Computer Communications 149 (2020) 44–50

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

Multi-level cluster-based satellite-terrestrial integrated communication in Internet of vehicles

Kai Lin [a,*], Chensi Li [a], Pasquale Pace [b], Giancarlo Fortino [b]

[a] School of Computer Science and Technology, Dalian University of Technology, Dalian, China
[b] Department of Informatics, Modeling, Electronics, and Systems (DIMES), University of Calabria, Italy

ARTICLE INFO

Keywords:
Satellite-terrestrial integrated communication
Internet of vehicles
Mobile tracking
Multi-level clustering

ABSTRACT

The growth of vehicles and mobile services has brought new challenges to Internet of vehicles (IoV). In order to accurately locate vehicles and quickly process data for alleviating the pressure of communication, clustering method based on data analysis are studied in this paper for vehicles in the Internet of vehicles, so as to make sure that mobile stations and terminals on the road are able to access maximum transmission opportunity and suffer minimal communication interference from adjacent vehicles. A satellite-terrestrial integrated communication framework for the IoV is designed to support analysis of data from vehicles, mobile tracking of vehicles, and unsupervised adaptive communication scheduling; then, a multi-level cluster-based satellite-terrestrial integrated communication model (MCSIC) is designed to realize the effective real-time data communication based on data analysis. Finally, the performance of MCSIC is analyzed through several simulations.

1. Introduction

The dramatically growing number of vehicles causes more serious traffic congestion and traffic accidents, which leads to the rising number of traffic casualties. According to [1], there are 4.2 billion hours wasted in traffic congestion and more than 30,000 fatalities caused by traffic accident only in the United States each year. In order to alleviate these realistic traffic problems, vehicle auxiliary devices combined with roadside infrastructure constitute the Internet of vehicles (IoV) to improve the safety, security, and efficiency of roadway transportation, and provide various services to different types of vehicles, including private cars, taxies, buses and so on. Wireless communication technology enables vehicles to communicate with the surrounding neighboring vehicles, roadside infrastructure, and traffic control centers, and it is the basis for IoV to collect information [2]. Despite the extraordinary potential of IoV, it still faces the problem of increasing data transmission computing resources and increasing data analysis time consumption [3]. How to effectively achieve high quality mobile services in the IoV faces many challenges. In particular, the ever-increasing number of vehicles and the complexity of their motions have led to a decline in the quality of service for the IoV. Moreover, the communication is the essential part of the network [4]. Therefore, it is important to improve the communication capacity and real-time transmission performance of the IoV.

Location service as one of the primary services in smart automated systems of Internet of Things (IoT) is also an important part of IoV [5]. Many researches have focus on vehicle-to-infrastructure (V2I) and vehicle-to-vehicle (V2V) communications to collect and exchange local information, such as road conditions and traffic information. Wang et al. [6] proposed a dynamic clustering and cooperative scheduling algorithm for data services in the bidirectional road scenario based on the analysis of the SINR in V2V communication, which enables vehicles dynamically join or leave clusters according to their realtime velocities and positions. Won et al. [7] designed a phantom jam control protocol to mitigate phantom jams by capturing the dynamics of traffic jams. Per-lane speed difference under traffic congestion is also taken into account in their work. Shieh et al. [8] presented a method to locate the position of the vehicle and track its trajectory for infrared V2I communication systems, which is able to locate the position of the vehicle in a communication area of 6 m in width and 20 m in length. Yin et al. [9] designed a bicycle sharing system by using a series of data mining tasks based on real data sets. Emura et al. [10] proposed a road-to-vehicle communication system with relaxed anonymity by considering time-dependent linking properties, where a vehicle is unlinkable unless it generates multiple signatures in the same time period. These methods improve the communication performance of the IoV from different angles, but they are limited to local short-range communication and cannot be extended to long-distance real-time data transmission.

As a typical long-distance wireless communication technology, satellite communication can provide mobile services for a variety of

* Corresponding author.
E-mail addresses: link@dlut.edu.cn (K. Lin), licheni@mail.dlut.edu.cn (C. Li), p.pace@dimes.unical.it (P. Pace), g.fortino@unical.it (G. Fortino).

https://doi.org/10.1016/j.comcom.2019.10.009
Received 1 May 2019; Received in revised form 10 July 2019; Accepted 2 October 2019
Available online 4 October 2019
0140-3664/© 2019 Elsevier B.V. All rights reserved.

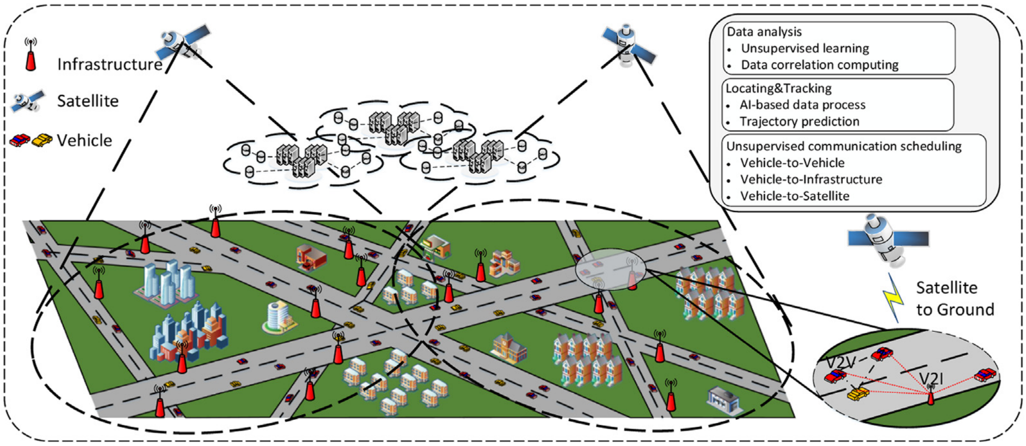


Fig. 1. Satellite-terrestrial integrated communication framework.

