Topic Based Clustering of Vehicles for Information Retrieval and Sharing

Lijun Dong
Huawei Technologies Co., Ltd.
10180 Telesis Ct, Suite 220
San Diego, CA, U.S.A
lijun.dong@huawei.com

Richard Li
Huawei Technologies Co., Ltd.
2330 Central Expressway
Santa Clara, CA, U.S.A
renwei.li@huawei.com

***Abstract*—The paper proposes a topic based clustering mechanism, which achieves information retrieval and sharing by delegating such responsibilities to those vehicles which have infrastructure access and are willing to provide the service to other neighboring vehicles. The paper leverages the name based transport protocol of Named Data Networking (NDN), builds the request and data routing according to the information topic names through the corresponding cluster head. The paper illustrates the details on how the topic clustering is set up, dissolved and split as needed. It also presents the procedures for a vehicle without infrastructure access to detect, join and leave a topic based cluster. The proposed mechanism has the advantages of preventing the Forwarding Information Base (FIB) from expanding too large and reducing the network bandwidth consumption by avoiding transporting the same content multiple times in serving different neighboring vehicles.**

Keywords—clustering, information topic, vehicular network, NDN, FIB, PIT, decision tree classifier

I. INTRODUCTION

In the connected vehicle ecosystem [1], a high volume of information-rich data is exchanged by roadside units and onboard transceivers for applications such as safe driving, entertainment, location based services, etc. Different applications in vehicular networks could exchange different types of information with various topics. The communication pattern could be proactive pull by requesters or passive push from content providers. Thus applications in vehicular networks are information oriented. Users know what information they want, but do not care who provides the information, how and from where the information will arrive. Thus information-centric networking (ICN) is well suited for exchanging information in vehicular networks [2]. The ICN approach assumes the integration of naming, name-based routing, in-network processing and caching as part of the network architecture. ICN also has remarkable support for the communication patterns [3][4] (pull or push) to realize the applications in the vehicular networks.

In the rest of the paper, the Named Data Networking (NDN) [5] is used as one of the most popular and well recognized ICN architectures in the community to illustrate the proposed ideas. A NDN router maintains three data structures: the Forwarding Information Base (FIB) that associates the content names to the forwarding face(s) towards the producer(s), the Pending Interest Table (PIT) that records the incoming faces where the interests came from and have not replied by producer, the Content Store (CS) that caches content from a producer when it is forwarded to the users. It has been proposed in [6] that each vehicle embeds a NDN module with FIB, PIT and CS databases.

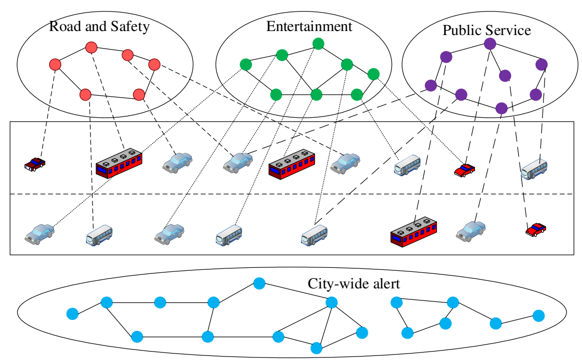

Fig. 1. Vehicle clustering based on information topic

It is noticeable that some vehicle may not have access to the infrastructure, if it does not have any network interface that can connect itself to the base station or access point (i.e. without cellular or WiFi interface). Thus it needs other vehicle with such network interface to relay its data request to the infrastructure and the corresponding data through the peer to peer communication interfaces such as DSRC [7], WAVE [8], 802.11p [9][10]. In our former work [11], we briefly proposed that the content provider can delegate the information exchange responsibilities to those vehicles which have infrastructure access and are willing to provide the service to other neighboring vehicles. The vehicles are clustered by the information topic(s) that the members are requesting. One vehicle may be in multiple clusters, due to the multiple information topics it may be interested in. The topic could be the general description of the data, which is the attribute/semantics information associated with the data and published along with the data. Such clustering can be of multiple layers as shown in the example in Fig. 1. In [11], we discussed the possibility of carrying the clustering based content request in the IP packet extended blocks [12] and in NDN interests. In this paper, we further propose the details on how to set up a topic based cluster for information retrieval and sharing in vehicular networks, how a vehicle detects, joins, leaves a topic-based cluster, and how to dissolve and split a topic based cluster, which will be described in Section II. With the NDN as the transport protocol, the vehicles only need to maintain the topic name based FIB table, and all the interests requesting the same topic name will be forwarded to one of the corresponding cluster headers. This can significantly simplify the routing and processing in the intermediate vehicles, which will be illustrated in the Section III with performance discussions. Section IV concludes the paper.

