IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received April 14, 2018, accepted May 13, 2018, date of publication May 21, 2018, date of current version June 19, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2837870

SDN-Enabled Social-Aware Clustering in 5G-VANET Systems

WEIJING QI[1], QINGYANG SONG[1], (Senior Member, IEEE), XIAOJIE WANG[2], (Student Member, IEEE), LEI GUO[1], (Member, IEEE), AND ZHAOLONG NING[2], (Member, IEEE)

[1]School of Computer Science and Engineering, Northeastern University, Shenyang 110819, China
[2]School of Software, Dalian University of Technology, Dalian 116024, China

Corresponding authors: Qingyang Song (songqingyang@cse.neu.edu.cn) and Zhaolong Ning (zhaolongning@dlut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grants 61401082, 61501104, 61501105, 61775033, and 61771120, and in part by the Fundamental Research Funds for the Central Universities under Grants N161606001, N150401002, and N161608001.

ABSTRACT Nowadays, vehicular applications for traffic safety, traffic efficiency, and entertainment have put forward higher requirements to vehicular communication systems. The fifth-generation (5G) cellular networks are potential to provide high-capacity low-latency communications for vehicles in highly mobile environments. In addition, the IEEE 802.11p-based dedicated short-range communication technology is proposed to form vehicular ad hoc networks (VANETs). Integrating cluster-based VANETs with 5G cellular networks is beneficial for saving scarce spectrum resources, preventing network congestion, and reducing packet loss. However, it is a challenging problem to find an effective clustering algorithm that has high stability adapting to dynamic VANETs. Considering the advantages of software-defined networking (SDN), in this paper, we present an SDN-enabled social-aware clustering algorithm in the 5G-VANET system, which exploits a social pattern (i.e., vehicles' future routes) prediction model in order to enhance the stability of clusters. Each vehicle's movement is modeled as a discrete time-homogeneous semi-Markov model, where the state transition probability and sojourn time probability distribution are inputs and each vehicle's social pattern is output. The predicted social patterns are subsequently used to create clusters so that vehicles in the same cluster tend to share the same routes. Cluster heads are selected based on the metrics of inter-vehicle distance, relative speed, and vehicle attributes. We evaluate our algorithm and the results show that it performs better in terms of cluster lifetime and clustering overhead compared with traditional clustering algorithms.

INDEX TERMS 5G-VANET, software-defined networking, social pattern prediction, clustering.

I. INTRODUCTION

With the rapid development of automotive industries, modern vehicles are able to feel their surroundings through onboard sensors such as radar, lidar, GPS, laser scanners etc., and hence generate a huge amount of data. By coupling transportation systems with wireless communications, Intelligent Transportation Systems (ITSs) support vehicles to share the vehicle data with each other for improving traffic safety and efficiency in both highway and urban environments. In addition, people in cars have an increasing demand for mobile Internet access in business and entertainment applications [1], which can be extended to real-time interactions such as video conferencing or multiplayer gaming. Further, the future trend of autonomous driving and vehicular cloud-based services [2] presents higher requirement to the vehicular communication system's performance in terms of latency, reliability, capacity, etc.. The Fifth-Generation (5G) cellular networks, which can provide high-capacity low-latency communications for vehicles in highly mobile environments, are potential to meet the application requirements in ITSs. Nevertheless, it is still impractical for future 5G networks to handle the growing in-car data alone owing to its scare spectrum resources. Imagining that if vehicles in dense urban areas all build connections with a Base Station (BS), traffic congestions and packet losses will easily occur, besides high signaling overhead and expensive channel resource.

On the other hand, in order to support sharing the large volumes of in-vehicle data, industry, and academia have been seeking a series of short-range radio technologies specifically for inter-vehicle communications. A promising approach is to utilize the IEEE 802.11*p*-based Dedicated Short-Range Communication (DSRC) to form Vehicular Ad hoc Networks (VANETs). Occupying a spectrum of 75 MHz at 5.9 GHz, communication range around 300 meters and data rate ranging from 6 to 27 Mbps are provided in VANETs. To overcome the above-mentioned problems brought by using 5G network to carry vehicular data alone, one vehicle is selected as a mobile gateway from a group of vehicles for connecting with BS, thus the other vehicles in this group send and receive data through this gateway using IEEE 802.11*p* interface. This group of vehicles is a cluster, where the gateway is a Cluster Head (CH) and the other vehicles are Cluster Members (CMs). Jia *et al.* [3] analyzed the influence of Floating Car Data (FCD) over other LTE traffic. It is proved that FCD transmission through a CH decreases the blocking rate of traditional cellular traffic compared with no cluster scenario. Collectedly, it brings the following benefits to integrate cluster-based VANETs with 5G cellular networks : a) improving DSRC spectrum utilization and saving scarce cellular spectrum resources, b) reducing traffic congestion and packet loss, c) reducing control overhead caused by highly dynamic topology, d) increasing network scalability, e) providing opportunities for data aggregation.

Numerous clustering protocols have been proposed in Mobile Ad Hoc Networks (MANETs) and sensor networks, most of which are with the goal of reducing energy consumption. These algorithms are not appropriate for the VANET environment because energy consumption is not a problem for vehicles with unlimited power. By contrast, VANETs own higher topology dynamics caused by vehicles' high-speed mobility compared with MANETs. How to improve cluster stability is one of the most important issues for clustering design in VANETs. Concerning this issue, much work has been done on CH selection utilizing weighting network metrics such as connectivity degree, link stability, and inter-vehicle distance. However, these algorithms are road network topology-agnostic, making short-run optimal clustering decisions that will harm both cluster stability and effectiveness. Some protocols, such as those reported in [4], incorporate the mobility of the nodes into the clustering procedure. However, they exploit a localized manner of mobility such as vehicles' speed and direction but ignore human sociological factors because the true intentions of the people in vehicles are the primary reasons of their movements [5]. At each intersection, vehicles may have significant variations on moving direction and speed, easily causing the CH's leaving or losing its superiority. To further realize dynamic clustering as road conditions change, some of this category of algorithms emphasize that their proposed clustering algorithms are executed periodically. However, deciding the length of the re-clustering interval still remains a problem, might adding the computing burden on CHs. In the future 5G-VANET system with a coexistence of multiple infrastructures, the limited information of a single base station might make it difficult to predict the arriving traffic and execute clustering algorithms adaptively [6]. Beyond that, the great majority of these current clustering algorithms are designed to implement in a distributed way, where a large number of control messages on neighboring discovering, CH announcement and cluster maintaining occupy more bandwidth resources. Also, without a global network view, the clustering decisions are possible to plunge into local optima.

To overcome the aforementioned shortcomings, in this paper, we put forward an effective vehicle clustering algorithm in the 5G-VANET system, which utilizes a social pattern prediction model under a Software-Defined Networking (SDN) architecture and hence it is called SDN-Enabled Social-Aware Clustering (SESAC) algorithm. The social pattern here is defined as a set of road segments that the vehicle will travel successively in a future period and also the corresponding sojourn time on each road segment. The SDN architecture decouples the control and data forwarding planes, enabling external control of data through a logical software entity. The programmability and flexibility of SDN realize the rapid service deployment and simplified resource management. Specifically, we model each vehicle's movement as a discrete time-homogeneous semi-Markov model, where state transition probability and sojourn time probability distribution are inputs and each vehicle's social pattern is output. The prediction is performed in the SDN controller after it obtains the historical statistical information of vehicles. Subsequently, nodes with the same social pattern that potentially have similar routes in the future are grouped into one cluster in order to improve cluster stability. CHs are elected based on the metrics of inter-vehicle distance, relative speed, and vehicle attributes. Exploiting the SDN controller's global view, our SESAC executes clustering only when needed. We evaluate our SESAC algorithm and the results show that it performs better in terms of cluster lifetime and clustering overhead compared with the traditional clustering algorithms. The key contributions are listed as follows:

- It adopts a semi-Markov model to predict each vehicle's social pattern which includes the future moving route and the corresponding sojourn time on each road segment.
- Based on the social pattern prediction model, the SESAC algorithm is developed, which exploits the SDN controller's global view and computing power. A CH election method considering different network metrics is presented to improve cluster stability.
- The performance of our clustering algorithm is evaluated and compared with the traditional clustering methods. The evaluation results show the effectiveness of our method.