terrestrial applications with its wide-area coverage, such as positioning and navigation services for vehicles supported by global positioning system (GPS), BeiDou navigation satellite system (BDS), global navigation satellite system (GLONASS), etc [11]. However, satellite communication also has its own shortcomings, especially for it being easily blocked by obstacles such as buildings, plants and transportations. Recently, there is a novel trend to integrate satellite systems with terrestrial networks to provide high quality communication services for emerging large-scale application [12,13]. The development of satellite-terrestrial integrated communication is in line with the urgent need for further development of the IoV, but the scheduling strategy is necessary to be redesigned to accommodate the IoV environment. In particular, satellite-terrestrial integrated communication enables vehicles in the IoV to selectively communicate with satellites, roadside units (RSUs), or other vehicles as needed [14]. Here, the RSU is a typical infrastructure in the IoV. Although this choice flexibility increases the network capacity and the diversity of data transmission mode, it is also need to take into account the stability of communication and optimization of global performance [15].

In this paper, we focus on how to utilize satellite-terrestrial integrated communication to improve the quality of mobile communication services in the IoV. Firstly, we proposed a satellite-terrestrial integrated communication framework for the IoV (STICI), assisted by intelligent perception–interaction–control and constrained by satellite-terrestrial mobile communication, which explores integrative cooperative and interactive new theory and method for communication resource scheduling. The STICI framework integrates satellite communications, terrestrial backbone communications, and short-range communications, including V2V, V2I and vehicle-to-satellite (V2S), to provide flexible, high-quality, full-coverage mobile services for the IoV. The whole set of effective platform environment and solution is built to provide targeted approaches and theoretical foundation for mobile service of satellite-terrestrial network, and further improve practicality. Then, comprehensive consideration for multi-modal perceptive information is used to realize full perception for mobile tracking information through various types of methods. In addition, interaction data training is conducted to make satellite-terrestrial network gain experience of completing communication tasks in different environments, and make it enable to unsupervised schedule tasks and independently optimize resource allocation. Finally, a multi-level cluster-based satellite-terrestrial integrated communication model (MCSIC) is designed to realize the effective real-time data communication based on data analysis.

The rest of the paper is organized as follows. Section 2 introduces the STICI framework and the mobile tracking of vehicles. The proposed MCSIC for the IoV is described in Section 3 and the performance analysis of our method based on experiments, is discussed in Section 4. Finally, Section 5 concludes the paper.

2. STICI framework and vehicle tracking

In this section, the STICI framework with multi-level communication is described in detail; then both tracking and behavior identification of vehicles are designed.

2.1. STICI framework

The STICI framework is designed as shown in Fig. 1 to support different functions, such as data analysis, locating and tracking of vehicles, and unsupervised communication scheduling of mobile stations and terminals. The STICI integrates the satellite communication, terrestrial backbone communication, V2I and V2V communication. Considering the communication coverage of satellite, roadside unit and vehicles, a vehicle can select the most suitable communication mode when it has data to be transmitted. The architecture mainly includes a mode and state identification module, a trajectory tracking module, a data analysis module, and a communication scheduling module for the IoV. Vehicles are equipped with various sensors and devices to get the surrounding environment information and exchange data with other vehicles. Based on their real-time and historical generated data, unsupervised learning can be adopted to learn and build data analysis model to explore the potential rules, then to optimize the data transmission and resource allocation by adjusting the communication modes. Vehicle automatic cruise system, GPS, BDS, GLONASS, etc. are employed to assist in the positioning and tracking of the vehicles. Nowadays the artificial intelligence technology based geo-spatial information science can realize massive data acquisition of geo-spatial information, intelligent data analysis, geo-spatial data-driven application [16–19]. For example, Yairi et al. [20] proposed a new artificial satellite data-driven health monitoring and anomaly detection method based on probability dimension reduction and provided valuable information for satellite operators. Liu et al. [21] designed a multi-scale deep feature learning method for high-resolution satellite image scene classification, which is proved to be superior to other allocation algorithms. The STICI also applies artificial intelligence method into satellite-terrestrial integrated communication, which is divided into three stages: data acquisition, data analysis and feedback adjustment. Moreover, the unsupervised learning and reinforcement learning are used to improve mobile communication service of the IoV. The historical information from vehicles such as direction, location, and speed can be analyzed by using self-learning techniques, which are able to realize predicting the future position of mobile vehicles, then quickly adjusting the vehicles' communication mode while changing their positions, to improve the overall communication efficiency of the IoV [22]. In addition, by using clustering based on data analysis, vehicles of the IoV are able to complete data communication by accessing maximum communication opportunity and suffering minimal communication interference

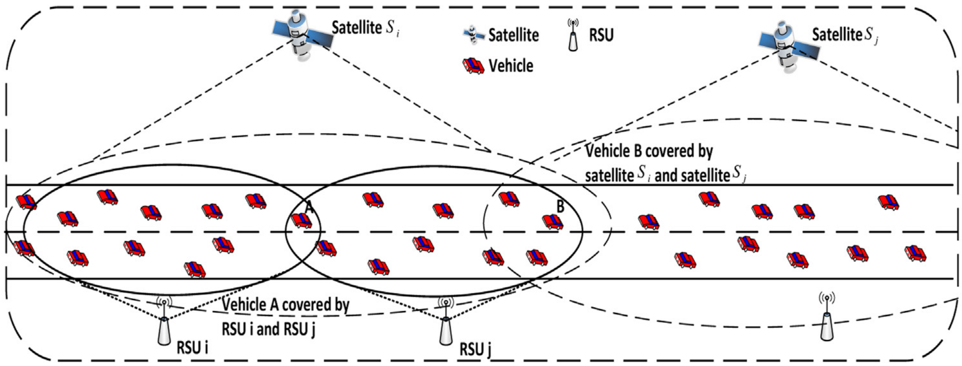


Fig. 2. Vehicles communication covered by satellites and roadside units.

of mobile communication service from adjacent vehicles within their communication coverage [23].

