Vehicular Knowledge Networking and Mobility-Aware Smart Knowledge Placement

Seyhan Ucar*, Takamasa Higuchi*, Chang-Heng Wang*, Duncan Deveaux†, Onur Altintas* and Jérôme Härri†
*InfoTech Labs, Toyota Motor North America R&D, Mountain View, CA, USA
†EURECOM - Communication Systems Department, Sophia-Antipolis, France
{seyhan.ucar, takamasa.higuchi, chang-heng.wang, onur.altintas}@toyota.com
{duncan.deveaux, jerome.haerri}@eurecom.fr

***Abstract*—It is estimated that the data volume between connected vehicles and edge/cloud server(s) will be about 100 petabytes per month by 2025. The networking framework we have, on the other hand, is the existing cellular network in which the most connected vehicles function today. However, such a network suffers from several issues and may not work under this predicted data demand. To address such a dilemma, a new paradigm, Vehicular Knowledge Networking (VKN), is recently introduced. In VKN, the data is transformed into knowledge and it is distributed with various lifetimes/relevance. To benefit from the knowledge, on the other hand, it should be placed intelligently such that a high number of vehicles can access and consume it. In this paper, we tackle this issue and propose mobility-aware smart knowledge placement. In the proposed method, vehicle mobility is analyzed to measure the centrality degree of a region. The computed centrality degrees are then further analyzed to identify the most central zones. The knowledge is placed on these zones to increase availability. We demonstrate the benefits of the proposed method through a simulation. Our preliminary result has shown that the mobility-aware smart knowledge placement makes knowledge accessible from vehicles over short range communication. Through such short-range availability of knowledge, vehicles can use the free spectrum to download it which decreases the cellular communication cost significantly.**

I. Introduction

Today, vehicles have a rich set of resources that includes sensing, communication, and computation. Such increasing resource capabilities pave the way for vehicles to be the building blocks of future smart cities and intelligent applications [1]. Through diverse communication and sensing technologies, vehicles become connected, and they not only sense the environment but also share their observation with edge/cloud server(s) to benefit from unique mobile services. According to Automotive Edge Computing Consortium (AECC), the data volume transmitted back and forth between connected vehicles and edge/cloud will be about 100 petabytes per month by 2025 [2]. The more connected vehicles get involved, the more data is generated. The networking framework we have, on the other hand, is the existing cellular network in which the most connected vehicles function today. However, such a network suffers from several limitations and may not scale well under this predicted data demand. Multiple vehicles under the same traffic incident may execute the same computation in parallel to interpret similar sensor measurements. Delivering all sensor data in such a situation is highly redundant or sometimes unnecessary. Having more networking infrastructures with rich resources seems to be the solution at first glance. However, it may not be feasible since the increased data may consume the available resources rapidly.

In light of these facts, an urgent need for a unique solution and a brand-new network architecture would rise. For example, ETSI standardized the information sharing through Cooperative Awareness Message (CAM) and Local Dynamic Maps (LDM) [3]. In this standard, the information is defined as a group of one or more pieces of raw data that are processed to be meaningful. However, there still exists redundant computation when vehicles share a set of similar information (e.g., a group of vehicles is calculating the risk of an intersection based on the speed and position information of vehicles in the vicinity) [4]. To address this problem, a new paradigm, VKN, is recently introduced in [5]. In VKN, a mechanism is designed to transform data into searchable knowledge and it is distributed with various lifetimes and relevance. There is a hierarchy among data, information, and knowledge as demonstrated in Figure-1. Knowledge is created through analysis of multiple instances of information and it is a fact or a belief that represents the hidden relationship among them.

