Scalable Data Dissemination for Inter-Vehicle-Communication: Aggregation versus Peer-to-Peer

Skalierbare Informationsverbreitung für die Fahrzeug-Fahrzeug-Kommunikation: Aggregation versus Peer-to-Peer

Christian Lochert, Jedrzej Rybicki, Björn Scheuermann, Martin Mauve, Heinrich Heine Universität Düsseldorf

Summary This paper investigates the scalable dissemination of data between vehicles. The application context of this work is traffic information systems where cars are not only consumers but also producers of information. The key challenge in those systems is to ensure scalability in an environment where data is provided and requested by all participating vehicles in a large area. We discuss two fundamentally different approaches to this problem: using direct communication between cars and compressing the data via aggregation versus relying on infrastructure. The latter approach can further be divided into client-server and peer-to-peer systems. We outline all three approaches and highlight their advantages and disadvantages.

►►► **Zusammenfassung** Im Rahmen dieses Beitrages wird die Verbreitung von Informationen zwischen Fahrzeugen untersucht. Im Vordergrund stehen dabei Verkehrsinformationssysteme, bei denen Fahrzeuge sowohl Produzenten als auch Konsumenten von Informationen sind. Die zentrale Herausforderung solcher Systeme ist ihre Skalierbarkeit. Wir stellen hier zwei fundamental verschiedene Herangehensweisen vor: Einerseits die direkte Kommunikation zwischen Fahrzeugen, andererseits den Einsatz von Infrastruktur für den Nachrichtenaustausch zwischen Fahrzeugen. Letzeres kann weiter in Client-Server- und Peer-to-Peer-Ansätze unterteilt werden. In diesem Artikel geben wir Beispiele für alle drei Ansätze und untersuchen diese hinsichtlich ihrer Vor- und Nachteile.

KEYWORDS C.2.1 [Computer Systems Organization: Computer-Communication Networks: Network Architecture and Design]; C.2.4 [Computer Systems Organization: Computer-Communication Networks: Distributed Systems]; H.4.3 [Information Systems: Information Systems Applications: Communications Applications]; H.3.5 [Information Systems: Information Storage and Retrieval: Online Information Services]

1 Introduction

Recently a whole community consisting of public authorities, academia, and car manufacturers [1; 4; 12] has formed in pursuit of improving driving security and comfort by enabling inter-vehicle communication. Proposed applications span from the reduction of road casualties by means of brake warning, intersection assistance, or collision avoidance systems [14] to offering guidance to available parking lots [2], discovering the traffic situation on a planned route [13], and coordinating car flow and traffic lights [3; 12].

In this paper, we focus on non-safety applications that can be subsumed under the term distributed *traffic information systems* (TIS). Such applications aim to improve the driving comfort by enhancing a driver's awareness of the traffic conditions. The participating cars are not only consumers of information but at least some of them also produce information by sharing their observations. An observation is, e. g., a local measurement of the current traffic conditions or the number of currently free parking slots, which is then distributed to other vehicles. This results in a characteristical communication pattern. TIS usually rely on communication among many participants over relatively large distances that can span some ten kilometers in the case of a city scenario up to some hundred kilometers on highways. Thus, the communication requirements of TIS applications are quite challenging: continuously updated data measured by a large number of network nodes is to be made available to many vehicles in a relatively large area.

This paper provides an overview of different paradigms that may be used to build a traffic information system. In Section 2 we look at a solution that is based purely on infrastructure-less VANET technology. To cope with the existing capacity limits we discuss an aggregation technique that is tailored to the needs of a TIS. Section 3 deals with a completely different, infrastructure-based concept. First we sketch a solution based on a client/server architecture showing the advantages of using an existing network technology. Subsequently, we contrast it with a peer-to-peer approach that exhibits quite appealing properties regarding scalability and robustness. We summarize the advantages and disadvantages of each approach in Section 4 and give hints for developers of TIS applications which paradigm might fit their specific needs.