Since the data analysis of the massive data is particularly important in ensuring the performance of communication scheduling, the STICI framework utilizes the cloud processing platform to assist the data analysis. It means that the processing and analysis of the massive data is realized by programming in the cloud processing platform. Combined with the self-learning capability of artificial intelligence, results of data analysis can be feedback to the corresponding computers and controllers to achieve communication scheduling between vehicles and satellites, roadside units, or other vehicles, as well as selecting suitable communication mode for vehicles.

2.2. Vehicle tracking in the STICI

The collaborative strategy of satellite multi-perceptual data analysis in satellite-terrestrial network, considering the perceived performance and resource optimization, can improve network communication performance and operation efficiency. During communicating among satellites, roadside units and vehicles in STICI, it is necessary to precisely locate and track the vehicles because the movement of vehicles requires adjusting the communication mode in time to obtain the best communication condition. For mobile services in IoV, the location of vehicles change rapidly over time; thus, it is required to implement locating and tracking of the vehicles, so as to ensure that the vehicles access stable communication services to improve the overall communication performance of the IoV. The vehicle tracking of STICI integrates multiple technologies, such as data acquisition and signal processing technology, automatic navigation technology, precision machinery design technology, sensor technology, simulation technology, satellite communication technology, system engineering technology, and so on. To achieve the locating and tracking purpose, data filtering and correction methods are adopted for the readings generated from these technologies to eliminate the effects of uncertainty on perceived data caused by interference in the communication between satellites, roadside units and vehicles. Moreover, the statistical-based learning methods and reasoning mechanisms can be used for the motion mode of mobile vehicles, and establishing classification and identification model to improve their moving identification. In addition, the uncertainty of self-organizing mapping and data dependency in the process of perceptual computing is eliminate to get effective perceptual decision.

The satellites are required to be arranged firstly for ensuring that the satellites which work together in orbit around the earth can cover all the vehicles in the IoV. That is, the vehicles on the road can only communicate with those satellites whose coverage area contain them, which also applies to V2I communication. Considering one vehicle may be covered by more than one satellite or roadside unit, as shown in Fig. 2, the vehicle can adaptively choose more suitable satellite or roadside unit for communication.

2.3. Motion state and behavior identification

With the development of motion capture technology, the acquisition of vehicles' motion data becomes simpler and more convenient. The three-dimensional motion posture data of the mobile stations and terminals can be directly calculated through the acquired information by means of different sensing devices, such as compass, GPS, monocular camera, and so on. When tracking the vehicles, the acquired motion data can be analyzed through artificial intelligence to predict their future trajectory, which makes the vehicle tracking more rapid and accurate. Therefore, the STICI gives a behavior intention understanding method based on vehicles' motion data. The analysis of historical evaluation data can be done through the artificial intelligence system, and the results can be summarized and used for the judgment of future trends. In this way, it is able to realize the rapid adjustment of vehicles, as well as rapidly select the most suitable communication mode to improve the overall communication efficiency of the IoV.

In order to recognize the motion state of mobile vehicles, the representative key frames in the set of motion data are extracted firstly. Due to the high dimensionality of the motion data, the overall performance is nonlinear, although the locality is linear. For mapping the data to the low-dimensional manifold space and preserve the local linear features, the STICI draws on the idea of local linear embedding, which extracts the key frames while tries to retain motion features during reducing the dimensions of motion capture data. The key frame set is trained based on the support vector machine (SVM) to establish a mapping relationship between the key frames and the basic motion coding set, so that a set of motions can be converted into a set of motion coding. In this way, the motion state identification can be realized according to the mapping relation of motion coding, motion mode and state set.

The vehicle tracking needs to ensure the accurate and complete extraction of the information from vehicles, but the current vehicle tracking methods have the problem of low accuracy and cannot use for long-term prediction. Therefore, the vehicle tracking of the STICI establishes transformational model and realizes vehicle locating through extracting vehicle motion features and description matching, then improves vehicle tracking performance through average locating error and standard deviation to ensure the accuracy of tracking result.

The main purpose of behavior perception is to predict the movement intention of the vehicles, and conduct logical deduction for the following tracking execution. In the STICT, various motions are not isolated from each other, there is some causal relationship between successive motions, which can be simulated through probability net. The inference method based on probability net is adopted in the STICI to make inference about behaviors of the vehicles. Firstly, the motion recognition is performed according to the motion that has already occurred, and the expected result is estimated based on the recognition result. Then, according to the probability estimation value of the expected result, probabilistic reasoning is performed on the action that may occur in the previous step, so as to realize the reverse probability reasoning of the