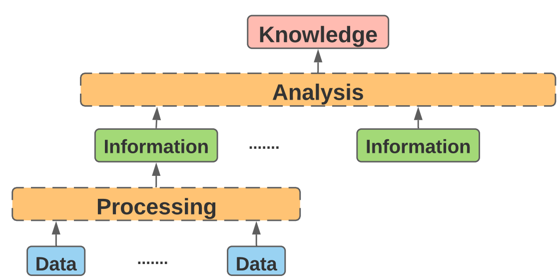

Fig. 1: The hierarchy in data, information and knowledge

The creation of knowledge, on the other hand, is not cheap. It may require computationally hungry algorithms and a set of information and/or data that are potentially from multiple sources (e.g., vehicles). To maximize the benefit, knowledge should be placed intelligently when it is created. How to place data and/or contents in mobile environments is one of the well-explored topics in existing literature [12]. For example, possible trajectories of the vehicle are determined and data/content popularity is estimated. The data/contents

TABLE I: Related Work on Knowledge Networking and Knowledge Oriented Applications

Ref.	Purpose	Knowledge	Method(s)	Application
[6]	Resource Utilization	Network Selection	Optimization	Access Network Selection
[7]	Network Management	Optimal Links and Paths	Optimization	UAV Network Management
[8]	Caching Utilization	Content Election	Machine Learning	Edge Caching
[9]	Parking Lot Monitoring	Parking Availability	Machine Learning	End-to-End Parking
[10]	Understanding Human Mobility	Socioeconomic Activities	Clustering	Urban Area Planning
[11]	Vehicular Risk Assessment	Maneuver Conflict Zones	Clustering	Vehicular Path Planning

are prefetched according to this estimation to minimize the delay and maximize the retrieval rate in [13], [14]. In addition to these IP network-based solutions, another body of research focuses on Named Data Networking (NDN) in which the existing placement and dissemination strategies are well discussed in [15]. However, both IP and NDN solutions rely on an accurate prediction in which inaccurate estimation degrades their performance and wastes all of the placement effort. Knowledge placement, on the other hand, is in a different setting in which the key goal is to increase the availability and make a high number of vehicles benefit from it. When knowledge is created, it is mutable compared to information and should stay alive, being stored, updated, and forwarded to nearby entities. In this paper, we tackle this issue and propose mobility-aware smart knowledge placement. The vehicles' mobility characteristics are leveraged to mine the centrality degrees of a region. The computed centrality degrees are then further analyzed to identify the most central zones. The connected vehicles in the most central zones are instructed to form Vehicular Micro Clouds (VMCs) [16], and VMCs behave as a virtual edge server to store, update and forward the knowledge. Through the virtual edge servers, knowledge is accessible in short-range. Vehicles can use the free spectrum (e.g., Vehicle-to-Vehicle (V2V) communication) to download the knowledge rather than requesting it from an edge/cloud server(s) which reduces the cellular communication cost significantly.

The rest of the paper is organized as follows. Section II presents the recent effort in VKN applications. Section III describes the system model of mobility-aware smart knowledge placement. Section IV presents the proposed method. Section V provides the performance evaluation of proposed method via simulations. Finally, concluding remarks and future works are given in Section VI.

II. Related Work

Table I summarizes the recent research with some key features in knowledge oriented applications. Optimization methods are applied to intelligently select the network interface for better resource utilization in [6]. A monitoring platform is integrated with Software Defined Networking to find the optimal links and paths for Unmanned Aerial Vehicles in [7]. Content election strategies are examined to proactively decide which content to cache at the edge in [8]. [9] studied the parking availability prediction in parking lot monitoring system to provide end-to-end parking. Human mobility characteristics are mined to identify the socioeconomic activities in [10]. From the existing literature, we observed that there exists an increasing interest in knowledge oriented applications. Knowledge is created and utilized in management and feedback systems for better resource utilization and achieving user's objectives, respectively. Recently, we looked at the knowledge networking from vehicle perspectives and proposed vehicular knowledge creation and knowledge networking in [11]. In [11], vehicular maneuver conflicts (e.g., merging, crossing conflicts) are leveraged to identify the conflicting zones in which the maneuver conflicts occur the most.

