A Structure-free Aggregation Framework for Vehicular Ad Hoc Networks

Stefan Dietzel, Elmar Schoch, Boto Bako, Frank Kargl
Institute of Media Informatics, Ulm University
`{firstname}.{lastname}@uni-ulm.de`

Abstract— **One of the major difficulties for cooperative, decentralized information dissemination in vehicular networks is the heavily varying node density, which can lead to capacity issues of the wireless channel when many vehicles are driving or standing closely together. At the same time, a number of applications do not require exact information from all participating nodes, but higher-level aggregated information. For example, reports on road conditions or on flow of traffic can be aggregated before further dissemination, since remote drivers just need to know a coarse-grained picture of the situation. In this paper, we propose an information aggregation framework using the example of cooperative traffic congestion detection. The difference of our aggregation framework compared to other approaches is that it completely abandons any predefined structures such as grids and any group establishment. First evaluation results show that our approach works well for average speed dissemination on a highway.**

I. INTRODUCTION

Inter-vehicle communication (IVC) has a big potential to enhance driving safety and comfort. Applications for IVC require several specific and also novel ad hoc communication paradigms. For example, continuous cooperative awareness applications are commonly implemented with periodic beacon messages for status updates. Other applications require multi-hop dissemination of messages, e.g. to inform drivers in a larger area about a hazardous situation such as an accident. A comprehensive classification of applications and communication patterns can be found in [1].

One application which requires a more sophisticated form of information dissemination is the collaborative detection and dissemination of traffic congestions. To be able to use alternate routes and avoid congestions, drivers need to be aware of them already several kilometers in advance. Simple flooding strategies do not scale to high vehicle densities found on multi-laned highways in congestion situations. Available bandwidth and frequent collisions do not allow disseminating individual, exact speed reports of all involved vehicles over large distances. Moreover, drivers do not need exact individual reports, but only an overview of the general average speed on the road ahead.

Therefore, aggregation is an appropriate way to reduce the communication overhead by combining multiple, similar reports of different vehicles. The challenge of aggregation lies in the proper decision if two single data items should be aggregated or not.

We investigated a new approach for aggregation for detection of traffic congestions. In contrast to previous work, our solution operates completely structure-free. It is not necessary to define aggregate structures such as hierarchical grids or fixed road segments or to establish a processing structure like a hierarchy or a group of nodes.

The rest of this paper is organized as follows. In Section II we elaborate on existing approaches for data aggregation both in the Wireless Sensor Networks domain and in the Vehicular Ad Hoc Networks domain. In Section III we give an abstract view the basic concept of our framework. Section IV builds upo n that abstract view to explain the system architecture in more detail. The presentation of our framework is followed by an analysis of our simulation results in Section V. Finally, Section VI concludes this paper.

II. RELATED WORK

Data aggregation is an essential information dissemination paradigm in Wireless Sensor Networks (WSNs). The primary goal is to minimize energy consumption by combining data from different sensor nodes, and thus lowering redundancy and communication overhead [2]. Unfortunately, these aggregation mechanisms cannot be applied directly for VANETs for several reasons.

First, most mechanisms assume stationary sensor nodes in order to realize hierarchical aggregation. For example, data is aggregated over a chain of nodes [3], a tree hierarchy [4], [5] or a cluster based scheme [6]. In a highly dynamic network like VANETs, building such hierarchies is not feasible or would at least result in high communication overhead.

The second reason is the on demand nature of WSN applications. Many applications need a base station (sink) which queries sensor nodes for specific data. After such a query, sensor nodes propagate collected information towards the base station. Protocols which use this pull paradigm ([7] for example) can't be deployed for many VANET applications because there are no dedicated sinks. Instead, all vehicles (for example in a specific region) are supposed to be receiver. Moreover, VANET applications like collaborative detection and dissemination of traffic congestions require periodic status information over road conditions in order to work properly.

Therefore, some new aggregation schemes were proposed by the research community, which better fit the characteristics of VANETs. For example, SOTIS [8] uses a fixed length road segmentation that serves to average traffic data. Traffic information is periodically exchanged via beacons. The problem of this scheme is the fixed road segmentation, which requires all vehicles to have a common knowledge about the segmentation. In addition, is is difficult so select an appropriate segment size. If segments are very short, the scheme delivers more accurate traffic information but also the communication complexity grows. On the other hand, large segments lower the amount of data but also decrease accuracy.

