hummel
Communication Systems Group
Computer Engineering and Networks Lab.
ETH Zurich, Switzerland
karin.hummel@tik.ee.ethz.ch

Fabio Ricciato
Telecommunications Research Center (FTW)
Vienna, Austria
fabio.ricciato@ftw.at

Helmut Hlavacs
Research Group Entertainment Computing
Faculty of Computer Science
University of Vienna, Austria
helmut.hlavacs@univie.ac.at

ABSTRACT
Road traffic can be monitored by means of static sensors and derived from floating car data, i.e., reports from a sub-set of vehicles. These approaches suffer from a number of technical and economical limitations. Alternatively, we propose to leverage the mobile cellular network as a ubiquitous mobility sensor. We show how vehicle travel times and road congestion can be inferred from anonymized signaling data collected from a cellular mobile network. While other previous studies have considered data only from active devices, e.g., engaged in voice calls, our approach exploits also data from idle users resulting in an enormous gain in coverage and estimation accuracy. By validating our approach against four different traffic monitoring datasets collected on a sample highway over one month, we show that our method can detect congestions very accurately and in a timely manner.

Author Keywords
Cellular Floating Car Data, Mobility Sensor, Large Mobility Data Sets, Travel Time Estimation, Congestion Detection

ACM Classification Keywords
H.4.0 Information Systems Applications: General.

General Terms
Algorithms, Experimentation, Measurement, Performance

UbiComp '12, Sep 5-Sep 8, 2012, Pittsburgh, USA.

INTRODUCTION
Improving road traffic management by exploiting smart sensors as well as information and communication technologies is one of the key challenges of an Intelligent Transportation System (ITS). One of the primary goals of any ITS is to reduce the occurrence of congestion events and mitigate their effects by means of Real-time Traffic and Travel Information (RTTI) as detailed, e.g., in [4, 17, 18]. RTTI systems have the potential to reduce overall travel times by suggesting appropriate trip start times, preferable routes, alternative modes of transport, etc. to users and smart vehicles. In order to provide this service, accurate and timely traffic information is required. However, obtaining such information for an entire road network should be cost-effective.

Our approach leverages the signaling data observable in the mobile cellular network to infer road traffic status – and specifically congestion events – in real-time. Thus, instead of a costly deployment of new sensors, we propose to treat the cellular network as a large-scale mobility sensor that estimates physical movement based on mobility management events observed in the network. While almost all previous studies on the topic have considered only mobile phone data related to "active" terminals, i.e., terminals engaged in a voice call or data connection, we contribute with a novel method that exploits the more complete signaling data captured from the network links near the Radio Access Network (RAN) of the cellular network. In this way, also the position of "idle" terminals can be observed, though at a coarser spatial granularity, i.e., Location Area (LA) level instead of cell level. Since "idle" terminals are the overwhelming majority of the mobile terminal population, this approach allows to reach much better coverage. To benefit both from the large set of idle terminals and the more accurate positions of active users, we propose a two-stage approach. First, large-scale

coarse-grained signaling data from all the terminals are used to estimate travel times and *detect* a congestion. Then, fine-grained data from the subset of active terminals is leveraged to *localize* – and possibly *classify* – the congestion events more accurately from the rate of cell handovers.

We present experimental results from a case-study based on an unprecedented set of diverse data sources. The considered anonymized mobile cellular dataset covers the signaling traffic from 2G/3G cells of a real operational network. Additionally, four different road monitoring sources are considered: static sensors, a toll system, floating taxi data, and radio broadcasts. In a case study, we analyze one full month of data and 74 different congestion events on a selected highway segment. The results indicate that our solution is able to detect road congestions before any other monitoring technology and with an acceptably low rate of errors.

BACKGROUND AND RELATED WORK

Leveraging cellular networks for ITS is a recent interdisciplinary field of research. We summarize fundamental concepts of vehicular traffic estimation, state-of-the-art methods for traffic monitoring, and mobility management signaling in cellular networks. Then, we present related work on using cellular data for mobility and vehicular traffic estimation.

Traffic patterns

Congestion events on highways can be delimited by two fronts: the *upstream front*, where vehicles enter the congestion zone and slow down, and the *downstream front*, where they leave it and accelerate. Depending on the movement of these fronts and other characteristics of the traffic flow, further differentiations of traffic congestions are possible [16]. One of the main models in traffic theory, namely the *three-phase* model [10], describes three states of a highway segment:

- *Free flow:* No congestion is present.

- *Synchronized flow:* congestion with a significant "synchronization" of traffic is observed (i.e., all vehicles proceed at similar, lower speeds at all lanes) and the downstream traffic front usually remains at the bottleneck. Road constructions and lane reductions typically fall in this category.

- *Wide moving jam:* congestion showing flow "synchronization", but characterized by a sharp change of vehicles speed and the fronts may move upstream.

The latter two situations are not mutually exclusive and may occur simultaneously, e.g., in case of accidents occurring in heavy traffic situations. We will describe how to discern the three states of a highway by investigating cellular data.

Traffic monitoring systems

Road traffic congestions can be monitored and estimated by means of various technologies which can be categorized into:

- *Point-based:* Stationary recording systems such as traffic counters or video systems

- *Distance-based:* Pairs of static sensors computing travel times between them often based on tagged probe vehicles, e.g., by plate recognition or RFID transmission

- *Dynamic:* Floating car data, i.e., data generated by vehicles reporting continuously their location and speed to a central server via short-range wireless links or via cellular networks.

A detailed survey on latest developments of data-driven ITS can be found in [21]. The traditional approach to road traffic monitoring involves the use of static infrastructure sensors working *point-based* or *distance-based*, such as visual sensors (cameras), magnetic induction loops, and microwave radars. This approach provides fixed-point or short-section traffic information extracted from vehicles passing the detection zone but suffers from high investment and installation costs. In order to gain a realistic and complete view of traffic conditions, a large quantity of sensors must be installed. Monitoring short-section traffic information, such as travel time between two sensors, may be prone to privacy constraints since private vehicles are tracked based on their license plate number. In other cases they might be limited to specific types of vehicles, e.g., trucks using toll transmitters.