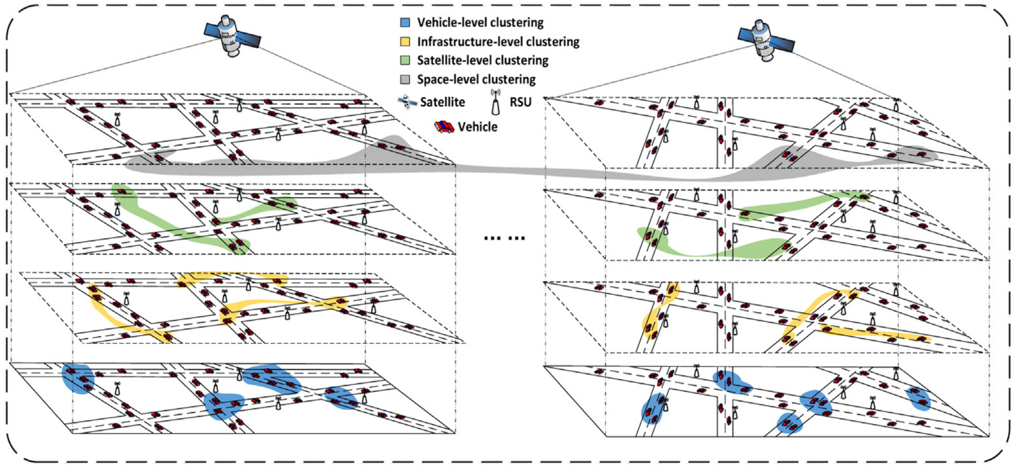


Fig. 3. Multi-level clustering.

remaining of the whole process. Finally, a moving route that maximizes the likelihood of expected results is selected according to the result of probabilistic reasoning; thus, the trajectory tracking and prediction of the vehicles is completed.

3. Multi-level cluster-based satellite-terrestrial integrated communication model

In IoV, the maximum communication coverages of satellites, roadside units and vehicles, are assumed to be unchanged; moreover, each vehicle can only communicate with satellites, roadside units or other vehicles within its communication coverage. Considering that the position of the vehicle will constantly change, vehicle tracking is necessary to select the most suitable communication mode in order to establish stable mobile communication towards the vehicles while the vehicles need to select the satellite or roadside unit with the strongest peaking signal. In particular, the correlation between data is very valuable for data transmission and processing. We combine the above three factors to classify vehicles into multi-level clusters, and on this basis, realize communication scheduling to improve mobile service quality.

3.1. Multi-level clustering

In order to achieve the multi-level clustering in the IoV, this paper uses the unsupervised learning and takes the vehicles' transmitted data as input for data training. Considering a transport packet may include multiple attribute data, some of the them are important, such as movement speed, location information, etc., while others can be ignored. The importance of these transmitted data can be quantified using the vehicle preference function $u(x) = \sum_{i=1}^{n} w_i * u_i(x)$, where w_i is the vehicle weight and $\sum_{i=1}^{n} w_i = 1$, and $u_i(x)$ is the preference value. Therefore, the input data is processed and classified through an unsupervised learning process to obtain valuable information. Then, the valuable information $V = \{v_1, v_2 ... v_n\}$ of the IoV is extracted from a large amount of useful and useless information and input into the vehicle preference function $u(x)$ to estimate the real-time position and trajectory of the mobile vehicles and terminals. After that, the feature extraction is performed on the valuable vehicle information for searching similarity detection, and the motion information is classified and described by using the pattern matching and discriminant functions according to some characteristics of the historical information.

In the presented study, the MCSIC consists of the following four levels shown in Fig. 3: vehicle-level, infrastructure-level, satellite-level and space-level.

3.1.1. Vehicle-level clustering

Vehicle-level clustering is limited to the maximum communication range between vehicles and is affected by three factors: data correlation, distance between vehicles, channel noise of V2V.

Data correlation analysis can be used to handle a tremendous number of multi-modal data in various scenarios and extract valuable information in the IoVs. In this paper, the data package is able to be decomposed into multiple single modal data, such as a dynamic voice, dynamic image, and dynamic text. The data correlation calculation considers both high-dimensional data and low-dimensional data. For the former, the transmitted data dimension of the vehicles is first reduced using principal component analysis (PCA) [24] and then, transmitted multimedia data correlation among vehicles, are analyzed via canonical correlation analysis (CCA) [25]. On the contrary, low-dimensional data can be directly processed using CCA.

Considering there are two data package sets $X = \{x_1, x_2, \dots, x_{|X|}\}$ and $Y = \{y_1, y_2, \dots, y_{|Y|}\}$ from two vehicles including different data model, data correlation between X and Y can be calculated as [26]:

$$\Theta(X,Y) = \mathrm{corr}(a^T x, b^T y) = \frac{a^T \omega(X,Y) b}{\sqrt{a^T \omega(X,X) a}\sqrt{b^T \omega(Y,Y) b}} \tag{1}$$

where $\omega(X,X)$ is the weighted covariance matrix between the matrix X and itself while $\omega(Y,Y)$ is the covariance matrix between the matrix Y and itself. $\omega(X,Y)$ is the weighted covariance matrix of the matrices X and Y, and a^T and b^T is the transformation of the linear coefficient vector of X and Y, respectively.

Although it is significant to coordinate the V2V communication based on data correlation to maximize the function of data process, such as data sharing and data fusion, the interference of V2V communication due to the wireless channel contention and high mobility of vehicles should be taken into account. In vehicle-level clustering, a SINR model is adopted to quantitatively evaluate the interference.

Assuming the set $V^s = \{s_1, s_2, s_3, \dots, s_n\}$ represents the sending vehicles while the set $V^r = \{r_1, r_2, r_3, \dots, r_n\}$ represents the receiving vehicles, $d(i,j)$ is the distance between sender s_i and the receiver r_j. Supposing that the s_i wants to send data to the r_j and all the sending vehicles in the set V^s are broadcasting at the same time, the Signal to Interference plus Noise Ratio (SINR) value [6] is adopted to represent the interference from the receiver vehicles V^r to the sender vehicle, which is calculated as follows:

$$I_{(V^s)\to(V^r)} = \frac{P_t |d|^{-\alpha}}{M_0 + \sum_{s_i \in V^s} P_t |d|^{-\alpha}} \tag{2}$$

where P_t is the transmission power of the sender vehicles. M_0 means the background noise while α is the pass-loss exponent. For the threshold ρ, if $I_{(V^s)\to(V^r)} > \rho$, the communication between two vehicles can be successfully performed.