A more advanced aggregation scheme is applied in the TrafficView system [10], also based on a fixed road segmentation. The authors introduce a ratio-based algorithm which can be parameterized to weight road segments according to the importance of the region. This aggregation ratio can be used for example to lower the accuracy of the broadcast information for remote regions. Another approach is used in [11] for the decentralized discovery of free parking places. Vehicles use periodic broadcasts (beacons) to disseminate information about free parking slots. In that work aggregation is performed over a hierarchical quad-tree. Thus, the map is divided into non-overlapping areas and information is organized in different levels. This way the communication overhead can be reduced drastically because aggregated information is disseminated into remote regions with increasing aggregation level (i.e. with decreasing accuracy).

Although these approaches are designed for VANETs and they reduce the communication overhead by reducing the accuracy for regions which are farer away, they still are based on a pre-defined road segmentation or quad structure. This requires that vehicles have a common knowledge about the road network and its segmentation which is problematic regarding the variety of VANET applications. We argue that a reduction of the information accuracy with increasing distance is not always an appropriate solution. For example a driver on a highway might be interested where the end of a (distant) traffic jam is located exactly for being able to decide which exit he can take. On the other hand a driver located in the middle of a traffic jam is not interested on accurate (atomic) information in his neighborhood. But a predefined segmentation is not able to adapt to the current conditions flexibly, e.g. to reflect very short or very long parts of the road with similar average speed. Therefore, we propose in this work a completely structure-free aggregation framework which provides the flexibility needed by VANET applications.

III. BASIC CONCEPT

The following example will explain, why the question is challenging whether two bits of information are similar enough to be aggregated, if one does not want to rely on any grid or tree structures. Taking our example application, traffic congestion propagation, we want to combine speed reports that are reasonably similar, but we want to keep reports separate which do not fit together, e.g. because the report locations are too distant from each other. One of the intuitive criteria for such a decision is the speed difference of two reports. Clearly, it is necessary to set a limit on the maximum speed difference of two reports to ensure a certain quality of the aggregated data. On the other hand however, different types of vehicles and trucks driving on the same location of a highway can exist, particularly when position information is not fine-grained enough to distinguish different lanes. This leads to reports originating from almost the same location and time with quite different speeds. Yet, they should still be combined for an average representation of that highway part's traffic. Therefore, complex decision rules are necessary. Neither are strict decisions possible whether to aggregate or not, nor is it likely that the required rules and influences are fully known at design-time of the system.

Following this rationale, we propose to employ fuzzy reasoning systems to make aggregation decisions. These are based on the fuzzy set theory [12]

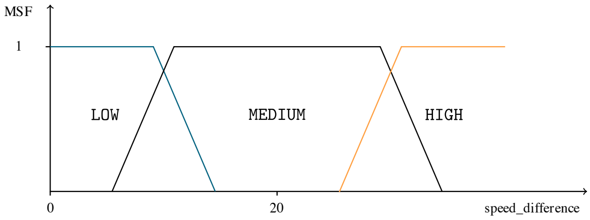


Fig. 1. A sample fuzzy variable with several membership functions.

which extends the classic understanding of sets with the notion of an element being *somewhat* part of a set. More formal, a fuzzy set is a tuple (A, m), A being a normal set and $m : A \rightarrow [0, 1]$ a function that assigns each member of the set a degree of membership.

Fuzzy reasoning systems [13] build upon the idea of fuzzy sets and extend them to a logical reasoning system. We will use our aggregation framework and its application to collaborative congestion detection as an example to elaborate the fuzzy reasoning principles in the following.

The decision process comprises three parts. The first step is the so called *fuzzyfication* of input values. In our case, input values are e.g. the difference of two aggregates' mean speeds measured in km/h. Such an input value is assigned to fuzzy variables. Each fuzzy variable is in turn defined by several membership functions $f_{var,adj} : \mathbb{R} \rightarrow [0, 1]$ that assign natural language adjectives to the actual input value. Figure 1 shows an example for the aforementioned variable speed_difference . If e.g. the input value is 10.5, then the evaluation of the variables membership functions results in the variable being LOW with a degree of 0.8 and at the same time MEDIUM with a degree of 0.8. In addition to the variables representing the input values, one ore more output variables are defined in a similar fashion, but without the assignment of a value. In our use case, only one output variable, representing the aggregation decision, is defined. It has two membership functions "yes" and "no" which are equal in the area they enclose with the abscissa and which do not intersect.