Other systems rely on mobile devices equipped with positioning technologies such as the Global Positioning System (GPS) that report continuously (e.g., via 3G network access) the vehicle location and speed to a central server. In this *dynamic* approach, the information can be used to estimate traffic speed and intensity for areas where the density of the probe vehicles is high enough. Several recent studies focus on traffic monitoring using GPS equipped probe vehicles [5, 11, 13, 19]. Similar to the taxi validation data used in our work, the study in [19] uses floating car data based on GPS receivers installed in taxis. The requirements of processing floating car data in the large-scale are summarized in [12]. In [20], patterns are derived from taxi cab trajectories to suggest routes of estimated shortest travel time while incorporating day of the week, time of day, weather conditions, and driving strategies. Disadvantages of these approaches are privacy concerns, limited representativeness of the traffic situation and limited GPS accuracy in particular nearby high buildings (urban canyons).

Mobile cellular networks

The cellular infrastructure is composed of a Core Network (CN) and a Radio Access Network (RAN). The CN is divided in two distinct domains: Circuit-Switched (CS) and Packet-Switched (PS). Mobile terminals can "attach" to the CS for voice call services, to the PS for packet data transfer, or to both. Radio communication occurs between a mobile terminal and a fixed base station serving one or more radio cells (sometimes also called "sectors"). Cells are the smallest spatial entities in the cellular network. The geographic area covered by each radio cell, which often includes or is close to the corresponding base station location, has a radius that varies from a few dozen meters (microcells) up to several kilometers (macrocells) depending whether the area is an urban, suburban, or a country-side area. Depending on the radio bearer, radio cells can be classified as 2G (GSM/EDGE), 3G (UMTS/HSPA) or 4G. Our dataset comprises data from both 2G and 3G terminals.

At any time each mobile terminal can be in *active* or *idle* state. During voice calls and data transfers, i.e., while send-

ing and receiving IP packets, the terminals are in active state. When the voice call or the data connection (i.e., the so-called PDP-context in 3G terminology) is closed [open], the terminal switches to the idle [active] state. It is important to remark that having an open data connection is a necessary, but not a sufficient condition for a terminal attached to the PS domain to be in active state. In fact, the terminal will switch to idle after a certain timeout from the last data packet sent or received (e.g., 5 seconds). When a new packet arrives, the terminal switches again to active. Therefore, also "always-on" terminals with permanently open data connections remain in idle state most of the time, and switch to active only during packet bursts, e.g., during downloads. The overwhelming majority of mobile terminals are in idle state at any generic instant. Note that state transitions are not necessarily communicated to the network, unless they occur together with other events that require signaling exchange.

The instantaneous position of each terminal is known by the cellular network – and can be inferred from the network signaling data – with different spatial granularity. Cells are grouped into larger logical entities, i.e., Routing Areas (RAs) and Location Areas (LAs) for the PS and CS domains respectively (note that one LA can contain one or more RAs and any RA is always contained within one LA). In order to remain reachable, active and idle terminals always inform the CN whenever they change LA and/or RA. Terminals in active state reveal also cell changes within the RA/LA to the network. In other words, the position of active users is known by the network at the cell level, while the position of idle users is known only at RA/LA level.

Traffic estimation schemes based on cellular data

Mobility information available in the cellular network can be extracted and used as a basis for mobility studies. We distinguish two main approaches: "CDR" and "passive monitoring". The latter includes sub-variants with very different characteristics in terms of costs, software and hardware complexity, and achievable accuracy.

Call Data Records (CDRs) are tickets generated by certain network devices and sent to the billing system. They are produced whenever the user initiates or terminates a voice call, data connection or SMS/MMS envoy. The CDR contains summary information about the call/connection, including the mobile user identifier and the start/end timestamps. The CDR format is not standardized, and the amount and quality of the additional information that is contained in a CDR may differ across networks and operators. CDR always include the starting cell where the call/connection was initiated and often (but not always) also the final cell where it was terminated. In some settings CDR data contain also some information about the intermediate cells visited during the call. CDR are stored in dedicated databases from which they can be easily retrieved. Therefore, extracting CDR data does not impose any particular additional cost, and for this reason CDR have been the first source of data for human mobility studies based on mobile cellular data [7, 15, 2]. Dedicated to traffic estimation, in [2], a CDR dataset with cell handover information is used for measuring traffic speed and travel time across a highway segment of 14 km for several weeks. The results indicate a good correspondence between the cellular data and validation data from magnetic loop detectors. Still, the study is limited to active users which are only a small fraction of the terminal population.

The *passive monitoring* approach is based on the observation of the signaling messages exchanged between the mobile terminals and the network. It requires a large monitoring infrastructure to tap the network links and parse the signaling protocols. The cost of the monitoring installation, as well as the achievable accuracy and coverage, depend heavily on which network interfaces are monitored. Monitoring the links within the PS-CN (as done, e.g., by [9]) is the simplest option but allows to monitor only the terminals with an open data connection, i.e., only a small fraction of the total terminal population, especially on highways. Instead, by monitoring the links between the CN and the RAN, one can observe RA/LA changes of *all* users, including idle ones [6]. Our dataset is based on this approach, and contains data from CS and PS users from both 2G and 3G cells. Finally, a third approach can be followed, namely monitoring at sub-cell level via power measurement reports of the links within the RAN which requires additional monitoring installations, as done, e.g., in [4] for a limited geographic area.

Only few studies exist on the specific problem of real-time road traffic estimation from cellular network signaling. An early attempt is found in [1] based on double-handovers, i.e., pairs of cell handovers. In [14], the feasibility of using mobile phones as traffic probes is analyzed. The authors mention that, compared with available alternatives, mobile phones offer some appealing characteristics such as sample size, coverage, and cost. Different monitoring approaches are surveyed and classified in [17], while in [6] possible steps towards the extraction of vehicular mobility patterns from passively monitored 3G signaling data is detailed. Our work extends this approach by introducing detailed strategies for traffic congestion estimation and a large-scale evaluation against other data sources.

DATA SETS

In this section we describe the cellular data and all datasets that have been used for validating our results.