II. DETAILED DESCRIPTIONS

A. Setup a Topic Based Cluster

It is proposed in this section that a vehicle may be triggered to set up a new cluster (given the cluster does not exist yet or the vehicle cannot reach the cluster although it exists) when there are multiple interests of the same topic or there are multiple pushed messages of the same topic that have been forwarded by the vehicle. Such statistics is maintained and analyzed by the vehicle, which is capable of performing the operations due to the fairly large storage and computation power equipped by the vehicle. The following conditions should be satisfied before the vehicle sets up a new cluster:

- It has been verified by the vehicle that there is no existing cluster of the same topic in the reachable distance or the existing cluster of the same topic does not accept new member.
- The vehicle can reach the infrastructure directly and stably through one or more multiple network interfaces.
- The vehicle is willing to forward and aggregate other vehicles' interests with the same topic. In NDN, the interests with the same content name are aggregated in one entry in PIT, which are dropped by the vehicle from being further forwarded. The vehicle should also be willing to distribute the returned data to the corresponding requesters.

Each cluster head broadcasts the existence of its cluster in the vehicular network periodically by using a beacon message. The beacon message includes the cluster's information topic, the expiration time, as well as the distance between the current vehicle and the cluster head (which is initially set to 0 by the cluster head). The beacon message may have a time-to-live field, which is decreased by 1 when it is broadcast in the network by one hop. It is used to prevent the beacon message from overloading the network bandwidth. And it is also likely that the cluster head wants to the keep the cluster in a small adjacent area. On the other hand, the beacon message can optionally contain a tag which is unique among all the clusters for the same topic name. If the tag is included in the beacon message, the vehicles receiving this beacon can associate the tag with the next hop in the FIB. If the same beacon is received by the vehicle again, the vehicle can easily identify that the beacon is duplicate by looking at the tag value and then simply discard it. Later the tag can also be used for other vehicles to join this specific cluster.

It is assumed that if the tags of clusters for a topic are not known to the vehicles, by default the network will allow the vehicles to try to join all the clusters (i.e. the "*Cluster Joining Request Message*" will be forwarded to all the next hops corresponding to the topic). When a vehicle receives a beacon message, it will extract the topic name and expiration time and carry out the procedures as shown in Fig. 2.

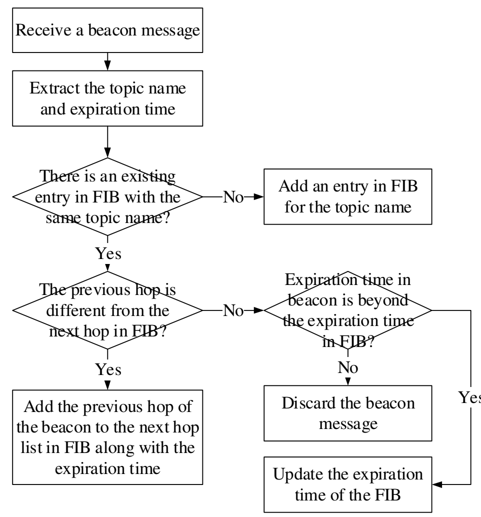

Fig. 2. Processing of beacon message by a vehicle

If there is no existing entry in the FIB for the topic name, then the vehicle creates an entry in the FIB, with the topic name recorded, next hop set to the previous hop of the beacon message to reach the cluster head, the expiration time is set up based on the beacon message, the hop count is increased by 1. The beacon message is broadcast by the current vehicle since it was not received before.

If there is an existing entry for the topic name in the FIB, the expiration time and the previous hop of the beacon message are used to determine whether the beacon message has been received before and whether there is a new cluster for the same topic name to be recorded.