The remaining paper is organized as follows. We summarize the related work in Section II, and the network architecture of SDN-enabled 5G-VANET is described in Section III. A semi-Markov model for social pattern prediction for each vehicle is introduced in Section IV. Section V describes our designs of the SESAC algorithm. We demonstrate simulation results in section VI, before concluding this paper in Section VII.

II. RELATED WORK

A. CLUSTERING IN VANET

Clustering algorithms were first studied in MANETs and sensor networks [7], [8] and most of the current methods were designed to achieve the energy-saving objective, which are apparently not appropriate for VANETs consisted of vehicles with unlimited power. Instead, for the high-dynamic VANET, the primary aim of clustering is to improve cluster stability. Utilizing the unique node IDs, the Lowest ID and Highest Degree (LID/HD) clustering algorithms [7] proposed for MANETs are the earliest schemes adapted to the VANET environment, which has a significant influence on the subsequent algorithms. The Mobility Based Clustering (MOBIC) method [8] incorporates mobility considerations into the clustering phase by introducing inter-node distance into the CH election. Similarly, many researchers have proposed clustering algorithms for VANETs, most of which are based on a weighted sum of several network values including inter-vehicle distance, degree of connectivity, link stability, node uptime, etc..

Different from MANET nodes which are distributed randomly in the space, vehicles move along the road infrastructures in a geographical direction, providing advantageous conditions for clustering algorithm design. Much work on clustering schemes specifically for VANETs has been done [9]. In particular, the Connected Dominating Set-Stable Virtual Backbone (CDS-SVB) was proposed in [10], where vehicles' speed, direction, and relative distance are used for CH election to stabilize the backbone structure. The Affinity PROpagation for VEhiclar networks (APROVE) [11] adopts the idea of affinity propagation, forming clusters with both minimum distance and minimum relative velocity between CH and CMs. A Density-Based Clustering (DBC) scheme is developed in [12], taking the density of connection graph, link quality, and road traffic conditions into account. Ni *et al.* [4] proposed a Mobility Prediction-Based Clustering (MPBC) scheme which uses relative speed estimation for stable cluster formation. Ucar *et al.* [13] proposed a multi-hop-cluster-based IEEE 802.11*p*-LTE hybrid architecture for the first time. A CH is selected based on the metric of average relative speed with the neighboring vehicles. Using vehicle speed to predict link duration is conducive to reduce network overhead.

However, they exploit in a localized manner of the mobility such as vehicles speed and direction but ignore human sociological factors. And also, without an SDN controller's global view, cluster maintenance, and updating operate harder only using Hello packets between vehicles.

B. SDN-ENABLED VANET ARCHITECTURE

SDN is a layered network structure, where the control layer provides efficient centralized management over the underlying infrastructures through software modules. Much work on architecture, application and resource management of SDN-enabled VANETs has been done.

Ku *et al.* [14] proposed an SDN-based VANET architecture, whose benefits and available services were discussed. They also proved that the SDN-based routing is advantageous in packet delivering ratio compared with traditional routing protocols. A VANET architecture supporting SDN and fog computing technologies was proposed in [15]. Resource management and fog coordination for both safety and non-safety services were also analyzed in this architecture. Merits of this architecture include reducing latency and improving resource utilization rate. A fog-enabled real-time traffic management system is proposed in [16]. A hierarchical SDN-based vehicular architecture is presented in [17], solving the problem of connectivity loss between forwarding switches and the controller. Simulation results showed that the approach performed better than traditional solutions without SDN controllers.

Moreover, a new vehicular network architecture combined with 5G mobile cellular network and SDN technology was proposed recently [18]. For vehicles beyond the reach of any RSU, data delivering is implemented through several vehicle relays and a gateway; that is, a cluster is formed. Simulation results showed that this SDN-based 5G-VANET could provide enough flexibility and compatibility for future vehicular applications. An SDN-enabled vehicular network architecture integrating IEEE 802.11*p* and 5G radio access technologies was proposed in [19]. It was validated that the proposed architecture can meet Quality of Service (QoS) requirements of diverse applications such as low latency, high reliability and also maintain network scalability. To overcome the high dynamics of network topology and high complexity of infrastructures in 5G-VANET systems, Duan *et al.* [22] presented an adaptive clustering scheme and also a beamformed transmission method. In their scheme, all the aggregated traffic data is delivered to BS.

Our work is inspired by these work on SDN-enabled 5G-VANET architecture and to solve the clustering problem.

C. VEHICULAR SOCIAL-AWARE NETWORK

Drivers' behavior with social characteristics produces great influences on vehicles' mobility, e.g., people tend to go to the same places, at the same day periods, through the same trajectories [23]. This stimulates the study of the vehicles' mobility from a social perspective in order to improve QoSs in VANETs. Vehicular Social Networks (VSNs) hence arise [24]–[27]. In the past few years, a wide range of routing and clustering algorithms based on social metrics including degree centrality, closeness centrality, betweenness centrality and bridging centrality [28] have been proposed in VSNs.

Additionally, researchers exploit the social mobility patterns of vehicles to improve the performance of routing and clustering. For example, the tendency of vehicles to move along the same routes was recognized in [29] and [30]. Due to the complexity of human relationships, the social characteristics are unknown in most cases. Mobility prediction models play important roles in obtaining social patterns and analyzing social ties. The current methods for mobility prediction mainly include machine learning, Kalman filtering, data mining [31], [32] and Markov model. Specifically, Ghouti [33] proposed a neural learning-based solution to predict the mobility of MANET nodes. The proposed predictor achieves satisfying accuracy scores. Machine learning is simple, but it needs appropriate penalty/reward policies, and have slow convergence to the correct actions. Liu *et al.* [34] proposed an accurate location prediction method using extended self-learning Kalman filter. A mobile motion equation relying on velocity, acceleration and direction of movement is constructed for each user, and the performance of the prediction method largely depends on the stabilization time of the Kalman filter and the estimation of the system parameters. Data mining techniques have been used for location prediction [35] in mobile environments, where user mobility patterns are mined from the history of mobile user trajectories. Mobility predictions are accomplished using mobility rules extracted from mobility patterns. However, it needs large amounts of historical trajectory data to extract the mobility pattern. The effective and efficient algorithms are those based on Markov chains [36], for they can be applied to any problem domain as long as the state-space of the prediction problem can be converted into one of discrete-sequence prediction.

III. NETWORK ARCHITECTURE OF SDN-ENABLED 5G-VANET SYSTEM

SDN decouples control plane and data plane, realizing the flexibility and programmability of network management and resource deployment [20], [21]. It is widely accepted as a key technology in 5G networks to realize network resource virtualization and customize applications according to users' demands. SDN is introduced into the 5G-VANET system [22] in order to enable the coordination and information sharing between BSs to guarantee adaptive and efficient clustering. The SDN-enabled 5G-VANET system logically includes application plane, control plane, and data plane, as shown in Fig. 1. More details are given below:

1) APPLICATION PLANE

in the application plane, network functions for diverse application scenarios are written as software modules by developers through the NorthBound Interface (NBI). In general, the applications in vehicular networks are related to three aspects: traffic safety, traffic efficiency and entertainment. The way of packing applications into software modules and providing open interface is advantageous to simplify network function optimization and updates with the development of new technology.

2) CONTROL PLANE

the control layer is the core part of the network architecture, which provides global unified management over the underlying network infrastructures through the SouthBound Interface (SBI). It monitors and predicts the location of vehicles based on the historical information. After performing adaptive and efficient clustering, flow rules are distributed to BSs by fiber links and then to vehicles by cellular links.

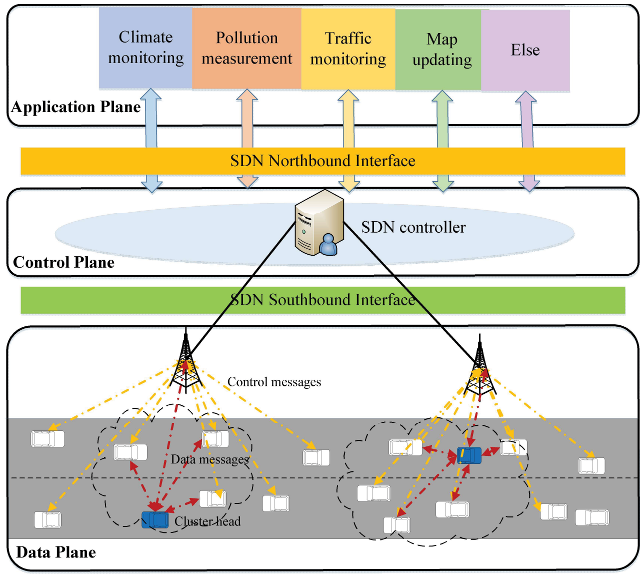


FIGURE 1. SDN based 5G-VANET architecture.

