A Cluster-Based V2V Approach for Mixed Data Dissemination in Urban Scenario of IoVs

Shujuan Wang, *Member, IEEE*, Guangchao Chen, Yanxiang Jiang, *Senior Member, IEEE*, and Xiaohu You, *Fellow, IEEE*

***Abstract*—Data dissemination in Internet of Vehicles (IoVs) is attracting more and more attention. However, few efforts have been made on mixed data dissemination in urban scenario. The rigid Quality of Service (QoS) requirements of Emergency Messages (EMs), the varied QoS tolerant degrees of Service Messages (SMs), and the complex transmission environment in urban scenario pose great challenges. In this paper, a Cluster-based Vehicle-to-Vehicle (V2V) approach for Mixed Data Dissemination (CMDD) is proposed to tackle the aforementioned challenges in urban scenario of IoVs. The proposed approach consists of two main algorithms: the bus-based clustering and the mixed data scheduling. To deal with the complex and harsh transmission environment in urban scenario, we propose to use buses, which have fixed driving routes and schedules as well as sufficient resources, to act as cluster heads. Meanwhile, multiple characteristics are deeply analyzed and artfully utilized in the design of the clustering algorithm, to achieve efficient data dissemination from buses to ordinary vehicles. To meet the rigid and personalized QoS requirements of EMs/SMs under stringent resources, a new priority calculation method is proposed and mixed data are scheduled accordingly. Simulation results show that the proposed approach achieves remarkable performance advantage in terms of data download delay, emergency warning ratio and service response ratio.**

***Index Terms*—Internet of vehicles, mixed data dissemination, clustering, bus-based, intersections.**

I. Introduction

With the continuous development of vehicular technologies, wireless communications [1], sensor networks and Big Data technologies [2], vehicles are becoming more intelligent and autonomous than ever before [3]. The concept of Internet of Vehicles (IoVs) is born to support the realization of a future Intelligent Transportation System (ITS), in which seamless interconnections between vehicles, pedestrians, roadside infrastructures and other transportation components can be realized. Appealing applications and services, such as accident alert [4], cooperative collision avoidance [5], road condition information [6], entertainment services [7] and so on, can be deployed and provided effectively in ITS. For this reason, IoVs have attracted more and more attentions in the aim of improving traffic efficiency and guaranteeing traffic safety.

However, great challenges are faced when transmitting data in IoVs. The fast-moving vehicles introduce unstable communications links and dynamically changing topologies, thus greatly harming the data transmission performance and hinder the development of IoVs. In IoVs, a vehicle can communicate and exchange data packets with other vehicles through on-board units(OBUs) deployed on the vehicles (Vehicle-to-Vehicle, V2V), or with roadside infrastructures (Vehicle-to-Infrastructure, V2I), such as RoadSide Units (RSUs). However, the high cost for the deployment and maintenance of RSUs prevents its pervasive deployment and thus limiting its availability and effectiveness. On the contrary, V2V communications can be easily deployed with low cost, and it is suitable for real-time information interaction between vehicles.

Compared with the highway scenario of IoVs, the urban scenario of IoVs has more complex transportation and communications environments. There are plenty of buses on the roads in urban environment, which have regular traffic characteristics and sufficient computation and communications resources. Meanwhile, numerous vehicles run in the urban environment with arbitarily changing routes. For urban scenario of IoVs, it is natural to use V2V communications as the main way for data dissemination. For one thing, the routes of vehicles in urban area are irregular and complex, making it practically and financially impossible to implement enough RSUs to cover all areas. For another, given the presence of buses in the network, it will be a great waste not utilizing the easily accessible resources of buses.

The data transmitted in IoVs can be classified into two types. One is Emergency Messages (EMs) which are related to safety services, such as accident alert and road condition information. The other is Service Messages (SMs) which are related to non-safety services, such as entertainment service and information query. A great number of efforts have been put on the design of satisfying transmission strategies for pure EMs or SMs in IoVs, whereas little work has been done for transmitting EMs and SMs together. Unfortunately, the nature of IoVs requires

Manuscript received 8 October 2022; accepted 16 October 2022. Date of publication 20 October 2022; date of current version 14 March 2023. This work was supported in part by the National Natural Science Foundation of China under Grants 61962032, 61561029, and 61971129, in part by Yunnan Ten Thousand Talents Plan Young & Elite Talents Project, and in part by Yunnan Province Fund for Excellent Young Scholars under Grant 202001AW070003. The review of this article was coordinated by Dr. Shahid Mumtaz. *(Corresponding authors: Yanxiang Jiang; Xiaohu You.)*

Shujuan Wang and Guangchao Chen are with the School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, Yunnan 650500, China (e-mail: wangsj@kust.edu.cn; chenguangchao0722@163.com).

Yanxiang Jiang and Xiaohu You are with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China (e-mail: yxjiang@seu.edu.cn; xhyu@seu.edu.cn).

Digital Object Identifier 10.1109/TVT.2022.3216236

transmitting EMs and SMs together in a reliable and efficient way, since this transmission scene corresponds to practical needs and regulations on IoV applications.

As a result, it is of vital importance to realize reliable and efficient, V2V-based, mixed EMs and SMs transmission for IoVs, especially in the urban scenario. The key issues that must be solved are as follows. First, since the travel routes and times of different vehicles are arbitrary, how to establish and maintain quality channels between vehicles is essential in the design of V2V-based strategies. Second, the communications and computation capabilities of vehicles are limited, as well as their storage and power resources. Therefore, how to guarantee that vehicles can accomplish the corresponding data transmissions with limited resources is a challenging problem. Third, due to the above mentioned reasons, it is already hard to achieve successful SMs or EMs transmission solely in urban scenario. It is even harder to transmit SMs and EMs together. The fundamental is that the need for transmitting EMs usually comes from an emergency such as traffic accidents, which cannot be predicted. Meanwhile, EMs' transmission needs more stringent performance than SMs, to alert vehicles nearby about the emergency as fast as possible so that proper actions can be taken. As a result, how to efficiently schedule EMs and SMs together to realize timely alert and notification of emergencies, and meet the usual service needs of vehicles at the same time, is the most vital issue.

To address the above problems, this paper is dedicated to studying the V2V-based mixed EMs and SMs transmission in an urban environment for IoVs. The main contributions of this paper can be summarized as follows:

1) To overcome the transmission challenges brought in by the fact that vehicles travel arbitrarily and are subject to limited resources, the characteristics of vehicular communications in urban scenario are studied thoroughly and a bus-based clustering algorithm is designed. Multiple characteristics, such as path similarity, link reliability, average relative velocity, average relative distance, etc., have been deeply analyzed and artfully utilized in the design of the clustering algorithm. Efficient and reliable data dissemination from buses to ordinary vehicles is accomplished with this algorithm.
2) To efficiently solve the problems resulting from the complex mixed EMs and SMs transmission, the important factors affecting the transmission performance of EMs and SMs are analyzed comprehensively, and a tailored scheduling algorithm is proposed. The algorithm focuses on the timeliness characteristics of both EMs and SMs by exploring the implicit time constraints of safety and non-safety applications in IoVs. Moreover, various factors such as data popularity, link reliability and distance, are analyzed and smartly integrated in the algorithm. Correspondingly, both EMs and SMs transmissions can be delivered with satisfying QoS for the considered applications.

The rest of this paper is organized as follows: Section II reviews related works. Section III introduces the system model. Section IV explains the proposed approach in detail. Section V provides extensive simulation results and evaluates the proposed scheme. Finally, Section VI concludes this paper.

II. Related Works

A. Cluster-Based Data Dissemination for IoVs in Urban Scenario

To improve the efficiency of V2V-based data transmission, the authors in [8] proposed an efficient clustering algorithm based on particle swarm optimization in Vehicular Ad Hoc Networks (VANETs). In [9], a clustering based data propagation protocol was proposed, where the most reliable vehicle was selected as the cluster head. To solve the problem of short service connection time under fast vehicle movement, the authors in [10] proposed a new data transmission optimization strategy under Vehicular Cyber-Physical Systems (VCPS) based on random alliance game. The authors in [11] proposed a 5G architecture to deal with the problems of frequent topology change and overlong transmission delay in VANETs, where all vehicles in the area were grouped into clusters using an Adaptive Motion-Aware Path Similarity (A-MAPS) algorithm. To establish reliable V2V communications, a Bayesian Rules Fuzzy Logic (BRFL) based algorithm was introduced to determine the best V2V forwarding node, which can achieve better performance in packet transmission rate and transmission delay, compared with previous works.

In summary, the existing work often focus on improving the transmission reliability and efficiency through designing better clustering algorithms, which select cluster heads or members with more appropriate and intelligent methods. However, in doing so these schemes increase the overhead and complexity of clustering, posing big challenges to the tough V2V transmission environment. Meanwhile, the high dynamic mobility of ordinary vehicles can lead to frequent re-clustering and decreasing cluster stability, thus degrading the overall performance.

B. EMs/SMs Dissemination in IoVs

1) EMs Dissemination: Safety is the core of IoVs and it relies heavily on the reliability and efficiency of EMs dissemination. In [12], the authors proposed a reliable transmission mechanism based on road layout information according to the characteristics of the intersection area, which solved the problem of EMs transmission failure in a certain direction caused by obstacles. The authors in [13] proposed an EMs communications scheme, which determines the region of interest according to the information contained in EMs. To ensure the transmissions of vehicles that have the maximum number of EMs, the authors in [14] proposed an effective emergency information transmission scheme, which selected the appropriate relay vehicles according to the mobility index of vehicle nodes and link reliability. In [15], the authors proposed a link quality-based scheme for urban VANET safety information transmission, which estimated the connectivity between vehicles based on the physical channel connectivity calculation method. In order to increase EMs propagation reliability and reduce packet loss rate, the authors in [16] proposed a RBEM/CBEM handshake mechanism to enhance broadcast performance, which showed good feasibility and effectiveness in EMs transmission process. The authors in [17] proposed a two-way cooperative transmission strategy for transmitting accident video EMs in the highway scenario of IoVs. Vehicles in the same direction of accident vehicle are formed into clusters and communicate within clusters, while vehicles in the opposite direction select relay vehicles to help spreading video fast and reliably.