- If the previous hop of the beacon message is the same as one in the next hop list, and the expiration time is beyond the corresponding expiration time in the FIB, then it is a new beacon message for an existing cluster but with a updated expiration time. Thus the expiration time in the FIB is updated accordingly. In this scenario, it is likely the beacon message is from the cluster head of a new cluster for the same topic name, which reaches the current vehicle through the same previous hop. As shown in Fig. 3, the beacon message from the new cluster head is sent to the vehicle 3. Before the forwarding, the vehicle 3 adds a new entry in the next hop list for the "video clip" topic name. The new cluster head's beacon message is forwarded to the vehicle 4.The vehicle 4 does not need to know that there is a new cluster formed for the same topic, but only needs to update the expiration time. In the later sections, we will discuss how the FIB will be used for cluster discovery as well as interest forwarding/routing.

- If the previous hop of the beacon message is the same as one in the next hop list, and the expiration time is not beyond the corresponding expiration time in the FIB, then the beacon message was received by the current vehicle before. The beacon is discarded and not forwarded further.

- If the previous hop of the beacon message is not the same as anyone in the next hop list, then there must be a new cluster formed for the same topic name. The previous hop of the beacon message is added to the next hop list and the corresponding expiration time is calculated based on the expiration time in the beacon and the current time.

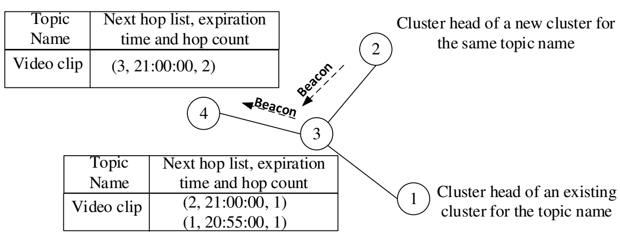

Fig. 3. Two clusters for the same topic name

B. Detect Topic Based Cluster Presence

Before a vehicle initiates creating a new cluster for a topic name, it is likely to detect whether there is any cluster existing for the same topic. The cluster presence detection is carried out by the vehicle by firstly checking its FIB if there is an existing entry for the topic name.

- If yes, there is a cluster existing in the network that is reachable. It means that the vehicle has received the periodic beacon message from the cluster head. Thus the vehicle can choose not to create a new cluster but to join the existing cluster.

- If no, the vehicle broadcasts the "*Cluster Presence Inquiry Message*" to its neighboring vehicles, which shall include the topic name. When a neighboring vehicle receives the Cluster Presence Inquiry Message, and if there is an existing entry for the topic name in its FIB, it will return a confirmation to the source vehicle, which includes the hop count between itself and the cluster head and the expiration time. Otherwise, the Cluster Presence Inquiry Message is broadcast further. The "*Cluster Presence Confirmation Message*" can record the intermediate vehicles between the answering vehicle and the source vehicle of the inquiry message. The source vehicle thus can calculate the distance between itself and the existing cluster head(s). It can also help the source vehicle determine the direction to send the "*Cluster Joining Request Message*", which will be discussed in Section C.

After the cluster presence detection procedures, the vehicle can decide whether to create a new cluster for the topic name or to join an existing cluster. To create a new cluster for the topic name, the vehicle will declare itself as the cluster head, and start to broadcast a beacon message periodically in the network.

C. Join a Topic Based cluster

When a vehicle needs the other vehicles to forward its interest message or notifications for itself, it needs to join a cluster for the topic it is interested in. The vehicle can firstly carry out cluster presence detection in the network, which will make itself be aware of whether there is such cluster in its reachability, the distance between itself and the cluster head, as well as the unique tag of the cluster. If there are multiple clusters, the vehicle can choose one or multiple to join.

The vehicle can also skip the cluster presence detection step, and let the network determine whether there are alternative clusters existing for the topic name and decide which one for the vehicle to join. In the following of this section, we will discuss this scenario. Fig. 4 shows the corresponding message flow.