Cellular data

Our system collects anonymized traces of signaling traffic from the operational 2G/3G network of a major Austrian mobile operator. On average, the traces consist of 400–500 million events per day, and contain signaling messages for both the PS and CS domains captured on specific interfaces of the cellular network (for details on the 3GPP architecture refer, e.g., to [8]). To preserve user privacy, all sensitive identifiers are removed from the traces. Distinct users are discriminated only by means of pseudonyms computed via a one-way hash function. The pseudonyms are changed every 24 hours. The traces consist of event-based tickets containing the following fields:

- `anonymous_ID` of the user generating the event
- `timestamp` of the event

- cell information including the `LAC` (Location Area Code) and `Cell_ID`, the `coordinates`, `type`, `beamwidth`, and `direction` of the base station antenna[1]

- information about the event such as the `type_of_event` and the `input_source`, i.e., the capturing interface where the signaling message was observed.

Validation data

Four datasets are used to validate our estimation approach where each dataset originates from a different source. The description of the validation data-sets includes particularities of the sample target highway used in our study.

Sensors - *point-based road monitoring data:*
Fixed sensors measure speed and traffic at stationary points. These sensors are either placed under the road (inductive, magnetic, etc.), or aside/above the road (e.g., radar, laser or ultrasound). Our target highway is covered by nine stationary sensors in every direction which report number and speed of vehicles passing this sensor every 60 seconds.
- *Advantages*: Very detailed information is available like the type of vehicle, speed/capacity per lane, road surface conditions etc. These sensors are especially useful when installed at on/off-ramps or highway-intersections. The temporal information is very accurate (updated every 60 sec).

- *Disadvantages*: The sensors provide only point-based information on traffic conditions at few sections, in a sort of spatial sampling. A dense deployment of such sensors across the whole highway network is economically unfeasible. Incidents occurring shortly before a sensor cannot be detected.

Toll - *distance-based road monitoring data:*
On our target highway, trucks are charged based on the actual travel distance. Each truck is identified via a mandatory RFID-based electronic car toll transponder. This system can be used to calculate *speed-over-distance* between two stationary sensors by computing the average travel time of trucks between two toll gantries over bins of 15 minutes. It has to be mentioned that the data provided to us is already aggregated without providing any user information (neither RFID tags nor license tags are available).

- *Advantages*: Incidents between or directly before sensors (toll gantries) can be detected.
- *Disadvantages*: The number of probe vehicles is limited to trucks, which are not allowed to travel during night, weekends, and holidays. Moreover, the speed limit for trucks is often different than for other vehicles. Temporal granularity is limited (update every 15 minutes).

Taxi - *dynamic floating car data:*
This data source consists of a floating car data repository based on GPS. The probe vehicles are part of a taxi fleet and equipped with GPS devices that periodically transmit the vehicle speed and location to a central system.

- *Advantages*: The data is dynamic and not limited to the location of sensors or gateways.

- *Disadvantages*: As taxis are used as probe vehicles, the coverage area is usually limited to urban areas. Moreover, taxi drivers are not representative for all driver types.

Radio - *event-data base:*
This event data-base extracted from a radio broadcast station contains all road incidents on our target highway for the considered period. The broadcast traffic news are partly based on a cooperation with highway maintenance authorities, and partly on a campaign where registered users can report road incidents directly to the radio station.

- *Advantages*: In many cases, broadcast information is particularly precise and faster than road monitoring data, e.g., a listener reporting an accident right after its occurrence.

- *Disadvantages*: The accuracy heavily depends on subjective grading of users, e.g., five minutes in a traffic jam may feel awfully long for an individual, but this congestion may be only temporary or already regressive; this may lead to false alarms (false positives).

METHODOLOGY

In order to monitor traffic conditions as accurately and timely as possible, our system adopts a two-stage approach for the estimation of travel times from cellular data:

- *Stage 1* aims at detecting congestion events from LA/RA changes reported by active *and* idle mobile terminals

- *Stage 2* aims at further localizing (and possibly classifying) the congestion event detected at Stage 1 from the cell changes reported by active terminals.

The stages differ in type and amount of cellular network signaling events that are considered, and in terms of length and granularity of the considered road section. Stage 1 offers a continuous estimation of the travel times at large-scale with good terminal coverage. However, its spatial accuracy is limited by the large radius of LAs (up to several kms). Stage 2 offers higher spatial accuracy (up to few hundred meters) but less terminal coverage. It can be used to further localize the area of congestion, but only after that congestion has been detected. We remark that in general it is not possible to rely exclusively on Stage 2 in isolation, since the number of active terminals in the region of interest is not always sufficiently large to carry out a reliable estimation. We will show later that the combination of Stage 1 and Stage 2 has also the potential to distinguish between congestion events, namely *wide moving jam* and a *synchronized flow*.

LA and RA updates – Stage 1

Despite the wide-spread use of smart-phones often involved in background data activity, the vast majority of mobile terminals remain most of the time in "idle state" and emit only LA and RA updates. In [6, Section 4] we have shown that the exclusive observation of "active" terminals produces a biased picture of the overall human mobility. In our highway study, Stage 1 is able to track up to 20 times more terminals than Stage 2. This factor depends on the road section, the size of the LA/RA, and the number of cells in the proximity. Since in urban and semi-urban areas the size of LAs is relatively small (5-10 km on our target highway) the congestion

[1]We use antenna parameters to determine more exactly the cell coverage area.

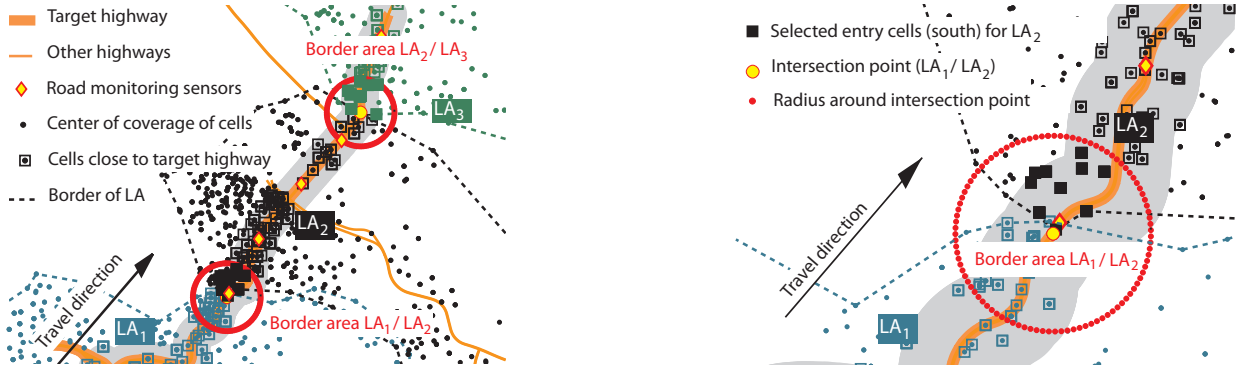


Figure 1. Tracking devices over the distance of LAC 2 based on events triggered in (1) the entry cells (south) of LAC 2 in Start area, and (2) the entry cells (south) of LAC 3 in Arrival area.

detection delay for Stage 1 is small. The same consideration applies when the analysis focuses on RA updates (recall that RA are always contained within LAs). For simplicity, in the remainder of this paper we always refer to LA updates instead of LA/RA updates.

