Ad Hoc Networks 8 (2010) 518–530

Contents lists available at ScienceDirect

Ad Hoc Networks

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

A probabilistic method for cooperative hierarchical aggregation of data in VANETs

Christian Lochert, Björn Scheuermann *, Martin Mauve

Heinrich Heine University Düsseldorf, Computer Science Department, Universitätsstr. 1, 40225 Düsseldorf, Germany

A R T I C L E I N F O

Article history:
Available online 13 January 2010

Keywords:
VANET
Aggregation
Sketch
Information dissemination
Arithmetic coding
Soft-state sketch

A B S T R A C T

We propose an algorithm for the hierarchical aggregation of observations in dissemination-based, distributed traffic information systems. Instead of transmitting observed parameters directly, we propose soft-state sketches—an extension of Flajolet–Martin sketches—as a probabilistic approximation. This data representation is duplicate insensitive, a trait that overcomes two central problems of existing aggregation schemes for VANET applications. First, when multiple aggregates of observations for the same area are available, it is possible to combine them into an aggregate containing all information from the original aggregates. This is fundamentally different from existing approaches where typically one of the aggregates is selected for further use while the rest is discarded. Second, any observation or aggregate can be included into higher-level aggregates, regardless if it has already been previously—directly or indirectly—added. Those characteristics result in a very flexible aggregate construction and a high quality of the aggregates. We demonstrate these traits of our approach by a simulation study.

1. Introduction

Cooperative information gathering and sharing forms a prominent class of VANET applications. For instance, approaches that disseminate traffic information like TrafficView [1] or SOTIS [2,3] as well as a system to exchange information on free parking places [4] have been designed. These applications have in common that they distribute measurement results obtained by the participating cars in a comparatively large area.

Typically, this is accomplished in the way schematically shown in Fig. 1. Each car makes observations. An observation is essentially some measured value (traffic density, free parking places, road condition, etc.), related to a position in space (typically a road segment or a small area) and a point in time when the observation has been made. All or part of the locally stored information is periodically single-hop broadcasted in beacon packets. Upon reception of such a beacon, a node incorporates the received data into the local knowledge base. By comparing the timestamps of observations, it can ensure that always the most up-to-date value for each position is stored and redistributed. However, if we assume that the spatial density of points for which observations are made is approximately constant, the amount of data increases quadratically with the covered radius. Thus, the amount of data to be broadcasted by each car will likewise increase quickly. This is fatal for the scalability of such a system [5].

To overcome this problem, the use of hierarchical data aggregation has been proposed: with increasing distance, observations concerning larger and larger areas (or road segment lengths) are combined into one single value. Such an aggregated value could, e.g., be the average speed on a longer road segment [1–3], or the percentage of free parking places in a part of a city [4]. Coarse aggregates are made available at greater distances, more detailed data is kept only in the near vicinity. However, even though the idea

* Corresponding author. Tel.: +49 211 81 11692; fax: +49 211 81 11638.
E-mail address: scheuermann@cs.uni-duesseldorf.de (B. Scheuermann).

1570-8705/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.adhoc.2009.12.008

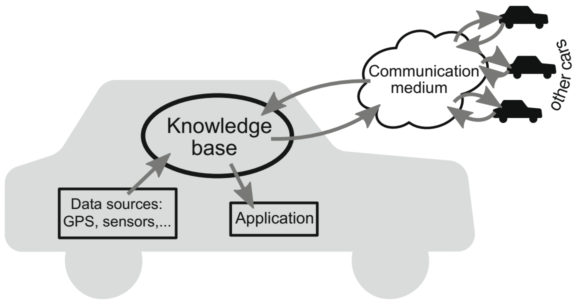

Fig. 1. System model of a typical VANET data dissemination application.

has often been mentioned [1–4,6–8], central problems of such an approach have, so far, remained unsolved.

