

Applications of intelligent computing in vehicular networks

Daxin Tian

Beihang University, Beijing, China

Weiqliang Gong

Nanjing LES Information Technology Co., Ltd, Nanjing, China

Wenhao Liu, Xuting Duan, Yukai Zhu and Chao Liu

Beihang University, Beijing, China, and

Xin Li

Department of Computer Science and Technology, Tsinghua University, Beijing, China

Abstract

Purpose – This paper aims to introduce vehicular network platform, routing and broadcasting methods and vehicular positioning enhancement technology, which are three aspects of the applications of intelligent computing in vehicular networks. From this paper, the role of intelligent algorithm in the field of transportation and the vehicular networks can be understood.

Design/methodology/approach – In this paper, the authors introduce three different methods in three layers of vehicle networking, which are data cleaning based on machine learning, routing algorithm based on epidemic model and cooperative localization algorithm based on the connect vehicles.

Findings – In Section 2, a novel classification-based framework is proposed to efficiently assess the data quality and screen out the abnormal vehicles in database. In Section 3, the authors can find when traffic conditions varied from free flow to congestion, the number of message copies increased dramatically and the reachability also improved. The error of vehicle positioning is reduced by 35.39% based on the CV-IMM-EKF in Section 4. Finally, it can be concluded that the intelligent computing in the vehicle network system is effective, and it will improve the development of the car networking system.

Originality/value – This paper reviews the research of intelligent algorithms in three related areas of vehicle networking. In the field of vehicle networking, these research results are conducive to promoting data processing and algorithm optimization, and it may lay the foundation for the new methods.

Keywords Intelligent computing, Vehicular *ad hoc* networks

Paper type Research paper

1. Introduction

The advancement of computer technology in recent years has allowed researchers to develop efficient optimization techniques for solving large-scale problems in various fields and helped practitioners to incorporate some of these techniques into their planning activities through specific integrated decision support systems (Jin and Meng, 2010; Chiong and Weise, 2011). Despite this, many real-world problems still cannot be solved within reasonable computational time using exact algorithms because of the complexities associated with the large number of decision variables involved and the extent of the constraints imposed (Chetty *et al.*, 2010). As a consequence, non-exact methods such as custom designed heuristics, search algorithms and meta-heuristic approaches

have become increasingly popular due to algorithmic efficiency, ease of implementation, flexibility for adaption and quality of solutions (Chiong *et al.*, 2012). Among them, intelligent computing approaches inspired by principles of nature such as evolutionary algorithms, swarm intelligence algorithms, artificial neural networks and fuzzy logic have emerged as a rapidly growing research area. These approaches have been applied to various problems in different fields (Catay *et al.*, 2013; Jin and Hammer, 2014).

© Daxin Tian, Weiqliang Gong, Wenhao Liu, Xuting Duan, Yukai Zhu, Chao Liu and Xin Li. Published in *Journal of Intelligent and Connected Vehicles*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

This research was supported in part by the National Natural Science Foundation of China under Grant Nos. 61672082, U1564212.

Received 11 January 2018

Revised 9 April 2018

12 April 2018

Accepted 13 April 2018

The current issue and full text archive of this journal is available on Emerald Insight at: www.emeraldinsight.com/2399-9802.htm



Journal of Intelligent and Connected Vehicles
1/2 (2018) 66–76
Emerald Publishing Limited [ISSN 2399-9802]
[DOI 10.1108/JICV-01-2018-0001]

With the rapid development of the key technology of wireless networks, intelligent computing and other intelligent technologies, networking industry makes the vehicular networks enter a new stage of research, research on cross car networking involves many subjects, including the automotive industry, transportation, information and communication technology (Liu *et al.*, 2010). The promising applications of vehicular *ad hoc* networks (VANET) are collision warning, location tracking, emergency brake notification, signal violation notification, parking lot availability, weather information and major information services on the road (Saravanan and Arunkumar, 2014; Saravanan and Arunkumar, 2016). At the same time, some of applications such as routing algorithms (Ratnasamy *et al.*, 2002; Golden *et al.*, 2010), data mining (Kargupta *et al.*, 2004; Kargupta, 2010), positioning enhancement (Golestan *et al.*, 2015; Du *et al.*, 2015) and others have been widely used in the vehicular networks on the basis of intelligent computing.