In addition, the distance between vehicles is also considered for vehicle-level clustering, it can be calculated using the longitude and latitude coordinates of the vehicle. $v_i = (\mu_i, \nu_i)$ and $v_j = (\mu_j, \nu_j)$ are the location of vehicle v_i and vehicle v_j, respectively, where μ means the longitude and ν is the latitude. Thus the distance $D_{(i,j)}$ between vehicle v_i and vehicle v_j can be calculated as follows:

$$D_{(i,j)} = 2R \arcsin \sqrt{H(\mu) + \cos(\mu_i)\cos(\mu_j)H(\nu)} \tag{3}$$

where R is the earth radius, $H(x) = \sin^2(\frac{x_j - x_i}{2})$.

In the process of vehicle-level clustering, function is integrates the above three factors and is described as follows:

$$L_V = \sum_{a=1}^{M_V} \sum_{i=1}^{|C_a|} \sum_{j=1}^{|C_a|} \sigma_V \frac{\Theta_{(i,j)} + I_{(V^s)\to(V^r)}^{V^s}}{D_{(i,j)}} \tag{4}$$

where M_V is the number of the vehicle-level clusters, $|C_a|$ represents the number of the vehicles in cluster C_a. σ_V is the regulatory factor to ensure the L_V ranges from 0 to 1

The initial matrix Θ_V about data correlation is calculated based on Eq. (1), the distance threshold D_V is set as the maximum communication distance of V2V. The largest element in the matrix L_V^g is selected out and its corresponding clusters are integrated into a new cluster, then the modularity Q_g is calculated where g represents the iteration times. If $Q_g > Q_{(g-1)}$ and $|Q_g - Q_{(g-1)}| < \varrho$, the cluster structure is considered as the final state. The new and integrated cluster is restored back to the original two clusters, the clustering process is terminated and $C_V = \{C_1, C_2, \dots, C_{|C_V|}\}$ is the cluster set of the final vehicle-level cluster structure. Otherwise, the current cluster structure is considered as the initial clustering of the next iteration and is used to re-calculate the matrix, the above operations are performed until the $|Q_g - Q_{(g-1)}| < \varrho$.

3.1.2. Infrastructure-level clustering

Infrastructure-level clustering is designed for those vehicles with far distance that cannot communicate using V2V mode but locate within the same roadside unit. Assuming that one roadside unit can allocate the resource to vehicles within its coverage at the same time, the distance $D_{(i,r)}$ between vehicle v_i and the infrastructure r can be calculated according to the equation (3), and the correlation $\Theta_{(i,j)}$ between vehicle v_i and vehicle v_j is still obtained through Eq. (1). Thus, the infrastructure-level correlation function is formed as follows:

$$L_I = \sum_{a=1}^{M_I} \sum_{i=1}^{|C_b|} \sum_{j=1}^{|C_b|} \sigma_I \frac{\Theta_{(i,j)}}{D_{(i,r)} + D_{(j,r)}} \tag{5}$$

where M_I is the number of the clusters in infrastructure-level clustering, $|C_b|$ means the number of vehicles in cluster C_b, and σ_I is the regulatory factor to ensure the L_I ranges from 0 to 1.

The infrastructure-level clustering process is similar with the vehicle-level clustering. The initial matrix Θ_I about data correlation is computed based on Eq. (1), D_I is set as the communication coverage of the roadside unit. The largest element in the matrix Θ_I is selected out and its corresponding clusters are integrated into a new cluster, then the L_I^g is computed where g represent the iteration times. If $L_I^g > L_I^{(g-1)}$ and $|L_I^g - L_I^{(g-1)}| < \varrho$, the cluster structure is considered as the final state. The new and integrated cluster is restored back to the original two clusters and the clustering process is terminated. Otherwise, the current cluster structure is considered as the initial state of the next iteration and is used to re-calculate the matrix Θ_I. The above operations are performed until the $|L_I^g - L_I^{(g-1)}| < \varrho$. $C_I = \{C_1, C_2, \dots, C_{|C_I|}\}$ is the cluster set of the final infrastructure-level cluster structure.

3.1.3. Satellite-level clustering

When the distance between the vehicles is not in the scope of a single infrastructure, but data correlation between the vehicles is still high. The satellite-level clustering is adopted to tackle the communication of these vehicles.

Apart from the data correlation between the vehicles in satellite-level clustering, the interference caused by several satellites should be taken into account. Considering a vehicle is covered by several satellites at the same time, the interference from different satellites should be taken into consideration. Let the set $S^i = \{S_1, S_2, \dots, S_r\}$ represents the satellites who cover the vehicle i, and the noise interference from satellites to the vehicle is computed through the following equation:

$$I_S = \frac{P_c |d|^{-\lambda}}{N_0 + \sum_{s_i \in V^s} P_c |d|^{-\lambda}} \tag{6}$$

where P_c is the transmission power of the sending vehicles. N_0 is the background noise while λ is the pass-loss exponent. In addition, The data correlation between vehicle i and other vehicles in the coverage of same satellite $S_r, S_r \in S^i$ is calculated as $\overline{\Theta_r^i}$. Thus, the satellite-level correlation function is formed as follows:

$$L_S = \sum_{a=1}^{M_S} \sum_{i=1}^{|C_b|} \sum_{j=1}^{|C_b|} \sigma_S (\Theta_{(i,j)} + I_S) \tag{7}$$

where M_S is the number of the satellite-level clusters, $|C_b|$ means the number of the vehicle in cluster C_b, and σ_S is the regulatory factor to ensure the L_S ranges from 0 to 1.