2 Vehicular Ad-Hoc Networks

2.1 Motivation

Considering the characteristics of a TIS application it turns out that many vehicles are interested in similar information. Hence, it is one approach to let the vehicles disseminate it themselves – by means of a VANET. We define a VANET to consist of all cars in a specific region which are equipped with some WiFi-like communication technology [7] and can communicate directly with each other in an ad-hoc fashion.

Typically, data is distributed in the following way. Each car makes observations like the traffic density, the number of free parking places, etc., related to a position in space (i. e., 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. Upon reception of such a broadcast, a node incorporates the received data into its 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. The amount of data to be broadcasted by each car will thus likewise increase quickly. Since the network's bandwidth is constant, this is fatal for the scalability of such a system.

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, or the percentage of free parking places in a part of a city. Coarse aggregates are made available at greater distances, more detailed data is kept only in the closer vicinity.

A fundamental issue that arises in such a system is that aggregates cannot, like single observations, be directly compared regarding 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.

2.2 Sketch Based Aggregation

The quality of the TIS information provided to the driver depends fundamentally on the quality of the aggregation technique. Based on the characteristics of the stored values we presented in [9] a soft-state sketch approach which is based on Flajolet-Martin sketches [5].

A soft-state sketch is a data structure for probabilistic counting of distinct elements. Its soft state component allows to account for values that change over time. It represents an approximation of the current element count by a vector $S = c_1, \ldots, c_w$ of time to live (TTL) counters c_i of n bits length. All counters in the vector are initialized to zero. To add an element x to the sketch, it is hashed by a hash function h with geometrically distributed positive integer output, where $P(h(x) = i) = 2^{-i}$. The counter $c_{h(x)}$ is then set to the maximum TTL, i. e., to $2^n - 1$. The TTL entries are periodically decremented if they are not already zero. Thereby, the information ages over time and old observations will be removed. From the number of consecutive non-zero entries on the left hand side of the sketch an estimate of the total number of distinct values inserted into the sketch can be derived.

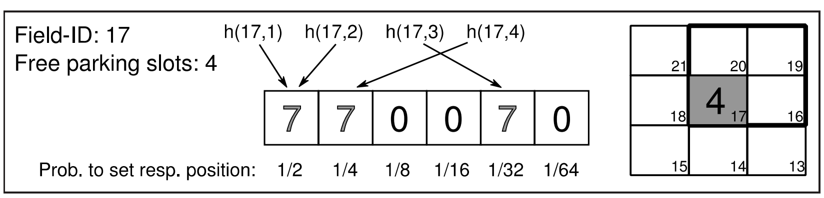

Figure 1 The four free parking slots of field 17 are hashed into a sketch.

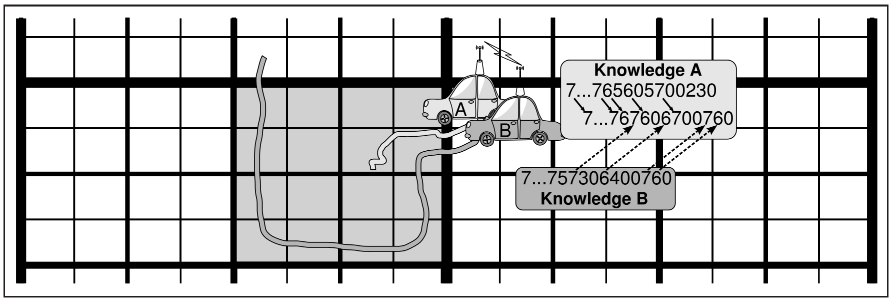

Figure 2 Merging car B's knowledge into the knowledge base of car A.

Figure 1 depicts an example application. Counters with a length of three bits are used. Here, four free parking slots have been observed by some car in an area with ID 17. They are hashed into the respective sketch for this area. It may well happen that the same position is "hit" multiple times. This is a desired feature that allows to estimate the number of distinct elements in a highly compressed form.

