Transportation Research Part C xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Accurate and cost-effective traffic information acquisition using adaptive sampling: Centralized and V2V schemes

Shiau Hong Lim*, Yeow Khiang Chia, Laura Wynter

IBM Research, 10 Marina Boulevard, 018983, Singapore

ARTICLE INFO

Keywords:
Probe vehicle
Traffic data acquisition
Real-time traffic estimation
Adaptive sampling
Vehicle-to-vehicle communication

ABSTRACT

The new generation of GPS-based tolling systems allow for a much higher degree of road sensing than has been available up to now. We propose an adaptive sampling scheme to collect accurate real-time traffic information from large-scale implementations of on-board GPS-based devices over a road network. The goal of the system is to minimize the transmission costs over all vehicles while satisfying requirements in the accuracy and timeliness of the traffic information obtained. The system is designed to make use of cellular communication as well as leveraging additional technologies such as roadside units equipped with WiFi and vehicle-to-vehicle (V2V) dedicated short-range communications (DSRC). As opposed to fixed sampling schemes, which transmit at regular intervals, the sampling policy we propose is adaptive to the road network and the importance of the links that the vehicle traverses. Since cellular communications are costly, in the basic centralized scheme, the vehicle is not aware of the road conditions on the network. We extend the scheme to handle non-cellular communications via roadside units and vehicle-to-vehicle (V2V) communication. Under a general traffic model, we prove that our scheme always outperforms the baseline scheme in terms of transmission cost while satisfying accuracy and real-time requirements. Our analytical results are further supported via simulations based on actual road networks for both the centralized and V2V settings.

1. Introduction

One of the most effective ways to acquire real-time road traffic information is by using vehicles as probes via mobile phones or dedicated onboard devices (De Fabritiis et al., 2008; Li et al., 2009; Shi and Liu, 2010; Ayala et al., 2010; Vandenberghe et al., 2012; Paulin and Bessler, 2013; Herrera et al., 2010; Seo and Kusakabe, 2015). In particular, dedicated onboard devices are at the heart of the new generation of GPS-based tolling systems (Velaga and Pangbourne, 2014; Qin et al., 2017). While their primary function is to enable GPS-based toll collection, an important secondary function of such devices is to collect pervasive traffic level information on the road network. Indeed, since the vehicles are able to transmit their location and speed, real-time traffic estimations can be obtained, and subsequently used in the estimation of the macroscopic fundamental diagram (Nagle and Gayah, 2014; Du et al., 2016) as well as in applications such as real-time routing and incident detection (Asakura et al., 2017). Furthermore, since GPS-based tolling schemes require placing devices in many, and, in some parts of the world, all of the vehicles traveling on the road network, they allow a far greater degree of road traffic sensing that has been available before.

Most onboard units (OBU) rely on cellular transmission for the bulk of their communications to the central server. In general, cellular transmission are costly, and hence reducing the number of such transmissions is an important criteria in the design of a traffic

* Corresponding author.
E-mail addresses: shonglim@sg.ibm.com (S.H. Lim), ykchia@sg.ibm.com (Y.K. Chia), lwynter@sg.ibm.com (L. Wynter).

https://doi.org/10.1016/j.trc.2017.10.017
Received 30 August 2017; Received in revised form 14 October 2017; Accepted 15 October 2017
0968-090X/

data collection scheme using onboard devices. The other requirement of any traffic data collection scheme is clearly to provide enough traffic data samples on every road link and every time period so that the real-time traffic level estimates are reliably accurate. Thus, the dilemma that the scheme must address is to ensure collection of enough traffic data samples, in both time and space, whilst minimizing the number of samples collected and thus minimizing the number of costly transmissions.

Accuracy of traffic level estimations can be measured as a function of the number of samples obtained and depends on the number of actual vehicles on the road link during each time period. The accuracy of the estimate also depends on the time difference between when the data was collected from the vehicle and when it is used in the traffic level estimate. The accuracy of the estimate can thus be evaluated theoretically and in simulation, under certain assumptions on the traffic generation process.