The initial matrix $\overline{\Theta_r^i}$ about data correlation is calculated based on Eq. (1). The largest element in the matrix $\overline{\Theta_r^i}$ is selected out and its corresponding clusters are integrated into a new cluster, then the L_S^g is calculated where g represents the iteration times. If $L_S^g > L_S^{(g-1)}$ and $|L_S^g - L_S^{(g-1)}| < \varrho$, the cluster structure is considered as the final state. The new and integrated cluster is restored back to the original two clusters, the clustering process is terminated and $C_S = \{C_1, C_2, \dots, C_{|C_S|}\}$ is the cluster set of the final satellite-level cluster structure.

3.1.4. Space-level clustering

Space-level clustering contributes to the vehicles that are not within the coverage of the same satellite. In the space-level clustering, the space-level clustering that integrates the satellite and the vehicle is adopted. Only the noise interference from satellites is used to calculate the space-level correlation function, which is formed as follows:

$$L_P = \sum_{a=1}^{M_P} \sum_{i=1}^{|C_b|} \sum_{j=1}^{|C_b|} \sigma_P * I_S \tag{8}$$

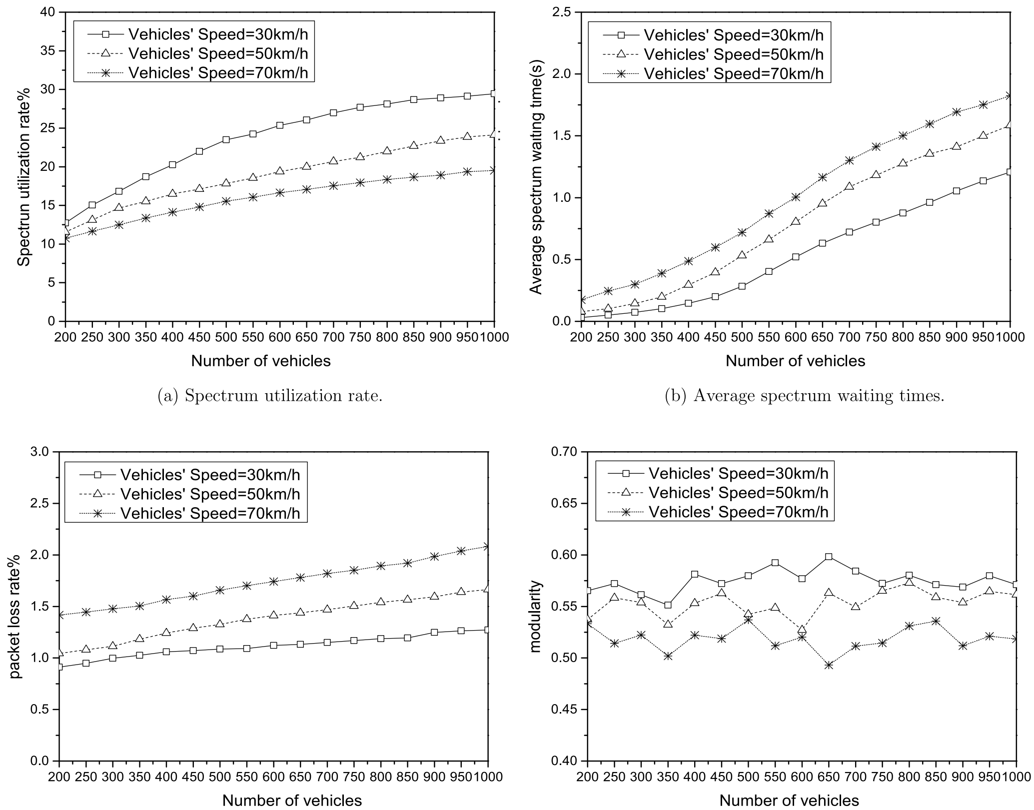
where M_P is the number of the satellite-level clusters, $|C_b|$ is the number of the vehicle in cluster C_b, and σ_P is the regulatory factor to ensure the L_P ranges from 0 to 1.

Setting ε as the space-level correlation threshold, if $L_P > \varepsilon$ firstly, then the communication between a vehicle and a satellite can be successfully performed, and the largest value corresponding the satellite is selected out to make communication with the vehicle v_i; moreover, the modularity Q_g of the cluster structure is calculated at gth iteration. If $Q_g > Q_{g-1}$ and $|Q_g - Q_{g-1}| < \varsigma$ (ς is a pre-set value), the cluster structure is stable. Otherwise, the current cluster structure is considered as the initial state of the next iteration and is used to re-calculate the matrix L_P. The above operations are performed until the cluster structure reach a stable state. The formed space-level cluster structure of the IoVs is represented as $C_P = \{C_1, C_2, \dots, C_{|C_P|}\}$.

3.2. Communication scheduling

The search for the best communication mode for vehicles to obtain stable mobile communication services, while considering network throughput and data correlation, is a challenging issue [27]. In this context, the machine learning method is used to search for the global optimized communication scheduling decision to ensure that the vehicles can adaptively access, also providing a good user experience for mobile services in the IoV.

Once the clusters of all levels are formed, the clusters which include vehicle v_i form the vector $E_i = \{C_V^i, C_I^i, C_S^i, C_P^i\}$. C_x^i is the cluster of



(a) Spectrum utilization rate.

(b) Average spectrum waiting times.

(c) Packet loss rate.

(d) Modularity of V2V clusters.