3) DATA PLANE

the data plane is the underlying physical infrastructures and devices in this network, mainly including BSs and vehicles. Each vehicle carries plenty of onboard sensors, continuously generating numerous data information such as speed, position, historical routes, etc.. Equipped with a cellular communication module and also a DSRC device, vehicles can establish links including Vehicle to BS (V2BS) links and Vehicle to Vehicle (V2V) links. In our design, to relieve the cellular burden and save licensed spectrum resources, the vehicles are grouped into clusters, where vehicles as CMs transmit data to their CH by V2V links and the CH will aggregate and deliver these data to the BS by V2BS links. Or vehicles can get service data from the BS through their CH. As shown in Fig. 1, V2BS links are utilized for control message transmission including the collection of vehicle information and dissemination of flow rules. As vehicles move through the network, they keep recording the road segments they traversed. Each BS broadcasts a short message to all vehicles in its vicinity, requests them to send their collected road segment sets. After that, the vehicles will create a packet containing the partial path as well as the node ID and send to the BS.

IV. VEHICLE SOCIAL PATTERN PREDICTION BASED ON SEMI-MARKOV MODEL

A semi-Markov model is an extension of Markov model, which predicts the future state transition direction according to the current state. In a semi-Markov model, the future state is related not only to the current state but also to the state transition time, i.e., the sojourn time in the current state. In this section, we model each vehicle's movement as a discrete time-homogeneous semi-Markov model, which takes the historical information including state transition probability and sojourn time probability distribution as inputs to predict the vehicle's social pattern. The social pattern is defined as a set of road segments that the vehicle will travel successively in a future period and also the corresponding sojourn time in each road segment. We will introduce our prediction model next, and the notation list is shown in Table. 1.

TABLE 1. Notation list.

Notations	Meanings
V	Set of vehicle nodes.
v_m	A vehicle node.
k	Number of vehicles in the network.
R	Set of road segments.
r_i	A road segment.
q	Number of road segments in the network.
S_n^m	The n^{th} state of vehicle v_m, i.e., the road segment where vehicle v_m is.
T_n^m	Time slot of state transition $S_{n-1}^m \to S_n^m$.
$G_{ij}^m(t)$	Probability that the state transition of vehicle v_m from road segment r_i to road segment r_j is completed in the period of time slot 0 to t.
t	A sojourn time threshold.
P_{ij}^m	Transition probability from state S_n^m to S_{n+1}^m
$W_{ij}^m(t)$	Sojourn time probability distribution for vehicle v_m.
l_i^m	Number of state transitions of vehicle v_m from road segment r_i, without considering which the next state is.
l_{ij}^m	Number of state transitions of vehicle v_m from road segment r_i to road segment r_j.
$U_{ij}^m(t)$	Probability that vehicle v_m moves into road segment r_j at the future time slot t on condition that vehicle v_m is currently on road segment r_i.
$G_{ir}^m(h)'$	Vehicle m has the first transition from the i^{th} state to the r^{th} state at time slot h.
$M^m(t)$	Location probability matrix of vehicle v_m at time slot t.

A. DISCRETE TIME-HOMOGENEOUS SEMI-MARKOV MODEL FOR VEHICLE MOVEMENT

We define $V = \{v_1, v_2, \ldots, v_k\}$ as the set of vehicles traveling on the road and $R = \{r_1, r_2, \ldots, r_q\}$ as the set of road segments in a given geographical area. The notations k and q represent the total numbers of vehicles and road segments in the whole network, respectively. A road segment here is an undirected edge between two consecutive intersections. In the following, we mention $S_n^m = r_i$ to represent the n^{th} state of vehicle v_m that it is on road segment r_i, and the state transition $S_{n-1}^m \to S_n^m$ means that the vehicle v_m moves into another road segment from the current road segment. We model the movement of vehicle v_m as a discrete time-homogeneous semi-Markov model: independent of the past state S_{n-1}^m, the state of vehicle v_m transits into the future state S_{n+1}^m from the current state S_n^m at time slot T_{n+1}^m. Note that we divide time into tiny discrete time slots, hence T_{n+1}^m is an integer. Therefore we can get that $(T_{n+1}^m - T_n^m)$ represents the sojourn time of vehicle v_m in the n^{th} state, i.e., the duration of maintaining the state S_n^m for vehicle v_m. The movement of each vehicle can be characterized as follows:

$$\begin{aligned} G_{ij}^m(t) &= P(S_{n+1}^m = r_j, T_{n+1}^m - T_n^m \le t | S_0^m, \ldots S_n^m; T_0^m, \ldots T_n^m) \\ &= P(S_{n+1}^m = r_j, T_{n+1}^m - T_n^m \le t | S_n^m = r_i) \end{aligned} \tag{1}$$

where t is a sojourn time threshold.

According to the time discreteness and conditional probability formula, Eq. (1) can be rewritten as follows:

$$\begin{aligned} G_{ij}^m(t) &= P(S_{n+1}^m = r_j, T_{n+1}^m - T_n^m \le t | S_n^m = r_i) \\ &= P(S_{n+1}^m = r_j | S_n^m = r_i) \\ &\quad \cdot P(T_{n+1}^m - T_n^m \le t | S_{n+1}^m = r_j, S_n^m = r_i) \end{aligned} \tag{2}$$

Assuming the current road segment of vehicle v_m is r_i and the next road segment is r_j, the transition probability from state S_n^m to S_{n+1}^m is as follows:

$$P_{ij}^m = P(S_{n+1}^m = r_j | S_n^m = r_i) = \lim_{t \to \infty} G_{ij}^m(t) \tag{3}$$

Given a sojourn time threshold t, we can obtain the probability that the duration of vehicle v_m staying at road segment r_i is less than the threshold. And the sojourn time probability distribution for vehicle v_m can be characterized by a function of t:

$$W_{ij}^m(t) = P(T_{n+1}^m - T_n^m \le t | S_{n+1}^m = r_j, S_n^m = r_i) \tag{4}$$

Therefore, Eq. (2) can also be written as follows:

$$G_{ij}^m(t) = P_{ij}^m \cdot W_{ij}^m(t) \tag{5}$$

Through this model, we can see that the movement of vehicles can be predicted by computing the state transition probability P_{ij}^m and sojourn time probability distribution $W_{ij}^m(t)$. In the following, we will give some details on these two parameters.

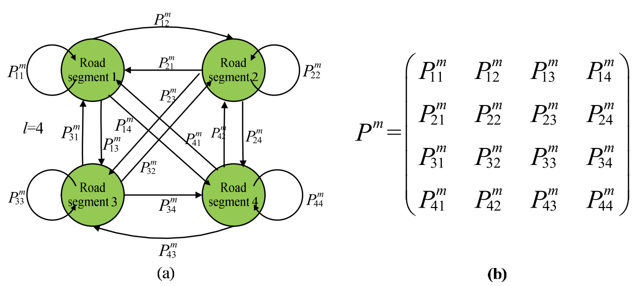


FIGURE 2. State transition probability matrix for vehicle v_m.

1) STATE TRANSITION PROBABILITY

an example of state transition probability matrix for a moving vehicle v_m is shown in Fig. 2, where P_{ij}^m represents the state transition probability of vehicle v_m from road segment r_i to r_j. When $i = j$, vehicle v_m stays in the current road segment. We use the following equation to calculate the matrix element P_{ij}^m:

$$P_{ij}^m = l_{ij}^m / l_i^m \tag{6}$$

where l_i^m represents the statistical number of vehicle v_m transiting from road segment r_i, without considering which the next road segment is; l_{ij}^m is the statistical number of vehicle v_m transiting from road segment r_i to r_j. Obviously, we have $l_{ij}^m \leq l_i^m$ and $P_{ij}^m \leq 1$. For example, in the historical database for vehicle v_m, if there are five state transitions from road segment 1 while one state transition from road segment 1 to 2, we can obtain $P_{12}^m = 1/5$. By keeping track of the historical values of l_i^m and l_{ij}^m, the state transition probability matrix of vehicle v_m is continuously updated.