In the next step, the *fuzzyfied* variables are evaluated using fuzzy logic rules. The following shows the basic structure of such a rule:

```
if var1 is adj1
    [and|or|not var2 is adj2
      [and|or|not ...]
    ]
then varN is adjN
```

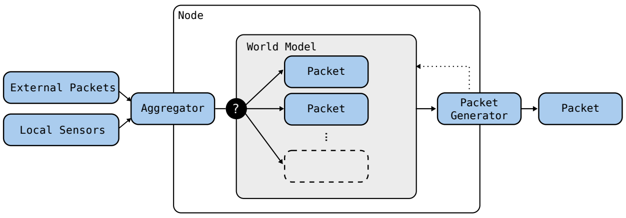


Fig. 2. Overview of the aggregation system architecture.

In the *condition* of the if-statement, fuzzy variables are combined using the boolean operators *and*, *or* and *not* which are therefore defined to mean $\min(a, b)$, $\max(a, b)$ and $1 - a$ respectively. The result is then assigned to the output variable in the *statement* of the rule. Note that several rules can be defined in this fashion assigning different values to the membership functions of the output variables.

Finally, the assignments to the output variable(s) have to be *defuzzyfied* to gain the result value(s). In the case of our aggregation framework, the defuzzyfication should result in a decision for or against the aggregation of two values. Recall that our only output variable therefore only has two non-intersecting buckets representing the two decision options. These buckets are filled by each fuzzy rule evaluation assigning to either of the two by filling it to the amount given by the rule evaluation's result. The decision of the fuzzy reasoning system is then given by the bucket which is filled the most.

For the example of different vehicle speeds located closely together as introduced in the beginning of this section, one could formulate the following, simplified rules:

```
if speed_difference is HIGH
  then decision is NO
if location_difference is SMALL
  then decision is YES
```

This fuzzy decision making process allows us to define all influences on the aggregation decision using independent fuzzy rules. Inside those rules, all complexity of the underlying influences is hidden because they are assigned to natural language adjectives beforehand.

IV. SYSTEM ARCHITECTURE

After outlining how we eliminate the need for predefined aggregation structures, we will now embed these ideas in a complete framework for aggregation in VANETs. Figure 2 shows an overview of the system architecture. The framework works with information that can either originate from local sensor readings or can be received from other vehicles. When a new information arrives, the *aggregator* decides, if the information should be added separately to the *world model* and which information should be aggregated with already existing information. Periodically, the *packet generator* evaluates information currently contained in the world in order to select a subset of the information that is best suited for further dissemination to nearby vehicles.

First, packets containing information atoms or aggregates enter the system. Local sensor readings are always atomic values, whereas information from other nodes may be either atomic or already aggregated. Information atoms contain their location and time as well as application specific information such as a vehicle's speed in case of congestion detection. Aggregates contain application specific information about time and location intervals.

Those incoming atoms and aggregates are then compared with the information already present in the node's world model. The world model is a complete view of all information available about the surrounding road network. The aggregator decides whether new information should be added to the world model as is or aggregated with already present information. For this decision, we apply the fuzzy reasoning system as described in the previous section. As a result of the fuzzy rules, we get a decision for or against aggregation of two atoms or aggregates. For each new information, these rules are evaluated to compare them to atoms or aggregates existing in the world model. Therefore, an aggregate's extent in time and location can freely adapt to the current road situation. There is no pre-defined segmentation that the aggregates need to adhere to.

To allow for multi-hop dissemination of both information atoms and aggregates in the world model, the packet generator periodically selects a certain subset of the information in the world model that best represents the current view of the surrounding road situation. For this selection, the available information is rated according to its time, location and application specific values. The highest rated information is then transmitted to the vehicles in communication range and added to their world model according to their aggregator's decision.

