SURVEY ON CLUSTER BASED DATA AGGREGATION FOR VANETs

[1]Sridevi Hosmani, [2]Dr. Basavaraj Mathapati
[1]Asst. Professor, [2]Professor
[1]*Information Science and Engineering*
[1]*Godutai Engineering College for Women, Kalaburagi, India.*

Abstract: *Vehicular adhoc network provides communication between vehicles and also with road side units. Data dissemination is an important aspect in VANETs as vehicles are in constant motion and the topology of the network changes frequently mainly increases the communication overhead. In this paper we review various data aggregation techniques available that has various advantages including the reduction of communication overhead and also present the data flow model for aggregation.*

Keywords: VANET, Data Aggregation, Clustering, Data dissemination, V2V, V2I.

I. Introduction

Vehicular adhoc network (VANET) is a derivative of mobile Adhoc network (MANET) that helps for information exchange among the vehicles. As vehicles are in continuous motion, the topology of the network changes frequently. In order to enable data transfer between the vehicles, VANET supports two types of communication.V2V- vehicle to vehicle communication and V2I- vehicle to infrastructure communication. In VANETs vehicles are equipped with On Board Units(OBU) that allow for communication and connection with other vehicles or infrastructure i.e., road side Units(RSU). The FCC (Federal Communications Commission) has assigned a frequency bandwidth of 75MHz in 5.855- 5.925 GHz for communication and the protocol is known as DSRC (Dedicated Short Range for Communication) using IEEE 802.11p. The frequency band is divided into seven, ten megahertz channel consisting of one control channel, six service channels and one five megahertz channel that is reserved [1].

Data aggregation is a technique where data packets from different sources are combined together to produce an aggregated packet. There are two types of aggregation Syntactic and Semantic aggregation. In syntactic aggregation, aggregation of data is done from multiple vehicles to fit into a single frame. Where as in Semantic aggregation the data collected from individual vehicles is summarized. E.g. The information of number of vehicles on the lane is propagated instead of providing the location information of vehicles as well. There are various advantages of data aggregation like reduction of redundant data, reduction of overhead, improves channel efficiency etc.

Clustering means grouping of vehicles based on certain parameters like speed, direction, location etc. Among the group of nodes one node acts as a co-ordinator called cluster head (CH) and other nodes co-ordinating with CH are called as Cluster members (CM). CH is normally elected by the CMs of the cluster. CH collects all information from the CMs and aggregates all information and only disseminates the merged information to next CHs or cluster .This technique makes the VANETs more robust and scalable. Cluster-based approaches may be the only viable solutions for supporting scalable multi-hop communication for high density VANETs [9].

II. Generic Model

Vehicles receive the information from other vehicles, the received information is processed and disseminated to other vehicles. To give the detailed insight of how the aggregation is performed, we present the generic model for data aggregation [2] that mainly presents how the information is represented and the different phases of how data is aggregated.

1. Representation of Information

(a)Aggregate

The information in an aggregation is represented by aggregates. An aggregate is a tuple that contains information of a specific region at a specific time period.

$$A = (L, T, v, Q) \in A \ldots\ldots. (1)$$

where L and T are the geographical and temporal locator of aggregate respectively. They represent the geographical location and a time period. v is the primary value of aggregate for example speed, temperature, parking slots etc. Q=q1, q2……qn are the auxiliary values of the aggregate. Auxiliary values can relate to one of primary values i.e. standard deviation of an average or relate to aggregate as a whole.

(b) Atomic observation

An atomic observation is a tuple that is composed of a vehicles local sensor values

$$o = (L, T, v, Q) \in O \subset A \; \ldots\ldots\ldots (2)$$

where L and T are the geographical and temporal locator of aggregate respectively. But in observation they identify a specific point in space and time where o was observed. V represents the observed value. Q represents the auxiliary value as in aggregates but is usually omitted.