2) SOJOURN TIME PROBABILITY DISTRIBUTION

we use the following equation to calculate the sojourn time probability distribution of vehicle v_m on road segment r_i on the premise that it will move to road segment r_j next:

$$W_{ij}^m(t) = P(t_{ij}^m < t) = \sum_{u=0}^{t-1} P(t_{ij}^m = u) \tag{7}$$

where t_{ij}^m is the sojourn time of vehicle v_m on road segment 1 before it moves into road segment 2. For instance, if there are 6 historical values of t_{12}^m in the database: 2, 3, 3, 6, 5, 4, we have $W_{12}^m(5) = P(t_{12}^m < 5) = 2/3$.

Based on the aforementioned continuously updated matrices, we can predict the future position of vehicles as described in the following.

B. PREDICTION OF VEHICLE POSITION

We let $X^m=\{X_t^m, t \in N^+\}$ represent a time-related state set of vehicle v_m. Here, X_t^m represents the state of vehicle v_m at time slot t. The transition probability Y_t^m can be described as follows.

$$U_{ij}^m(t) = P(X_t^m = r_j | X_0^m = r_i) \tag{8}$$

where $U_{ij}^m(t)$ is the probability that vehicle v_m will move into the road segment r_j at the future time slot t on condition that the vehicle v_m is currently on road segment r_i. Thus, we have two corollaries:

Corollary 1: For vehicle v_m with initial state of staying on road segment r_i. In the future time slot t, we have:

$$\sum_{j=1}^{q} U_{ij}^m(t) = 1, \quad \forall m, i, t. \tag{9}$$

Corollary 2: When $t = 0$, vehicle v_m must be on road segment r_i, thus we have:

$$U_{ij}^m(0) = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{10}$$

When $t > 0$, if we know the current state of vehicle v_m, we can obtain its location probability distribution at the future time slot t using $U_{ij}^m(t)$. We calculate $U_{ij}^m(t)$ in two cases:

- *Case 1:* vehicle v_m stays on road segment r_i from time slot 0 to t. Then we have:

$$\begin{aligned} U_{ij}^m(t) &= P(X_t^m = r_i | X_0^m = r_i, T_1^m - T_0^m \geq t) \\ &= P(T_1^m - T_0^m \geq t | X_0^m = r_i) \\ &= 1 - W_i^m(t) \end{aligned} \tag{11}$$

where $W_i^m(t)$ is the sojourn time probability distribution for vehicle v_m on road segment i, without considering which road segment it will move into next. According to Eq. (4), we obtain $W_i^m(t)$ as follows:

$$W_i^m(t) = P(T_{n+1}^m - T_n^m \leq t | S_n^m = i) = \sum_{j=1}^{q} G_{ij}^m(t) \tag{12}$$

- *Case 2:* There is at least one state transition for vehicle v_m from time slot 0 to t. Then we have:

$$\begin{aligned} U_{ij}^m(t) &= P(X_t^m = r_j | X_0^m = r_i, \text{at least 1 state transition}) \\ &= \sum_{y=1}^{q} \sum_{z=1}^{t-1} G_{iy}^m(z)' \cdot U_{yj}^m(t-z) \end{aligned} \tag{13}$$

where $G_{iy}^m(z)' = G_{iy}^m(z) - G_{iy}^m(z-1)$, which means that vehicle v_m has the first state transition from road segment r_i to r_y at time slot z.

With the consideration of the aforementioned two cases, we determine the following location probability for vehicle v_m:

$$U_{ij}^m(t) = [1 - W_i^m(t)] \cdot \delta_{ij} + \sum_{y=1}^{q} \sum_{z=1}^{t-1} G_{iy}^m(z)' \cdot U_{yj}^m(t-z) \tag{14}$$

From Eq. (10), we can see that when $i \neq j$, the first term of Eq. (14) becomes 0, which means that vehicle v_m has at least one state transition. When $i = j$, it indicates that vehicle v_m's initial state is in road segment r_i, and after that: (i) vehicle v_m may stay in road segment r_i without any state transition, as shown in the first term of Eq. (14); or (ii) vehicle v_m may have at least one state transition as shown in the second term of Eq. (14). Next, Eq. (14) can be converted into the following equation according to the time discreteness:

$$\begin{aligned} U_{ij}^m(t) = P(X_t^m = r_j | X_0^m = r_i) &= [1 - W_i^m(t)] \cdot \delta_{ij} \\ &+ \sum_{y=1}^{q} \sum_{z=1}^{t-1} f_{iy}^m(z) \cdot U_{rj}^m(t-z) \end{aligned} \tag{15}$$

where

$$f_{ij}^m(t) = \begin{cases} G_{ij}^m(1) & t = 1 \\ G_{ij}^m(t) - G_{ij}^m(t-1) & t > 1 \end{cases} \tag{16}$$

Given the value of $G_{ij}^m(t)$, the value of $U_{ij}^m(t)$ can be computed using an iteration method.

According to Eq. (15), once the value of $U_{ij}^m(t)$ is determined, we then obtain the location probability matrix $M^m(t)$ with the element $U_{ij}^m(t)$. As mentioned above, $U_{ij}^m(t)$ is the probability that vehicle v_m will move into road segment r_j in future time slot t on condition that vehicle v_m is currently on road segment r_i. For instance, we will have the following location probability matrix:

$$M^m(t) = \begin{bmatrix} U_{11}^m(t) & U_{12}^m(t) & U_{13}^m(t) & U_{14}^m(t) \\ U_{21}^m(t) & U_{22}^m(t) & U_{23}^m(t) & U_{24}^m(t) \\ U_{31}^m(t) & U_{32}^m(t) & U_{33}^m(t) & U_{34}^m(t) \\ U_{41}^m(t) & U_{42}^m(t) & U_{43}^m(t) & U_{44}^m(t) \end{bmatrix} \tag{17}$$

Obviously, the row number (i.e., i) of $M^m(t)$ represents the current road segment index of vehicle v_m, while the column number (i.e., j) represents the index of the road segment the vehicle v_m will move into at the future time slot t. Given the index of the current road segment i, we will find the maximal element value $U^m_{ij^*}(t)$ along the i^{th} row of $M^m(t)$, and the corresponding column number j^* is the index of the road segment the vehicle v_m will move into at the future time slot t. Only when $i \neq j$, a state transition occurs. So far, we have determined the future location of vehicles (i.e. the road segment where the vehicle will stay at a future time slot).

C. CONSTRUCTION OF VEHICLE SOCIAL PATTERN TABLE

We divide a day into several time periods since a vehicle with social attributes seems more like to have different routes in different periods. For example, a professor's vehicle may follow several routes in a day: from home to university between 8.30 and 9.00; from university to a restaurant between 12.00 and 12.30; from university back to home between 17.00 and 17.30. In a specific time period TP_a, the BS can predict the road segment where each vehicle is at a future time slot based on the above-mentioned process. Further, the future route (i.e. a series of road segments) and also the corresponding sojourn time for a certain vehicle can be obtained, which is defined as a social pattern. For instance, if in a time period TP_1, $U^m_{12}(5)$, $U^m_{24}(12)$, $U^m_{45}(18)$ and $U^m_{53}(21)$ are optimal values in Eq. (17) when $i \neq j$, which represents that there are state transitions for vehicle v_m at time slots 5, 12, 18 and 21. Then we can get a social pattern of vehicle v_m in TP_1: $1(5) \rightarrow 2(7) \rightarrow 4(6) \rightarrow 5(3) \rightarrow 3$. The figures in brackets are sojourn time in the road segments.

Once the social patterns for vehicles are obtained, a social pattern table is constructed in the BS, where each entry contains the social number, vehicle ID, social pattern and the specific time period. Note that there can be several social patterns in different time periods for one vehicle.