V. EARLY EVALUATION RESULTS

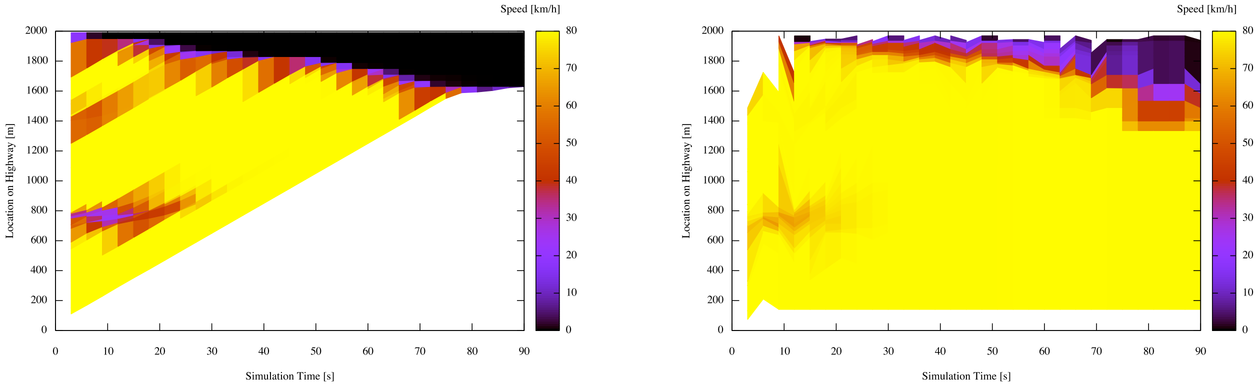
To evaluate our framework, we implement our aggregation scheme for the application of collaborative traffic congestion detection using the JiST/SWANS simulator [14]. We simulate a highway segment with 2 km length and three lanes, which is blocked by an obstacle at the end. A total of 30 vehicles are approaching the obstacle at different speeds, slowing down to avoid an accident. Thereby, a growing road congestion forms during the 90 seconds of the simulation. For the communication, we use broadcasts on an IEEE 802.11-based medium access model, with a transmission range of 250m. To visualize the situation, we created a 3D model of the situation showing both the real world traffic situation and the aggregated view of an approaching vehicle.

Fig. 3. A visualization of the aggregation framework.

Figure 3 shows the visualized simulation. The top half of the screen shows a bird's eye view of the vehicles approaching the obstacle. In the bottom half, both the current positions of all cars and the aggregated world view of one approaching vehicle is shown. It can be seen that both the growing congestion and the segments of normal traffic flow are well represented by the aggregated view. This is due to the aggregates being able to extend arbitrarily in time and space according to the application's quality requirements.

To get a more complete overview of the aggregated view's quality, we create a spectral view of the traffic situation during the whole simulation. For each point in time on the abscissa, the vertical line at that point on the x-axis gives the traffic situation at that time. The ordinate shows the location on the simulated highway and the mean speed at that location is color-coded with light, yellow colors meaning normal flow and dark colors meaning congestion or low average vehicle speeds. Hence, starting from the left, the average speed can be observed at every location on the road as it changes during simulation time.

Figure 4(a) shows the real traffic situation as it would be seen by an all-seeing observer that has all the speed data of all vehicles available at all times. To



(a) Real world view.

(b) Aggregated view.

Fig. 4. A spectral view of vehicles approaching an accident scene. The abscissa shows the simulation time, the ordinate shows the location along the highway. Colors indicate the vehicles' speed.

derive this real world view, which can be seen as the reference, the speeds of each vehicle in the simulation are taken every three seconds and are then connected to build the spectral view. The growing white area at the bottom right shows the parts of the highway where there is no traffic at all, i.e. no vehicles are driving on those segments. The growing black area on the upper left shows the growing total congestion due to the obstacle on the road.

For measuring the quality of the aggregation framework, we now compile a similar view from all aggregated views of vehicles in the simulation. Figure 4(b) shows this composite view. To allow for easy comparison of both the real world situation and the aggregated view, this graph averages all aggregated views of all vehicles on the simulated scene at each point in simulation time. For clarification, suppose the average view after t seconds of simulation time is to be calculated. Now, each vehicle reports its aggregated view of the whole highway at that time using our aggregation framework. Overlaying these views will result in some parts of the road where only one vehicle reports knowledge of the traffic situation and many parts where several vehicles report an estimated speed. For each of those parts with different speed information, the average of all reports is calculated. This results in an average aggregated view of the highway at one instant in time. Plotting these averages for the whole simulated timeframe results in a spectral view as seen in Figure 4(b), which is comparable to the accurate view by the global observer.

The growing congestion as the main feature of the simulated traffic situation is closely matched by the aggregation protocol. Colors – i.e. average speeds – are generally slightly lighter than in the real world situation, which shows the aggregation lag. This is because the aggregation takes into account older traffic reports that show a higher speed on the same highway part. In contrast, the real world reference always has the latest information available. After a short delay though, these older aggregates are not disseminated any more and the aggregated view adapts to the new situation. This feature of the proposed system allows to even out short-termed phenomena that are not relevant for vehicles approaching from further away.

The area on the lower left shows how the aggregation framework reacts to small road parts with lower speed that is due to faster vehicles overtaking slower ones. While the aggregation compensates the fast and short-termed fluctuations, it still reflects the presence of slower traffic flow in that region.

Both the aggregated view of the full congestion and the small low-speed periods again show an important feature of our completely structure-free approach. The predominant part of the road with normal traffic flow is represented by one large aggregate reflecting that state. The smaller congested sections however are represented by smaller aggregates covering only the area where the congestion actually happens. Grid based approaches would need several fixed-sized aggregates to cover the area of normal traffic flow and likely still not represent the smaller congestion situations properly.