2) SMs Dissemination: Traffic efficiency and travel comfort are also the goals of IoVs, which depend on the successful transmission of SMs. In [18], a bus track based advertisement publishing scheme was proposed, where the interference and data conflict in the process of advertisement transmission were avoided by using public transport and private vehicles. The authors in [19] used fuzzy logic to calculate the ability of vehicles to act as relay vehicles in RSUs' communication range, and determined the proportion of selected relay vehicles according to vehicle density. In [20], the authors proposed a cooperative data distribution algorithm based on multi-hop relay in IoV environments, which uses Device-to-Device (D2D) communications [21] to achieve flexible terminal communication between vehicles, and adaptively selects the optimal relay, route and transmission hops according to link quality.The authors in [22] proposed a model based V2V communications strategy, taking into account the real-time registration and exchange of travel time between vehicles. A bidirectional broadcast protocol was proposed in [23], which made a trade-off between broadcast reliability and coverage capability, and effectively improved the end-to-end delay.

3) Mixed Data Dissemination: As analyzed in Section I, the most common and practical data transmission scene in IoVs is that SMs and EMs are mixed and both have specific QoS requirements. There are few exploratory researches aiming at this direction.

The authors in [24] proposed two open-loop congestion control strategies, including DySch and TaSch strategies. Static and dynamic factors are calculated according to the information content and vehicle conditions respectively. Each message is assigned a priority and dynamic scheduling is carried out according to the priority. The authors in [25] proposed a scheduling strategy, which schedules messages according to their expiration date, size, static factors (message content) and vehicle speed. The authors of [26] proposed an Adaptive Priority data service Scheduling (AdPS) algorithm, which employs fuzzy logic for deadline estimation and priority ordering.

In summary, the existing work usually focuses on improving the transmission delay and reliability of EMs or SMs dissemination via carefully considering the characteristics of network and vehicles mobility, as well as analyzing the specific performance requirements of different types of data. However, these solutions are not suitable for mixed data dissemination in complex urban environment. First, most of the existing schemes rely on establishing clusters or selecting proper relay nodes to enhance the transmission performance of single type message (pure EMs or SMs). However, the establishment of clusters or the selection of relay vehicles are usually stimulated by the occurrence of emergency accident or the generation of service need. Therefore, extra delay brought by the process of clustering and relay selection is unavoidable. Second, due to the complex road environment, dense vehicles, limited communications resources and harsh link conditions in urban scenario, most of the current solutions lay particular emphasis on some part of the influencing factors and are not well applicable to solve the above problems in urban environment. Third, most of the existing works consider EMs and SMs transmission separately and thus are not suitable for practical applications in real urban environment. Although there are some works that have made attempts to transmit EMs and SMs simultaneously, they only focused on the design of scheduling algorithms which transfer messages in the light of priorities that are calculated according to several parameters. The lack of exploiting the implicit characteristics and the inherent QoS requirements of mixed EMs and SMs transmission limits their performance, keeping them from providing satisfactory QoS and becoming applicable to practical applications.

In view of this, we propose a cluster-based V2V approach for mixed data dissemination in urban scenario of IoVs. The network characteristics, vehicle movement properties, as well as the complex urban environment, are deeply analyzed. To resolve the problem of extra delay brought by the clustering process in previous works, a bus-based clustering algorithm is designed in which clusters are formed in advance. The Cluster Heads (CHs) are responsible for responding the non-safety service requests of Cluster Members (CMs) in normal times, and also for spreading out the emergency messages as soon as possible when necessary. To achieve fast emergency alert, while keeping acceptable QoS for SMs transmissions, the implicit and inherent characteristics of mixed data dissemination are thoroughly analyzed. The interactions between EMs and SMs transmissions are given full consideration and their specific QoS requirements are respected. The optimization of mixed data dissemination is accomplished through integrating the time constraints of EMs and SMs into the scheduling process.

III. System Model

In this paper, we focus on the V2V based mixed data dissemination in urban scenario of IoVs. The traffic situation at the intersections are typical and representative, thus, we select the intersection model to launch our study. The system model is illustrated in Fig. 1. The considered scenario is a typical urban area with gridded roads and multiple intersections. Each avenue in this area has two-way multiple lanes and dense traffic. Vehicles and buses are randomly distributed in the Range Of Interested (ROI).

It is assumed that each vehicle is equipped with an on-board unit, through which vehicles can communicate with each other within a range. Moreover, each vehicle is equipped with Global Positioning System (GPS) and multiple sensors to collect information such as real-time positions, speeds, surrounding road conditions of adjacent vehicles. These information are periodically exchanged between vehicles. Each vehicle works in half duplex mode, thus it cannot sends and receives data at the same time.

The mobility model used in this work is the full velocity difference model for car following [27]. Each vehicle is assigned a random speed according to a uniform distribution at the beginning, and overtaking is not allowed. Furthermore, each vehicle

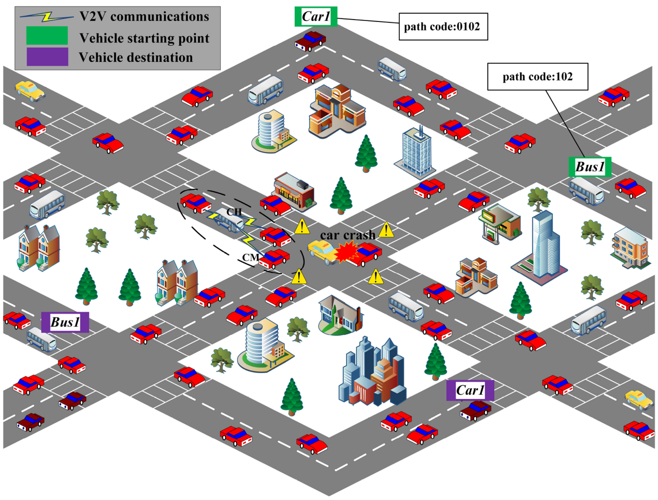

Fig. 1. V2V-based mixed EMs/SMs data dissemination in urban scenario of IoVs.

must keep a safe distance from the previous one, and proper actions (acceleration/deceleration) need to be taken to restore this distance once the safe distance is violated.

Under ordinary conditions, vehicles and buses can smartly form clusters and accomplish data services for specific applications, which is known as SMs transmission. For example, vehicle c_j wants to determine the best restaurant within 1 km and it sends out a query packet to the CH it connects with. After receiving the query packet, the CH will respond c_j the wanted information according to certain data scheduling policy. However, emergency events often occur and force immediate actions upon them. As shown in Fig. 1, two vehicles collide in the center of the ROI, and it is important to timely alert vehicles that are running toward this area about this emergency. If the emergency warnings fail to be delivered successfully to vehicles in the ROI within certain time constraints, such kind of accident may result in a series of troubles, such as rear-ended accidents, congestion and chaos. At the moment this accident occurs, the accident vehicles will generate an accident warning message, denoted as EM, to multicast and notify other vehicles about this emergency. Obviously, the vehicles in the ROI hope to be alerted by the EM as soon as possible so that they can have enough time to make the most appropriate actions. However, it is impossible to realize this goal without a highly efficient and reliable transmission mechanism. Meanwhile, the unfinished SMs data transmission still need to be taken care, despite the occurrence of the accident. In other words, upon the accident, it is urgent to alert c_j about the accident, while its restaurant query should also be responded before it runs too far away. The feasibility of IoVs in such scenario depends heavily on the transmission performance of both EMs and SMs.

IV. The Cluster-Based Mixed Data Dissemination in Urban Scenario of IoVs

To satisfy mixed EMs and SMs transmissions, we propose a Cluster-based Mixed Data Dissemination (CMDD) approach. The approach consists of two main strategies, namely the bus-based clustering algorithm and the mixed data scheduling algorithm. These two strategies work closely and complement each other. In the following, we introduce the approach in detail. The notations of symbols used in this paper are listed in Table I.

TABLE I
Notations of Symbols

R	Vehicles' communications radius
(x_0, y_0)	The coordinates of the accident spot
d_{ij}	The distance between b_i and c_j
$\mathcal{N}_i$	Set of b_i's neighboring nodes
$\overline{\Delta d_i}$	Average relative distance from b_i to its neighboring nodes
$\overline{\Delta v_i}$	Average relative speed between b_i and its neighboring nodes
$\overline{\Delta o_i}$	Average path similarity of b_i and its neighboring nodes
p_i	Successful probability of transmitting a message from b_i to a vehicle
$\Delta a_{ij}(t)$	Relative acceleration between b_i and c_j at time t
$\Delta v_{ij}(t)$	Relative velocity between b_i and c_j at time t
$T_{ij}^{connect}$	Accessible communication time between b_i and c_j
$\theta_{ij}(t)$	Directional angle of b_i and c_j at time t
$t_{ij}^{connect}$	The moment when b_i and c_j start to communicate
$t_{ij}^{disconnect}$	The moment when b_i and c_j disconnect
$\overline{l_i}$	Average link reliability of b_i and its neighboring nodes
T_j^{valid}	The valid time period of the EM message for a vehicle c_j
p_k	The data item a_k's priority
h_k	a_k's popularity
$\overline{l(a_k)}$	The average link reliability between the CH and all CMs who want a_k
$\overline{\Delta v(a_k)}$	The average relative speed between the CH and all CMs who want a_k
$t_{min}^{deadline}(a_k)$	The minimum time constraint for all CMs that ask for a_k
$d_{max}(a_k)$	The maximum distance between the CH and all CMs that ask for a_k

A. Bus-Based Clustering

In this work, we propose to use appropriate buses as cluster heads. Meanwhile, ordinary vehicles are promoted to choose the most suitable clusters to join in as their cluster members. As buses have regular routes and sufficient resources, the proposed algorithm can not only forms clusters before the need to transmit EMs arises, but also provides stable and uninterrupted channel conditions for achieving reliable V2V communications. The proposed algorithm considers the mobility characteristics of vehicles, communications characteristics, as well as network properties, to select the most appropriate buses as CHs.