Fig. 4. The performance of the proposed MCSIC.

level x that include vehicle v_i. Vector $K = \{K_V^{ij}, K_I^{ij}, K_S^{ij}, K_P^{ij}\}$ is the communication mode to be selected for end-to-end data transmission between vehicle v_i and v_j. K_x^{ij} is the element number of $E_i \cap E_j \cap C_x$, and the communication modes include V2V, VIV (vehicle–infrastructure–vehicle), VSV (vehicle–satellite–vehicle) and VSSV (vehicle–satellite–satellite–vehicle). As each vehicle only belong to no more than one cluster in each cluster level, the element in vector K is 0 or 1.

For these four communication modes, vehicles are likely to select the lower level communication if their K of that level is 1, but it will occur communication congestion. To avoid this situation, the machine learning is used to predict the communication competition status of each communication level by using the historical status such as the number of communication participants, the maximum traffic and SINR value in each level. Thus the vector $\eta = \{\eta_V, \eta_I, \eta_S, \eta_P\}$ of communication status of each level is generated.

The Eq. (9) is used to choose the communication level for two vehicles v_i and v_j that need to exchange their data.

$$\xi = \max\{\frac{K_V^{ij}}{\eta_V}, \frac{K_I^{ij}}{\eta_I}, \frac{K_S^{ij}}{\eta_S}, \frac{K_P^{ij}}{\eta_P}\} \tag{9}$$

According to the above equation, the communication mode with the maximum value is selected. For example, vehicle v_i and v_j adopt the V2V communication mode if $\xi = \frac{K_V^{ij}}{\eta_V}$.

4. Simulations

In order to evaluate the proposed MCSIC model, several simulations have been performed using NS-3 network simulator while the data correlation is calculated for clustering. In the simulation, three kinds of nodes are defined, which include vehicles, roadside units and satellites. The simulation area is fixed to 10 km $*$ 10 km, the number of satellites and roadside units are fixed to 2 and 20 respectively, while the number of vehicles is ranging from 200 to 1000. The comparison is executed under different vehicle speeds: 30 km/h, 50 km/h and 70 km/h.

Due to the multi-level framework, the vehicles can access the spectrum from either roadside units or satellites. With the assistance of multi-level communication, which includes V2V, VIV, VSV and VSSV, any vehicle has four choices for communication, thus the communication competition in single level is reduced. Moreover, the VSV and VSSV support long-distance end-to-end communication for vehicles through satellite relay. The evaluation of the proposed multi-level framework takes into account four aspects: spectrum utilization rate, average spectrum waiting time, packet loss rate and modularity. The spectrum utilization rate is the ratio of occupied spectrum to total spectrum. The average waiting time is the ratio of vehicles' total waiting time for spectrum to vehicles' number. The packet loss rate is the ratio of the failure transmission packet number to the total input packet number. The modularity is the modularity value of vehicle-level clustering.

The Fig. 4(a) shows the spectrum utilization rate, it increases with the increasing number of vehicles since the MCSIC can make full use of the free spectrum resources by multi-level communication. Although the increasing speed of vehicles leads to the decrease of spectrum utilization, the utilization rate is still over 15% when the amount of vehicles is over 500 and the speed of vehicles is 70 km/h. The Fig. 4(b) shows the evaluation of vehicles' average spectrum waiting time, although it increases with the increasing number of vehicles, it is still below 2 seconds due to multi-selection of communication modes. The packet loss rate is shown in Fig. 4(c). Since the communication competition is reduced by using multi-level framework and the redundant communication content is reduced by high correlation, the packet loss rate is quite low; always below 3% with slow increasing speed caused by the increasing number of vehicles. In Fig. 4(d), it is possible to observe that the modularity has a small fluctuation range with the increasing number and speed of vehicles, which proves the stability of the vehicle-level clustering.

5. Conclusion

In this paper, the effective mobile communication service in IoV is implemented by using satellite-terrestrial integrated communication. The STICI framework is proposed to integrate different communication modes, and support unsupervised vehicle tracking and communication mode selection. Furthermore, a multi-level cluster-based satellite-terrestrial integrated communication model is designed to improve real-time data communication based on data analysis. The communication mode selection is determined by data correlation, location, and channel competition. Finally, the proposed MCSIC model is evaluated through simulations to validate the obtained performances.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was partially supported by the Liaoning Province Higher Education Innovative Talent Support Program; and by the Fundamental Research Funds for the Central Universities under Grant No. DUT19JC22.