V. SDN-ENABLED SOCIAL-AWARE CLUSTERING IN 5G-VANET SYSTEM

In the SDN-enabled 5G-VANET system, the SDN controller has a global network view and can share it with heterogeneous BSs. In this section, we will present an SDN-Enabled Social-Aware Clustering (SESAC) method which creates clusters with nodes that will potentially have similar future routes. In addition, based on the global view, the clustering algorithm would be executed only when needed instead of periodically. the procedure will be elaborated in detail next:

A. INITIAL GROUPING

Vehicles report their position and speed information to the BS periodically, hence the BS has the knowledge of the vehicular network topology. Initially, the road is divided into fix-length rectangular areas, in each of which a list of vehicles are classified into a group, and then the group information is sent back to the vehicles in that area. Note that the rectangular area exists only at the beginning of the clustering and the vehicles in an initial group do not have to keep moving in a regular shape. The length of the road is generally not larger than the 802.11p transmission range, which is defined by:

$$d \leq R_t(1-\varepsilon) \tag{18}$$

where the notation R_t denotes the maximum transmission range and ε reflects the wireless channel fading conditions in the current location. The initial grouping result is shown in Fig. 3.

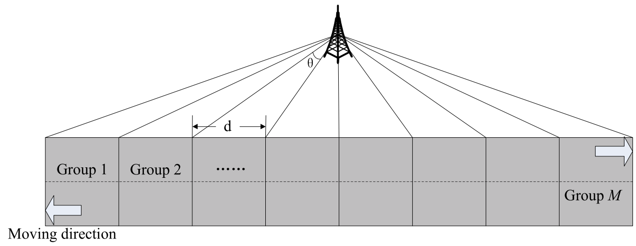


FIGURE 3. Initial grouping result.

B. NEIGHBOR IDENTIFYING

For a particular vehicle, any other vehicle that is within its transmission range is called a neighbor and the neighbor set is always changing since all nodes move fast. The vehicles in an initial group use IEEE 802.11p to identify their neighbors and verify the inter-vehicle distance. A moving vehicle v_i has a neighboring list N_i which keeps track of all neighbors' IDs, current locations, speed vectors as well as their current distances.

C. METRIC CALCULATION BETWEEN NEIGHBORS

Assume that each neighboring node v_j of vehicle v_i has an effect to vehicle v_i. Vehicle v_i will get a score from vehicle v_j, which is computed by a combined metric Me_{ij} related to their distance, relative velocity and also characteristics of the vehicle v_i itself. Each node v_i calculates the accumulated scores from its neighbors and the node which owns the highest scores is elected as a CH. Otherwise, it tries to join a cluster and become its CM. Note that when first entering the network, all nodes are in the non-clustered state. Related to the relative velocity vector, the metric Me_{ij} is a vector and can be decomposed along axis X and axis Y as follows:

$$Me^x_{ij} = k^x_{ij} \cdot \frac{C_i}{d^2_{ij}}, Me^y_{ij} = k^y_{ij} \cdot \frac{C_i}{d^2_{ij}} \tag{19}$$

where k^x_{ij} and k^y_{ij} are relative mobility parameters with signs depending on whether and how fast the vehicles are approaching (positive) or moving away (negative) along axis X and axis Y. Notation d_{ij} is the current distance between nodes v_i and v_j. Parameter C_i is related to some parameters on vehicle's status that can support it to be the best candidate for CH in the network (e.g. it follows a predefined route or has smooth driving habits.). All vehicles are assigned an initial value of C_{0i} and it varies as long-term behavior changes occur. Here we give some parameters related to C_i: Cp_i which represents the route stability of vehicles, Ct_i which is the height of the vehicle, and Cb_i which indicates whether the vehicle's driving behavior is statistically smooth. C_i is expressed as follows:

$$C_i = C_{0i} \cdot Cp_i \cdot Ct_i \cdot Cb_i \tag{20}$$

This makes it possible that a tall vehicle (like a truck), a fixed-route vehicle (like a bus) or a steady car is more likely to be selected to be a CH, which is advantageous to improve cluster stability.

In particular, we calculate the parameters k_{ij}^x and k_{ij}^y based on their distance at time t and their possible distance at time $t + \Delta t$ for each axis. According to the data from GPS, assume that the positions of vehicles v_i and v_j are: (x_i, y_i) and (x_j, y_j). The distances between these two nodes on axis x and y are $r_{ij}^{xc} = x_j - x_i$ and $r_{ij}^{yc} = y_j - y_i$, respectively. After time Δt, the positions become $(x_i + \Delta x_j, y_i + \Delta y_j)$ and $(x_j+\Delta x_j, y_j+\Delta y_j)$. And the distance between these two nodes become $r_{ij}^{xf} = x_j+\Delta x_j-x_i-\Delta x_i$ and $r_{ij}^{yf} = y_j+\Delta y_j-y_i-\Delta y_i$. We define the value of k_{ij}^x and k_{ij}^y as follows:

$$if\ r_{ij}^{xf} \geq r_{ij}^{xc}, \quad then\ k_{ij}^x = -(k_{ij}^{xc} - k_{ij}^{xf}). \tag{21}$$

$$if\ r_{ij}^{xf} \leq r_{ij}^{xc}, \quad then\ k_{ij}^x = \frac{1}{k_{ij}^{xc} - k_{ij}^{xf}}; \tag{22}$$

Eq. (21) represents the case that vehicles are moving apart from each other and $k_{ij}^x(k_{ij}^y)$ is a negative value and the faster it takes place the bigger the negative effect must be. In Eq. (22), vehicles are moving towards to each other and $k_{ij}^x(k_{ij}^y)$ is inversely proportional to the distance difference between the nodes. The reason is that the faster the two nodes move towards each other, the shorter their connected duration will be. Accordingly, vehicles that move towards each other in a relatively slow speed tend to stay connected for a longer time period, which is favored to create clusters.

At any time for vehicle v_i, positive or negative effects from their neighbors can be simultaneously applied, node v_i calculates:

$$Me_i^x = \sum_{j \in N_i} k_{ij}^x \cdot \frac{C_i}{d_{ij}^2}, Me_i^y = \sum_{j \in N_i} k_{ij}^y \cdot \frac{C_i}{d_{ij}^2} \tag{23}$$

$$Me_i = |Me_i^x| + |Me_i^y| \tag{24}$$

The node with the highest positive metric value Me_i is the most stable in its neighborhood and the best candidate to become a CH.

D. CLUSTERING BASED ON SOCIAL PATTERN

At the first stage, each vehicle tries to create a cluster with nodes that have the same social pattern. Specifically, at any time vehicle v_i recalculates its total scores gotten from its neighbors according to the metric Me_i. The node which gets the biggest positive score declares itself to be a CH, and nodes with the same social pattern in the neighbor list of vehicle v_i become CMs of v_i's cluster. When a CM moves out of the CH's transmission range, it is removed from the CM list maintained by the CH and becomes a free node again.

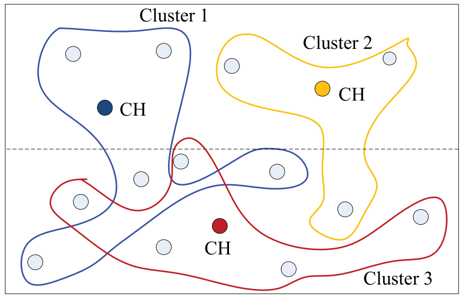


FIGURE 4. Example of clustering based on social pattern.

However, there may be some nodes could not join any cluster because they are surrounded by vehicles with different social patterns. With this situation, clustering is performed again using the total score applied to each vehicle. At this stage, clusters of vehicles with different social patterns are created. An example of the clustering result is shown in Fig. 4.

VI. EVALUATION RESULTS AND DISCUSSIONS

In this section, we evaluate the performance of our SESAC algorithm proposed for 5G-VANET systems. We assume that the VANET is in the urban environment. First, the superiority on processing latency of SDN controllers compared with traditional network architecture is proved. And then the performance of the SESAC in terms of cluster lifetime and network overhead is compared with previous clustering algorithms.

A. PARAMETER SETTINGS

In our simulation, all nodes are equipped with GPS receivers and OBUs. The movement information of all vehicles, which is needed for clustering, is collected using 5G links so that the SDN controller can get a whole view of the VANET topology. By default, the data rate of the IEEE 802.11*p* for inter-vehicle communication is 12 Mbps and the reliable communication range is 250 meters, in which area 25 vehicles moving forward and backward. We set the vehicles' velocities as 35 km/h on average in the urban environment. The communication transmission range, mean vehicle velocity and vehicle density (i.e. the number of vehicles in one-hop communication range) are taken as tunable parameters in order to verify its influence to the clustering performance, varying from 100 to 350 meters, 20 to 50 km/h and 10 to 40, respectively.