1) Vehicle Mobility Indicators: The distance between two vehicles is an important index to evaluate the communications success ratio. The closer the distance, the higher the probability that the communications can be successfully completed. Besides, it is known that when two vehicles move at close velocities, relative stable channel condition can be achieved between them which is conducive to transmission reliability. Denote the i^{th}

bus as b_i, the j^{th} ordinary vehicle c_j. Then, the distance between them can be expressed as follows:

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}. \quad (1)$$

Denote the set of neighboring nodes of b_i as $\mathcal{N}_i$, the average relative distance from b_i to its neighboring nodes can be calculated as follows:

$$\overline{\Delta d_i} = \frac{\sum_{j \in \mathcal{N}_i} d_{ij}}{|\mathcal{N}_i|}. \quad (2)$$

Similarly, the average relative speed of b_i and its neighboring nodes can be expressed as follows:

$$\overline{\Delta v_i} = \frac{\sum_{j \in \mathcal{N}_i} |v_i - v_j|}{|\mathcal{N}_i|}, \quad (3)$$

where v_i and v_j are the velocities of b_i and c_j, respectively.

2) Path Similarity: Path similarity is defined as the routes overlap degree between two vehicles. Obviously, forming vehicles with similar routes into one cluster is beneficial to cluster stability, thus mitigating the unpleasant performance degradation caused by frequent cluster update and maintenance. Assume that each vehicle is aware of its route to the destinations, under the support of navigation system on board. The route of a vehicle can be encoded based on the action taken by the vehicle in each intersection in the route, according to the following rules:[1]

0: the vehicle keeps going straight;

1: the vehicle turns left and keeps going straight;

2: the vehicle turns right and keeps going straight;

In doing so, the route is encoded as a sequence number, and it is further converted into binary codes one by one. The converted binary codes are then XORed to calculate the similarity. Here we give an example to explain how the path similarity is calculated.Take *Car1* and *Bus1* in Fig. 1 as an example. The starting points of *Car1* and *Bus1* are denoted as green blocks, and their destinations are painted in purple in Fig. 1. *Car1* wants to get to the place located in the right bottom corner in Fig. 1, and it makes a routing plan in advance, which is encoded as "0102". Each number of this code indicates the exact action that *Car1* will take when encountering an intersection. Thus, the specific route *Car1* will travel according to the code can be interpreted as follows: *Car1* firstly keeps running straight until entering the first intersection, it then turns on the left and keeps going, and at the second intersection, it keeps going straight. *Car1* turns on the right at the third intersection and it finally arrives at its destination. Similarly, we can get the route code for *Bus1* as "102". After that, the route code is converted into binary code. For the route code of *Car1*, its path code is "00010010," and the path code of *Bus1* is "010010". The path similarity between *Car1* and *Bus1* is then calculated as follows.

Firstly, the path codes are left aligned and the redundant bits are counted and removed. For the given example, the redundant bits are the last 2 bits in the path code of *Car1*, and they are counted and removed from the code. Then, the updated path code of *Car1* is "000100," and it is XORed with the path code of *Bus1*, the result is "010110". The number of "1" in the result is counted. Finally, the counted numbers are added and the sum is an indicator to describe the differences between two vehicles' paths. In this example, the length difference between the binary codes is counted as "2," and the number of "1" in the XORed result is counted as 3. Then the path difference between *Bus1* and *Car1* can be calculated to be $ps_{11} = 2 + 3 = 5$.

Similarly, we can get the path difference between bus b_i and each of its neighboring vehicles. Then, the average path similarity between b_i and its neighboring vehicles is calculated as:

$$\overline{\Delta o_i} = 1 - \frac{\sum_{j=1}^{|\mathcal{N}_i|} ps_{ij}}{|\mathcal{N}_i| \cdot L_i}, \quad (4)$$

where L_i is the length of the converted binary code of bus b_i's path.

3) Link Reliability: Due to the high mobility and the fact that a large number of reflective objects exist in urban environment, the communications between vehicles are probabilistic and the radio intensity changes with time in certain distances. In real wireless environment test, Nakagami-m distribution provides better matching degree with practical environment [28], thus we consider Nakagami-m distribution to describe the fast-fading vehicular channels. The successful probability of transmitting a message from b_i to c_j is [29]:

$$p_i = 1 - Fd\left(RT, m, \varphi\right) = e^{\frac{-mRT}{\varphi}\sum_{i=1}^{m}\frac{\left(\frac{-mRT}{\varphi}\right)^{i-1}}{(i-1)!}}, \quad (5)$$

where $Fd(RT, m, \varphi)$ represents the Cumulative Distribution Function (CDF) when the received signal power is smaller than RT, RT is the signal receiving threshold, φ is the average power of the received signal, and m is the fading parameter which is a function of the distance between vehicles. They can be expressed respectively as follows:

$$RT = \frac{P_T}{R^2} \cdot \frac{G_t G_r \gamma^2}{16\pi^2 L}, \quad (6)$$

$$\varphi = \frac{P_T}{(d_{ij})^2} \cdot \frac{G_t G_r \gamma^2}{16\pi^2 L}, \quad (7)$$

$$m = \begin{cases} 1, d_{ij} \geq 150\,m \\ 1.5, 50m \leq d_{ij} < 150\,m \\ 3, d_{ij} < 50\,m \end{cases}, \quad (8)$$

where P_T is the transmission power, G_t and G_r are the antenna gains of transmitter and receiver, and γ and L denote the wavelength and path loss, respectively.

The accessible communications time between b_i and c_j is another parameter to describe the link reliability. Denote the relative acceleration and velocity between b_i and c_j at time t as $\Delta a_{ij}(t) = a_i(t) - a_j(t)$, $\Delta v_{ij}(t) = v_i(t) - v_j(t)$. The accessible communications time can be calculated differently in three cases.

Case 1: $\Delta v_{ij}(t) = 0$, and $\Delta a_{ij}(t) = 0$. This is the simplest case when b_i and c_j keep running at the same speed and acceleration. As a result, the accessible communications time between them is infinite.

[1]If a vehicle turns around at an intersection, this intersection will be considered as its destination and the updated route will be encoded.

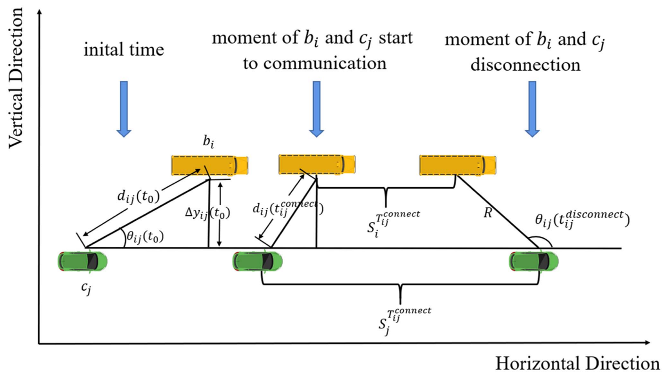

Fig. 2. The calculation of accessible communications time between b_i and c_j.

Case 2: $\Delta v_{ij}(t) \neq 0$, and $\Delta a_{ij}(t) = 0$. This means that b_i and c_j run with constant accelerations. In this case, according to Fig. 2, when the distance between b_i and c_j reaches R, we have:

$$\Delta v_{ij}\left(t_{ij}^{connect}\right) \cdot T_{ij}^{connect} + d_{ij}\left(t_{ij}^{connect}\right) \cdot \cos\theta_{ij}\left(t_{ij}^{connect}\right) = R \cdot \cos\theta_{ij}\left(t_{ij}^{disconnect}\right). \quad (9)$$

where $\Delta v_{ij}(t_{ij}^{connect})$ denotes the relative velocity between bus b_i and vehicle c_j when they establish a communication link between them. $d_{ij}(t_{ij}^{connect})$ stands for the distance between b_i and c_j at time $t_{ij}^{connect}$, and $\theta_{ij}(t_{ij}^{connect})$ and $\theta_{ij}(t_{ij}^{disconnect})$ represent the directional angles between b_i and c_j when they start to communicate and when they disconnect the link, respectively. Therefore,

$$T_{ij}^{connect} = \frac{R \cdot \cos\theta_{ij}\left(t_{ij}^{disconnect}\right) - d_{ij}\left(t_{ij}^{connect}\right) \cdot \cos\theta_{ij}\left(t_{ij}^{connect}\right)}{\Delta v_{ij}\left(t_{ij}^{connect}\right)}. \quad (10)$$

Case 3: $\Delta v_{ij}(t) \neq 0$, and $\Delta a_{ij}(t) \neq 0$. This is the most general case in practical scenario. b_i and c_j run with different initial speeds and varied accelerations.

Given the mobility model, the optimal velocity for c_j at time t is: $v_j^{opt}(t) = \frac{v_{max}}{2}[tanh(\Delta S_j(t) - S_{safe}) + tanh(S_{safe})]$, and the acceleration of c_j at time t is: $a_j(t) = k(v_j^{opt}(t) - v_j(t)) + \varepsilon \cdot \Delta v_j(t)$, where v_{max}, S_{safe}, $\Delta S_j(t)$, $\Delta v_j(t)$ are the maximum allowable velocity, the minimum safe car-following distance, the distance between c_j and its previous vehicle at time t and the velocity difference between c_j and its previous vehicle at time t, respectively, k is the driver reaction coefficient and ε is the sensitivity coefficient. Therefore, the driving distance of c_j in the following time slot $[t, t+\tau]$ can be calculated as follows:

$$S_j(t+\tau) = S_j(t) + v_j(t) \cdot \tau + 0.5 \cdot a_j(t) \cdot \tau^2, \quad (11)$$

and the driving distance of b_i in time slot $[t, t+\tau]$ can be calculated the same as that in equation (11).