In Section 2, this paper discusses the methods vehicular network platform and data quality assessment, and it also compares performance metrics. In Section 3, a routing and broadcasting methods in vehicular networks is drawn into this paper, and the effect has been improved obviously than before. We propose a method of vehicle positioning enhancement named CV-IMM-EKF in Section 4. Finally, this paper will sum up the three sections in Section 5.

2. Methods vehicular network platform and data quality assessment

Focus on our topic on internet of vehicle, vehicular network data need to be of high quality to reflect the operation pattern and running status of the vehicle, including vehicle fuel consumptions and traveling route (Tian *et al.*, 2015). However, due to abnormalities of sensors and the unreliability of wireless communication environment, some data of vehicular sensors are inaccurate or incomplete. Hence, the quality of vehicular data is difficult to be guaranteed, which has strong relation to the validity and accuracy of data mining results (Hou *et al.*, 2016; Chen *et al.*, 2017). Therefore, quality control and analysis on these data is indispensable, to assessment the quality level of the data or to detect the sensor error, filter out the vehicle with inferior data quality.

In this section, we give a brief introduction of vehicular network big data platform and its data quality problem, including the database and outlier situation. Then we introduced a classification framework based on data fusion and machine learning approach to assess the quality of vehicular data.

2.1 Vehicular network data platform and data quality problem

A brief system framework of integrated vehicular network contains cellular networks and vehicle-to-vehicle (V2V) communication (Biswas and Giaffreda, 2014). The topic and data set introduced in this paper is from a vehicular network platform, which contain the vehicle running state data from thousands of vehicles. These data consist of global positioning system (GPS) coordinates, speeds, driving direction angle, fuel level value and so on. Figure 1 shows the real-time big data

platform of our vehicular network and the data types of database. However, there are several types of abnormal situation during the process of fuel level data acquisition.

It is well-known that there should be positive correlation between the correct fuel consumption and the speed data, which is the fuel level should decline when the vehicle is running (Chen and Zhao, 2014). Contrary to the common sense, the fuel level of some vehicles remains unchanged while the vehicle is moving over an extended period. In some other conditions, same values appear frequently, while the uploaded fuel-level data are float numbers which should always fluctuate in a lesser range. We suppose that the possible causes of these abnormal situations are in the following items:

- The sensors of vehicles may have too much noise or their errors are too large.
- The terminal equipment may upload or read wrong values or error codes in communication with the sensors.
- The information processing system on the server side may read and update wrong or null values into the database.

2.2 Data quality analysis for vehicular network data

Data quality problem is a further topic of anomaly detection, which is generally statistical approach (Chen *et al.*, 2017). It could make use of the historical distribution of data. As a result, the classification approach is much stronger connected to their application area; this is often achieved by building rules. Multi-sensor data fusion generally provides significant advantages in data mining procession. In addition to the statistical advantage gained by obtaining an improved estimate of a physical phenomenon through redundant observations. The use of multiple types of sensors may increase the accuracy with which a phenomenon can be observed.

In this study, we tried to find the vehicles with large amounts of outliers which are defined as typical samples with abnormal data through observing the fuel level-time curve. The typical fuel level-time curves of two sample sets are as Figure 2. From the comparison of these figures, Figure 2(a) shows a typical fuel level curve of the sample with relatively credible data, while Figure 2(b) and (c) shows the opposite situation.