Although a social pattern probably exists for each vehicle, it is not completely deterministic due to the nature of driving. Even if a driver follows a standard route every day, it is still likely that he will deviate from it once in a while. For this reason, we assume vehicles are injected onto the map in a random sequence and follow their path according to their social pattern with a default probability of 85%.

In addition, we define the social pattern length as the number of road segments in a vehicle's social pattern. Taking

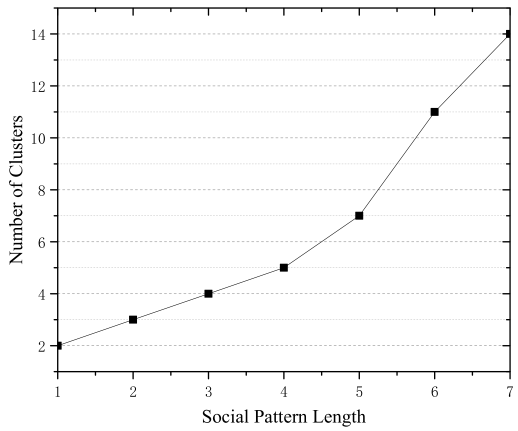


FIGURE 5. **Example of clustering based on social pattern.**

the case in Section IV-C as an example, the social pattern length of vehicle v_m is 5 since it has 5 road segments in $1(5) \rightarrow 2(7) \rightarrow 4(6) \rightarrow 5(3) \rightarrow 3$. The longer social pattern results in low-probability pattern matching among neighboring nodes and too many CHs elected in an initial one-hop-range group. The relationship of the average number of clusters in one-hop communication range and the social pattern length is shown in Fig. 5. We take the value of 3 as the social pattern length so that there is a reasonable number of clusters in each group.

In order to incorporate different characteristics in the method, we assign values to parameters C_i. These parameters represent a special role that a vehicle may have in the network due to its mobility behavior or physical characteristics. All the simulation parameters with their values are represented in Table 2.

TABLE 2. **Parameter settings.**

Parameter name	Default	Range
Data rate of IEEE Standard 802.11p	12 Mbps	＼
Communication range (m)	250	100∼350
Number of vehicles in one-hop range	25	10∼40
Mean vehicle velocity (km/h)	35	20∼50
Probability of following the social pattern(%)	85	＼
Social pattern length	3	1∼7
Vehicles height	1	1∼2
Route stability	1	0.5∼2
Driver's behaviour	1	0.5∼2

B. SCHEME COMPARISON

We compare our proposed clustering (SESAC) method with the Lowest-ID (LID) clustering algorithm [7] and Mobility Prediction-Based Clustering (MPBC) [4], respectively.

1) LOWEST-ID CLUSTERING (LID) ALGORITHM

In the LID algorithm, each node has an ID and it periodically broadcasts its ID to nodes within its two-hop communication range. A node declares itself as a CH when it hears a node with a higher ID unless it gives up this role (deferring to a lower ID node).

2) MOBILITY PREDICTION-BASED CLUSTERING (MPBC) ALGORITHM

The basic idea of MPBC is to estimate the relative velocity for each node with its neighbors, and nodes with the lowest relative speeds are selected as CHs. During the clustering stage, nodes broadcast periodically Hello packets in order to build their neighbor lists.

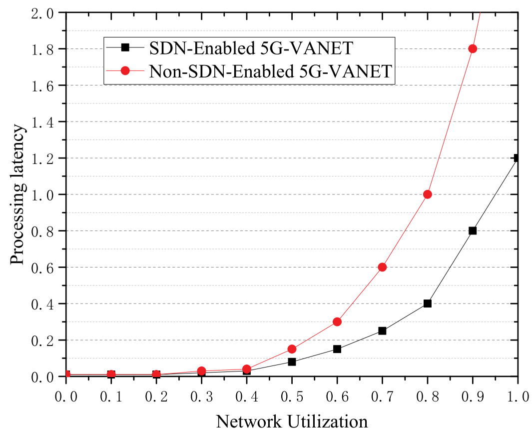


FIGURE 6. **Example of clustering based on social pattern.**

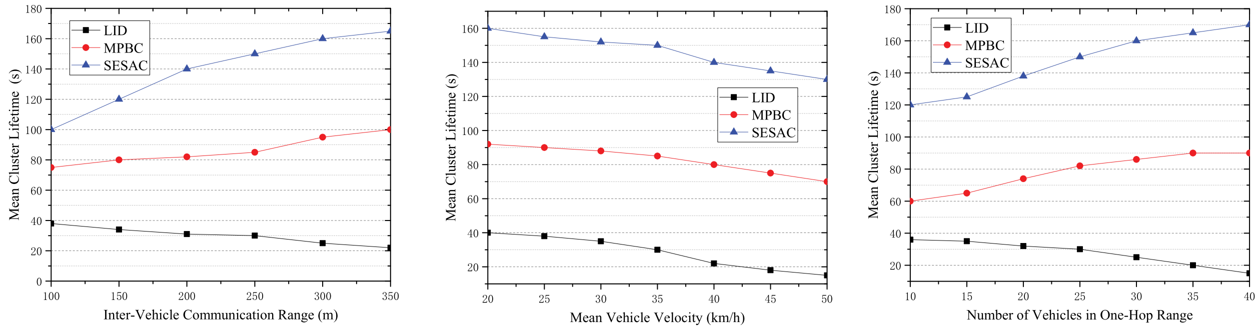
C. DISCUSSION OF RESULTS

1) SDN PROCESSING LATENCY

The SDN controller has a global VANET topology view and executes clustering when needed, while the traditional method periodically executes clustering. Despite the advantage of reduced control overhead compared with periodically clustering method, we have to make sure that there is no extra latency introduced by SDN-enabled data traffic processing. Fig. 6 shows the comparison of processing delay versus network utilization rate. Here network utilization rate is defined as the ratio of total data arrival rate and the controller processing rate. It can be seen from Fig. 6 that when the network load is fairly low, processing delay is not a problem for both SDN and non-SDN networks. With the increase of network load, SDN enabled 5G-VANET still keeps the latency under 1 ms most of the time, which meets the 5G latency requirement. It is obvious that SDN-enabled 5G-VANET system has better performance and maintains the flexibility, programmability of networks.

2) CLUSTER LIFETIME

In a highly dynamic VANET, nodes keep joining and leaving clusters. The lifetime of a cluster is defined as the time period from the moment when a vehicle becomes a CH to the time when it becomes CM. Good clustering algorithms should be designed to prolong the cluster lifetime. In order to investigate the stability of clusters that are created by each method, we measure variations of cluster lifetime along with vehicle communication range, mean velocity, and vehicle density. The results analysis is as follows.



(a) (b) (c)

FIGURE 7. Clustering stability. (a) Clustering stability vs communication range. (b) Clustering stability vs mean velocity. (c) Clustering stability vs vehicle density.

a: CLUSTER STABILITY VERSUS COMMUNICATION RANGE

Fig. 7 (a) shows the comparisons of the mean cluster lifetime varying with communication range among the three clustering algorithms. We can see that the cluster lifetime increases as the transmission range increases when the SESAC is used. This is because the SESAC creates clusters of nodes sharing common social patterns. As communication range increases, the duration that such nodes stay interconnected with their CH becomes longer. Communication range does negative impact on LID's performance. This is because in the LID only vehicles' IDs are used to elect CHs and that way, although increased communication range may have a positive impact on nodes' connectivity, it also has a negative impact since nodes are more likely to meet a neighbor with lower ID and perform re-clustering. The MPBC achieves longer average CH lifetime compared to the LID since the method considers the relative speeds. But its performance is still worse than the proposed method because it doesn't incorporate drivers' social patterns.

b: CLUSTER STABILITY VERSUS VEHICLE'S VELOCITY

Fig. 7 (b) shows the mean cluster lifetime varying with the mean vehicle's velocity in different clustering algorithms. We observe that with the increase of the mean vehicle's speed, the cluster's lifetime slightly decreases in the three clustering algorithm. This is due to the fact that high speed will accelerate vehicles' moving across intersections and driving on different road segments in the urban environment, and hence moving directions may change faster in the driving process. Even though, SESAC has much better performance compared to MPBC and LID because vehicles with similar social patterns are more likely to keep the same route for a long period.

c: CLUSTER STABILITY VERSUS VEHICLE DENSITY

Fig. 7 (c) shows the mean cluster lifetime varying with vehicle density which is denoted by the number of vehicles in one-hop communication range for different clustering algorithms. We can get that for the SESAC and the MPBC, the cluster lifetime is prolonged as vehicle density increases. This is because the cluster scale may be larger in dense areas, i.e., there are more CMs in a cluster. The probability that a CH connects with a CM is higher. In our definition, a cluster exists provided that the CH is connected with a CM. The SESAC has much better performance compared to the MPBC for the reason of social pattern predictions. The LID's performance gets worse as the vehicle density increases. This is because nodes are more likely to meet a neighbor with lower ID and perform re-clustering when there are more nodes in its communication range.