As shown in Fig. 2, denote the distance between b_i and c_j at time t_0 as $d_{ij}(t_0)$, the directional angle of b_i and c_j as $\theta_{ij}(t_0)$, the distance between b_i and c_j at time $t_0 + \tau$ can be calculated as follows:

$$d_{ij}(t_0+\tau) = \frac{S_i(t_0+\tau) - S_j(t_0+\tau) + d_{ij}(t_0) \cdot \cos\theta_{ij}(t_0)}{\cos\theta_{ij}(t_0+\tau)}. \quad (12)$$

Define the indicator functionblue I_{ij} as follows:

$$I_{ij} = \begin{cases} 1, b_i \text{ runs in front of } c_j \\ -1, \text{else} \end{cases}. \quad (13)$$

When the distance between b_i and c_j becomes longer than the communications radius R, the link disconnects and we have:

$$S_i(t_0+\tau) - S_j(t_0+\tau) + d_{ij}(t_0) \cdot \cos\theta_{ij}(t_0) = R \cdot I_{ij} \cdot \cos\theta_{ij}(t_0+\tau). \quad (14)$$

Denote the time when b_i and c_j start to communicate as $t_{ij}^{connect}$, and the time when they disconnect is denoted as $t_{ij}^{disconnect}$. Thus, the time period from $t_{ij}^{connect}$ to $t_{ij}^{disconnect}$ is the total time that b_i and c_j communicate with each other, which can be denoted as $T_{ij}^{connect}$. The distance that bus b_i can drive in the time period $T_{ij}^{connect}$ is calculated as $S_i^{T_{ij}^{connect}} = \int_{t_{ij}^{connect}}^{t_{ij}^{disconnect}} v_i(t)dt$, then we have:

$$S_i^{T_{ij}^{connect}} - S_j^{T_{ij}^{connect}} = 0.5\Delta a_{ij}(T_{ij}^{connect})^2 + \Delta v_{ij}(T_{ij}^{connect}), \quad (15)$$

where $T_{ij}^{connect} = t_{ij}^{disconnect} - t_{ij}^{connect}$. Further, if b_i is in front of c_j, then:

$$S_i^{T_{ij}^{connect}} - S_j^{T_{ij}^{connect}} + d_{ij}\left(t_{ij}^{connect}\right) \cdot \cos\theta_{ij}(t_{ij}^{connect}) = R \cdot \cos\theta_{ij}\left(t_{ij}^{disconnect}\right), \quad (16)$$

and if b_i is behind c_j,

$$S_i^{T_{ij}^{connect}} - S_j^{T_{ij}^{connect}} + d_{ij}\left(t_{ij}^{connect}\right) \cdot \cos\theta_{ij}\left(t_{ij}^{connect}\right) = -R \cdot \cos\theta_{ij}\left(t_{ij}^{disconnect}\right). \quad (17)$$

Applying formula (15), (16) and (17) simultaneously, we have formula (18) shown at the bottom of this page, where

$$T_{ij}^{connect} = \begin{cases} \frac{-\Delta v_{ij}\left(t_{ij}^{connect}\right) \pm \sqrt{\left(\Delta v_{ij}\left(t_{ij}^{connect}\right)\right)^2 + 2\Delta a_{ij}\left(t_{ij}^{connect}\right)\left(\left(R \cdot \cos\theta_{ij}(t_{ij}^{disconnect}) - d_{ij}\left(t_{ij}^{connect}\right) \cdot \cos\theta_{ij}\left(t_{ij}^{connect}\right)\right)\right)}}{\Delta a_{ij}(t_{ij}^{connect})}, \\ x_{b_i} > x_{c_j} \, at \, time \, t_{ij}^{connect} \\ \frac{-\Delta v_{ij}(t_{ij}^{connect}) \pm \sqrt{\left(\Delta v_{ij}\left(t_{ij}^{connect}\right)\right)^2 - 2\Delta a_{ij}\left(t_{ij}^{connect}\right)\left(\left(R \cdot \cos\theta_{ij}\left(t_{ij}^{disconnect}\right) + d_{ij}\left(t_{ij}^{connect}\right) \cdot \cos\theta_{ij}\left(t_{ij}^{connect}\right)\right)\right)}}{\Delta a_{ij}\left(t_{ij}^{connect}\right)}, \\ x_{b_i} < x_{c_j} \, at \, time \, t_{ij}^{connect} \end{cases} \quad (18)$$

$\theta_{ij}(t_{ij}^{connect} + T_{ij}^{connect}) = \theta_{ij}(t_{ij}^{disconnect})$ represents the directional angle between b_i and c_j at time $t_{ij}^{disconnect}$ (the moment when b_i and c_j disconnects).

Assume that vehicles are not allowed to change lanes during the driving period, thus the vertical distance between b_i and c_j is a constant. Then, we have:

$$\begin{aligned} \Delta y_{ij}\left(t_{ij}^{connect}\right) &= \Delta y_{ij}\left(t_{ij}^{disconnect}\right), \\ \theta_{ij}\left(t_{ij}^{disconnect}\right) &= \arcsin\left(\frac{\Delta y_{ij}\left(t_{ij}^{disconnect}\right)}{R}\right) \\ &= \arcsin\left(\frac{\Delta y_{ij}\left(t_{ij}^{connect}\right)}{R}\right). \end{aligned} \tag{19}$$

Therefore, for any bus B_i with its neighboring vehicles, the average link reliability is calculated as:

$$\overline{\Delta l_i} = \frac{\sum_{j\in\mathcal{N}_i}(p_i \cdot T_{ij}^{connect})}{|\mathcal{N}_i|}. \tag{20}$$

The average link reliability is a good indicator for describing the fast-fading vehicular channels between bus b_i and its neighboring vehicles. A big $\overline{\Delta l_i}$ refers to a reliable channel and a small one indicates a bad channel where packet loss happens frequently. Meanwhile, the average relative distance and the average relative velocity between b_i and its neighboring vehicles are important parameters to reflect the transmission reliability of the V2V connections among b_i and its neighboring vehicles. It is commonly known that stable channels can be established between vehicles with small average relative distance and low average relative velocity. Moreover, to improve the cluster stability so that clustering cost can be reduced, it is favored to select a bus that has the most similar path with its neighboring vehicles as CH. In summary, the proposed algorithm takes the transmission reliability, the cluster stability and the multicast gain (the number of neighboring vehicles) into the consideration of selecting the most suitable bus as the CH. The suitability of bus b_i for working as a cluster head is formulated as its utility, which can be calculated as follows:

$$u_i = \frac{|\mathcal{N}_i| \cdot \overline{\Delta o_i} \cdot \overline{\Delta l_i}}{\overline{\Delta d_i} \cdot \overline{\Delta v_i}}. \tag{21}$$

To achieve efficient and reliable data dissemination between a cluster head and it's members, it is preferred to choose a bus that has a large number of neighboring vehicles which shares same path, and has stable link conditions with its neighboring nodes, to be the CH. Meanwhile, it is favored that the bus and its neighboring vehicles move at similar speeds and closely with each other. If there exists many buses in a small area (with overlapped communications coverage), the suitability of each bus will be calculated according to formula (21), a bus with the largest suitability value is then selected as the cluster head, and all ordinary vehicles that have connections to the CH become the cluster members.

After the primary clustering process, if there are some vehicles that fail to join in any cluster (no bus can be connected with), a secondary clustering process will be executed. For any vehicle c_j which has no bus to connect with, its capability to work as a CH can be calculated as follows:

$$u_j = \frac{|\mathcal{N}_j| \cdot d_j}{v_j}, \tag{22}$$

where d_j denotes the distance between c_j's current location and the intersection (the accident spot). The vehicle with the largest u is selected as the CH, and its neighboring vehicles become members accordingly.

Once forming a cluster, CH will periodically multicast the Cluster Head Announcement (CHA) to its neighboring vehicles, and the vehicles which are running within its coverage will send back an acknowledgement message upon receiving the CHA packet, including their IDs, instant speeds, locations, etc. CH then checks the received acknowledgement packets, and updates the CM vehicles' attributes accordingly. Moreover, if an acknowledgement packet is sent by a vehicle that is not a CM yet, CH will recalculates its utility according to formula (21), and sends the updated utility value to this vehicle, to invite it to join in the cluster. Upon receiving the invitations (a vehicle can receive multiple invitations from multiple cluster's heads nearby), the vehicle compares their utilities and chooses the cluster with the largest utility to join in. Meanwhile, if a CH does not receive a CM's acknowledgement packet after the CHA is multicasted, it will consider the CM as a failure node and remove it from the CM list. The pseudocode of the proposed bus-based clustering algorithm is shown in Algorithm 1.

The utility update of bus b_i requires to calculate the path similarity, link reliability, relative distance and relative velocity between b_i and each of its neighboring vehicle. The complexity is of the order $\mathcal{O}(N)$ and N is the number of ordinary vehicles. Meanwhile, a vehicle c_j compares the updated utility value of b_i with other buses' utility values, which consumes time of the order $\mathcal{O}(M\log M)$, where M is the number of buses. As a result, the complexity of Algorithm 1 is of the order $\mathcal{O}(M\log M)$.

B. Mixed Data Scheduling

For the urban IoVs environment, vehicles can request specific data items anytime anywhere to get comfort and fun driving experience. Normally, a CH will respond to the requests from its CMs efficiently with specially designed purposes. However, if there is an emergency accident occurs in the network, every vehicle in the ROI should be alerted as fast as possible. Therefore, the dissemination of the EM alert becomes the first priority that may interrupt the execution of SM response process. Hence, the design of a smart algorithm that can handle mixed EM/SMs dissemination is very important.

The core of the EM dissemination process lies in the fact that the emergency alert message should be transmitted successfully to any vehicle before it arrives at the accident spot. In other words, there is a strict time constraint on the EM delivery delay for each related vehicle in the ROI. The valid time period of the EM message for a vehicle c_j, denoted as T_j^{valid}, is estimated as the time period that c_j takes to drive from the current location to the accident spot.