VI. CONCLUSION

Our studies have shown that efficient aggregation in VANETs is possible without the need for predefined road segments. Fuzzy reasoning enables us to employ natural language rules to make aggregation decisions. This allows for an easy representation of the application requirements, even if not all influences are exactly known. As a result of this fully structure-free aggregation approach, the aggregated view of the road directly represents the current traffic situation both as fine grained as necessary and as coarse as possible. The approach is not restricted to congestion detection but can e.g. also be used to report road conditions or fog situation.

Besides further assessing the framework's performance, we currently address security implications that arise when information from several different vehicles is aggregated to one single report. Early results of both research directions underline the promising observations of our first simulations.

REFERENCES

[1] Elmar Schoch, Frank Kargl, Tim Leinmüller, and Michael Weber, "Communication Patterns in VANETs," *IEEE Communications Magazine*, vol. 46, no. 11, pp. 2–8, Nov. 2008.

[2] Bhaskar Krishnamachari, Deborah Estrin, and Stephen Wicker, "The Impact of Data Aggregation in Wireless Sensor Networks," *Distributed Computing Systems Workshops, International Conference on*, vol. 0, pp. 575, 2002.

[3] S. Lindsey and C. S. Raghavendra, "PEGASIS: Power-Efficient Gathering in Sensor Information Systems," in *Aerospace Conference Proceedings, 2002. IEEE*, vol. 3, pp. 3–1125–3–1130 vol.3.

[4] Stephanie Lindsey, Cauligi Raghavendra, and Krishna Sivalingam, "Data Gathering in Sensor Networks using the Energy* Delay Metric," *Parallel and Distributed Processing Symposium, International*, vol. 3, pp. 30188b, 2001.

[5] Haowen Chan, Adrian Perrig, and Dawn Song, "Secure hierarchical in-network aggregation in sensor networks," in *CCS '06: Proceedings of the 13th ACM conference on Computer and communications security*, New York, NY, USA, 2006, pp. 278–287, ACM.

[6] Wendi Rabiner Heinzelman, Anantha Chandrakasan, and Hari Balakrishnan, "Energy-Efficient Communication Protocol for Wireless Microsensor Networks," in *HICSS '00: Proceedings of the 33rd Hawaii International Conference on System Sciences-Volume 8*, Washington, DC, USA, 2000, p. 8020, IEEE Computer Society.

[7] Chalermek Intanagonwiwat, Ramesh Govindan, and Deborah Estrin, "Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks," in *MobiCom '00: Proceedings of the 6th annual international conference on Mobile computing and networking*, New York, NY, USA, 2000, pp. 56–67, ACM.

[8] Lars Wischhof, Andre Ebner, Hermann Rohling, Matthias Lott, and Rudiger Halfmann, "SOTIS - A Self-Organizing Traffic Information System," in *Proceedings of the 57th IEEE Vehicular Technology Conference (VTC 03 Spring)*, Jeju, South Korea, 2003.

[9] Christian Lochert, Björn Scheuermann, and Martin Mauve, "Probabilistic aggregation for data dissemination in VANETs," in *VANET 2007: Proceedings of the Fourth ACM International Workshop on Vehicular Ad Hoc Networks*, Sept. 2007, pp. 1–8.

[10] Tamer Nadeem, Sasan Dashtinezhad, Chunyuan Liao, and Liviu Iftode, "TrafficView: Traffic Data Dissemination using Car-to-Car Communication," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 8, no. 3, pp. 6–19, 2004.

[11] Murat Caliskan, Daniel Graupner, and Martin Mauve, "Decentralized Discovery of Free Parking Places," in *VANET '06: Proceedings of the 3rd international workshop on Vehicular ad hoc networks*, New York, NY, USA, 2006, pp. 30–39, ACM.

[12] Lotfi A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, pp. 338–353, 1965.

[13] George J. Klir and Bo Yuan, *Fuzzy sets and fuzzy logic: theory and applications*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1995.

[14] R. Barr, Z.J. Haas, and R. van Renesse, "JiST: An efficient approach to simulation using virtual machines," *Software Practice & Experience*, vol. 35, no. 6, pp. 539–576, 2005.

[15] Maxim Raya, Adel Aziz, and Jean-Pierre Hubaux, "Efficient secure aggregation in vanets," in *VANET '06: Proceedings of the 3rd international workshop on Vehicular ad hoc networks*, New York, NY, USA, 2006, pp. 67–75, ACM Press.