Figure 1 Vehicular network big data platform

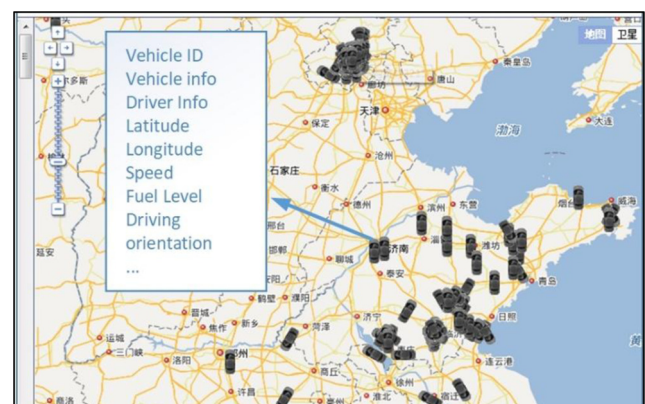
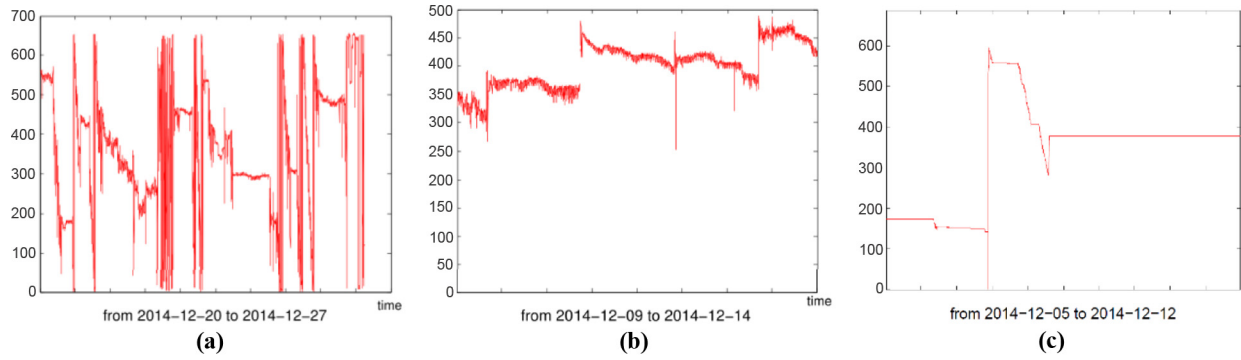


Figure 2 Typical fuel level curves

Notes: (a) Credible fuel-level; (b) superabundant noise; (c) abnormal invariable data

2.2.1 Features extraction based on data fusion and discrete wavelet transform

Data fusion with vehicular sensor data. To assess the data quality of the speed data, we compute the travel speed according to the GPS coordinates and compare it with the speed which is collected from running cars on real time (Smith et al., 2011). The travel speeds could be computed by latitudes φ and longitudes λ , and we compute the relative error ratio of the uploaded speeds v_u with the calculated travel speeds v_t . We use these calculated travel speeds to evaluate the accuracy of the uploaded speeds and the GPS coordinates, and found that these data are credible to some extent. In equation (1), t is the timestamp and i means the i th data point in the series:

$$v_i = \frac{111.199 \left[(\varphi_i - \varphi_{i+1})^2 + (\lambda_i - \lambda_{i+1})^2 \cos\left(\frac{(\varphi_i + \varphi_{i+1})^2}{2}\right) \right]}{t_i - t_{i+1}} \quad (1)$$

Classification algorithm needs the input of the feature attributes before training. There is a relationship between fuel level and speed value; hence, we use standard deviation (STD) of these data and correlation coefficient between these multi-dimension data as features for classifier.

In the preprocessing procedure of many types of sensor data, the original series data may contain variety of noise and anomaly pattern. With the decomposition with discrete wavelet transform (Rodgers and Nicewander, 1988), the set of approximation coefficients A_k retained the main profile of the original data, which is a smooth series with less noise. On the other hand, the set of detail coefficients D_k represented the noise and slope which contains anomaly pattern in different scaling function. The signal is decomposed simultaneously using a high-pass filter H . The outputs give the detail coefficients from the high-pass filter and approximation coefficients from the low-pass. However, as half the frequencies of the signal have now been removed, half the samples can be discarded:

$$y_H[n] = [n](x * G) = \sum_{k=-\infty}^{\infty} x[k]H[2n-k] \quad (2)$$

$$y_L[n] = [n](x * G) = \sum_{k=-\infty}^{\infty} x[k]G[2n-k] \quad (3)$$

2.2.2 Data quality assessment with machine learning algorithm

Bayesian network is a useful and powerful algorithm for uncertainty reasoning. For a data set S with attribute set $X = \{X_1, \dots, X_n\}$ and class label $Y = \{y_1, \dots, y_n\}$, $t = \{x_1, \dots, x_n\}$ is a sample in S , the classifier predicts the class label of t by calculating probability. To overcome the limitation of conditional independent assumption, functional dependency is an important part of relational database, which represents the constraint relations among different attributes. Some existing research studies have found some similarities between relational database and probabilistic inference. Then we have:

$$\begin{aligned} & \arg \max_{\sum} P(y|x_1, \dots, x_n) \\ & = \arg \max_{\sum} P(y) \cdots P(\alpha|y) \cdots P(x_n|y) \end{aligned} \quad (4)$$