Algorithm 1: Bus-Based Clustering Algorithm.

input: ROI, N vehicles, M buses
output: Clusters
for each bus in the ROI **do**
  for each time slot **do**
    A bus b_i broadcasts a CHA message;
    for vehicles within b_i's communication radius **do**
      if A vehicle c_j receives the CHA message **then**
        c_j responses b_i with an acknowledgement message;
      end if
    end for
    for each received acknowledgement message **do**
      if the sending vehicle c_j is not on the neighboring nodes list of b_i **then**
        b_i calculates it's utility value according to formula (21) and broadcasts the updated utility;
        c_j receives the update utility and compares it with other buses' utility values;
        if b_i's utility value is the greatest **then**
          c_j recognizes b_i as the CH and joins in the cluster;
        end if
      end if
    end for
    for all received acknowledgement messages **do**
      b_i checks if all member vehicles have responded;
      if a member vehicle c_k fails to send the ACK message **then**
        b_i removes c_k from the cluster and updates it's neighboring nodes list;
      end if
    end for
  end for
end for

Given the vehicular mobility model and the network model, the time period for a vehicle to drive from location A to location B can be estimated as follows. Firstly, the motion of a vehicle in a very small time step τ is considered as an uniformly accelerated motion.Then, we have:

$$d_j^{A,B} = \sum_{n=0}^{C} \left[v_j\left(n\cdot\tau\right)\cdot\tau + 0.5\cdot a_j\left(n\cdot\tau\right)\cdot\tau^2 \right], \tag{23}$$

where C is the counts of time steps that c_j will consume to arrive at the spot. On the other hand, we have:

$$\begin{aligned} &v_j\left(C\cdot\tau\right) = v_j\left((C-1)\cdot\tau\right) + a_j\left((C-1)\cdot\tau\right)\cdot\tau,\\ &v_j\left((C-1)\cdot\tau\right) = v_j\left((C-2)\cdot\tau\right) + a_j\left((C-2)\cdot\tau\right)\cdot\tau,\\ &\qquad\vdots\\ &v_j\left(2\cdot\tau\right) = v_j\left(1\cdot\tau\right) + a_j\left(1\cdot\tau\right)\cdot\tau,\\ &v_j\left(1\cdot\tau\right) = v_j(0) + a_j(0)\cdot\tau,\\ &v_j(0) = v_j(0). \end{aligned} \tag{24}$$

therefore, $v_j(n\cdot\tau) = v_j(0) + \sum_{n=0}^{C-1} a_j(n\cdot\tau)\cdot\tau$. Then, we have:

$$d_j^{A,B} = \sum_{n=0}^{C} \left\{ \left[\left(v_j(0) + \sum_{n=0}^{C-1} a_j\left(n\cdot\tau\right)\right)\cdot\tau \right]\cdot\tau + 0.5\cdot a_j(n\cdot\tau)\cdot\tau^2) \right\}. \tag{25}$$

When an accident happens on the intersection, the moving speeds of all vehicles in the ROI decrease gradually and the vehicles stop moving finally in most cases. The acceleration of each vehicle in each time slot also decreases slowly. Moreover, in view of the fact that each vehicle during this period is driving on an uniformly accelerated motion, the value of c_j's acceleration in the following C time slots can be approximated as its acceleration in the current time slot t_0. Then, we have:

$$\begin{aligned} a_j(t_0) &= a_j(t_0 + 1\cdot\tau)\\ &= a_j(t_0 + 2\cdot\tau)\\ &= \ldots\\ &= a_j(t_0 + (C-1)\cdot\tau). \end{aligned} \tag{26}$$

Furthermore, we have:

$$\begin{aligned} &\sum_{n=0}^{C-1} \left(a_j\left(t_0 + n\cdot\tau\right)\cdot\tau\right)\\ &= a_j(t_0)\cdot\tau + a_j\left(t_0 + 1\cdot\tau\right)\cdot\tau + \cdots\\ &+ a_j\left(t_0 + (C-1)\cdot\tau\right)\cdot\tau\\ &\cong C\cdot a_j(t_0)\cdot\tau. \end{aligned} \tag{27}$$

Correspondingly,

$$C = \frac{-[a_j(t_0)\cdot\tau + v_j(t_0)\cdot\tau + 0.5a_j(t_0)\cdot\tau^2 \mp \lambda^{\frac{1}{2}}]}{2a_j(t_0)\cdot\tau^2}. \tag{28}$$

where $\lambda = [a_j(t_0)\cdot\tau + v_j(t_0)\cdot\tau + 0.5a_j(t_0)\cdot\tau^2]^2 - 4a_j(t_0)\cdot\tau\cdot\tau^2[v_j(t_0)\cdot\tau + 0.5a_j(t_0)\cdot\tau^2 - d_j^{A,B}]$. Therefore, the estimated valid time period of the EM for vehicle c_j is:

$$T_j^{valid} = floor(C\cdot\tau). \tag{29}$$

For the mixed data dissemination, several important characteristics should be taken into consideration. By considering transmission efficiency, a series of parameters are analyzed, including data priority, data popularity, and the valid time period of data items. Firstly, the EM data clearly is more urgent than the SM data, thus high priority should be assigned on the EM message. Secondly, the data popularity is another vital property to consider when data is scheduled. Data popularity is defined as the number of requests that ask for the same data item. Obviously, the multicast of data item with high popularity is more efficient than that of data item with low popularity. Thirdly, the valid time period of data item is critical for satisfying service requests of vehicle users, especially for EM transmission. In consideration of transmission reliability, multiple characteristics

Algorithm 2: Mixed Data Scheduling.

for each time slot **do**
  if CH bus b_i receives SM requests from member vehicles or EM alert from other CHs **then**
    b_i puts the requested data items/EM into the scheduling queue;
    b_i calculates the scheduling utilities of each item in the scheduling queue according to formula (30);
    b_i sorts the data item in the queue according to their utilities and broadcasts the one with the largest utility;
  end if
end for

are explored, including the channel quality, the relative speed, as well as the distance between a CH and a CM.

Denote p_k as the data item a_k's priority, h_k as a_k's popularity (hotness). Thus, the utility of scheduling data item a_k can be calculated as:

$$u_k = \frac{p_k \cdot h_k \cdot \overline{\Delta l(a_k)}}{t_{min}^{deadline}(a_k) \cdot \overline{\Delta v(a_k)} \cdot [(R - d_{max}(a_k))]}, \tag{30}$$

where $\overline{\Delta l(a_k)}$ and $\overline{\Delta v(a_k)}$ denote the average link reliability and average relative speed between the CH and all CMs who want a_k, respectively. $t_{min}^{deadline}(a_k)$ represents the minimum time constraint for all CMs that ask for a_k (when a_k stands for an EM, $t_{min}^{deadline}(a_k)$ represents the shortest valid time of this EM among all CMs in the current cluster, on current time slot), and $d_{max}(a_k)$ means the maximum distance between the CH and all CMs that ask for a_k. R is the transmission radius of the CH. Thus, the CH calculates the utility value of each data item in the queue and schedules the one with the highest utility to broadcast to all members. The pseudocode of the proposed mixed data scheduling algorithm is shown in Algorithm 2.

The scheduling utility calculation of each item in the queue requires to calculate the link reliability and relative velocity between the CH and all CMs who want the item, which is of the order $\mathcal{O}(N)$. Moreover, the calculations of $t_{min}^{deadline}(a_k)$ and $d_{max}(a_k)$ require to sort all CMs that ask for item a_k according to the time constraints for them and the distances between them and the CH. The sorting consumes time of the order $\mathcal{O}(N\log N)$. After the calculation, b_i sorts the data item in the queue according to their utilities, which takes time of the order $\mathcal{O}(X\log X)$, where X is the size of SMs data set. As a result, the complexity of Algorithm 2 is of the order $\max\{\mathcal{O}(N\log N), \mathcal{O}(X\log X)\}$.

In the two proposed algorithms, Algorithm 1 shows the procedure of the proposed bus-based clustering algorithm. In a traditional scenario, vehicles are grouped into clusters when there are messages need to be disseminated. The clustering process adds extra delay in disseminating messages. To reduce that delay, Algorithm 1 forms buses and vehicles into clusters in advance. Each running bus in the ROI will periodically broadcasts the CHA message to claim itself as a potential CH. Each vehicle that receives the CHA message will responds the bus with an acknowledgement message. Upon receiving this ACK message, the bus checks if the responding vehicle is already on its neighboring nodes' list. If not, the bus further invites the vehicle to become a CM of itself, by sending the updated utility of itself to the vehicle. On the vehicle's side, it can receive multiple invitations from different buses at the same time, and it chooses the one that has the greatest utility value as its CH. Once the cluster is formed, vehicles in this cluster maintain the communications links with the CH until they leave this cluster. Therefore, if an emergent event accurs, the EMs can be disseminated real quickly from a cluster to another cluster, also the EMs can be broadcasted within clusters fastly, thus minimizing the overall transmission delay. Moreover, to fully explore the potentiality of buses, multiple static and dynamic characteristics are analyzed to depict the capability of a bus to work as a CH. Specifically, we propose to use the path similarity between a bus and its neighboring vehicles as an effective characteristic in the process of evaluating the bus's capability of working as the CH. The path similarity characteristic is crucial for grouping the most matching bus and its neighboring vehicles into a cluster, as well as for achieving stable clusters so that the overhead of cluster maintenance and updation in traditional methods can be minimized.

After the proposed clustering process, Algorithm 2 enables each CH to schedule EMs and SMs efficiently by exploring the inherent timeliness characteristic of EMs, and explicitly integrating it with multiple characteristics of both EMs and SMs. In the existing work, EMs are given special preference over SMs to guarantee the timeliness requirements of EMs. The transmissions of SMs are paused until the transmissions of EMs are completed. What's worse, the transmissions of SMs can be interrupted by EMs determinedly. In other words, EMs enjoy an unconditional privilege in the scheduling process, at the sacrifice of SMs transmission performance. Although the rigid QoS requirements of EMs can be satisfied in this way, the transmissions of SMs are harmed and network resources are underutilized, thus in turn degrading the overall transmission performance. Distinct from traditional works, the proposed Algorithm 2 treats EMs and SMs equally under the unified standard given by formula (30). The utility values of EMs and SMs waiting in the queue are dynamically adjusted according to the instant situation in every time slot. As a result, the transmission opportunity is assigned to the most profitable data item in each time slot, so as to achieve good balance between satisfying the rigid QoS requirements of EMs and responding to CMs' service requests as much as possible.