2. Phases of Aggregation

(a) Decision

The responsibility of decision component is to decide whether the two information items that are similar can be aggregated. To do this the information received from current vehicle and all information from world model can be used [3]. Therefore the decision component compares the set of all aggregates i.e the power set of aggregates and groups similar items for aggregation

$$\text{Decision} :(P (A), C) \rightarrow \{\text{yes, no}\} \ldots\ldots.. (3)$$

Where P (A) represents power set of aggregates and C represents context such as current time, location, direction etc.

(b) Fusion

The fusion component does the merging of aggregates provided by decision component. Fusion can be lossless or lossy process. A lossless approach is nothing but concatenation of received information whereas lossy approach can reduce the size of packets.

$$\text{Fusion: } P (A) \rightarrow A \ldots\ldots\ldots (4)$$

The properties of fusion function are:

1. Hierarchical applicability: It should be hierarchically applicable in order to scale well and support propagation of data to large areas.
2. Duplicate Insensitivity: Same parameter/event will be sensed by more than vehicle. The fusion function should filter out the duplicate information.
3. Data quality tracking: Fusion function keeps track of quality of data while aggregating the data.

(c) World Model

World model of a vehicle represents the information received by a vehicle x at a time T that is a set of aggregates.

$$W(x, T) :=\{ A1, \ldots , An\} \subseteq A \; . \ldots\ldots. (5)$$

(d)Dissemination

The decision component selects a subset of world model for dissemination where S represents subset of world model. The most relevant data are selected for dissemination considering the context C.

$$\text{Dissemination: } (W, C) \rightarrow S \subset W \ldots\ldots. (6)$$

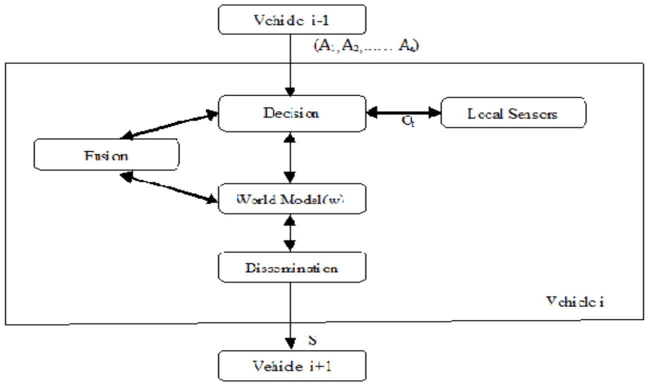

Figure 1: Generic Model for Aggregation

Figure 1 represents the generic model that is most commonly used for data aggregation. The information sensed by a vehicle is sent to decision component and is compared with the previous observations. The decision whether the new information can be aggregated or no is done here .If the information is to be aggregated, it is forwarded to fusion component otherwise is forwarded to world model. The fusion component performs the aggregation and forwards the resultant to world model. Lastly the dissemination component selects the subset of world model and disseminates them to other vehicles.

There are certain requirements that need to be considered for data aggregation. Some of the requirements are listed below [4]

1. The data aggregation scheme should be scalable so that it provides good performance.
2. The data aggregation scheme should provide accurate data considering the application.
3. The data aggregation scheme should be able to compare the data to one another so that it provides high efficiency.
4. The data aggregation scheme should consider security of data.

III. Literature survey

In this section some of the data aggregation techniques found in the literature are reviewed.

Authors of paper [5], presented an approach called data aggregation in VANETs: the VESPA approach which considers peer to peer (P2P) architecture. In this approach vehicles exchange all types of events for example urgent braking, accident, parking space etc. The vehicles consider fresh data as well as old aggregates for extracting additional knowledge. In order to achieve this, tables are used knowingly space index, time index and aggregation boundaries.