The calculation of the travel time t of one mobile terminal is processed as follows: Every time the mobile terminal attaches to a cell belonging to a new LA, it emits an LA update event. In order to identify the terminals traveling along a target road segment, a suitable set of cells is pre-selected meeting the following criteria: (i) the cells are located at the border of two LAs and (ii) they are in close proximity of the target highway. We term these cells *entry cells*. Let now LA_1, LA_2, and LA_3 be three adjacent LAs crossed in sequence by a mobile terminal. The set of entry cells of LA_2 indicating a change from LA_1 to LA_2 is termed *border area* LA_1/LA_2while the set of entry cells of LA_3 indicating a change from LA_2 to LA_3 is termed LA_2/LA_3.The mobile terminal (vehicle) traveling from LA_1 to LA_3 will generate at least one event in both LA_1/LA_2 and LA_2/LA_3. The travel time estimate of the terminal is now simply calculated as $t = t_a - t_s$, where t_a is the time of the first event in LA_2/LA_3 and t_s is the time of the last event in LA_1/LA_2. The border areas have been introduced to increase the chances of capturing cell events at the borders and that the selection strategy of using the last and the first event, respectively, minimizes the length of the road segment under investigation between the two border areas. Further, we assumed in this calculation that the highway is the fastest connection between the start and arrival area, i.e., the fastest mobile device users are all traveling on the target highway.

Figure 1 shows an example that depicts the portion of the target highway used in our study. Consider a terminal traveling from south to north. The entry cells of LA_2 belonging to border area LA_1/LA_2 are depicted in the right part of the figure. The radius around the intersection point (red circle, 1 km) is chosen to be larger than the radius around the target highway (gray area, 0.55 km) because in many cases terminals traveling at high speeds do not attach to the first cell of the LA but emit the LA update in one of the subsequent cells. After extracting the set of all users who generate an event in the southern entry cells of LA_2, and later in the southern entry cells of LA_3 (border area LA_2/LA_3 in the left part of Figure 1) we are able to compute the travel time for each vehicle and deviations in travel time.

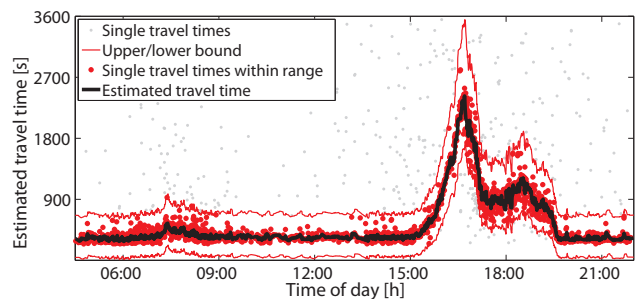


Figure 2. Travel times through LA_2 using floating cellular data

The single travel time of all terminals tracked over one day using the Stage 1 approach are visualized as gray dots in Figure 2. Obviously, the mean travel time of all users is not a good congestion indicator since at any time (and most prominently between 6:00 and 9:00 am) there are users that travel on other, slower side roads and/or take a break along their journey, regardless on whether congestion is present on the highway or not. We found also some very fast users during the traffic jam in the afternoon, that might correspond to motorcycles or vehicles driving on emergency lanes. The estimation scheme has to consider both artifacts.

Congestion is detected by evaluating the travel time of fastest users. Therefore, a threshold is introduced and the travel time of fastest users is evaluated against this threshold. However, it is challenging to create the set of fastest users in a robust and reliable way. Therefore, in a pre-processing step, we remove "ultra-fast" outliers corresponding to, e.g., helicopters and other exceptional phenomena. Then, we remove the devices whose travel times are not in the interval $[t_{est} - \Delta, t_{est} + \Delta]$ centered around the current estimation of the average travel time t_{est} (see Figure 2). The width of the interval Δ is chosen as $\Delta = t_{min} + t_{est} \cdot r$, wherein r is an adjustable parameter and t_{min} is the fixed minimum travel time as defined below (after some trials we found that $r = 0.2$ is a good setting). The new estimated travel time is calculated as the mean travel time of all devices tracked within the last 60 seconds whose travel time was within the interval $[t_{est} - \Delta, t_{est} + \Delta]$; t_{est} is set to this new estimated travel time.

Δ increases during congestion (due to large t_{est}) and decreases in free-flow situations (when t_{est} is small). The minimum travel time between two border areas is defined as the travel time at the maximum speed limit. However, it is difficult to select exact locations for estimating the length of the highway segment between these areas. In fact, due to the nature of the cellular network and to the heterogeneity of mobile terminals, one cannot predict exactly at which position a cell change occurs. Instead of using a possibly imprecise static value for the minimal travel time we introduce an adjustable parameter κ which defines the minimum travel time through an LA as the travel time of the user at the κ-quantile of all users tracked for this LA over one day. In other words, we set the minimum travel time to the maximum travel time of the fraction κ of fastest users over one day. As a result the minimum travel time increases with increasing κ. We will later discuss the impact of the value of κ on the detection performance (cf. when discussing Table 3).

Cell handovers – Stage 2

In Stage 2, we exploit the more accurate location data of terminals in active state available on cell granularity. However, identifying the cells that serve mobile phones on the highway is not straightforward. For example, our target highway is placed in a semi-urban region. Given the high density of mobile device customers, these areas are usually characterized by redundant cellular coverage, i.e., many overlapping cells. In addition, 3G networks are affected by the so-called *cell-breathing effect*, causing the coverage of a cell to change dynamically depending on the level of cellular traffic load and interference. Hence, in a first step, we identify all cells that are in proximity of the target highway. Second, we detect all pairs of these cells that present a large number of cell handovers. By considering the rate at which consecutive cell handovers occur, we are able to finally define sequences of cells representing a given highway travel direction. An example of our target highway segment and sequence of cells and cell pairs for LA_2 is shown in Figure 3. Cells are denoted with capital letters (A, B, C, ...) and cell pairs are represented by arrows. It can be seen that by using cell handover events, LA_2 can be further divided. Five different sections are marked, which correspond to the location of road sensors on this highway sector.