V. Performance Evaluation

In this section, we evaluate the performance of the proposed approach and compare it with referenced algorithms as follows: (1) Clustering and Probabilistic Broadcasting (CPB) based data dissemination algorithm [30]. In CPB, the selected CHs are responsible for data transmission and they adopt probabilistic forwarding to disseminate data. (2) Time Barrier-based Emergency Message Dissemination (TBEMD) [31]. In TBEMD, the EM transmission is given higher scheduling priority over SM

TABLE II
SIMULATION PARAMETER SETTINGS

Parameter	Value
Simulation area	3.2 km × 3.2 km
Number of vehicle nodes	[60 – 300]
Proportion of bus nodes	0.2
Communications radius of vehicles	150 m
Proportion of SM requests	[0.2 – 1]
vehicle speed	[0 – 15] m/s
P_T	20
G_t	23
G_r	23
L	1000
γ	1
λ	[0.2 – 1]
θ	[0 – 1]

transmission. The time barrier is inversely proportional to the distance between the sending vehicle and the receiving vehicle, that is, the longer the distance is, the smaller the time barrier value is. Thus, remote vehicles are more likely to receive the required data. These methods are chosen as the baseline algorithms because they represent the latest or widely used EMs/SMs data dissemination algorithms in the urban scenario of IoVs, and they adopt clustering methods similarly as the CMDD does.

A. Simulation Setup

The simulation scenario is a two-way four lane intersection in urban scenario. The size of the ROI is set to 3.2 km*3.2 km. Full Velocity Difference (FVD) car following model was adopted as the vehicular mobility model. The number of vehicles in the ROI is N_{total}, and the number of buses in this area is approximately $0.2N_{total}$ [32]. Vehicles in the ROI will randomly generate requests one by one with exponentially distributed inter-arrival time. The mean request arrival rate is set to be a constant λ, which is actually a scaling factor of the request arrival rate. The data access pattern is shaped with the Zipf distribution with skewness parameter θ, $0 \leq \theta \leq 1$ [33]. The smaller the value of θ is, the less different items are accessing frequency among all data items in the database. In particular, when $\theta = 0$, the distribution becomes uniform, whereas the Zipf distribution becomes extremely skewed as θ increases to 1. Time is slotted. The specific parameter setting are shown in Table II. The results are obtained when the system is in a steady state and all data points are based on the average of over 5000 simulation runs.

We evaluate the data dissemination performance, as well as the cluster stability of the proposed mechanism, respectively. Below we give the specific analysis on simulation results. In the following, we also present the statistical analysis and 95% confidence intervals for the simulation results in all figures.

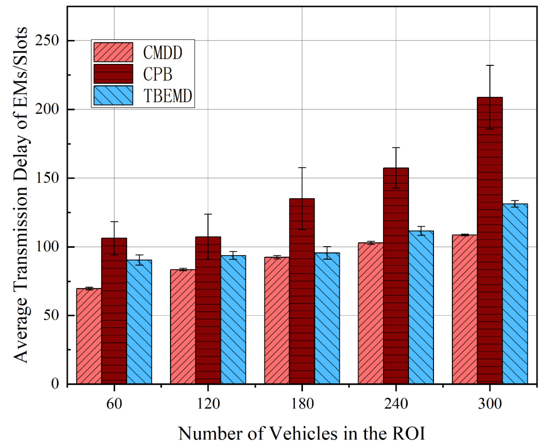

Fig. 3. Average transmission delay of EMs VS number of vehicles in the ROI.

B. Data Dissemination Performance Analysis

In this subsection, we analyze the data dissemination performance of the proposed approach from various aspects. The performance metrics are as follows:

1) Average Transmission Delay (ATD): the average transmission delay of EMs stands for the average time period from the moment that an emergency event occurs to the moment that all vehicles in the ROI have received the EMs successfully. Similarly, the average transmission delay of SMs represents the average time period over the sum of response time of requesting vehicles in the ROI, and the response time of a requesting vehicle is measured from the moment the vehicle sends out a request for a particular service data item, to the moment that the vehicle has received the requested data item completely. ATD is a sensitive parameter to describe the efficiency of the data transmission mechanism, and shorter ATD indicates higher efficiency.
2) Successful Warning Ratio (SWR): SWR is defined as the ratio of the total number of vehicles that have received EMs timely to the total number of vehicles in the ROI. It is an important parameter to show the effectiveness of the proposed approach in realizing traffic safety.
3) Successful Response Ratio (SRR): SRR is defined as the ratio of the total number of vehicles that have received their requested data items timely to the total number of vehicles that have asked for data items. It is another important parameter to show the effectiveness of the proposed approach in accomplishing comfort and smart travel experience.

Fig. 3 and Fig. 4 give the simulation results of average transmission delay for EMs and SMs under different number of vehicles in the ROI, respectively. The skewness parameter $\theta = 0.6$ and the mean request arrival rate is set to 0.5. A small λ indicates that a vehicle will send out a data item request at a low frequency and thus introducing a light load for the CH. Contrarily, a big λ means that a vehicle will request data from the CH frequently, and brings a heavy burden to the CH. It clearly can be seen that ATD increases with the number of vehicles in

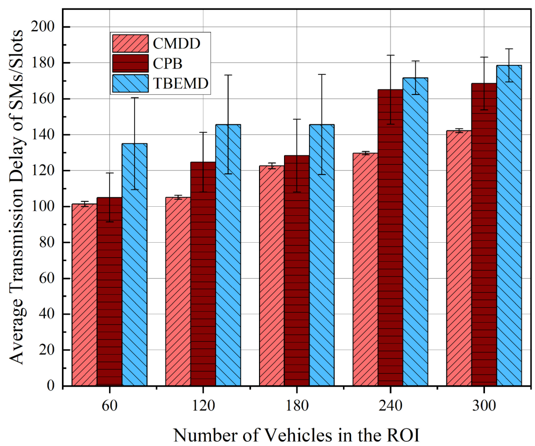

Fig. 4. Average transmission delay of SMs VS number of vehicles in the ROI.

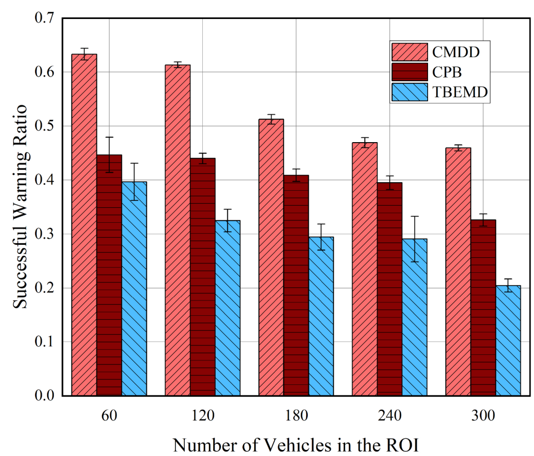

Fig. 5. Successful warning ratio VS number of vehicles in the ROI.

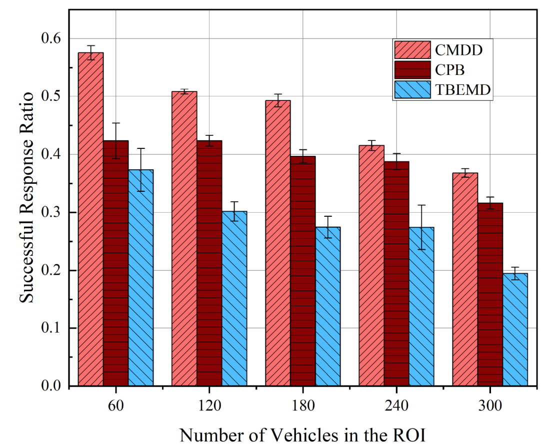

Fig. 6. Successful response ratio VS number of vehicles in the ROI.

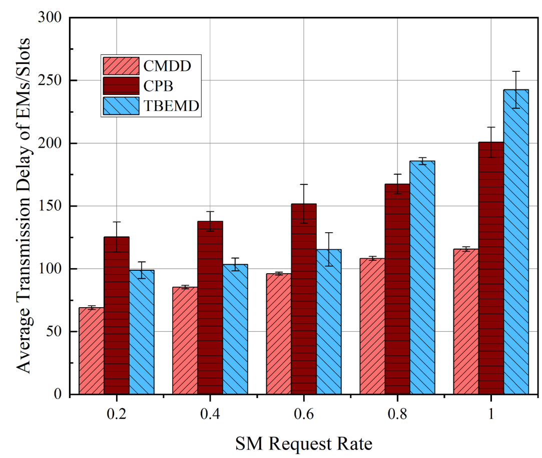

Fig. 7. Average transmission delay of EMs VS SM request rate.

the ROI for both EMs and SMs. Moreover, the proposed CMDD approach achieves the lowest ATD for both EMs and SMs under various number of vehicles in the ROI. Meanwhile, CPB treats EMs and SMs equally, as a result, the ATDs of EMs and SMs are almost the same with CPB. TBEMD gives high priority to EMs and the ATD of EMs with TBEMD is better than that with CPB. However, the better performance of TBEMD is achieved at the sacrifice of the ATD of SMs as shown in Fig. 4, the ATD of SMs with TBEMD is the worst among all compared algorithms. The reason is that TBEMD assigns high priority to EMs and simply schedules mixed data according to their priorities, neglecting the fact that EMs might take too much unnecessary resources and thus degrading the transmission of SMs. On the contrary, the proposed CMDD algorithm jointly considers EMs and SMs transmission requirements, and quantities their priorities appropriately through multiple characteristics. In doing so, the timely requirement of EMs is satisfied and the transmission quality of SMs is also valued, so that both EMs and SMs transmission performance can be guaranteed.

Fig. 5 shows the successful warning ratios of compared algorithms under different number of vehicles. The proposed CMDD achieves the highest SWR among all compared algorithms. It can be seen that SWR decreases with the increment of the number of vehicles, the reason is that with more vehicles in the ROI, the number of clusters increases, as well as the number of CHs. With more clusters and CHs, there are more SMs requests to be answered which may occupy a lot resources, thus degrading the transmission of EMs. Meanwhile, the transmission hops are increased as well, and the transmission conflict occurs more often, introducing extra longer transmission time that some vehicles are not tolerant with.