3) NORMALIZED CLUSTERING OVERHEADS

The clustering overhead includes overhead for cluster creation and maintenance, which is an important metric providing an indication of the extra bandwidth consumed to deliver data traffic. The lower overhead, the better performance. Note that long cluster lifetime does not mean low overhead here because the former represents the lifetime of a CH and there may be frequent changes of CMs which induces high cluster maintenance overhead. Fig. 8 shows the variations of clustering overhead with different network parameters.

a: CLUSTERING OVERHEAD VERSUS COMMUNICATION RANGE

Fig. 8 (a) shows the variations of the normalized clustering overhead with communication range. We can see that the SESAC has less overhead than the LID and MPBC. This can be explained by the fact that the SESAC organize nodes with the similar future route in one cluster, which could reduce cluster recreation and maintenance. As communication range increases, CMs potentially stay connected with their CH for a longer period and the overhead for cluster maintenance is hence cut down. The number of cluster changes is reduced which means the number of cluster creations is low. In LID, clusters are formed only using vehicles' IDs. With the communication range being expanded, re-clustering occurs as long as the CH meets a neighbor with lower ID, resulting in increased overhead. The MPBC achieves better overhead performance than LID but still worse than our proposed the SESAC since it predicts vehicles' short-term

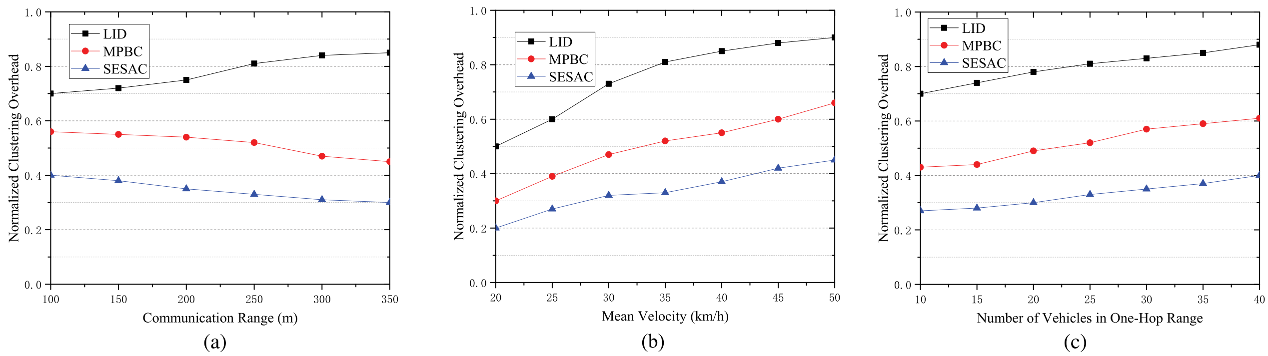


FIGURE 8. **Network overhead. (a) Normalized overhead vs communication range. (b) Normalized overhead vs mean velocity. (c) Normalized overhead vs vehicle density.**

movements and clusters become invalid in the long-term run.

b: CLUSTERING OVERHEAD VERSUS VEHICLE'S VELOCITY

Fig. 8 (b) shows the clustering overhead varying with mean vehicle velocity in different algorithms. With the increase of mean vehicle speed, the clustering overhead increases in the three clustering algorithms since the lifetime of clusters decreases and frequent re-clustering is needed. SESAC performs much better than MPBC and LID because vehicles with similar social patterns are more likely to keep the same route for a period.

c: CLUSTERING OVERHEAD VERSUS VEHICLE DENSITY

Fig. 8 (c) shows the clustering overhead with vehicle density for different clustering algorithms. As is expected, with the increasing vehicle density, the clustering overhead in LID is in an uptrend since it owns shorter cluster lifetime as shown in Fig. 7 (c). However, high vehicle density induces more clustering overhead for the SESAC algorithm and the MPBC algorithm. This is because the cluster scale may be larger in a dense area and it needs more overhead for cluster maintenance.

VII. CONCLUSION

In this paper, we have presented a social pattern prediction model to predict vehicles' future routes and the corresponding sojourn time on each road segment. Based on the social pattern prediction model, we have designed the SDN-Enabled Social-Aware Clustering (SESAC) algorithm, aiming at improving cluster stability. Nodes with the same social pattern that potentially have similar future routes are grouped into one cluster. A CH is elected based on the metrics of inter-vehicle distance, relative speed, vehicle attributes, etc.. We have evaluated the SESAC algorithm and compared it with traditional clustering methods in VANETs. The evaluation results have shown the effectiveness of our proposed method.

REFERENCES

[1] Z. Ning *et al.*, "A cooperative quality-aware service access system for social Internet of vehicles," *IEEE Internet Things J.*, to be published, doi: 10.1109/JIOT.2017.2764259.

[2] W. Hou *et al.*, "Service degradability supported by forecasting system in optical data center networks," *IEEE Syst. J.*, to be published, doi: 10.1109/JSYST.2018.2821714.

[3] S. Jia, S. Hao, X. Gu, and L. Zhang, "Analyzing and relieving the impact of FCD traffic in LTE-VANET heterogeneous network," in *Proc. IEEE ICT*, May 2014, pp. 88–92.

[4] M. Ni, Z. Zhong, and D. Zhao, "MPBC: A mobility prediction-based clustering scheme for ad hoc networks," *IEEE Trans. Veh. Technol.*, vol. 60, no. 9, pp. 4549–4559, Nov. 2011.

[5] X. Kong, X. Song, F. Xia, H. Guo, J. Wang, and A. Tolba, "LoTAD: Long-term traffic anomaly detection based on crowdsourced bus trajectory data," *World Wide Web*, vol. 21, no. 3, pp. 825–847, 2018, doi: 10.1007/s11280-017-0487-4.

[6] W. Hou, R. Zhang, W. Qi, K. Lu, J. Wang, and L. Guo, "A provident resource defragmentation framework for mobile cloud computing," *IEEE Trans. Emerg. Sel. Topics Circuits Syst.*, vol. 6, no. 1, pp. 32–44, Jan./Mar. 2018, doi: 10.1109/TETC.2015.2477778.

[7] M. Gerla and J. T.-C. Tsai, "Multicluster, mobile, multimedia radio network," *Wireless Netw.*, vol. 1, no. 3, pp. 255–265, 1995.

[8] P. Basu, N. Khan, and T. D. C. Little, "A mobility based metric for clustering in mobile ad hoc networks," in *Proc. WNMC*, Apr. 2001, pp. 413–418.

[9] Z. Ning, X. Wang, X. Kong, and W. Hou, "A social-aware group formation framework for information diffusion in narrowband Internet of Things," *IEEE Internet Things J.*, to be published, doi: 10.1109/JIOT.2017.2777480.

[10] M. A. Togou, A. Hafid, and P. K. Sahu, "A stable minimum velocity CDS-based virtual backbone for VANET in city environment," in *Proc. LCN*, Sep. 2014, pp. 510–513.

[11] C. Shea, B. Hassanabadi, and S. Valaee, "Mobility-based clustering in VANETs using affinity propagation," in *Proc. GLOBECOM*, Nov./Dec. 2009, pp. 1–6.

[12] S. Kuklinski and G. Wolny, "Density based clustering algorithm for vehicular ad-hoc networks," *Int. J. Internet Protocol Technol.*, vol. 4, no. 3, pp. 149–157, 2009.