In semi-supervised learning, as the joint likelihood of the labeled and unlabeled data is not in closed form, usual solutions to this would be to use expectation maximization (Chen and Zhao, 2014; Bishop, 2006). The algorithm initials model parameters from limited amount of labeled data and uses these to get probabilistic labels for each unlabeled sample in E-step. M-step gets the parameters using these labels for unlabeled instances. A regulating variable λ is used in the proposed method. This parameter moderates the unlabeled data by reduce the learning rate η and the weight of the unlabeled samples in step M . θ_t is current estimator of the model parameters. f_i^x gives the count of feature in instance x . Conditional estimates of $P(y|x)$ from labeled data improve the accuracy of Bayesian approach:

$$\begin{aligned} & \text{Initial: } \theta_L = \arg \max_{x \in D_l} \log p_{\theta}(x, y) \\ & \text{Estep: } \forall x \in D_u \cup D_l \text{ compute } p_{\theta_t}(y|x) \\ & \text{Mstep: } \theta_{t+1} = \arg \max_{x \in D_u \cup D_l} \log p_{\theta_t}(x, y) \end{aligned} \quad (5)$$

$$N_{yi} = \sum_{x \in D_u} f_i^x P_{\theta_t}(y|x) \quad (5a)$$

In the vehicular network data quality problem, the quality states of this deployment are adopted from the flagging scheme. There are flags associated with particular processing tasks. The quality flags are designed to indicate the level of uncertainty involved in the sensor reading. Sometimes, sensor readings associated with abnormal events may be incorrectly labeled. These labels belong to four different classes of the classification problem. The data that we used in the experiments were labeled into four classes:

- 1 good data;
- 2 probably good data;
- 3 bad data which are potentially correctable; and
- 4 bad data.

Figure 3 shows a labeled data series after separated into many pieces of length; thus, a long series of data can be evaluated for each period.

The blue lines denoted the probably good or bad curves, and the yellow one is series with bad flag.

This work presents our experiences for big data analytics based on a vehicular network big data testbed, in terms of sensors data management, multi-dimension data fusion and data quality assessment for the vehicular sensor data. In this section, we have investigated the problem of multidimensional analysis of vehicular network testbed data. Some statistical indicators are introduced and applied in data quality evaluation, and a novel classification-based framework is proposed to efficiently assess the data quality and screen out the abnormal vehicles in database. Our experiments on large real data sets show the feasibility and utility of proposed methods.

3. Routing and broadcasting methods in vehicular networks

With the development of wireless communication technologies and increasing demand for high-data-rate communication, the integration of the cellular network with other access networks has become a popular trend. Device-to-device (D2D) communication is an example of an integrated network technology that appears to be a vital component in next generation communication technologies (5G). The D2D technology allows for direct message delivery between terminals that are near each other to lighten the load of node

base stations, improve spectral reuse and enhance system capacity. Therefore, mobile *ad hoc* networks (MANETs) are expected to be popularized in the future based on D2D communication (Mumtaz et al., 2015). VANETs, as a special application of MANETs in transportation, have attracted increasing academic and industry attention in recent years. These networks can help to realize a variety of vehicular safety and intelligent applications, such as cooperative collision avoidance, road obstacle warning, self-adapted cruise control system, multimedia streaming and other telematics services (Panichpapiboon and Pattara-Atikom, 2012). Nevertheless, VANETs have a highly dynamic characteristic due to the fast movement of vehicles, which results in a number of great challenges such as dynamic network topology, frequent disconnections, changing node density, and a lack of centralized control. For example, the urban road net is a complex network system. The vehicle's movement appears as a frequent entrance and departure to the area networks, which results in a lack of instantaneous end-to-end paths; hence, the VANET is considered as a type of delay and disruption-tolerant networks (DTNs). As a common phenomenon in large cities, traffic congestion causes the vehicular network to become very dense, which has an influence on the subsequent data dissemination. In addition, traffic scenes or phenomena such as freeways, platoons or cluster driving and tide traffic may lead to many specific network changing modes. The integrated vehicular network is shown in Figure 4, including internet, cellular network and D2D networks.