Fig. 6 shows the successful response ratio of all compared algorithms, and CMDD achieves the highest SRR. Similarly, it is seen that SRR decreases with the increment of number of vehicles in the ROI. It is also seen that CPB achieves the second best in SRR with different number of vehicles, while TBEMD is the worst of all compared algorithms.

Fig. 7 and Fig. 8 depict the ATDs of compared algorithms with different SM request rate. The number of vehicles in the ROI is set to 100 and $\theta = 0.6$. It is clearly seen that the ATD of EMs and SMs increase with the increment of the SM request rate. This is reasonable since with higher SM request rate, there will be more SM requests waiting to be answered and the load of each CH is heavy, thus degrading the transmission of both EMs and SMs. However, it can be seen from Fig. 7 and Fig. 8 that

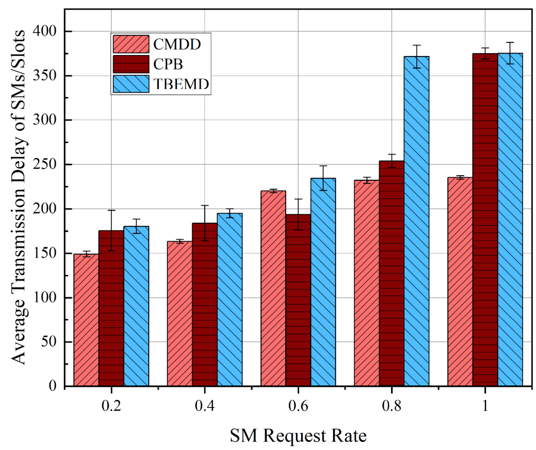

Fig. 8. Average transmission delay of SMs VS SM request rate.

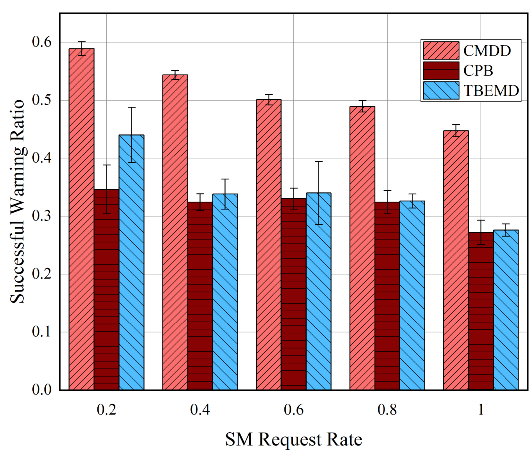

Fig. 9. Successful warning ratio VS SM request rate.

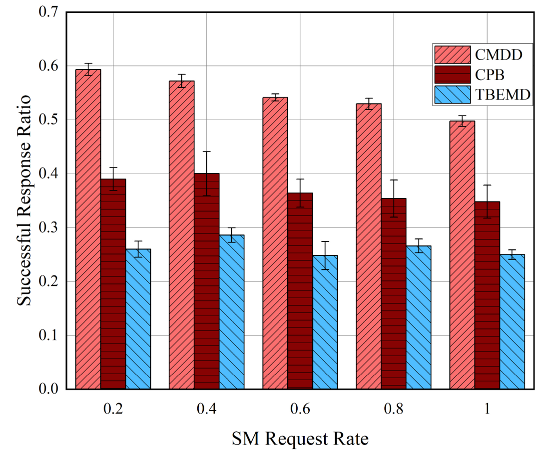

Fig. 10. Successful response ratio VS SM request rate.

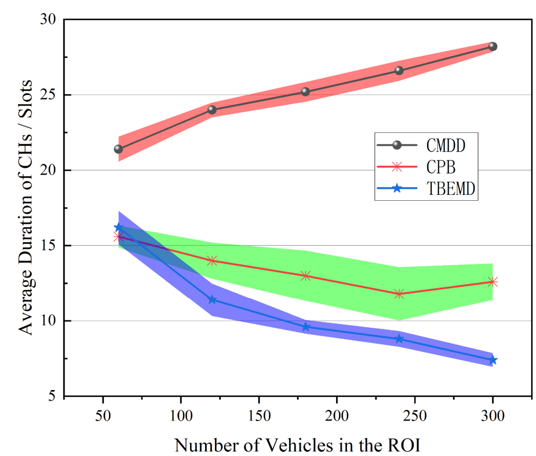

Fig. 11. Average duration of CHs VS number of vehicles in the ROI.

the proposed CMDD achieves the best performance among all algorithms under different SM request rate, and the increasing trend of CMDD is more stable than that of the rest algorithms, indicating good adaptability of CMDD in dealing with varied SM request rates.

Fig. 9 shows the successful warning ratio under different SM request rates. It can be seen clearly that the SWR of CMDD achieves the highest among all compared algorithms, and the performance gain is remarkable. When the system is heavy loaded (e.g, $\lambda = 0.6$), the successful warning ratio of CMDD is still above 50%, revealling its high adaptability for varying loads and complex network environments.

Fig. 10 shows the successful response ratio under different SM request rates. It can be seen that the SRR of CMDD achieves the highest among all algorithms, and the ratio decreases slightly with the increment of SM request rate. With the increment of λ, the frequency that a vehicle requsts SM data items from a CH increases, and thus degrading the responsiveness of buses. However, the proposed CMDD policy dynamically adjusts the communications resources allocated among EMs and SMs. The transmission of SMs is promoted on the premise that the deadline constraints of EMs are not violated. As a result, both EMs and SMs transmission requirements can be responded satisfyingly.

C. *Cluster Stability*

In this subsection, we evaluate the cluster stability of the proposed bus-based clustering algorithm. The performance metrics are as follows:

1) Average Duration of CHs: the average time duration of the selected buses that are working as CHs.
2) Average Duration of CMs: the average time duration of the vehicles that are working as CMs.

Obviously, both metrics represent the stability of clusters, and longer average durations of CHs/CMs reflect more stable clusters.

Fig. 11 and Fig. 12 show the average duration of CHs and CMs, of three compared algorithms, under different number of vehicles in the ROI. It is noticed that for a specific number of vehicles in the ROI, the proposed bus-based clustering algorithm has the longest average duration of CHs, and the performance advantage keeps growing with the increment of the number of vehicles. Similar observations can be obtained for the average duration of CMs from Fig. 12. The reason behind this lies in the fact that CPB and TBEMD select CHs according to multiple parameters such as the probability of successful transmitting packet, the channel conditions, the average relative distance, the

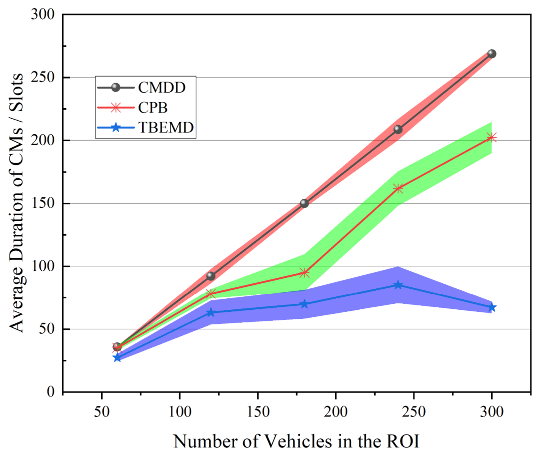

Fig. 12. Average duration of CMs VS number of vehicles in the ROI.

average relative speed, etc.. Obviously, these characteristics are very important for finding appropriate vehicles to take responsibility for data dissemination and forwarding in the environment of IoVs. However, all these characteristics are dynamically and quickly changing with time and situations. Especially when vehicles are densified, there might be multiple vehicles with similar characteristics which are all eligible for competing for the CHs even in a small area. Thus, a vehicle that works as a CH in a single time slot will be replaced by another vehicle whose typical characteristics become better in the next time slot, thus resulting the update of other involved vehicles. In other words, CM vehicles are also updated and they need to decide whether to join in the new cluster with the updated CH. As a result, clusters with CPB or TBEMD update frequently and introduce relatively short duration of CHs and CMs. On the contrary, the proposed CMDD selects proper CHs from a different angle, where we propose to use buses as CHs to take fully advantage of buses that have regular routes and greater capability, comparing to normal vehicles. Moreover, the dynamic characteristics that are emphasized by CPB and TBEMD are also valued by CMDD, and more parameters that can reflect the characteristics of vehicle mobility and urban scenario, such as path similarity, link reliability and so on, are jointly considered. In summary, when selecting CHs, CMDD not only considers the typical vehicular and mobility characteristics that are dynamic and quick-changing, but also values particular attributes that depict the vehicle moving behavior in a long term. In doing so, CMDD achieves longer average duration of selected CHs and CMs, and more stable clusters.

VI. Conclusion

Efficient and reliable data dissemination in IoVs is the basis of realizing the vision of intelligent transportation systems. In this paper, we have proposed an improved cluster-based V2V approach for mixed data dissemination in urban scenario of IoVs. We have proposed to use buses, the most common vehicles in urban scenario, as cluster heads, to provide reliable data communications between vehicles. We have designed a bus-based clustering algorithm which considers not only mobility characteristics of vehicles, such as vehicles' speeds and locations, but also environment and communications characteristics of urban scenario, such as path similarity and link reliability, to help vehicles find the most appropriate buses as CHs and establish connections with them. Moreover, we have proposed a novel priority-based data scheduling algorithm which emphasizes multiple parameters in the priority calculation process, such as message importance, data popularity, time constraint, distance, etc. These parameters are carefully integrated into the calculation so that the QoS requirements of both EMs and SMs are reasonably appreciated and respected, to guarantee that all vehicles in the ROI can receive EMs timely and get satisfied with their SMs requests at a certain level. Simulation results have shown that our proposed approach can achieve obviously better transmission performance, in terms of transmission delay, successful warning ratio and successful response ratio, compared with the considered previous methods. In addition, the proposed bus-based clustering algorithm performs stably under varied vehicle densities, indicating its good adaptability for complex urban environment.