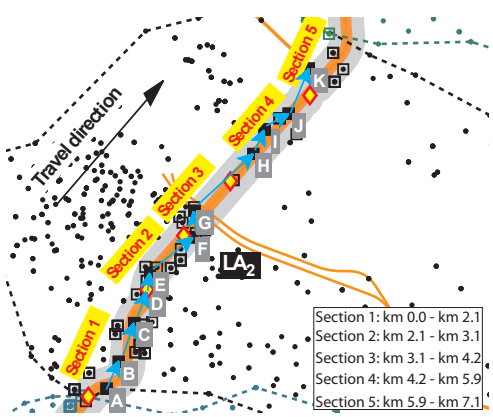


Figure 3. Sequence of cells used for tracking cell handover events of active users in Stage 2.

The time between two cell handovers by the same terminal represents the time interval the terminal stays within the coverage area of the cell, i.e., the *dwell time*. In order to perform two consecutive cell handovers, a mobile phone must be active during the cell handovers, e.g., remain in a voice call or data transfer during the entire dwell time within the cell. However, the slower is the user speed, the larger is the dwell time, and the lower is the probability that the user will remain in active state until the next cell transition. Therefore, traffic congestions and slow traffic situations are characterized by a drop in the number of handovers, rather than by an increase of the inter-handover time. On the other hand, the lack of handover events could be also due to the opposite scenario, i.e., very low road traffic intensity (few users driving along the road, hence few handover events). That explains why Stage 2 alone can not be taken reliably as a detector for congestion, but serves well the purpose of locating traffic events more precisely in combination with Stage 1.

ANALYSIS RESULTS

We now evaluate our approach by analyzing the congestion estimation accuracy and delay of estimation for Stage 1 (all terminals). Further, we exemplify how incidents can be further localized and types of congestions can be differentiated. Hereby, we compare our results with the estimates derived from traditional sensory and broadcast radio data. We conducted all experiments as real-time emulations, i.e., although all data are available in batch, we process the data in an online stream fashion that emulates the real progress of time.

The sample period stretches over 31 days. While the mobile cellular data are available for all days, the validation data are only partially available due to the nature of the sensors or temporary faults. Data availability is summarized in Table 1. *Toll* data are partly missing on weekends and holidays due to a ban on truck vehicles, *sensor* data are missing for some days due to a problem in the recording system, and concerning the *taxi* data we do not have access to the first part of the sample period. Radio data are available throughout the whole period. The target highway stretches over 32 km from a rural area into the center of Vienna, Austria. It is covered by four different LAs; the considered users are traveling northbound.

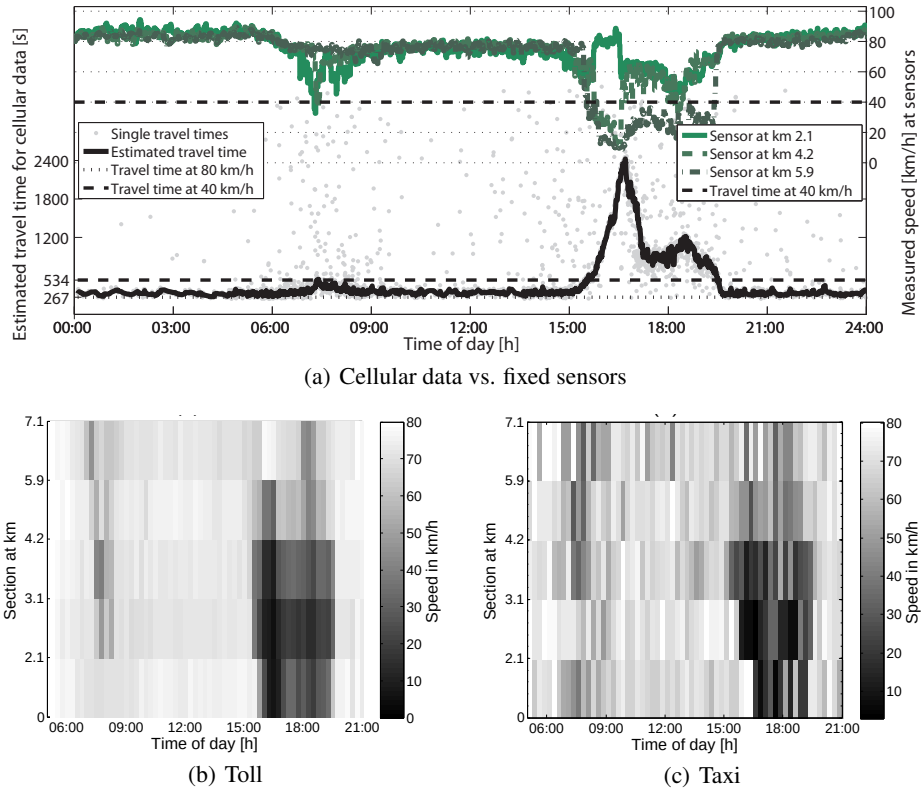
Table 1. Available data during target period (June 1st to Juli 1st, 2011).

Mobile:	available for all days
Toll:	partly not available on weekends
Sensors:	available except June 3 - June 10
Taxi:	available except June 1 - June 17
Radio:	available for all days

Detecting congestions

As there is no general agreement in the literature about the definition of a traffic congestion [3], we adopt the following definition.

> **Definition of congestion:** A road segment is marked as congested if the estimated speed of the fastest vehicles considered falls below half the speed limit (i.e., the travel time is doubled).



(a) Cellular data vs. fixed sensors

(b) Toll

(c) Taxi

Figure 4. Comparison of estimated travel times on June 30th for cellular data and speed measured with fixed sensors, toll data and taxi data over the distance of LA_2 (see Figure 1) using Stage 1. The incident in the afternoon was broad-casted on radio at 15:33 h ("heavy traffic").