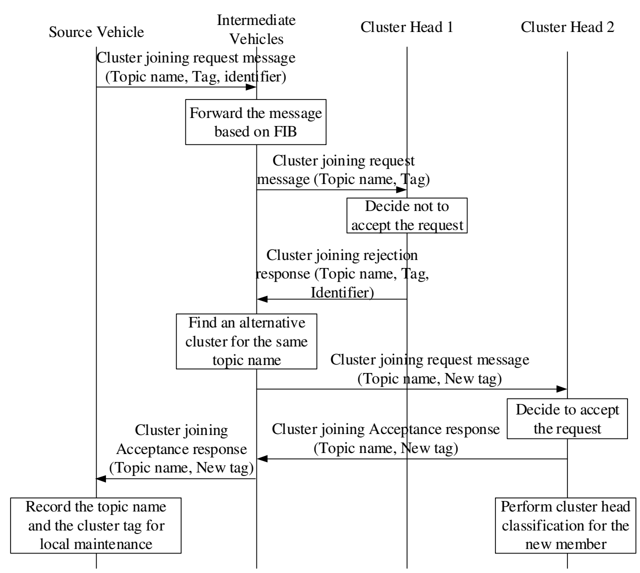

Fig. 4. Message flow of joining a cluster

The Cluster Joining Request Message includes the topic name that the vehicle is interested in, the unique tag of the cluster that the vehicle wants to join, as well as the identifier of the vehicle. If the vehicle has the topic name maintained in the FIB, it matches the tag and forwards the message to the next hop. Such forwarding is proceeded by the intermediate vehicles until the message reaches the cluster head of the cluster. The cluster head decides whether to accept the vehicle's request. Factors could be the estimation of its current load, the setting of the maximum number of members and the current number of members, etc. In the message flow as shown in Fig. 4, the cluster head 1 rejects the source vehicle's cluster joining request, and the rejection response is returned back to the vehicle hop-by-hop through the intermediate vehicles. One of the intermediate vehicles finds an alternative cluster for the same topic name in its FIB. We propose that the intermediate vehicle has the intelligence to reconstruct the Cluster Joining Request Message by changing the tag to the alternative one on behalf of the source vehicle. The new cluster joining request is then forwarded to the new cluster head 2, which eventually accepts the request. The "*Acceptance Response Message*" contains the new tag and is returned back to the source vehicle. The source vehicle records the topic name the cluster tag in the Topic Cluster Table (TCT) for local maintenance. Each time when there is a new member joining a cluster, the cluster head needs to perform cluster head classification [11] to determine whether the new member can be classified as the new cluster head in replacing the current one.

D. Leave a Topic Based cluster

A vehicle can leave its participated cluster any time it wants by sending a "*Leaving Cluster Notice Message*" to the cluster header. The notice message shall include the topic name, the cluster tag and the vehicle's identifier, which can be routed based on the FIB maintained by the intermediate vehicles. After the cluster head receives the notice message, it removes the vehicle from the member list and updates the attributes related to the cluster, such as number of members, overall average load etc. The cluster head replies a confirmation to the requesting vehicle. The requesting vehicle will update its TCT by removing the corresponding topic name and cluster tag.

E. Dissolve a Topic Based Cluster

A cluster may be dissolved when the current cluster head does not want to be cluster head anymore, but there is no other vehicle in the cluster taking over the role. The cluster head firstly perform the cluster head candidate classification on all the members to determine whether any member can be the candidate of the new cluster head. Fig. 5 shows the decision tree for the cluster head candidate classification.

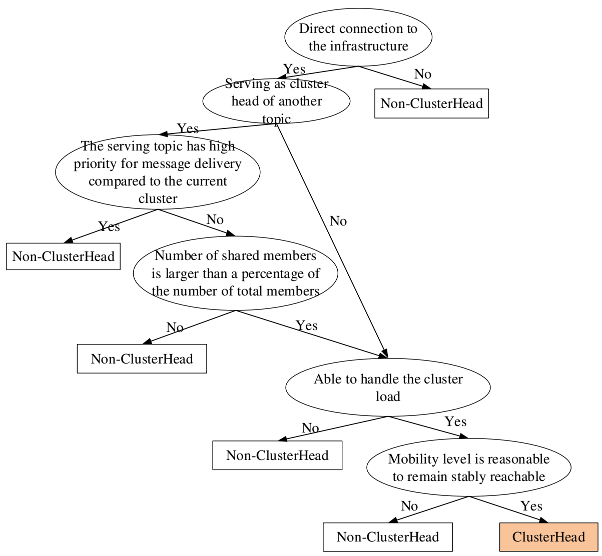

Fig. 5. Decision tree for cluster head candidate classification

The direct connection to the infrastructure through network interface(s) is mandatory to become a cluster head. Thus it is chosen as the root node in the decision tree. If a member vehicle does not satisfy this condition, then it is classified as Non-ClusterHead immediately.