Sketches are particularly well suited for data aggregation in VANETs due to their ability to be easily merged and their duplicate insensitivity: to obtain a soft-state sketch for the union of two groups of elements, their individual sketches can be merged by a position-wise maximum operation. This allows to form aggregates of larger areas from smaller components. By their construction, repeatedly combining the same sketches or adding already present elements again does not change the results, no matter how often or in which order these operations occur. So, information that arrived multiple times over different paths is never duplicated.

Figure 2 depicts a scenario where two cars meet and exchange their aggregated knowledge about a certain geographical region highlighted by the light grey rectangle. Both cars have made observations about different sub-areas. During their trip some counters in the sketch have been decremented several times while others have been refreshed due to newly inserted observations. The merged aggregate for the gray area contains information from both contributors.

2.3 Discussion

The VANET approach incurs no additional cost for a driver in order to receive and use up-to-date information for her navigation or parking guidance system. However, a difficult task when building such a system is to equip enough cars with the technology. We showed in [8] that the speed and reliability of information dissemination in a VANET city environment with delay tolerant network-like opportunistic data exchange are limited. One conclusion is that sufficient performance is hard to reach, if not impossible, with penetration ratios that are realistic within the near future or during a roll-out phase. This might be mitigated by leveraging the VANET approach and using additional infrastructure support.

3 Infrastructure Based Communication

3.1 Motivation

Realizing that due to a low equipment density communication within a VANET might become a major hurdle when developing a useful traffic information system we propose to also consider other communication networks. For instance, 3G cellular networks already offer affordable and widespread mobile Internet access. It is reasonable to assume that cheap mobile Internet access will soon be common, long before VANET technology becomes a reality.

3.2 Client/Server

The usage of existing infrastructure based communication networks reduces the problem of implementing a working traffic information system to the question on how to make collectively gathered data available to all interested parties. One way of achieving this is to use a client/server architecture. In this approach one server (or a server farm) on the Internet stores a central knowledge base consisting of all the collaboratively gathered data. Cars make observations, e. g., on the current traffic situation, while driving. These observations are sent directly to the server.

A car may then request this information. For the support of navigation the best out of many alternative routes needs to be found, which requires information on substantial parts of the road network. To fulfill this task two approaches are conceivable:

(1) The car downloads (and regularly updates) traffic data for all possible alternative routes. It can then compute the fastest route to the destination based on current traffic information.

(2) As the downloads require a high communication effort, the route computation might alternatively be performed by the server. This, however, requires substantial computational resources in a central location.

3.3 Discussion

When using a client/server architecture combined with an infrastructure based communication approach the TIS application can neglect the network layer problems that typically arise in VANETs. The greatest advantage of infrastructure based communication is the fact that the density of the equipped cars needed for a working application is much smaller than in the case of a VANET.

The major technical challenge of a centralized system is to deal with the huge amount of simultaneous updates and queries [6] – recall that each car is a source of queries and sends own measurements regularly. Furthermore, due to the high degree of the centralization the server can become a bottleneck or even a single point of failure. Most importantly, a central authority is needed to build up and maintain the server with substantial financial investments.

3.4 Peer-to-Peer

To avoid the drawbacks of a client/server solution we have shown in [10] how infrastructure based communication can be used to build a distributed peer-to-peer network for traffic information systems. This approach combines the central advantages of VANETs – robustness resulting from the decentralization and independence from a central authority – with the good connectivity provided by infrastructure-based communication.

In this system each participating car acts as a peer of the distributed network. The data access interface of the central server is substituted by a distributed hash table (DHT). Similar to its non-distributed counterpart, the DHT associates keys with values. For TIS one may use geographical coordinates as keys and the respective observations as values. Each peer is thus responsible for a specific subset of keys (e. g., streets) and associated data (observations). The unambiguous mapping of the keys to peers is done by means of a random and uniform hash function. The main part of the DHT is the implementation of a *lookup* algorithm in a decentralized fashion. For a given key, the lookup returns the responsible peer. This is done in the following way. Participants form an overlay network. Each peer is connected to a number of "neighbors". Either a peer possesses a given key or it knows a peer offering a progress towards the responsible one in the overlay.