We mark a congestion event in our dataset if *at least one data source* out of *toll*, *sensors* or *taxi* triggers the above condition. In this way, 74 congestion events could be identified over the sample period, and 58 of them were sent as broadcast also via radio. All events in the *radio* dataset could be detected by at least one of the other three validation sources.

Now, we evaluate our approach (Stage 1) and refer to the approach simply as *mobile* when comparing the results to the other technologies. Table 2 shows the congestion events that were identified by each type of validation data. Some events could not be detected due to temporary unavailability of the corresponding data (NA in Table 2). Some other events were missed by some source although the corresponding data was available; such *False Negative* events are marked as FN in Table 2. It can be seen that *taxi* misses 11 congestions, followed by *sensors* (7 FN) and *toll* (4 FN). Our approach, *mobile*, could identify between 68 and 74 congestions correctly out of 74 (see Table 3, bottom).

Table 2. Numbers of identified congestions.

	Correctly identified	No data available (NA)	Not detected (FN)
Toll:	66	4	4
Sensors:	25	42	7
Taxi:	25	38	11
Radio:	58	16	0
Mobile:	depends on parameter κ, see Table 3		

Congestion detection via different technologies

We analyze the quality of congestion estimation based on cellular data vs. other data sources in terms of the *number of detected congestions* and the *estimation delay*.

Let us first exemplify a comparison of a single congestion. Figure 4 shows the estimated travel times through a given LA over a whole day for all sensor-based traffic data sources. The estimated travel time for cellular data is shown in Figure 4(a) (lower curve, left y-axis) as well as the speed measured at stationary road sensors for the same highway segment (upper curves, right y-axis). Figure 4(b) contains similar information for the toll data where dark regions refer to sections and periods of lower speed. The placement of stationary sensors and toll gantries differ which causes different detection characteristics of the technologies. Taxi data were aligned to the segments defined by toll gantries (Figure 4(c)). A huge traffic jam in the afternoon is easily detectable by all datasets. A less severe congestion is present also in the morning which is detected by all sensor-based validation data sources as well as our approach. However, *toll* identified the congestion with a delay of about 30 minutes compared to *sensors* and to our approach (Stage 1).

Table 3 reports the average advance ("-") or delay ("+") of *mobile* against each validation data source. The differences are given in seconds and the number of comparable events (identified by both sources) is given in brackets. Different values of κ, which determines the fraction of users consid-

ered as being in the fastest κ quantile, impact the performance of congestion detection (at Stage 1). For small κ (0.03 or 0.04) our approach is very aggressive as it identifies congestions much faster than all other ITS sources (e.g., 533 seconds faster than *toll* on average), but it identifies also False Positive (FP) events, i.e., congestion events that are not confirmed by any other dataset. Conversely, for larger κ the number of FP decreases, but the advance over other data sources is also reduced. Moreover, a few congestions are missed (False Negatives, FN). Setting $\kappa = 0.05$ is a good trade-off between FPs, FNs and advance over other ITS sources. All experimental results further reported are obtained with $\kappa = 0.05$.

Table 3. Delay of *mobile* vs. validation data (in s).

κ	0.03	0.04	**0.05**	0.06	0.07	0.08	0.09
Toll:	-533 (66)	-282 (66)	**-152** (65)	0 (62)	+72 (62)	+175 (61)	+196 (61)
Sensors:	-586 (25)	-377 (25)	**-259** (24)	-93 (22)	-15 (22)	+37 (22)	+79 (22)
Taxi:	-583 (25)	-383 (25)	**-311** (25)	-182 (24)	-113 (24)	-11 (23)	+7 (23)
Radio:	-451 (58)	-195 (58)	**-177** (57)	+5 (57)	+77 (57)	+197 (56)	+217 (56)
Numbers of identified congestions for *mobile* data							
Correct:	74	74	**73**	70	70	68	68
FN / FP:	0 / 65	0 / 19	**1 / 3**	4 / 2	4 / 1	6 / 1	6 / 0

Figure 5 compares all available data sources in terms of detection delay (or advance). *Mobile* (Stage 1) is always faster than all other sources in detecting congestion. The average (from Table 3) and median (from Figure 5) delay for *mobile* vs. other sources are always negative.

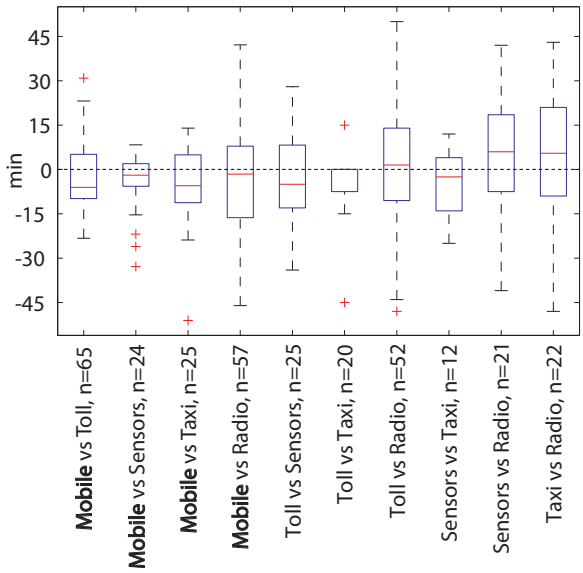


Figure 5. Boxplots of detection delays between data sources (edges of the box are the 25th and 75th percentiles). n is the number of congestions that could be compared for each dataset pair.

Differences of detection times between our approach and the validation data are plotted as histogram functions in Figure 6. As example, most intervals between *mobile* and *sensors* are evenly distributed around 0. For some congestions, the sensor data has a significant delay which leads to the average advance of our approach. For *toll* and *taxi*, the density functions are rather similar, for *radio* the high standard deviation is the reason for the heavy tailed distribution.

Detection delay by type of incident

When an incident is broadcasted on the radio, its record includes the cause of a congestion which can be used for classification (accident, broken vehicle, heavy traffic, and unknown). We classify incidents into two groups, i.e., suddenly appearing events without any prior indication such as accidents and broken vehicles (denoted as 'accidents' in Figure 6), and all other events. These results confirm the fast detection by *mobile* for different types, but also show that *radio* is much faster when focusing only on accidents and broken vehicles which can be easily identified by users. Surprisingly, *sensor* identifies suddenly appearing congestions also fast, however, only four congestions could be compared.