A fundamental issue that arises is that aggregates cannot, like single observations, be directly compared with respect to the up-to-dateness and completeness of the contained data. They are created by cars that will typically not have the most up-to-date measurements for all underlying points available. Therefore, multiple aggregates for the same area may exist, based on different, but likely overlapping knowledge. To decide which one is based on "better" underlying data is hard, if not impossible. While this problem has been mentioned before in the literature (see, e.g., [6,1]), no fundamental solution has so far been proposed.

In this paper—a revised and extended version of [9]—we propose an algorithm that solves this issue. To the best of our knowledge, this is the first approach that can handle aggregates of overlapping entries. We achieve this by a special data representation: both single observations and aggregates in our scheme do not carry the value of, e.g., the number of free parking places directly, but instead contain an approximation of it in form of a so-called soft-state sketch. The herein introduced soft-state sketches are a data structure based on modified Flajolet–Martin sketches [10].

While soft-state sketches do still not provide a way to compare the quality of two aggregates directly, they allow for something even better: in our scheme, multiple aggregates for the same area can be merged, yielding a new one that incorporates all the information contained in any one of the aggregates. This is fundamentally different from all existing dissemination approaches, where two aggregates describing the same area cannot be merged.[1] In our scheme there is no need to decide which aggregate contained more up-to-date information since the resulting aggregate comprises all the information from all aggregates that have been merged.

Our approach also allows observations or lower-level aggregates to be integrated into an already existing higher-level aggregate at any time. This, too, is not possible with any previously existing hiearchical aggregation approach, because it cannot be determined which data is already present in the aggregate and interesting aggregates like sums or averages are typically duplicate sensitive.

Apart from making decisions regarding the aggregate quality unnecessary, the proposed scheme also largely eases the generation of good aggregates. A node would usually have to collect data on a significant fraction of the covered area before an aggregate that likely constitutes a good representation can be formed. With our scheme, the aggregate can instead be maintained while being passed around in the network, always incorporating new information on-the-fly.

The remainder of this paper is structured as follows. In the next section we review aggregation as it has been proposed for VANET applications, as well as some previous uses of Flajolet–Martin sketches in the networking area. Thereafter, we quickly recapitulate Flajolet–Martin sketches in Section 3, to set the stage for the introduction of soft-state sketches and the detailed discussion of our algorithm in Section 4. We subsequently propose two extensions in Section 5. In Section 6, we present and discuss the results of a simulation-based evaluation of the algorithm in a VANET city scenario. Finally, we conclude this paper with a summary in Section 7.

2. Related work

Recently, many convenience applications for VANETs have been discussed, and often they use some form of data dissemination. In the Self-Organizing Traffic Information System (SOTIS) [2,3], information on the traffic situation is distributed opportunistically, by sending periodic beacons containing the knowledge of the sending node on the traffic situation in a larger surrounding. The authors also outline a (non-hierarchical) aggregation scheme, combining all the known information on each fixed-length road segment to one average value. Upon reception, a node considers an aggregate to be "better" if it has a newer timestamp. But since these timestamps are assigned when the aggregate is computed, this system exhibits the problems outlined in the introduction: a newly computed aggregate with a new timestamp is not necessarily based on the most up-to-date information, and aggregates representing largely disjoint knowledge can neither be identified as such nor can they be merged. For these reasons, a system like SOTIS could largely benefit from the aggregation scheme introduced here.

TrafficView [1] is another system for disseminating traffic information, similar to SOTIS in both aims and mechanisms. The authors of TrafficView also introduce a data aggregation scheme. Different from SOTIS, TrafficView distributes information on position and speed of single vehicles. The aggregation mechanism combines a number of "similar" vehicles in an adaptive way, aiming to minimize the introduced errors. Again, each aggregate is assigned a timestamp: TrafficView uses the minimum information generation time of the combined measurements. Consequently, similar problems as for SOTIS arise. A definite decision about the relative up-to-dateness of stored and received values can not reliably be made. Again, this could be overcome by using our probabilistic aggregation scheme.

[1] Any hierarchical aggregation scheme will provide a way to merge lower-level aggregates into a higher-level aggregate. This is not what we refer to. The point here is that two aggregates describing the same area are merged.