If a member vehicle is already a cluster head for another topic (denoted as serving cluster in the following), and if this topic is of higher priority for message delivery compared to the current topic, then it is wise to classify it as Non-ClusterHead. The reason is that the vehicle cannot take advantage of aggregating the interest message or data message since the high priority topic message needs to be delivered immediately to the members. Otherwise, if the number of shared members is more than certain percentage of the total number of members in the serving cluster and the current cluster, the member vehicle can be chosen as the candidate cluster head for the current cluster. It can aggregate or concatenate the data requested by those shared members to save the controlling overhead in the intermediate vehicles. In order to become a candidate for the cluster head, the member vehicle is also required to be able to handle the cluster load and remain stably reachable to other members. The model does not consider the distance to the centroid as a factor. But this attribute can be used in sorting the cluster head candidacy.

After the cluster head candidate classification finished on all members, the candidates can be ordered for candidacy preference by the factors or combined factors, e.g. more number of shared members, higher capability in handling the cluster, lower mobility level, shorter distance to the centroid.

The cluster head will send the "*Cluster Head Replacement Request*" to the candidates in the order of preference. If one of the candidates accepts the request, then the cluster can be rescued. If none of the candidates accepts the request, then the cluster will be dissolved. The cluster head broadcasts the "*Cluster Dissolving Notice*" in the network. All the vehicles maintains the cluster in the FIB will remove its information. Meanwhile, all the cluster members will remove the topic and cluster tag from its TCT. The procedure is shown in Fig. 6.

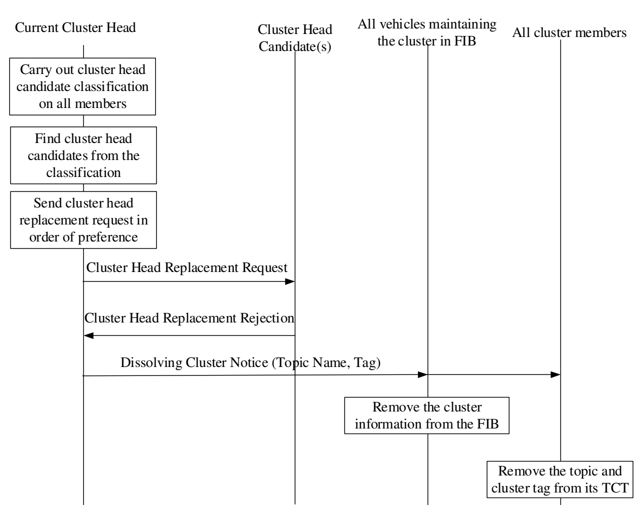

Fig. 6. Message flow of dissolving a cluster

F. *Split a Topic Based Cluster*

A cluster may be split into two or multiple clusters: (1) when there are too many members in the cluster such that the cluster head wants to make the cluster smaller, the split-out cluster(s) remains to be associated with the same topic; (2) The cluster head sees frequent requests on different sub-topics belonging to the topic which the cluster is currently associated with. For example, the entertainment topic can have the sub-topics, such as movie, TV show, music etc. The sub topic information may be embedded in the content name, or a field in the interest message.

For scenario (1), the procedure of splitting a cluster involves the following major steps:

- The current cluster head carries out the cluster head candidate classification on the members and finds the cluster head candidates for new split-out cluster.
- The current cluster head picks one candidate that confirms its willingness to be a cluster head.
- The current cluster head selects the members to be transferred to the new cluster with the same topic name.
- The new cluster head takes over the selected members and notify the members to update the cluster tag in their TCT. The new cluster head broadcasts cluster beacons in the network.