References

[1] Linjun Zhang, Gábor Orosz, Beyond-line-of-sight identification by using vehicle-to-vehicle communication, IEEE Trans. Intell. Transp. Syst. 19 (6) (2018) 1962–1972.
[2] Yin Zhang, Min Chen, Guizani Nadra, Di Wu, C.M. Leung Victor, SOVCAN: Safety-oriented vehicular controller area network, IEEE Commun. Mag. 55 (8) (2017) 94–99.
[3] Yin Zhang, Xiao Ma, Jing Zhang, M. Shamim Hossain, Ghulam Muhammad, Syed Umar Amin, Edge intelligence in the cognitive Internet of Things: Improving sensitivity and interactivity, IEEE Netw. 33 (3) (2019) 58–64.
[4] Kai Lin, Min Chen, Jing Deng, Hassan Mohammad Mehedi, Fortino Giancarlo, GVT: Green video transmission in the mobile cloud networks, IEEE Trans. Circuits Syst. Video Technol. 27 (1) (2016) 159–169.
[5] Kai Lin, Jeungeun Song, Jiming Luo, Wen Ji, Ahmed Ghoneim, Enhanced fingerprinting and trajectory prediction for IoT localization in smart buildings, IEEE Trans. Autom. Sci. Eng. 13 (3) (2016) 1294–1307.
[6] Junhua Wang, Kai Liu, Ke Xiao, Chao Chen, Weiwei Wu, Victor C.S. Lee, Sang Hyuk Son, Victor C.S. Lee, Sang Hyuk Son, Dynamic clustering and cooperative scheduling for vehicle-to-vehicle communication in bidirectional road scenarios, IEEE Trans. Intell. Transp. Syst. 19 (6) (2018) 1913–1924.
[7] Myounggyu Won, Taejoon Park, Sang H. Son, Toward mitigating phantom jam using vehicle-to-vehicle communication, IEEE Trans. Intell. Transp. Syst. 18 (5) (2017) 1313–1324.
[8] Wern-Yarng Shieh, Chen-Chien James Hsu, Ti-Ho Wang, Vehicle positioning and trajectory tracking by infrared signal-direction discrimination for short-range vehicle-to-infrastructure communication systems, IEEE Trans. Intell. Transp. Syst. 19 (2) (2018) 368–379.
[9] Yin Zhang, Haoyu Wen, Feier Qiu, Zie Wang, Haider Abbas, IBike: Intelligent public bicycle services assisted by data analytics, iBike Future Gener. Comput. Syst. 95 (2019) 187–197.
[10] Keita Emura, Takuya Hayashi, Road-to-vehicle communications with time-dependent anonymity: A lightweight construction and its experimental results, IEEE Trans. Veh. Technol. 67 (2) (2018) 1582–1597.
[11] Zhenyu Zhou, Junhao Feng, Chuntian Zhang, Zheng Chang, Kazi Mohammed Saidul Huq, SAGECELL: Software-defined space-air-ground integrated moving cells, IEEE Commun. Mag. 56 (8) (2018) 92–99.
[12] Kang An, Min Lin, Tao Liang, On the performance of multiuser hybrid satellite-terrestrial relay networks with opportunistic scheduling, IEEE Commun. Lett. 19 (10) (2015) 1722–1725.
[13] Chao Qiu, Haipeng Yao, F. Richard Yu, Fangmin Xu, Chenglin Zhao, Deep q-learning aided networking, caching, and computing resources allocation in software-defined satellite-terrestrial networks, IEEE Trans. Veh. Technol. 68 (6) (2019) 5871–5883.
[14] Ning Zhang, Shan Zhang, Peng Yang, Omar Alhussein, Weihua Zhuang, Xuemin Sherman Shen, Software defined space-air-ground integrated vehicular networks: Challenges and solutions, IEEE Commun. Mag. 55 (7) (2017) 101–109.
[15] Gang Xiong, Fenghua Zhu, Xisong Dong, Haisheng Fan, Bin Hu, Qingjie Kong, Wenwen Kang, Teng Teng, A kind of novel ITS based on space-air-ground big-data, IEEE Intell. Transp. Syst. Mag. 8 (1) (2016) 10–22.
[16] Yin Zhang, Yongfeng Qian, Di Wu, M. Shamim Hossain, Ahmed Ghoneim, Min Chen, Emotion-aware multimedia systems security, IEEE Trans. Multimed. 21 (3) (2019) 617–624.
[17] Angadh Nanjangud, Peter C. Blacker, Saptarshi Bandyopadhyay, Yang Gao, Robotics and ai-enabled on-orbit operations with future generation of small satellites, Proc. IEEE 106 (3) (2018) 429–439.
[18] Karim Miftah Al, Currie Jonathan, Lie Tek Tjing, A machine learning based optimized energy dispatching scheme for restoring a hybrid microgrid, Electr. Power Syst. Res. 155 (2018) 206–215.
[19] Yin Zhang, Xiao Ma, Shaohua Wan, Haider Abbas, Mohsen Guizani, Cross-Rec: Cross-domain recommendations based on social big data and cognitive computing, Mob. Netw. Appl. 23 (6) (2018) 1610–1623.
[20] Takehisa Yairi, Naoya Takeishi, Tetsuo Oda, Yuta Nakajima, A data-driven health monitoring method for satellite housekeeping data based on probabilistic clustering and dimensionality reduction, IEEE Trans. Aerosp. Electron. Syst. 53 (3) (2017) 1384–1401.
[21] Qingshan Liu, Renlong Hang, Huihui Song, Zhi Li, Learning multiscale deep features for high-resolution satellite image scene classification, IEEE Trans. Geosci. Remote Sens. 56 (1) (2018) 117–126.
[22] Kai Lin, Jiming Luo, Long Hu, M. Shamim Hossain, Ahmed Ghoneim, Localization based on social big data analysis in the vehicular networks, IEEE Trans. Ind. Inf. 13 (4) (2017) 1932–1940.
[23] Kai Lin, Chensi Li, Daxin Tian, Ahmed Ghoneim, M. Shamim Hossain, Syed Umar Amin, Artificial-intelligence-based data analytics for cognitive communication in heterogeneous wireless networks, IEEE Wirel. Commun. 26 (3) (2019) 83–89.
[24] Weiwei Sun, Qian Du, Graph-regularized fast and robust principal component analysis for hyperspectral band selection, IEEE Trans. Geosci. Remote Sens. 56 (6) (2018) 3185–3195.
[25] Davud R. Hardoon, Sandor Szedmak, John Shawe-Taylor, Canonical correlation analysis: An overview with application to learning methods, Neural Comput. 16 (12) (2004) 2639–2664.
[26] Kai Lin, Di Wang, Fuzhen Xia, Hongwei Ge, Device clustering algorithm based on multimodal data correlation in cognitive Internet of Things, IEEE Internet Things J. 5 (4) (2018) 2263–2271.
[27] Park Seungil, Yoon Hoyoung, Kim Byungjun, Choi Sunghyun, FAGA: Feedback-aided greedy algorithm for periodic messages in LTE V2V communications, IEEE Trans. Veh. Technol. 67 (11) (2018) 11062–11068.