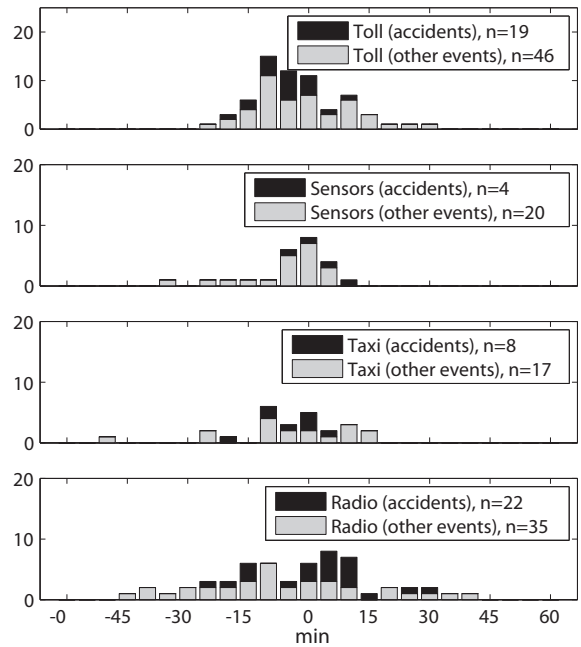
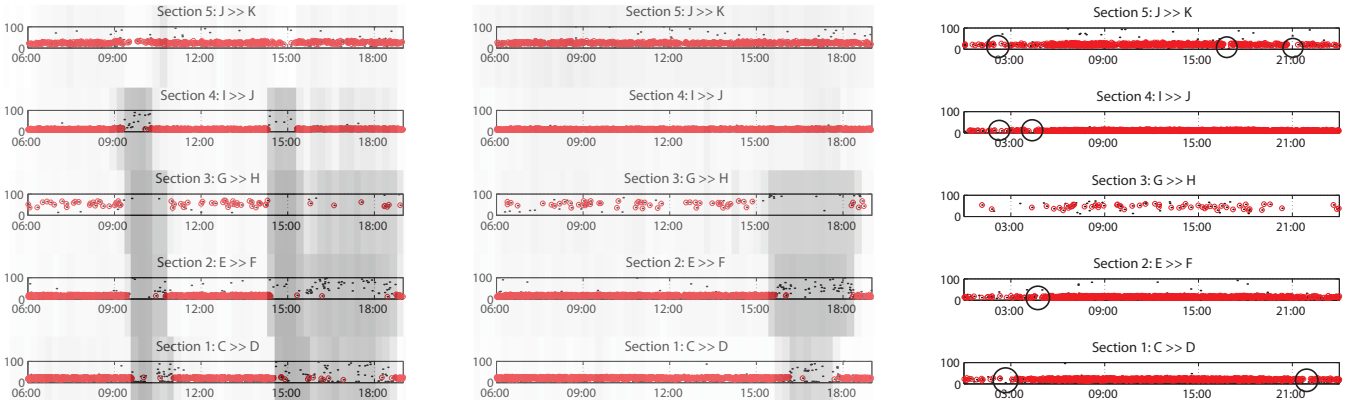


Figure 6. Detection delay *mobile* vs. validation data separated by type of incident.

On further localizing the area of a congestion

We will now demonstrate how Stage 2 can help to further localize a congestion within an LA up to a few hundred meters while Stage 1 can only localize a congestion at a granularity of several kilometers. The segment we are using for demonstration is LA_2 which covers our target highway for about 6.4 km (cf. Figures 1 and 3). For comparison, we use the most accurate sensor data available for this segment, i.e., *toll*, which is measured at six toll gantries in the area of LA_2.

In Figure 3, LA_2 has been divided into five smaller sections defined by the toll gantries in LA_2. Active terminals moving along these sections will perform multiple cell handovers. In order to perform a fair comparison among the mobile cellu-



(a) Two congestion caused by accidents (b) Congestion caused by heavy traffic (c) No congestion

Figure 7. Left: Two congestion caused by accidents (June 16). Middle: Congestion caused by heavy traffic (June 9). Right: No congestion (June 3). Black dots refer to cell handovers, red edges refer to terminals performing handovers close to the average rate, gray overlay refers to toll data. Mobile terminals travel from section 1 to section 5.

lar network and the toll sensors, we selected a representative pair of cell-handovers for each section. Figures 7(a) and 7(b) show pairs of cell handover events selected for each section (cf. Figure 3) and – as an overlay – the toll data (darker regions refer to lower speed). Each terminal performing the selected cell handovers is plotted as a black dot. Terminals with a cell handover rate that does not deviate too much from the average rate are additionally edged in red. The absence of handover events is an indicator for a traffic congestion in the specific section. The traffic jam in the morning on the left side of Figure 7(a) started shortly after 9 am in section 4 and the traffic jam in the afternoon started around 2:30 pm, also in section 4. Section 5 (down-stream of section 4) was not influenced by the congestions – neither in the morning nor in the afternoon. The congestion in Figure 7(b) started in section 3. It can be seen that for most sections the absence of red edged dots overlaps with the congestion detected by the *toll* data. We conclude, that Stage 2 of our approach indeed can provide sufficient fine-granular congestion estimates compared to recent monitoring technology. However, further evaluation is necessary to see whether the availability of active terminals is sufficient for reliable detection. For example, in section 3, G $\rightarrow$ H, the number of cell handovers is too low for an effective estimation.

To demonstrate the lack of available active terminal data, we present Figure 7(c) showing a whole day without any congestion (note the different time ticks on the x-axis). The high number of possible false positives are highlighted in circles which occur when no pair of cell handovers can be detected for a given amount of time. By combining Stages 1 and 2, this problem can be eliminated and congestions can be identified fast and reliably (only few false positives and false negatives) because of Stage 1, and with high accuracy in terms of location because of Stage 2.