In Fig. 3 the considered scenario is divided into multiple areas with unique IDs used as keys. In this example Chord [11] is used as the lookup protocol. In Chord both peers and keys are hashed to an identifier. These Chord-IDs form a ring structure. Each peer has a link to its direct successor (the peer with the next higher Chord-ID). The keys are also hashed and stored on their successors. In the example area 16 is hashed to Chord-ID 14. So, the peer with Chord-ID 35 (successor of Chord-ID 14) is responsible for this key.

To find a peer responsible for a given key one has to send a lookup query along the ring. By passing it on to a neighbor it will reach its destination. In order to perform such queries more efficiently, each peer maintains a finger table with fingers pointing to peers that are at least 2^i IDs away with increasing i (depicted by the dotted lines). With this extension, the complexity for locating the responsible peer becomes logarithmic in the number of network participants.

Peers can join and leave the network freely. Upon a join the new participant takes the responsibility for some set of keys, reducing the average workload in the network. When leaving, the peer hands over the stored data to some other responsible peer. The random allocation of the data results in a fair load distribution and thus contributes to the robustness of the system.

In order to use the peer-to-peer system as a basis for a TIS application, participants send their observations to the responsible peers. Thereby, the information is made accessible to all other drivers. To gain information about current traffic conditions on a planned route, the inquiring application has to perform a number of lookups. For each segment of the route the responsible peer needs to be found and queried for the data. Due to the random allocation of the keys, a typical usage pattern involves many queries for neighboring segments that form a route.

However, we propose a possible solution that aims to deal with this problem. If the observations of all segments of a route are stored "close together" in the system, the number of independent lookups can be reduced. In order not to harm the Chord structure we preserve the according neighborship relations by means of a second finger table that consists of *semantic* fingers. Such fingers point at peers responsible for street segments adjacent to the ones stored at the participant. Therefore the requesting node will only have to find a peer responsible for the first segment of the planned path and then "follow" the route in the peer-to-peer structure without performing further lookups.

A traditional challenge in peer-to-peer networks is to ensure fair load balancing. This becomes even more vital when dealing with data exhibiting a high skew, i. e., when some data is more demanded than other, which is also the case for TIS applications. For example, the responsibility for a segment on a highway means much more data to store and more queries to serve than managing a small street in the suburbs. The simplest way to ad-

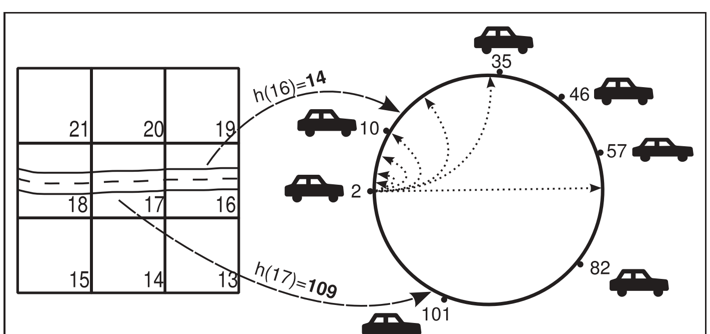

Figure 3 Distribution of street segments among peers.

dress this problem and to obtain a fair load balancing is to increase the number of peers responsible for each segment beyond one, thereby introducing some redundancy and resulting in a fairer workload distribution.

So far we presented the simplest way of implementing a TIS over a peer-to-peer network. This allows to offer the same functionality as the client/server architecture, while it does not require dedicated, centralized infrastructure. However, up to now, each car needs to obtain the whole information about planned and alternative routes. In order to keep track of changes of the traffic situation on the planned route and to react to possible changes, the application has to perform periodic queries to other peers. This proactive solution will cause a substantial amount of redundant network traffic.

