Estimating Traffic Performance in Road Networks from Anonymized GPS Vehicle Probes

Carlos Morán (carlos.moran@abe.kth.se) Division of Traffic and Logistics School of Architecture and the Built Environment. KTH, Stockholm, Sweden
Győző Gidófalvi (gyozo.gidofalvi@abe.kth.se) Division of Geoinformatics, School of Architecture and the Built Environment. KTH, Stockholm, Sweden

The rapid growth of demand for transportation, high levels of car dependency caused by the urban sprawl have exceeded the slow increments in transportation infrastructure supply in many areas causing severe traffic congestion. In dense urban areas, adding capacity through construction of new facilities is difficult due to lack of space and prohibitive costs. Policy makers and transport planners need to monitor the development of congestion and they consider for this purpose the use of congestion indicators. For estimating these indicators, transportation authorities have developed and devised different kinds of sensor technologies. Earlier efforts have used fixed location sensors. In this way, gantry portals have been built in highways for collecting information such as flow and punctual speed at certain locations. However, using this method for data collection would require be huge building cost as well as have a severe impact of visual intrusion when applied for dense urban areas. In comparison, the use of GPS for traffic data collection in these dense areas does not require large investments and does not create any other external impact.

Current methods estimating congestion indicators either consider *travel-time based* indicators that relate to the time-delay or consider *bottleneck based* indicators that relate to the time stopped or number of stops caused by congestion [1]. Currently, traffic monitoring centres consider only *travel-time based* indicators. The estimation uses fixed and mobile data collection methods for estimating the traffic parameters (such as flow or travel time) required for the aggregation and calculation the congestion indicator. Studies using empirical data [2] have analysed data provided by two methods for estimating of travel-time: Automatic Travel Time System –ATTS (that identifies number plates at intersections and later matches them) and, Floating Car Surveys – FCS (where specialized vehicles cover a route back a forth during the survey period). It was observed that ATTS provide in large part inaccurate or unfeasible data and FCS have a low sample rate and high cost. All these shortcomings have an incremental effect on the estimations of the congestion indicators and degrade their accuracy and reliability [2]. Such inaccurate and unreliable congestion indicators can mislead transport authorities in their decisions and measures for coping with congestion problems.

In comparison, positioning technologies (primarily GPS) embedded in on-board navigation systems of both private and commercial vehicles and mobile devices that are carried by the drivers are becoming increasingly available and accurate. These positioning technologies have the potential allow real time access to the positions of a large fraction of the vehicles in traffic. Periodically obtained positions of vehicles, often termed as vehicle probes, have a great potential to be efficiently used to provide accurate support for real time management of traffic systems. In the past, a number of methods have been proposed to extract different types of mobility/traffic patterns from vehicle probes, including finding dense areas [3], finding frequent routes[4], and fining clusters of similarly moving objects [5]. Such patterns individually or combined with other data sources can provide information for estimating the congestion indicators on road networks. However, obtaining *accurate* mobility/traffic patterns requires accurate, near-continuous, near-real-time tracking of the vehicles. It has been recognized by many that the location traces of these vehicles are highly sensitive and can pose sever threats to the privacy of the drivers of the vehicles if the data is not anonymized adequately [6]. Specifically, the location trace of an individual contains private locations, e.g., home, work, which can easily be identified based on the frequency and time of visits to these locations. Subsequently, a subset of these private locations can be linked to publicly available external data sources, e.g., Yellow pages, etc, to reveal the identity of the individual.

A number of privacy protection frameworks have been proposed to protect against the above described threat. A common approach is to generalize the exact locations of individuals to *cloaking region*. Following the traditional notion of *k-anonymity*, most privacy preserving frameworks construct cloaking regions such that at the time of the location report there are at least *k* objects in the given region. As identified in [6], there are a number of problems with this method. First, determining the *k-anonymity* based cloaking region requires trusted components, e.g., a server, often termed as the anonymizer, that is aware of the exact positions of the objects. Second, as the cloaking regions depend on the positions of nearby objects, for a given location the cloaking region varies over time depending on the density of the objects, allowing an attacker infer the locations of the private location to be within the intersection of the reported cloaking regions for the location. Finally, *k-anonymity* based cloaking regions are likely to be over-protective in low density rural areas, and under-protective in high density, sensitive hot spots, e.g., red light district. To overcome the above shortcomings, prior work [6] proposed an alternative privacy preserving framework in which users specify their requirements of location privacy, based on the notions of *anonymization rectangles* and *location probabilities*, intuitively saying how precisely they want to be located in given areas. Experimental results on realistically simulated data sets have demonstrated that using the proposed privacy preserving framework probabilistic grid-based data mining methods can efficiently extract (with high precision and recall) accurate mobility/traffic patterns from anonymized location traces.

Based on the positions expressed in the present paper and the prior research of the authors the following collaborative research agenda is proposed:

1. Estimate congestion indicators based on mobility/traffic patterns that are extracted by data mining methods.
2. Derive accurate location probabilities used in the privacy preserving framework in [6] from a number of publically accessible geospatial data sources such as road networks, traffic flows, toll stations, surveillance cameras, ATTS, etc.
3. Adapt the grid-based privacy preserving framework in [6] to link-based privacy preserving framework, which is suitable for transportation research.
4. Using the results from 1.–3., derive traffic performance estimates in road networks from anonymized GPS vehicle probes.

References

[1] Morán, C. and K.-L. Bang. *Area wide analysis of urban road traffic congestion: Analysis of travel time based measures*. In Proc. of the 5th International Symposium on Highway Capacity and Quality of Service. 2006: Japan Society of Traffic Engineers (JSTE).

[2] Morán, C. and K.-L. Bang. *Reliability of Congestion Performance Measures.* In Proc. of the Institution of Civil Engineers - Transport, 2010 (Accepted December 2009).

[3] Jensen, C.S., et al. *Effective Density Queries on Continuously Moving Objects*. In Proc. of the 22nd International Conference on Data Engineering, ICDE. 2006: IEEE Computer Society.

[4] Gidófalvi, G. and T.B. Pedersen, *Mining Long, Sharable Patterns in Trajectories of Moving Objects.* Geoinformatica 2009. 13(1): p. 27-55.

[5] Pelekis, N., et al. *Clustering Trajectories of Moving Objects in an Uncertain World*. In Proc. of the Ninth IEEE International Conference on Data Mining, ICDM. 2009: IEEE Computer Society.

[6] Gidófalvi, G., X. Huang, and T.B. Pedersen. *Privacy–Preserving Data Mining on Moving Object Trajectories*. In Proc. of the Eighth International Conference on Mobile Data Management, MDM. 2007: IEEE Computer Society.