Cost reduction can be achieved in general via two techniques. The first is random sampling, where only a fraction of the vehicles transmit their traffic data at a given point in time. Clearly, this leads to reduced accuracy when the sample size is low, and higher sampling frequencies may be unnecessarily costly. The second technique is onboard data aggregation: due to typical communications charging schedules which use a minimum charging unit (e.g. 1 kB of data), cost savings can be achieved by aggregating traffic data on the device and aligning the size of each transmission with the minimum charging unit. The latter approach means that some on-vehicle traffic observations (e.g. a vehicle's speed at a particular point in space and in time) will be transmitted late. The tradeoff is then in the timeliness of the data, and hence its relevance to the real-time traffic level, versus the number of transmissions made by each vehicle's onboard device.

We address this problem of designing a system to collect accurate real-time traffic information from large-scale implementations of onboard GPS-based devices over a road network. The goal of the system is to minimize the transmission costs over all vehicles while satisfying requirements in the accuracy and timeliness of the traffic information obtained. The system we design should be effective and accurate using purely cellular communication but should be able to leverage additional communication technologies. In particular, the scheme should be able to take advantage of roadside units equipped with WiFi as well as vehicle-to-vehicle (V2V) dedicated short-range communications (DSRC) where available.

We thus propose an adaptive sampling traffic data collection scheme that incorporates both random sampling and data aggregation with an underlying traffic model so as to provide accuracy guarantees. As opposed to fixed sampling schemes, which transmit at regular intervals, the sampling policy we propose is adaptive to the road network and the importance of the links that the vehicle traverses. We consider first the setting in which vehicles can communicate only with the central server directly. Since cellular communications are costly, in the basic centralized scheme, the vehicle is not aware of the road conditions on the network. Then, we consider the case where vehicles are equipped with DSRC and are able to communicate in a V2V manner, thus transmitting some of their traffic data amongst themselves at no cost as well as to the central server via WiFi-equipped roadside units. We further extend the scheme to handle non-cellular communications via roadside units.

Our scheme makes use of a priority queue concept – i.e., assigning different priority classes to different links in the network. The motivation is that links with different priorities typically have different accuracy and real-time latency requirements depending on the characteristics of the links. By arbitrating across the priority classes we are able to achieve a greater degree of transmission cost reduction while satisfying the accuracy requirements of the most critical road links.

We analyze and compare our proposed scheme with a class of baseline schemes commonly used in mobile traffic data collection, in which transmissions take place on a fixed schedule (Vandenberghe et al., 2012; De Fabritiis et al., 2008; Li et al., 2009; Shi and Liu, 2010; Drira et al., 2016; Fusco et al., 2016; Asakura et al., 2017). Vehicle-to-vehicle communication technologies as part of a vehicular ad hoc network (VANET) have also been discussed for traffic data collection and dissemination in Zhang et al. (2016), Turcanu et al. (2016), Baiocchi et al. (2015), He and Zhang (2016), and Wongdeethai and Siripongwutikorn (2016) but without the adaptive sampling method that we propose here. We assume that non-cellular communications are cost-free relative to cellular communications and do not address issues such as bandwidth limitation in V2V or V2I communications. Our proposed scheme, however, can be used in conjunction with other techniques that address such issues. For instance, recent work by Dai et al. (2016) considers bandwidth allocation schemes between vehicles and roadside units that take into consideration both data freshness and timeliness.

For our analysis and model derivation, we focus on the class of traffic models with Poisson arrival rates. In particular, we assume that the time interval between visits to two links is exponentially distributed. Under a general traffic model, we prove that our scheme always outperforms the baseline scheme in terms of transmission cost while satisfying accuracy and real-time requirements. Our analytical results are supported via simulations based on actual road networks for both the centralized and V2V settings. In particular, while the time interval between simulated link visits does not follow an exponential distribution, we observe experimentally that the Poisson model used for the model derivations is adequate in the sense that the theoretically-optimized control parameters work well.

The remainder of the paper is as follows. We formulate the problem in the next section as well as the traffic generation model used in the evaluation of our scheme. Section 3 defines our proposed adaptive sampling scheme itself and its key properties, along with the baseline scheme to which we shall compare it. Section 4 presents the V2V extension of our proposed scheme. We provide numerical simulation results in Section 5, and Section 6 presents some areas worthy of further study that have arisen from this work.

2. Problem formulation

We assume that the traffic data produced by each vehicle at each time point is collected as a packet. Each packet is associated with a particular road link before transmission. Specifically, the packet contains the vehicle's GPS coordinates and instantaneous speed along with header information required for data transmission. In addition, we assume that a table is maintained in the onboard unit (OBU) such that when the vehicle passes through each road link, according to a given network description, the ID of that road link