For scenario (2), the procedure of splitting a cluster involves the following major steps:

- The current cluster head collect statistics on the potential members that would be in the sub-topic clusters. The members could be overlapped.
- The current cluster head carries out the cluster head candidate classification on the members and finds the cluster head candidates for the new clusters for the different sub-topics.
- The current cluster head picks one candidate that confirms its willingness to be a cluster head for each sub-topic.
- The potential members for each sub-topic are transferred to the cluster.
- The new cluster head takes over the selected members and notify the members to update the cluster tag in their TCT. The new cluster head broadcasts cluster beacons in the network. The old cluster may still exist, which does not need to be dissolved.

III. PERFORMANCE DISCUSSION

The proposed topic based clustering allows the cluster members to retrieve information with the help of the cluster header for the same topic. Thus the interests for the content with the same topic name are all forwarded to the same cluster header that the requesting vehicles are associated with. On the other hand, the intermediate vehicles only need to maintain the topic names in the FIB for interest forwarding, instead of prefixes of the content names.

Fig. 7 shows FIB size ratio (which is defined as the ratio of the FIB sizes in the legacy NDN approach and in the proposed mechanism) linearly increases with the average number of name prefixes in each topic, which indicates that the more number of name prefixes associated with each topic, the more dramatically the FIB size could be reduced by the proposed clustering mechanism. Although the above observations seem to be simple and straightforward, they actually point a direction other than performing aggregation on the prefixes themselves [13] to solve one of the most challenging problems lingered around in NDN networks, which is how to prevent FIB explosion.

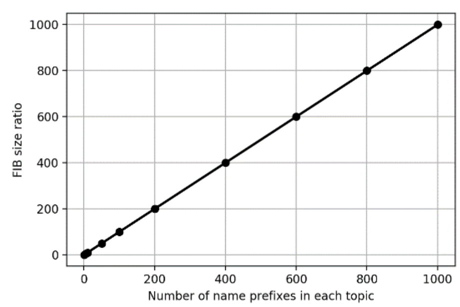

Fig. 7. FIB size ratio versus number of name prefixes in each topic

In the legacy NDN implementation, the interest message could be restrained by one of the en-route NDN routers if the router finds that the interests requesting the same content name have been forwarded before and are waiting in the PIT for the content to be returned. However, the same interests will be forwarded in the network and the same content would be returned multiple times to different requesters if the interests take different paths towards the producer(s) (it is likely there are surrogate servers deployed by CDN providers) without any intersecting NDN router. With the proposed topic based clustering, the interests for the same topic will be forwarded to the cluster head, the same content would not be transported multiple times in order to reach the different requesters. As shown in an example in Fig. 8, all the cars are interested in retrieving the same content. We compare the procedures that will be taken with and without the proposed topic based clustering.

Without the proposed topic based clustering, the car 1 and car 7 sends a separate interest to the access point 1, the interest that arrives later will not be forwarded further. Similarly, only one interest from the car 6 and car 2 will be forwarded by the base station 1, and only one interest from the car 3, car 4, and car 5 will be forwarded by the access point 2. The NDN protocol already does good job in preventing the interests for the same content from propagating in the network and increasing the network bandwidth usage and the content server' load. However, the three interests which take different paths towards the producer still will travel all the way till they reach an intersecting NDN router of the three paths. After the content is returned to the intersecting router, the same content will be transported along the three paths to the different groups of requesters, which wastes the network bandwidth between the intersecting router and the edge nodes (i.e. the access point 1, base station 1, and access point 2).

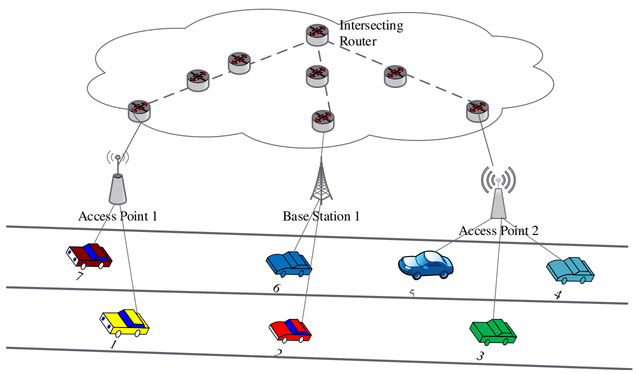

Fig. 8. Example topology

With the proposed topic based clustering, if all the cars join the same cluster with one of them as the cluster head, then the interests will be aggregated at the cluster head. Only one copy of the content will be transported in the network until it reaches the cluster head, which in turn will deliver the content to each car sharing the same interest.