Identifying the type of incident

Combining Stages 1 and 2 can further be used to differentiate between different types of incidents as it allows to localize congestion fronts. Both congestions in Figure 7(a) are present in the broadcast data as 'accident'. When looking at the upstream fronts and downstream fronts of the congestions it can be seen that they show the typical behavior of a *wide moving jam*, i.e., both fronts can move upstream. Especially the traffic jam in the morning is an interesting example: We can see that the accident happened at around 9:15 am in section 4 and that the upstream front moved and reached section 1 (C $\rightarrow$ D) around 20 minutes after the accident. Around 60 minutes after the accident the downstream front moved from section 4 to section 3, i.e., the area of the accident in section 4 was cleared. However, in section 1 the congestion remained until 11:00 am, 45 minutes longer than in the area of accident. Figure 7(b) shows a 'heavy traffic' as sent over radio broadcast and shows a that the downstream front remains at the bottleneck (as typical for a *synchronized flow*). In this example only the upstream front is moving and the congestion has the form of an ice cone. Hence, by applying Stage 2 of our approach, the improved accuracy in localizing congestions turns mobile cellular data into a valuable input to further traffic congestion analysis.

CONCLUSION

In this paper we presented a novel approach for estimating vehicular travel times based on anonymous mobility signaling collected from an operational mobile cellular network. In order to provide an accurate estimation of travel times and a timely detection of congestions, we combined two different strategies. Spatially coarse mobility data from all users (most of which in idle state) is exploited to capture speed deviations in long road sections and detect congestion events. Subsequently, the finer-grained mobility data produced by active users is used to refine the location accuracy and classify the type of event.

Experiments on a major traffic route in Austria showed that the presented system is able to detect road incidents reliably and in a timely manner. Compared to other road monitoring sensory technologies, our approach presents a higher detection success and at least a similar detection delay. Hereby, we conclude that cellular network data enable an accurate

and wide-spread monitoring of traffic status. At the same time, our approach does not require investments in new infrastructure but leverages the mobile cellular network as a large-scale mobility sensor. We further demonstrated that by including the localization step based on active terminal mobility data, such a system also allows to analyze the development of congestion fronts which can help in differentiating between different types of road congestions.

We plan to extend our work by including signaling data from the Radio Access Network, e.g., power measurement reports from the active terminals, which would allow to further refine the location accuracy and derive advanced traffic characteristics also for minor roads or urban downtown areas or highways where the geometric characteristics of the LAs may impair the estimation results.

Acknowledgments

We acknowledge the support of the FTW project RoadCell by the partners ASFiNAG, A1 Telekom Austria, Nast Consulting, and Kapsch CarrierCom. Furthermore, we acknowledge the support of the Austrian Government and the City of Vienna under the COMET program, and the Commission of the European Union under the FP 7 Marie Curie project MOVE-R, contract PIEF-GA-2010-276336.

REFERENCES

1. M. Alger, E. Wilson, T. Gould, R. Whittaker, and N. Radulovic. Real-time traffic monitoring using mobile phone data. Online: `http://www.maths-in-industry.org/miis/30` Vodafone Pilotentwicklung GmbH, 2004.
2. H. Bar-Gera. Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: A case study from Israel. *Transportation Research Part C: Emerging Technologies*, 15(6):380–391, 2007.
3. R. Bertini. You are the traffic jam: an examination of congestion measures. Technical report, Department of Civil and Environmental Engineering, Portland State University (2005).
4. F. Calabrese, M. Colonna, P. Lovisolo, D. Parata, and C. Ratti. Real-time urban monitoring using cell phones: A case study in Rome. *IEEE Trans. on Intelligent Transportation Systems*, 12(1):141 –151, 2011.
5. C. de Fabritiis, R. Ragona, and G. Valenti. Traffic estimation and prediction based on real time floating car data. In *International IEEE Conference on Intelligent Transportation Systems (ITSC'08)*, pages 197 –203, oct. 2008.
6. P. Fiadino, D. Valerio, F. Ricciato, and K. A. Hummel. Steps towards the extraction of vehicular mobility patterns from 3G signaling data. In *Traffic Monitoring and Analysis (TMA) Workshop*, 2012.
7. M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.
8. H. Holma and A. Toskala. *WCDMA for UMTS : Radio access for third generation mobile communications*. John Wiley and Sons, Ltd., 2000.
9. I. Trestian *et al.* Measuring serendipity: Connecting people, locations and interests in a mobile 3G network. In *Internet Measurement Conference (IMC'09)*, 2009.
10. B. Kerner. Empirical macroscopic features of spatial-temporal traffic patterns at highway bottlenecks. *Physical Review*, E 65(4), 2002.
11. B. Kerner, C. Demir, R. Herrtwich, S. Klenov, H. Rehborn, M. Aleksic, and A. Haug. Traffic state detection with floating car data in road networks. In *International IEEE Conference on Intelligent Transportation Systems (ITSC'05)*, pages 44–49, 2005.
12. C. Oberauer, T. Stottan, and R. Wagner. Requirements of processing extended floating car data in a large scale environment. *Advanced Microsystems for Automotive Applications 2011*, pages 335–342, 2011.
13. A. Poolsawat, W. Pattara-Atikom, and B. Ngamwongwattana. Acquiring road traffic information through mobile phones. In *ITS Telecommunications 2008*, pages 170–174, 2008.
14. G. Rose. Mobile phones as traffic probes: Practices, prospects and issues. *Transport Reviews*, 26(3):275–291, 2006.
15. C. Song, Z. Qu, N. Blumm, and A. L. Barabási. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021, 2010.
16. M. Treiber, A. Kesting, and D. Helbing. Three-phase traffic theory and two-phase models with a fundamental diagram in the light of empirical stylized facts. *Transportation Report Part B*, (44):983–1000, 2010.
17. D. Valerio, A. D'Alconzo, F. Ricciato, and W. Wiedermann. Exploiting cellular networks for road trafc estimation: a survey and a research roadmap. In *IEEE 69th Vehicular Technology Conference (IEEE VTC '09-Spring)*, 2009.
18. F.-Y. Wang. Parallel control and management for intelligent transportation systems: Concepts, architectures, and applications. *IEEE Trans. on Intelligent Transportation Systems*, 11(3):630–638, 2010.
19. T. Wang, T. Fang, J. Han, and J. Wu. Traffic monitoring using floating car data in Hefei. In *Intelligence Information Processing and Trusted Computing (IPTC)*, pages 122 –124, 2010.
20. J. Yuan, Y. Zheng, X. Xie, and G. Sun. Driving with knowledge from the physical world. In *17th ACM SIGKDD int. conference on Knowledge Discovery and Data mining*, pages 316–324. ACM, 2011.
21. J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen. Data-driven intelligent transportation systems: A survey. *IEEE Trans. on Intelligent Transportation Systems*, 12(4):1624 –1639, 2011.