A possible solution would be to use communication paradigms for group communication like, e. g., publish/subscribe. In this approach, the application, instead of periodically requesting new information, can register its interest for a given data on the responsible peer ("subscribe") once and will be informed about any important changes on the planned route. The peer that is responsible for this route notifies all interested parties of substantial changes in the traffic conditions.

3.5 Discussion

A peer-to-peer overlay is based upon existing infrastructure-based communication networks. So, similar to the client/server case, the approach is not as dependent on a high equipment density as a VANET application.

Probably the most compelling feature of the peer-to-peer approach is its self-scalability. As opposed to the client/server architecture the amount of resources available in the network is proportional to the number of participants. Hence, a peer-to-peer network scales gracefully with an increasing number of users. Since the data are distributed among the participants there is no single point of failure and the financial and organizational costs of dedicated servers can be avoided.

A drawback of peer-to-peer is the higher overall bandwidth usage compared to client/server. This consists of the costs to relay queries and updates of others and to maintain the overlay.

4 Summary

In this paper we presented possible paradigms for building a traffic information system that takes the input of vehicles into account. Because all three designs rely on different techniques it is hard to put them in any kind of order. In this section we summarize the advantages and disadvantages of these approaches.

VANET-based solutions profit from the efforts of the car manufacturers to equip cars with communication technology to improve the safety of the passengers. If there is a VANET communication unit built in the car anyway, a TIS application comes at no additional cost. Nevertheless, there exists a lower bound of the equipment density which is needed to allow a dissemination of messages into a large area within reasonable time. This poses a limiting factor on the introduction of this approach.

The infrastructure-based communication paradigms are less affected by a low equipment density. It is sufficient that enough vehicles are equipped to report the current traffic status. In a server-based architecture, however, the system has to deal with a huge amount of simultaneous updates and queries. Due to the centralization the server might become a bottleneck and a single point of failure.

In contrast to the client/server architecture the amount of resources available in the peer-to-peer network is proportional to the number of participants. Since the data are distributed over all users, there is no single, central bottleneck. The distribution, however, makes accessing the data more challenging.

We envision that the preferred solution might change over time. It seems almost certain that client/server-based solutions will be the first to become commercially available. This is due to two reasons: they can be built right away with existing technology and there is a clear business plan. In fact it is a rather minor change from existing navigation systems. Peer-to-peer systems could become a strong competitor once this idea is taken up by developers of navigation systems and mobile service providers. The big advantage would be the ability to avoid all costs and overheads associated with maintaining a central server(-farm). Finally, as the density of VANET equipped cars increases, users might turn to a solution that is completely free of charge.

References

[1] Car 2 Car Communication Consortium. http://www.car-2-car.org/.

[2] M. Caliskan, D. Graupner, and M. Mauve. Decentralized discovery of free parking places. In: *Proc. of the 3rd ACM Int'l Workshop on Vehicular Ad Hoc Networks (VANET '06)*, pp. 30–39, Sep 2006.

[3] F. Doetzer, F. Kohlmayer, T. Kosch, and M. Strassberger. Secure communication for intersection assistance. In: *Proc. of the 2nd Int'l Workshop on Intelligent Transportation (WIT '05)*, May 2005.

[4] W. Enkelmann. Fleetnet – applications for inter-vehicle communication. In: *Proc. of the IEEE Intelligent Vehicles Symp. (IV '03)*, pp. 162–167, June 2003.

[5] P. Flajolet and G. N. Martin. Probabilistic counting algorithms for data base applications. In: *J. of Computer and System Sciences*, 31(2):182–209, Oct 1985.

[6] B. Hull, V. Bychkovsky, Y. Zhang, K. Chen, M. Goraczko, A. Miu, E. Shih, H. Balakrishnan, and S. Madden. Cartel: A distributed mobile sensor computing system. In: *Proc. of the 4th Int'l Conf. on Embedded Networked Sensor Systems (SenSys '06)*, pp. 125–138, Oct 2006.

[7] IEEE P802.11, Task Group p. WAVE: Wireless access for the vehicular environment. http://grouper.ieee.org/groups/802/11/Reports/tgp_update.htm.