Authors of paper [6], proposed a probabilistic aggregation approach called probabilistic aggregation for data dissemination in VANETs. This scheme used Flajolet- Martin sketch to represent the aggregation using hash function. For fusing the data bit-wise OR function is used and to remove old data counters are set that represent TTL. The evaluation is performed using NS2 and VSSIM simulator considering 10000 vehicles covering 500km. The parameters considered for evaluation are accuracy for local large aggregates, distant medium sized aggregates and distant large sized aggregates.

Authors of paper [7], presents an adaptive aggregation framework named data aggregation in VANETs. It mainly deals with aggregation at RSU and dissemination of aggregated data to other RSUs towards sink. Adaptive aggregation is achieved with the help of 3 controllers, aggregation level controller, Dissemination period controller and requirement controller. Three aggregation schemes are presented 1. Flat aggregation, where data is stored in data nodes that are direct children of root nodes 2. Binary interval aggregation scheme, here interval nodes are used to represent the metric that requires minimum quality 3. Free Hierarchical aggregation, here the tree consists of only data nodes. The data nodes that are neither root nor leaf represent only one metric. The proposed three schemes are evaluated for CBR and aggregation level changes.

Authors of paper [8], present a framework for message aggregation, which mainly consist of 2 components 1. Resource2.Application. It mainly focuses on different resource of data for aggregation and different applications of VANETs. The paper provides a detailed idea about mandatory parameters like vehicle ID, position and non- mandatory parameters such as road surface etc. Message aggregation is used to improve efficiency reducing overhead.

In paper [9], authors presented a structure free aggregation framework for vehicular ad hoc networks. Structure free aggregation indicates that the algorithm does not rely on any of the grid or tree structures. When structure free aggregation is used it is difficult to decide on the sensed parameter to be aggregated or no. so here fuzzy reasoning systems are used to make decision on aggregation. The parameters that packet consists are location, time and application specific information like speed. In this work, spectral view of the traffic situation is represented using JIST/SWANS simulator.

Authors of paper [10] proposed cluster based accurate syntactic compression of aggregated data in VANETs (CASCADE) protocol, which mainly focus on accuracy of aggregated data. In this scheme every vehicle periodically broadcasts its position information which is called as primary record. The primary records of the vehicles that are ahead of a particular vehicle form the local view of that vehicle. Here the local view is divided into clusters based on the distance information present in the received primary record corresponding to the receiving vehicle. Then an aggregated cluster record is obtained by concatenating the compressed data records of the vehicles in the cluster. The aggregated cluster records are then concatenated into a single frame and are broadcasted. This scheme ensures low delay and high reception rate.

The authors of the paper [11], presented an optimization of CASCADE data aggregation for VANET's. This scheme tries to reduce the size of aggregated frame while increasing the local view. Vehicles construct compact records that contain location and speed information by performing aggregation and compression of received information. Then the aggregated cluster record for each cluster is obtained. In order to achieve this, a cluster size of 16m wide and 126m long is considered.

In [12] authors proposed a cluster based data aggregation protocol for vehicular Ad Hoc Networks, Here autonomous clustering is used and aggregation is performed by CH and data packet is disseminated considering there are n different parameters. Weighted Geometric mean algorithm is used for data aggregation where in weights on values are function of time. The results depict reduction in number of packets due to aggregation compared to non aggregation.

Authors of paper [13] proposed Vehicular stable cluster based data aggregation (VeSCA) protocol, where in multilevel aggregation is performed by the nodes/vehicles until the packet reaches CH and from CH the aggregated data is disseminated. This algorithm demonstrates high stability of clusters and improved performance achieving low delay and high delivery ratio.

Authors of paper [14] proposed a protocol, Cluster based semantic data aggregation (CBSDA). The vehicles moving in the same direction are grouped into clusters. CH is chosen based on relative velocity of all CMs and the distance between the vehicles to back end of the cluster. Aggregation is performed in CH and data is disseminated to next CH. Decision component is based on weighted deviation scheme considering velocity and position as parameters for weight calculation. This paper also introduces two methods for aggregation control that are CBR and Density control.