In summary, the network bandwidth saving depends on the number of paths along which the interests will travel towards the producer, how far the intersecting node of those paths is away from the edge nodes if there is an intersecting router, and how large the content size is. The network bandwidth saving ratio of the proposed topic based clustering can be represented in the equation (1), in which we only consider the network bandwidth consumption in transporting the content (the interest message size is relatively small, which can be neglected).

$$bandwidthSavingRatio = \frac{total\_hop}{single\_path\_hop} - 1 \quad (1)$$

In the equation, the variable $total\_hop$ denotes the total number of hops in the tree with the farthermost intersecting node as the root in the legacy NDN approach. It is likely that the farthermost intersecting node could be the content producer itself, if the paths along which the interests travel do not overlap. The variable $single\_path\_hop$ denotes the number of hops between the cluster head and the content producer in the proposed topic based clustering scheme, which makes sure the intersecting node is one of the edge nodes. With more number of paths involved in the interest forwarding, $total\_hop$ would have higher probability to become larger in the legacy NDN approach. Thus the proposed topic based clustering scheme consumes significantly less bandwidth by avoiding transporting duplicate content copies in the network.

IV. CONCLUSION

The paper proposes a topic based clustering mechanism for information retrieval and sharing among connected cars. The detailed procedures and messages of cluster setup, detection, and participation are illustrated. On the other hand, the proper adjustments to an existing cluster, such as dissolution, splitting are also discussed in the paper. The NDN transport protocols are perfect for the information retrieval with the proposed topic based clustering, which successfully prevents the FIB size from exploding. The network bandwidth consumption is dramatically reduced due to the interest aggregation at the cluster head compared to the legacy NDN approach.

REFERENCES

[1] C. Cooper, D. Franklin, M. Ros, F. Safaei, and M. Abolhasan, “A Comparative Survey of VANET Clustering Techniques,” IEEE Communications Surveys and Tutorials, Vol. 19, No. 1, first quarter 2017.

[2] L. Dong, G. Wang, “INADS: In-network aggregation and distribution of IoT data subscription in ICN, 2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW).

[3] L. Dong, G. Wang, Enhanced In-Network Capabilities of Information-Centric Networks for Emerging IoT Applications, 2016 IEEE International Conference on Physical and Social Computing (CPSCom).

[4] L. Dong, Enhanced IoT Data Acquisition in Information Centric Networks, 2016 International Wireless Internet Conference.

[5] L. Zhang, D. Estrin, J. Burke, V. Jacobson and etc., "Named Data Networking (NDN) Project, NDN Technical Report NDN-0001."

[6] Ryuji Wakikawa et al.,"NDN Applicability to V2V and V2R Networks," in Proc. CCNxCon2012, September 2012.

[7] “Standard Specification for Telecommunications and Information Exchange Between Roadside and Vehicle Systems - 5 GHz Band Dedicated Short Range Communications (DSRC) Medium Access Control (MAC) and Physical Layer (PHY)Specifications”, ASTM DSRC STD E2313-02, 2002.

[8] IEEE 1609, WAVE: Wireless Access in Vehicular Environments.

[9] “IEEE Std. 802.11-2007, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications,” IEEE Std. 802.11, 2007.

[10] “IEEE P802.11p/D3.0, Draft Amendment for Wireless Access in Vehicular Environments (WAVE),” July 2007.

[11] L. Dong, R. Li, Information Exchange Oriented Clustering for Collaborative Vehicular System, WOCC, 2018.

[12] R. Li, A. Clemm, U. Chunduri, L. Dong, K. Makhijani, A New Framework and Protocol for Future Networking Applications, in submission to ACM Sigcomm NEAT workshop, 2018.

[13] H. Dai, J. Lu, Y. Wang, and B. Liu, “A Two-layer Intra-domain Routing Scheme for Named Data Networking,” In Proceeding of IEEE Global Communications Conference (GLOBECOM), 2012, pp. 2815–2820.