[8] C. Lochert, B. Scheuermann, M. Caliskan, and M. Mauve. The feasibility of information dissemination in vehicular ad-hoc networks. In: *Proc. of the 4th Annual Conf. on Wireless On-demand Network Systems and Services (WONS '07)*, pp. 92–99, Jan 2007.

[9] C. Lochert, B. Scheuermann, and M. Mauve. Probabilistic aggregation for data dissemination in VANETs. In: *Proc. of the 4th ACM Int'l Workshop on Vehicular Ad Hoc Networks (VANET '07)*, pp. 1–8, Sep 2007.

[10] 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. of the 13th Annual Int'l Conf. on Mobile Computing and Networking (MobiCom '07)*, pp. 215–221, Sep 2007.

[11] I. Stoica, R. Morris, D. Liben-Nowell, D. R. Karger, M. F. Kaashoek, F. Dabek, and H. Balakrishnan. Chord: a scalable peer-to-peer lookup protocol for Internet applications. In: *ACM/IEEE Trans. on Networking*, 11(1):17–32, 2003.

[12] U.S. Department of Transportation. ITS: Federal intelligent transportation systems. http://www.itsoverview.its.dot.gov.

[13] L. Wischhof, A. Ebner, H. Rohling, M. Lott, and R. Halfmann. SOTIS – a self-organizing traffic information system. In: *Proc. of the 57th IEEE Vehicular Technology Conf. (VTC '03-Spring)*, pp. 2442–2446, Apr 2003.

[14] Q. Xu, T. Mak, J. Ko, and R. Sengupta. Vehicle-to-vehicle safety messaging in DSRC. In: *Proc. of the 1st ACM Int'l Workshop on Vehicular Ad Hoc Networks (VANET '04)*, pp. 19–28, New York, USA, Oct 2004.


1


2


3


4

1 Dipl. Wirtsch.-Inf. Christian Lochert received his Diploma degree in Information Systems in 2003 from the University of Mannheim, Germany. Currently, he is a PhD student at the Heinrich Heine University Düsseldorf, Germany.
Address: Computer Networks Research Group, Heinrich Heine University Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf, Germany, Tel.: +49-211-8111634,
Fax: +49-211-8111638,
E-Mail: lochert@cs.uni-duesseldorf.de

2 M. Sc. Jedrzej Rybicki received a Masters degree in Physics and Computer Science from the Adam Mickiewicz University, Posen, Poland, in 2005. Currently, he is a PhD researcher in the field of Computer Networks at Heinrich Heine University Düsseldorf, Germany.
Address: Computer Networks Research Group, Heinrich Heine University Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf, Germany, Tel.: +49-211-8111635,
Fax: +49-211-8111638,
E-Mail: rybicki@cs.uni-duesseldorf.de

3 Dr. Björn Scheuermann received a B. Sc. in Mathematics and Computer Science and a Diploma degree (M. Sc. equivalent) in Computer Science from the University of Mannheim, Germany, both in 2004. In 2007 he received a PhD in Computer Science from the Heinrich Heine University Düsseldorf, Germany, where he now is a postdoctoral researcher.
Address: Computer Networks Research Group, Heinrich Heine University Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf, Germany, Tel.: +49-211-8111692,
Fax: +49-211-8111638,
E-Mail: scheuermann@cs.uni-duesseldorf.de

4 Prof. Dr. Martin Mauve received the M. Sc. and PhD degree in Computer Science from the University of Mannheim, Germany, in 1997 and 2000, respectively. From 2000 to 2003, he was an assistant professor at the University of Mannheim. In 2003, he joined the Heinrich Heine University Düsseldorf, Germany as a full professor and head of the research group for computer networks and communication systems.
Address: Computer Networks Research Group, Heinrich Heine University Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf, Germany, Tel.: +49-211-8111636,
Fax: +49-211-8111638,
E-Mail: mauve@cs.uni-duesseldorf.de