The authors of the paper [15], presented a generic algorithm, location aware data aggregation for efficient message dissemination in vehicular adhoc networks. Since it is a generic algorithm the messages can contain information regarding accident, parking space etc. The vehicles use location awareness (LA) algorithm to get the information regarding street section i.e., location. Every time a vehicle changes the street section, the database is refreshed for current information. Kalman filter is used to fuse the data together. In order to disseminate the data, passive clustering is used. Cluster us formed based on current direction obtained from LA algorithm. The proposed algorithm is evaluated for efficiency and accuracy of aggregation.

Authors of paper [16], presented an algorithm called data relationship degree based clustering data aggregation for VANETs based on data relationship degree (DRD). The algorithm consist of 3 steps. 1. Vehicle type calculation, where each vehicle examines and decides whether it can become core sensing vehicle. 2. Local cluster formation, here cluster of vehicles are formed based on road segment. 3. Universal delegate sensing vehicle selection, here local cluster are combined together according to maximal DRD information stored in each sensing vehicle and a delegate sensing vehicle is selected. The protocol is evaluated using NS2 and the parameters considered are Packet Delivery Ratio (PDR), data overhead, average delay, network reachability varying traffic density and data sensing rate.

In [17] authors presented a game theory approach for intra cluster data aggregation in VANETs. In this approach the algorithm is presented in 3 phases. 1. Cluster initialization. 2. Cluster stabilization. 3. Cluster reconstruction. In Cluster initialization stage, CH is selected based on direction, relative speed and leader duration. In Cluster stabilization stage, data aggregation is performed. Nash equilibrium and interruption process are used. If the cluster detaches and enters into Cluster reconstruction stage, cluster head is to be selected by repeating cluster initialization phase. The performance parameters demonstrated are accuracy ratio, stability ratio, compression ratio.

In [18] authors presented a context aware data aggregation (CADA) in vehicular ad Hoc networks, where clustering technique is used. CH performs data aggregation. Parameters considered for data aggregation are speed, direction and number of cluster members. In this protocol aggregation of velocity information of vehicles is done. Here data is aggregated if there is any change in the context and it needs to be disseminated. This protocol demonstrates better usage of bandwidth, reduction of number of packets to be disseminated.

Below table shows the comparison of various cluster based data aggregation scheme in VANET.

Ref. No	Parameters	Simulator	Performance Parameters
[10]	Location, Speed	ASH(Application aware SWANS with highway mobility)	MAC delay, Reception rate, Throughput, Visibility
[11]	Location, Speed	ASH(Application aware SWANS with highway mobility)	Local view length, Aggregated frame size
[12]	-	NS2	Bandwidth utilization, Number of packets
[13]	Relative mobility	Ns 3	Average cluster head duration, Average cluster member duration, Control overhead, Aggregated data packet delivery ratio, Data Aggregation rate, Average Delay.
[14]	Relative velocity, Position	Ns 3.18	Data precision
[15]	Location, Direction	OMNET++	Accuracy of Aggregation
[16]	Spatial correlation of sampled data	NS2	PDR, data overhead, Average Delay, Network reachability
[17]	Relative speed, Direction	NS2	Accuracy ratio, Stability ratio, Compression ratio
[18]	Speed, Density	Ns 3.18	Bandwidth utilization, No. of packets sent, Time interval

Table 1: Comparison of cluster based data aggregation routing protocols

V. Conclusion

In this paper, various concepts regarding data aggregation are studied, providing the insight of generic model used for data aggregation. This paper reviews various data aggregation techniques present in the literature emphasizing mainly on cluster based techniques. Data aggregation provides various advantages like reduction of overhead, improves scalability, reduces redundant data etc.

References

[1] The FCC DSRC (Dedicated Short Range Communications), web site. http://wireless.fcc.gov/ services/its/dsrc/.