On the other hand, there exist some recent works in the literature that aim to place the vehicular content for better utilization. The main goal of these works is to intelligently cache the contents and route them to improve the overall data downloading performance. [12] highlights the key principles in the content caching and placement strategies adopted in the existing literature. However, these proposed methods are not directly applicable to vehicular knowledge networking applications and have the following drawbacks. First, existing placement strategies select the contents to be elected/evicted according to some predefined criteria (e.g., popularity). However, a change in the environment could leave the placement strategy unable to respond adequately as its static election/eviction criteria become obsolete (e.g., non-recurring traffic congestion and mitigation strategy to improve the traffic efficiency). Second, although there exist some other methods that use prediction mechanisms to determine what content and where to place, inaccurate estimation degrades their performance and wastes all of the placement effort. The key goal of the knowledge placement, on the other hand, is to feed necessary part of knowledge to relevant vehicles beforehand so they can benefit from it. Through smart placement, knowledge should stay alive, being stored, updated and forwarded to other nearby vehicles.

III. System Model

Figure 2 shows the overview of the proposed mobility-aware smart knowledge placement system. We assume that connected vehicles have the V2V communication module in addition to Vehicle-to-Cloud (V2C) cellular communication functionality. Connected vehicles upload their position, speed information to the edge/cloud server and the edge/cloud server runs mobility analytics to identify the most central zones. In these most central zones, connected vehicles are instructed to form a VMC. VMC then keeps the knowledge alive through VMC-to-Cloud communication and knowledge is updated via two-way communication with edge/cloud server. VMC behaves as a collaborative data storage in which vehicles approaching the most central zones can obtain the kept knowledge over V2V networks (detailed in Section IV).

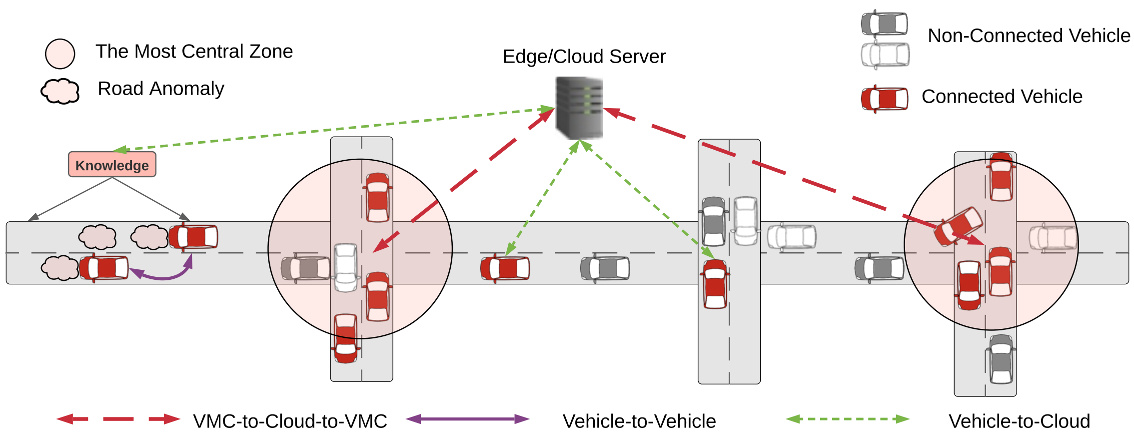

Fig. 2: Overview of the Mobility-Aware Smart Knowledge Placement

IV. Mobility-Aware Smart Knowledge Placement

The features of the proposed mobility-aware smart knowledge placement method are as follows.

- It analyzes vehicle trajectory data to measure the centrality degree of a region. The centrality degree is an important metric, which identifies a central region in a given road network.
- It not only mines the centrality degree of the region but also clusters computed centrality degrees to identify the most central zones. Connected vehicles in these zones share the common vicinity and can collaborate on tasks.
- It forms VMCs on these most central zones, and VMC is responsible for keeping the knowledge up-to-date via VMC-to-Cloud communication. VMC then shares knowledge with other nearby vehicles over the V2V network that significantly decreases the cellular cost.