In the context of this era, cutting-edge technologies such as 6G, vehicles situational awareness and demand forecasting technologies are advancing rapidly, revealing great potential for enhancing mixed data dissemination in IoVs. An interesting future research direction for this work is to consider combining 6G, deep learning and multi-sensor fusion technologies, to realize quick and accurate feature matching and association based on the running status and trajectory of vehicle, for mixed data dissemination in urban scenarios. Cross-regional data sharing and information interaction will be considered in the future work to reduce the cost of data dissemination and to improve the Quality of Experience (QoE) of drivers and passengers.

References

[1] X. H. You et al., "Towards 6G wireless communication networks: Vision, enabling technologies, and new paradigm shifts," *Sci. China Inf. Sci.*, vol. 64, no. 1, pp. 1–74, Jan. 2021.

[2] H. Zhou, W. Xu, J. Chen, and W. Wang, "Evolutionary V2X technologies toward the internet of vehicles: Challenges and opportunities," *Proc. IEEE*, vol. 108, no. 2, pp. 308–323, Feb. 2020.

[3] R. Hussain and S. Zeadally, "Autonomous cars: Research results, issues, and future challenges," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 2, pp. 1275–1313, Apr.–Jun., 2019.

[4] S. Wang, Y. Hu, and G. Qi, "Blockchain and deep learning based trust management for internet of vehicles," *Simul. Modelling Pract. Theory*, vol. 120, pp. 1–13, Nov. 2022.

[5] Y. Yuan, R. Tasik, S. S. Adhatarao, Y. Yuan, Z. Liu, and X. Fu, "RACE: Reinforced cooperative autonomous vehicle collision avoidance," *IEEE Trans. Veh. Technol.*, vol. 69, no. 9, pp. 9279–9291, Sep. 2020.

[6] Y. Wang, Y. Ding, Q. Wu, Y. Wei, B. Qin, and H. Wang, "Privacy-preserving cloud-based road condition monitoring with source authentication in VANETs," *IEEE Trans. Inf. Forensics Secur.*, vol. 14, no. 7, pp. 1779–1790, Jul. 2019.

[7] Y. Jiang, M. Ma, M. Bennis, F. -C. Zheng, and X. You, "User preference learning-based edge caching for fog radio access network," *IEEE Trans. Commun.*, vol. 67, no. 2, pp. 1268–1283, Feb. 2019.

[8] X. Bao et al., "Efficient clustering V2V routing based on PSO in VANETs," *Measurement*, vol. 152, pp. 1–24, Feb. 2020.

[9] S. Shafi and D. Venkata Ratnam, "A cross layer cluster based routing approach for efficient multimedia data dissemination with improved reliability in vanets," *Wireless Pers. Commun.*, vol. 107, no. 4, pp. 2173–2190, Apr. 2019.

[10] N. Kumar et al., "Optimized clustering for data dissemination using stochastic coalition game in vehicular cyber-physical systems," *J. Supercomputing*, vol. 71, no. 9, pp. 3258–3287, May 2015.

[11] S. A. Alghamdi et al., "Novel path similarity aware clustering and safety message dissemination via mobile gateway selection in cellular 5G-based V2X and D2D communication for urban environment," *Ad Hoc Netw.*, vol. 103, pp. 1–16, Jun. 2020.

[12] J. Yang, C. Huang, and X. Fan, "Reliable broadcast mechanism for emergency message in urban vehicular ad hoc networks," *IET Intell. Transport Syst.*, vol. 13, no. 9, pp. 1383–1393, Jun. 2019.

[13] S. Ullah et al., "Position-based vehicular clustering for emergency messages dissemination in zone of interest," in *Proc. 2nd Asia Symp. Signal Process.*, 2021, pp. 61–66.

[14] G. Abbas et al., "EEMDS: An effective emergency message dissemination scheme for urban VANETs," *Sensors*, vol. 21, no. 5, pp. 1–19, Feb. 2021.

[15] X. Zhang, Q. Miao, and Y. Li, "An adaptive link quality-based safety message dissemination scheme for urban VANETs," *IEEE Commun. Lett.*, vol. 22, no. 10, pp. 2104–2107, Oct. 2018.

[16] W. Zhu, D. Gao, C. H. Foh, W. Zhao, and H. Zhang, "A collision avoidance mechanism for emergency message broadcast in urban VANET," in *Proc. IEEE 83 rd Veh. Technol. Conf.*, 2016, pp. 1–5.

[17] S. Wang, Q. Zhang, and G. Chen, "V2V-CoVAD: A vehicle-to-vehicle cooperative video alert dissemination mechanism for internet of vehicles in a highway environment," *Veh. Commun.*, vol. 33, pp. 1–12, Jan. 2022.

[18] L. Zhang, S. Jia, Z. Liu, Y. Wang, and Y. Liu, "Bus-ads: Bus trajectory-based advertisement distribution in VANETs using coalition formation games," *IEEE Syst. J.*, vol. 11, no. 3, pp. 1259–1268, Sep. 2017.

[19] D. Zhang et al., "A new method of content distribution based on fuzzy logic and coalition graph games for VEC," *Cluster Comput.*, pp. 1–17, 2022, doi: 10.1007/s10586-022-03711-2.

[20] X. Chen and G. Chu, "Data cooperative distribution mechanism of internet of vehicles using D2D technology," *Adv. Multimedia*, vol. 2022, pp. 1–10, May 2022.

[21] Y. Jiang, Q. Liu, F. Zheng, X. Gao, and X. You, "Energy-efficient joint resource allocation and power control for D2D communications," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6119–6127, Aug. 2016.

[22] M. A. Lebre, F. Le Mouël, and E. Ménard, "Efficient vehicular crowdsourcing models in VANET for disaster management," in *Proc. IEEE 91st Veh. Technol. Conf.*, 2020, pp. 1–5.

[23] L. Khamer et al., "Enhancing video dissemination over urban VANETs using line of sight and QoE awareness mechanisms," *Ann. Telecommun.*, vol. 76, no. 9, pp. 759–775, Jul. 2021.

[24] N. Taherkhani and S. Pierre, "Prioritizing and scheduling messages for congestion control in vehicular ad hoc networks," *Comput. Netw.*, vol. 108, pp. 15–28, Oct. 2016.

[25] A. Agarwal, R. Pal, and A. Prakash, "A scheduling algorithm including deadline of messages in vehicular ad hoc network," in *Advances in VLSI, Communication, and Signal Processing*. Singapore: Springer, Dec. 2020, pp. 115–123.

[26] A. Sharma and L. K. Awasthi, "AdPS: Adaptive priority scheduling for data services in heterogeneous vehicular networks," *Comput. Commun.*, vol. 159, pp. 71–82, Jun. 2020.

[27] R. Jiang, Q. Wu, and Z. Zhu, "Full velocity difference model for a car-following theory," *Phys. Rev. E*, vol. 64, no. 1, pp. 1–4, Jun. 2001.

[28] M. Nakagami, "The m-distribution-a general formula of intensity distribution of rapid fading," in *Statistical Methods in Radio Wave Propagation*. New York, NY, USA: Elsevier, Jun. 1960, pp. 3–36.

[29] X. Zhang, X. Cao, L. Yan, and D. K. Sung, "A street-centric opportunistic routing protocol based on link correlation for urban VANETs," *IEEE Trans. Mobile Comput.*, vol. 15, no. 7, pp. 1586–1599, Jul. 2016.

[30] L. Liu et al., "A data dissemination scheme based on clustering and probabilistic broadcasting in VANETs," *Veh. Commun.*, vol. 13, pp. 78–88, Jul. 2018.

[31] S. S. Shah, A. W. Malik, A. U. Rahman, S. Iqbal, and S. U. Khan, "Time barrier-based emergency message dissemination in vehicular ad-hoc networks," *IEEE Access*, vol. 7, pp. 16494–16503, 2019.

[32] Z. Khan et al., "Street-centric routing scheme using ant colony optimization-based clustering for bus-based vehicular ad-hoc network," *Comput. Elect. Eng.*, vol. 86, pp. 1–12, Sep. 2020.

[33] X. Gan, Z. Qin, L. Fu, and X. Wang, "Unraveling the impact of users' interest on information dissemination in wireless networks," *IEEE Access*, vol. 6, pp. 32687–32699, 2018.

Shujuan Wang (Member, IEEE) received the B.S. and Ph.D. degrees from Beijing Jiaotong University, Beijing, China, in 2007 and 2013, respectively. She is currently an Associate Professor with the School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, China. From 2010 to 2011, she was a Visiting Scholar with the Department of Electrical Engineering, Columbia University, New York, NY, USA. Her research interests include Internet of Vehicles, vehicular communications, and intelligent transportations.

Guangchao Chen received the B.S. degree in 2019 from the Kunming University of Science and Technology, Kunming, China, where he is currently working toward the M.S. degree. His research interests include VANETs, wireless communication, and intelligent transportation systems.

Yanxiang Jiang (Senior Member, IEEE) received the B.S. degree in electrical engineering from Nanjing University, Nanjing, China, in 1999, and the M.S. and Ph.D. degrees in communications and information systems from Southeast University, Nanjing, China, in 2003 and 2007, respectively.

Dr. Jiang was a Visiting Scholar with the Signals and Information Group, Department of Electrical and Computer Engineering, University of Maryland at College Park, College Park, MD, USA, in 2014. He is currently an Associate Professor with the National Mobile Communications Research Laboratory, Southeast University, Nanjing, China. His research interests include broadband wireless mobile communications, covering topics, such as machine learning for wireless communications, edge intelligence, and B5G and 6G mobile communications.

Xiaohu You (Fellow, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Southeast University, Nanjing, China, in 1982, 1985, and 1988, respectively. Since 1990, he has been with National Mobile Communications Research Laboratory, Southeast University, where he holds the ranks of a Professor and the Director. He is currently the Chief of the Technical Group of China 3G/B3G Mobile Communication R & D Project. His research interests include mobile communications, adaptive signal processing, and artificial neural networks, with applications to communications and biomedical engineering.

Dr. You was the recipient of the Excellent Paper Prize from the China Institute of Communications in 1987, Elite Outstanding Young Teacher award from Southeast University in 1990, 1991, and 1993, and National Technological Invention Award of China in 2011. He was also the recipient of the 1989 Young Teacher Award of Fok Ying Tung Education Foundation, State Education Commission of China.