[2] Stefan Dietzel, Jonathan Petit, Frank Kargl, and Bjorn Scheuermann, "In-network aggregation for vehicular ad hoc networks. Communications Surveys & Tutorials," IEEE, 16(4):1909–1932, 2014.

[3] S. Dietzel, F. Kargl, G. Heijenk, F. Schaub, "Modeling in-network aggregation in VANETs, "In Proc. IEEE Communications Magazine, IEEE,2011.

[4] Pim van der Toolen, "Data aggregation in V2V & V2I communication infrastructure," 13th twente student conference on IT, 2010.

[5] Bruno Defude, Thierry Delot, Segio llari, Jose Luis, Nicolas cenerario,"Data aggregation in VANETs : the VESPA approach," Mobiquitous '08 Proceedings of the 5th Annual International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services.

[6] C. Lochert, B. Scheuermann, and M. Mauve, "Probabilistic aggregation for data dissemination in VANETs," In 4th ACM Int. Workshop on Vehicular Ad Hoc Networks, pages 1–8, 2007.Data aggregation in VANETs.

[7] Josef Jiru, Lars Bremer, and Kalman Graffi, "Data aggregation in vanets a generalized framework for channel load adaptive schemes," in Local Computer Networks (LCN), 2014 IEEE 39th Conference on, Sept 2014, pp. 394–397.

[8] Sanket Desai, Rabee Elhdad and Naveen Chilamkurti, "Message aggregation in VANETs for delay sensitive applications," International Journal of Smart Home Vol. 9, No. 10, (2015), pp. 215-222.

[9] S. Dietzel et al., "A Structure-free Aggregation Framework for Vehicular Ad Hoc Networks", Proc. 6th Int'l. Wksp. Intelligent Transportation, Hamburg, Germany, (2009), pp. 61–66.

[10] Ibrahim, K., & Weigle, M. C, "CASCADE: Cluster-based accurate syntactic compression of aggregated data in VANETs," In Proceedings of IEEE GLOBECOM workshops, New Orleans, LA (pp. 1–10). IEEE, 2008.

[11] Ibrahim, K., & Weigle, M. C., "Optimizing CASCADE data aggregation for VANETs," In Proceedings of 5th IEEE international conference on mobile ad hoc and sensor systems (MASS), Atlanta, GA (pp. 724–729). IEEE, 2008.

[12] Muhammad Shoaib, Wang-Cheol Song, and Keun Hyung Kim, "Cluster based data aggregation in vehicular adhoc network," In Communication Technologies for Vehicles, pages 91–102. Springer, 2012.

[13] Seyhan Ucar, Sinem Coleri Ergen and oznur ozkasap, "VeSCA: Vehicular stable cluster based data aggregation," Inretnational Conference on Connected vehicles and Expo(ICCVE), 2014.

[14] Aboobeker Sidhik Koyamparambil Mammu ; Josef Jiru , "Cluster based semantic data aggregation in vanets," 29th International Conference on Advanced Information Networking and Applications, IEEE, 2015.

[15] Milos Milojevic ,Veselin Rakocevic, "Location aware data aggregation for efficient message dissemination in VANETs," IEEE Transactions on Vehicular Technology (Volume: 64 , Issue: 12 , Dec. 2015)

[16] Rakesh Kumar, Mayank Dave, "Data relationship degree based clustering data aggregation for vanets," Pages 485-503, International Journal of electronics, 2015.

[17] Yuzhong Chen, Shining Weng, Wenzhong Guo, and Naixue Xiong, "A game theory algorithm for intra cluster data aggregation in VANET," Sensors (Basel), Feb,2016.

[18] Zubair Amjad, Wang Cheol Song ; Khi Jung Ahn ; Muhammad Shoaib, "Context aware data aggregation in vehicular Ad Hoc network," NOMS 2016 - 2016 IEEE/IFIP Network Operations and Management Symposium.

-