Figure 2 illustrates the use case where the mobility-aware smart knowledge placement is investigated. Vehicles are traveling from right to left and two connected vehicles at the left-most part of the road analyze their sensor readings and conclude that the road has an anomaly (e.g., pothole, traffic incident, black ice, etc.). Such knowledge is very valuable and needs to be kept alive. Connected vehicles passing through this road section need to be guided and they should update the knowledge regarding the current state of the anomaly. According to the current status, knowledge should be delivered to other nearby vehicles immediately. Otherwise, it may jeopardize the safety of other vehicles. When such knowledge is created, it should be placed intelligently, so that the other vehicles get prepared prior to hitting the abnormal part of the road. To solve this issue, we propose to mine the centrality degree of regions and place the knowledge in the most central zones. Knowledge in these zones stays alive through the edge/cloud server communication and it is consumed by the other nearby vehicles.

When the knowledge is created, the edge/cloud server runs the steps presented in Figure 3 to place the knowledge intelligently. We assume that connected vehicles upload their

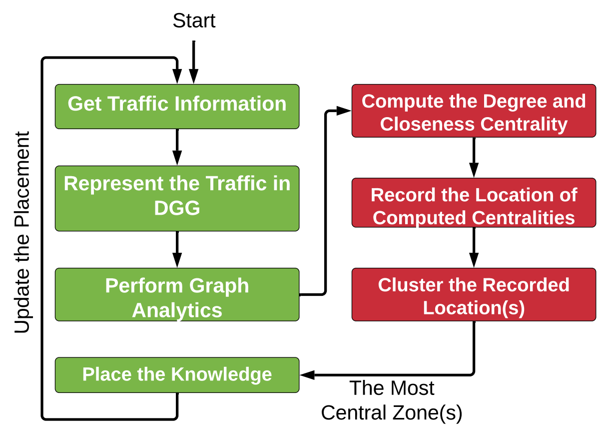

Fig. 3: Flowchart of Smart Knowledge Placement

data to the edge/cloud server periodically and/or when they are instructed. The steps of smart knowledge placement are explained in detail next.

Get Traffic Information: Knowledge placement starts with getting the traffic information of both V2V and V2C connected vehicles. The edge/cloud server instructs connected vehicles to share their speed and location information. Vehicles with V2C communication functionality can upload their information to the edge/cloud server directly. Other V2V communication-only vehicles, on the other hand, share their data with V2C enabled vehicles. V2C communication-enabled vehicles act as a gateway to link the data to the edge/cloud server. The overall traffic is then represented at each time instance with N road agents that are capable of V2V and V2C communications.

Represent the Traffic in DGG: After getting the traffic information, Dynamic Geometric Graph (DGG) [17] is constructed. DGG is an undirected graph with a set of vertices (Vs) and a set of edges (Es). In this paper, DGG illustrates the state of the network at each time instance with N connected vehicles where vehicles are denoted as Vs and the distance between each pair of Vs is represented as an E. Considering that vehicular network is not static, analysis of DGG is important as it implicitly contains the knowledge about how vehicles move over time and which regions are more central compared to others. We leverage such topological knowledge to place the knowledge itself so a high number of vehicles can benefit from it.

Perform Graph Analytics: After the construction of DGG, movements of connected vehicles are mined to explore the centrality degree of regions. Centrality generally indicates characteristics of a central node in a given graph. In this paper, we use the degree and closeness centrality of vehicles to measure the centrality degree of a region. The degree centrality for a vehicle V is the fraction of nodes it is connected to. Closeness centrality ($C_i(t)$), on the other hand, stands for the reciprocal sum of the shortest path ($distance(u_i, u_j)$) from node V to all other nodes as shown in Equation 1.

$$C_i(t) = \frac{1}{\sum_{j \neq i} distance(u_i, u_j)} \quad (1)$$

The edge/cloud server mines the DGG and computes the centrality degrees for each vehicle at each time step. The vehicles that have the highest degree and/or the lowest closeness centrality are defined as the most central node, and the locations of the most central nodes are recorded. Following that, these locations are further grouped to identify the most central zones. Via mining the DGG at each time step, the edge/cloud server quantifies how long it will take for knowledge to spread from a given location to other vehicles in the network. We propose to leverage such a piece of knowledge while placing the knowledge itself.