[13] S. Ucar, S. C. Ergen, and O. Ozkasap, "Multihop-cluster-based IEEE 802.11p and LTE hybrid architecture for VANET safety message dissemination," *IEEE Trans. Veh. Technol.*, vol. 65, no. 4, pp. 2621–2636, Apr. 2016.

[14] I. Ku, Y. Lu, M. Gerla, F. Ongaro, R. L. Gomes, and E. Cerqueira, "Towards software-defined VANET: Architecture and services," in *Proc. MED-HOCNET*, Jun. 2014, pp. 103–110.

[15] N. Truong, G. M. Lee, and Y. Ghamri-Doudane, "Software defined networking-based vehicular adhoc network with fog computing," in *Proc. IEEE IM*, May 2015, pp. 1202–1207.

[16] X. Wang, Z. Ning, and L. Wang, "Offloading in Internet of vehicles: A fog-enabled real-time traffic management system," *IEEE Trans. Ind. Informat.*, to be published, doi: 10.1109/TII.2018.2816590.

[17] X. Ge, Z. Li, and S. Li, "5G software defined vehicular networks," *IEEE Commun. Mag.*, vol. 55, no. 7, pp. 87–93, Jul. 2017.

[18] J. Liu, J. Wan, B. Zeng, Q. Wang, H. Song, and M. Qiu, "A scalable and quick-response software defined vehicular network assisted by mobile edge computing," *IEEE Commun. Mag.*, vol. 55, no. 7, pp. 94–100, Jul. 2017.

[19] S. Correia, A. Boukerche, and R. I. Meneguette, "An architecture for hierarchical software-defined vehicular networks," *IEEE Commun. Mag.*, vol. 55, no. 7, pp. 80–86, Jul. 2017.

[20] X. Zhang *et al.*, "Experimental demonstration of an intelligent control plane with proactive spectrum defragmentation in SD-EONs," *Opt. Express*, vol. 25, no. 20, pp. 24837–24852, 2017.

[21] X. Zhang, L. Guo, W. Hou, Q. Zhang, and S. Wang, "Failure recovery solutions using cognitive mechanisms based on software-defined optical network platform," *Opt. Eng.*, vol. 56, no. 1, p. 016107, 2017.

[22] X. Duan, Y. Liu, and X. Wang, "SDN enabled 5G-VANET: Adaptive vehicle clustering and beamformed transmission for aggregated traffic," *IEEE Commun. Mag.*, vol. 55, no. 7, pp. 120–127, Jul. 2017.

[23] A. Rahim *et al.*, "Vehicular social networks: A survey," *Pervas. Mobile Comput.*, vol. 43, pp. 96–113, Jan. 2018, doi: 10.1016/j.pmcj.2017.12.004.

[24] Z. Ning, F. Xia, N. Ullah, X. Kong, and X. Hu, "Vehicular social networks: Enabling smart mobility," *IEEE Commun. Mag.*, vol. 55, no. 5, pp. 49–55, May 2017.

[25] X. Kong *et al.*, "Mobility dataset generation for vehicular social networks based on floating car data," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 3874–3886, May 2018, doi: 10.1109/TVT.2017.2788441.

[26] Z. Ning, F. Xia, X. Hu, Z. Chen, and M. Obaidat, "Social-oriented adaptive transmission in opportunistic Internet of smartphones," *IEEE Trans. Ind. Informat.*, vol. 13, no. 2, pp. 810–820, Apr. 2017.

[27] Z. Ning, L. Liu, F. Xia, B. Jedari, I. Lee, and W. Zhang, "CAIS: A copy adjustable incentive scheme in community-based socially aware networking," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3406–3419, Apr. 2017.

[28] A. M. Vegni and V. Loscrí, "A survey on vehicular social networks," *IEEE Commun. Surveys Tut.*, vol. 17, no. 4, pp. 2397–2419, 4th Quart., 2015.

[29] G. Pallis, D. Katsaros, M. D. Dikaiakos, N. Loulloudes, and L. Tassiulas, "On the structure and evolution of vehicular networks," in *Proc. IEEE/ACM MASCOTS*, Sep. 2009, pp. 502–511.

[30] D. Naboulsi and M. Fiore, "On the instantaneous topology of a large-scale urban vehicular network: The cologne case," in *Proc. ACM MOBIHOC*, 2013, pp. 167–176.

[31] W. Hou, Z. Ning, L. Guo, and X. Zhang, "Temporal, functional and spatial big data computing framework for large-scale smart grid," *IEEE Trans. Emerg. Sel. Topics Circuits Syst.*, to be published, doi: 10.1109/TETC.2017.2681113.

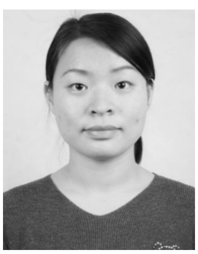
[32] W. Hou, Z. Ning, L. Guo, Z. Chen, and M. S. Obaidat, "Novel framework of risk-aware virtual network embedding in optical data center networks," *IEEE Syst. J.*, to be published, doi: 10.1109/JSYST.2017.2673828.

[33] L. Ghouti, "Mobility prediction in mobile ad hoc networks using neural learning machines," *Simul. Model. Pract. Theory*, vol. 66, pp. 104–121, Aug. 2016.

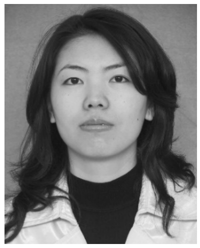
[34] T. Liu, P. Bahl, and I. Chlamtac, "Mobility modeling, location tracking, and trajectory prediction in wireless ATM networks," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 6, pp. 922–936, Aug. 1998.

[35] G. Yavaş, D. Katsaros, Ö. Ulusoy, Y. Manolopoulos, "A data mining approach for location prediction in mobile environments," *Date Knowl. Eng.*, vol. 54, no. 2, pp. 121–146, 2005.

[36] D. Katsaros and Y. Manolopoulos, "Prediction in wireless networks by Markov chains," *IEEE Wireless Commun.*, vol. 16, no. 2, pp. 56–64, Apr. 2009.



WEIJING QI received the M.S. degree in communication and information systems from Northeastern University, Shenyang, China, in 2015, where she is currently pursuing the Ph.D. degree. Her research interests are in network optimization and routing in highly dynamic ad hoc networks including UAV network and vehicular network.



QINGYANG SONG (SM'12) received the Ph.D. degree in telecommunications engineering from The University of Sydney, Camperdown, NSW, Australia. She is currently a Professor with Northeastern University, Shenyang, China. She has authored over 50 papers in major journals and international conferences. These papers have been cited more than 1300 times in scientific literature. Her current research interests are in radio resource management, cognitive radio networks, cooperative communications, ad hoc networks, heterogeneous cellular networks, and protocol design.

XIAOJIE WANG (S'16) received the master's degree from Northeastern University, Shenyang, China, in 2011. She is currently pursuing the Ph.D. degree with the School of Software, Dalian University of Technology, Dalian, China. From 2011 to 2015, she was a Software Engineer with NeuSoft Corporation. Her research interests include social computing and network security.

LEI GUO (M'06) received the Ph.D. degree from the University of Electronic Science and Technology of China in 2006. He is currently a Professor with Northeastern University, Shenyang, China. His research interests include communication networks, optical communications, and wireless communications. He has authored or co-authored over 200 technical papers in the abovementioned areas in international journals and conferences, such as the IEEE Transactions on Communications, the IEEE Transactions on Wireless Communications, the IEEE/OSA Journal of Lightwave Technology, the IEEE/OSA Journal of Optical Communications and Networking, the IEEE GLOBECOM, and the IEEE ICC. He is a member of the OSA and a Senior Member of CIC. He serves as an Editor for five international journals, such as the *Photonic Network Communications* and *The Open Optics Journal*.

ZHAOLONG NING (M'14) received the M.S. and Ph.D. degrees from Northeastern University, Shenyang, China. He was a Research Fellow with Kyushu University, Japan. He is currently an Associate Professor with the School of Software, Dalian University of Technology, China. His research interests include fog computing, vehicular network, and network optimization. He serves as an Associate Editor for the IEEE Access and the Lead Guest Editor for *The Computer Journal* and IEEE Access.