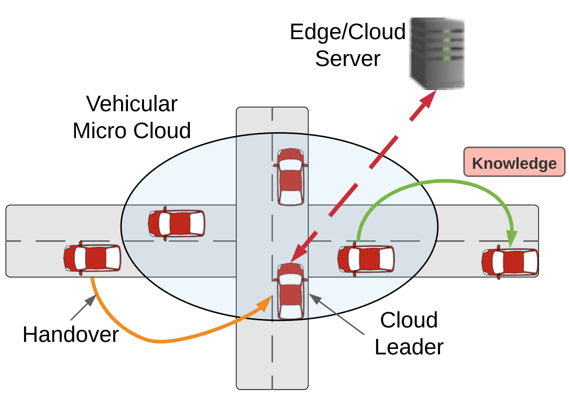

Fig. 4: Vehicular Micro Cloud (VMC)

Place The Knowledge: In the last step of the smart knowledge placement, connected vehicles at the most central zones are instructed to form a VMC as depicted in Figure 4. VMCs behave as a virtual edge server, and connected vehicles collaborate to keep the knowledge up-to-date at these zones through VMC-to-Cloud communication. The feasibility of such collaborative data storage by a VMC has recently been investigated in [18]. The basic idea of collaborative data storage is: vehicles hand over their data contents to other vehicles (e.g., vehicles closest to the center of VMC) before they leave the VMC. A cloud leader (e.g., vehicles with rich resources) is being elected to coordinate this process. We utilize such collaborative data storage service to make the knowledge accessible in a short-range. Vehicles approaching these zones can use the free spectrum (i.e. direct V2X) to download knowledge rather than requesting it from an edge/cloud server(s) individually.

V. Performance Evaluation

We have evaluated the performance of mobility-aware smart knowledge placement through simulation experiments. The simulations are performed in the Simulation of Urban Mobility (SUMO) [19]. SUMO is an open-source and discrete-time traffic simulator that is capable of simulating the micro-behavior of individual vehicles. For the network evaluation, we heuristically assume that vehicles use cellular (e.g., 4G) and IEEE 802.11p for V2C and V2V communication, respectively. The overall networking analysis is performed by counting the number of transmissions with each communication interface. To check the feasibility of smart knowledge placement, we model the El Camino Real road network located near Mountain View, California using real map data to emulate realistic traffic conditions. Table II lists other simulation parameters.

TABLE II: Simulation Parameters

Parameter	Value
Min/Max Speed	0/35 m/s
Number of Vehicles	300
Simulation Time	300 s
IEEE 802.11p Range	300 m
Most Central Zone Thresholds	10, 20, 30
Connected Vehicle Upload Period	1 sec

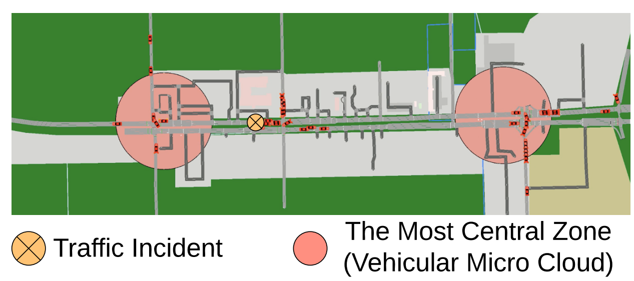

Fig. 5: El Camino Real Road Network, Mountain View, CA

The goal of the simulation is to compare the feasibility of proposed mobility-aware smart knowledge placement, denoted by *Proposed*, to the traditional individual request-based cellular network approach, denoted by *Baseline*. Figure 5 shows the use case in which a random traffic incident is placed on the left-most lane causing the lane closure for 8 minutes. We assume that connected vehicles encountering the traffic incident analyze their sensor readings and lane level traffic information to create the knowledge. The knowledge here is the mitigation strategy of such non-recurring traffic congestion that helps connected vehicles to pass the incident smoothly. Or, it can be the learned lane change suggestion and/or inferred game-theoretic lane-changing model.

Figure 6 demonstrates the average distance of all connected vehicles when they want to access placed knowledge in different thresholds under *Proposed* method. The threshold here refers to the number of times that the most central nodes (e.g., vehicle) appear in a given region. The simulation is warmed-up until simulation time reaches $t = 40s$, and the knowledge about the traffic incident is created at $t = 40s$. The edge/cloud server first randomly selects a subset of region and VMCs are formed in these zones to host the knowledge. However, V2V communication-enabled vehicles cannot access such knowledge because the average distance to access knowledge is higher than the V2V communication range (e.g., 300 meters). As connected vehicles upload their information, the edge/cloud server infers how the vehicle moves over time and which regions are more central compared to others. The edge/cloud server leverages such knowledge to refine initial placement and place the knowledge in the most central zones according to the vehicle's mobility. Such mobility-aware smart placement, on the other hand, makes the knowledge accessible in a short range. The knowledge is kept up-to-date over VMC-to-Cloud communication and connected vehicles approaching these zones can benefit from such knowledge before they reach the core area of the traffic incident.

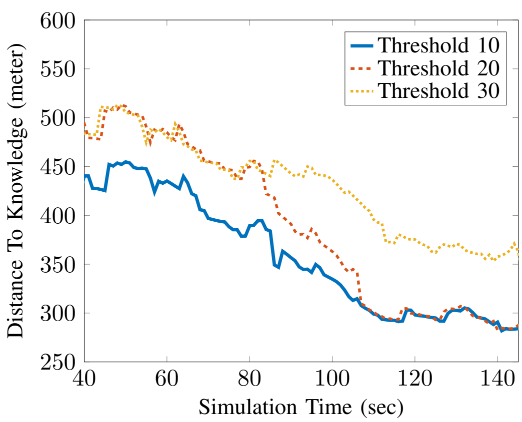

Fig. 6: Average Distance to Reach the Knowledge

When we consider the rate of decrease and the most central zone threshold in Figure 6, we observe that there is a tradeoff. The small threshold is good when we want to disseminate the knowledge quickly. However, it has a drawback in terms of management as a slightly larger number of VMCs need to be formed and managed. A straightforward extension could be to analyze the traffic incident and come up with dynamic thresholds that optimize the knowledge delivery and management accordingly.

Figure 7 illustrates the recorded location of the most central nodes (e.g. vehicles) (**X**) and the inferred most central zones (**O**) when the most central zone threshold (i.e., the number of times that the most central nodes (e.g. vehicle) appear in a given region) is 30. The edge/cloud server mines the DGG and records the location of the central nodes (e.g., the highest degree and the lowest closeness centrality). These locations are grouped to identify the most central zones. There are three central zones, and these zones shifted according to vehicle mobility around the traffic incident. The edge/cloud server instructs connected vehicles in these zones to form a VMC. VMC behaves as a virtual edge server, and the vehicles collaborate to keep the knowledge up-to-date through VMC-to-Cloud-to-VMC communication.

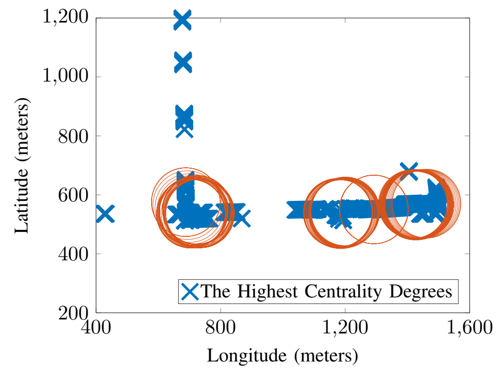

Fig. 7: The Change of The Most Central Zones

Figure 8 depicts the evaluated communication cost of *Baseline* and *Proposed* methods where the blue dashed bars compare the cost savings in the cellular portion. In *Baseline*, the inferred knowledge about the traffic incident is delivered to connected vehicles over a cellular network which is costly and could create a bottleneck on the edge/cloud server. The *Proposed* method, on the other hand, measures the centrality degrees of the region and selects a subset of regions (most central zone threshold equal or larger than 20) as the most central zones. The connected vehicles in these zones collaborate to not only keep the knowledge up-to-date through VMC-to-Cloud communications but also make it available in short range over vehicular networks. Therefore, connected vehicles approaching these zones download the knowledge over direct V2X networks which significantly decreases the cellular communication cost.

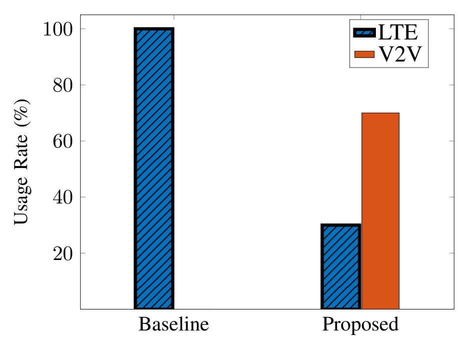

Fig. 8: Communication Cost

To ensure the feasibility of knowledge delivery in the short range, we analyze the knowledge disruption rate in the most central zones under different penetration rates of V2V-enabled vehicles. Knowledge disruption rate is defined as the ratio of the time duration that the number of V2V enabled connected vehicles drops to zero, causing temporary loss of knowledge which result in re-downloading the knowledge from the edge/cloud. Figure 9 shows the disruption rate analysis of *Baseline* and *Proposed* methods under different penetration rates of V2V-enabled vehicles. We observe that there is no knowledge disruption in *Baseline* where each connected vehicle relies on V2C communication and the knowledge is delivered over a cellular network. In the *Proposed* method, on the other hand, the disruptions occur more frequently under the lower penetration rates of V2V-enabled nodes. In such cases, VMC suffers from keeping the knowledge tied to the most central zone due to a lack of V2V connected vehicles. When V2V-enabled vehicle penetration rate reaches a certain degree of threshold (e.g., 30%), the VMC behaves as a regional distributed storage in which the knowledge is kept alive. Whenever the knowledge is updated on the edge/cloud side, it is delivered to VMC via VMC-to-Cloud communication. The other nearby connected vehicles can download the up-to-date knowledge over V2V based vehicular networks.

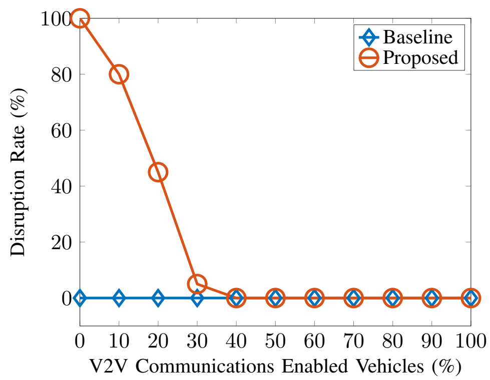

Fig. 9: Disruption Rate Analysis

VI. Conclusion and Future Work

In this paper, we propose a mobility-aware smart knowledge placement method in which mobility analysis is performed to measure the centrality degree of a region. The computed centrality degrees are then further analyzed to identify the most central zones. We propose to place the created knowledge in these zones to increase availability. The connected vehicles in these most central zones form a VMC and VMC is responsible for keeping the knowledge alive via VMC-to-Cloud communication. Through such a collaborative approach, knowledge is reachable in short ranges where vehicles can use the free spectrum to access it rather than requesting from the edge/cloud server individually. Extensive simulations under realistic traffic conditions demonstrate that the proposed method decreases the cellular cost significantly by making knowledge accessible in a short range under a certain degree of V2V communication penetration rate.

The promising results open a specific line of further works that should extend the validation of the proposed method by the sensitivity test on the performance under different road settings with various vehicle resource models and vehicle densities. We want to extend the proposed knowledge placement technique with a new knowledge querying system in which the placed knowledge is searched with given tags. Through such an extension, connected vehicles could submit knowledge queries and even they offload the knowledge creation to remote areas before they reach the core area.

References

[1] O. Altintas, F. Dressler, F. Hagenauer, M. Matsumoto, M. Sepulcre, and C. Sommer, "Making cars a main ICT resource in smart cities," in *IEEE Conference on Computer Communications Workshops*, April 2015.

[2] "Driving Data to the Edge: The Challenge of Data Traffic Distribution, Automotive Edge Computing Consortium (AECC)," https://aecc.org/driving-data-to-deliver-addressing-the-connected-vehicle-data-challenge/.

[3] *ETSI EN 302 895 V1.1.1 : Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Local Dynamic Map (LDM)*, 09 2014.

[4] A. Pierson, W. Schwarting, S. Karaman, and D. Rus, "Navigating Congested Environments with Risk Level Sets," in *IEEE International Conference on Robotics and Automation (ICRA)*, May 2018.

[5] D. Deveaux, T. Higuchi, S. Uçar, J. Härri, and O. Altintas, "A Definition and Framework for Vehicular Knowledge Networking: An Application of Knowledge-Centric Networking," *IEEE Vehicular Technology Magazine*, 2021.

[6] X. Li, Q. Huang, and D. Wu, "A repeated stochastic game approach for strategic network selection in heterogeneous networks," in *IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, April 2018.

[7] X. Zhang, H. Wang, and H. Zhao, "An SDN framework for UAV backbone network towards knowledge centric networking," in *IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, April 2018.

[8] H. Hao, C. Xu, M. Wang, H. Xie, Y. Liu, and D. O. Wu, "Knowledge-centric proactive edge caching over mobile content distribution network," in *IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, April 2018.

[9] D. Wu, Z. Li, J. Wang, Y. Zheng, M. Li, and Q. Huang, "Vision and Challenges for Knowledge Centric Networking," *IEEE Wireless Communications*, August 2019.

[10] J. Yang, J. Cao, R. He, and L. Zhang, "A unified clustering approach for identifying functional zones in suburban and urban areas," in *IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, April 2018.

[11] S. Ucar, T. Higuchi, C.-H. Wang, D. Deveaux, J. Härri, and O. Altintas, "Vehicular knowledge networking and application to risk reasoning," in *Proceedings of the Twenty-First International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing*, 2020.

[12] A. Kulkarni and A. Seetharam, "Model and machine learning based caching and routing algorithms for cache-enabled networks," *arXiv preprint arXiv:2004.06787*, 2020.

[13] D. Grewe, S. Schildt, M. Wagner, and H. Frey, "ADePt: Adaptive Distributed Content Prefetching for Information-Centric Connected Vehicles," in *IEEE 87th Vehicular Technology Conference (VTC Spring)*, 2018.

[14] Z. Hu, Z. Zheng, T. Wang, L. Song, and X. Li, "Game theoretic approaches for wireless proactive caching," *IEEE Communications Magazine*, 2016.

[15] Y. Meng, M. A. Naeem, R. Ali, and B. Kim, "EHCP: An Efficient Hybrid Content Placement Strategy in Named Data Network Caching," *IEEE Access*, 2019.

[16] T. Higuchi, J. Joy, F. Dressler, M. Gerla, and O. Altintas, "On the feasibility of vehicular micro clouds," in *IEEE Vehicular Networking Conference (VNC)*, 2017.

[17] G. Pallis, D. Katsaros, M. D. Dikaiakos, N. Loulloudes, and L. Tassiulas, "On the structure and evolution of vehicular networks," in *IEEE International Symposium on Modeling, Analysis Simulation of Computer and Telecommunication Systems*, 2009.

[18] S. Ucar, T. Higuchi, and O. Altintas, "Collaborative Data Storage by a Vehicular Micro Cloud," in *IEEE Vehicular Networking Conference (VNC)*. IEEE, 2019.

[19] "SUMO," http://sumo.sourceforge.net/.