As the high mobility of vehicles leads to increased uncertainty in the opportunistic transmission between vehicles, many of the traditional MANET routing protocols and message dissemination models that depend on the maintenance of an end-to-end transmission path are no longer suitable for VANETs. Therefore, the development of a feasible and efficient broadcasting mechanism for message dissemination is necessary. The fundamental ability of any kind of communication network is the transportation or routing of data from a source to a destination. Communication paths determine the communication quality and transmission delay to a great extent. Routing is the process to select the best communication paths in a network. There has been much research concerning the optimal routing methods in *ad hoc* networks. The simplest approach may be to let the source or a moving relay node carry or retransmit the message to the destination. A suitable method to perform routing in delay tolerant networks is epidemic routing through use of the “store-carry-forward” mechanism. This mechanism is referred to as an epidemic broadcast, as an analogy of the transmitting behavior of viruses.

In epidemic broadcasts, messages are spread among nodes. When a node with a message labeled as an infected node contacts other nodes without a message, the infected node sends a copy of the message to those susceptible nodes. For more detail, a mobile node stores a message in its local memory after receiving this message from a source and carries the message along the movement path until the next hop moves into the communication range. Once it encounters another node, this infected node duplicates and forwards the copy to the new node. This “store-carry-forward” process will be repeated until the message is forwarded to the destination. In

Figure 3 Processed fuel level data series

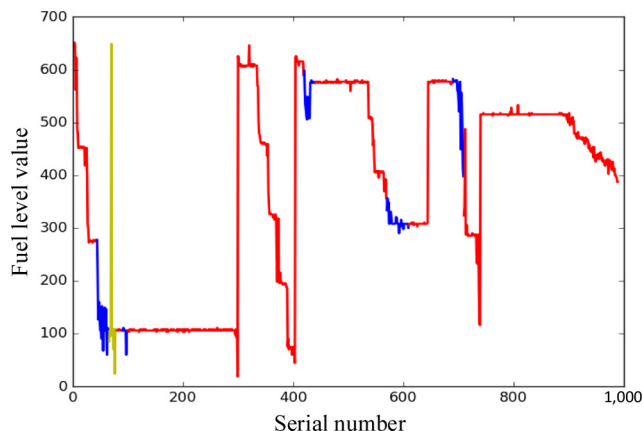
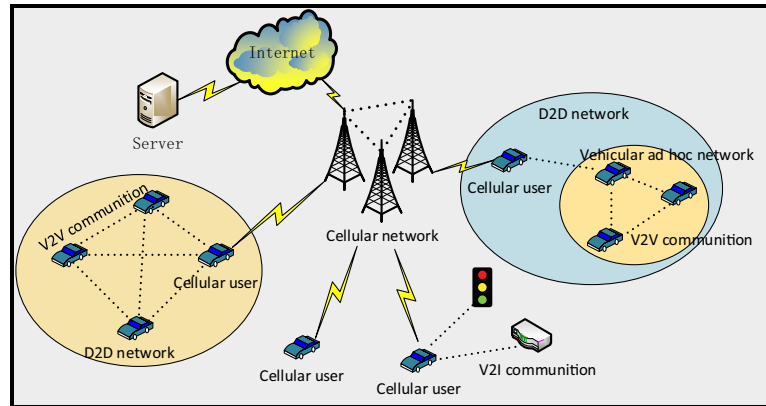
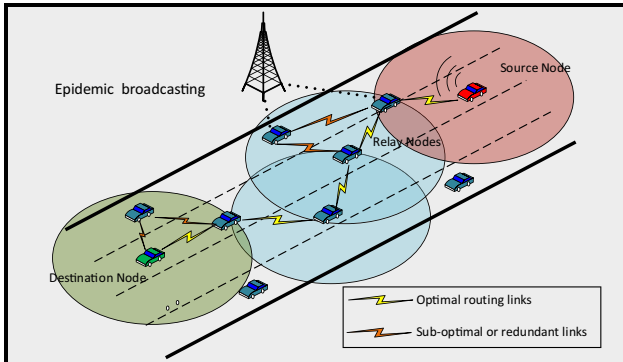


Figure 4 Integrated vehicular networks

VANETs, every vehicle can be such a mobile infected node, and thus the information packet will be able to propagate from the source to the other distant vehicles. In a dense road scene, a continuous communication path without cellular network can be found, this message dissemination process is shown in Figure 5.

3.1 Epidemic broadcasting simulation based on a realistic traffic scene

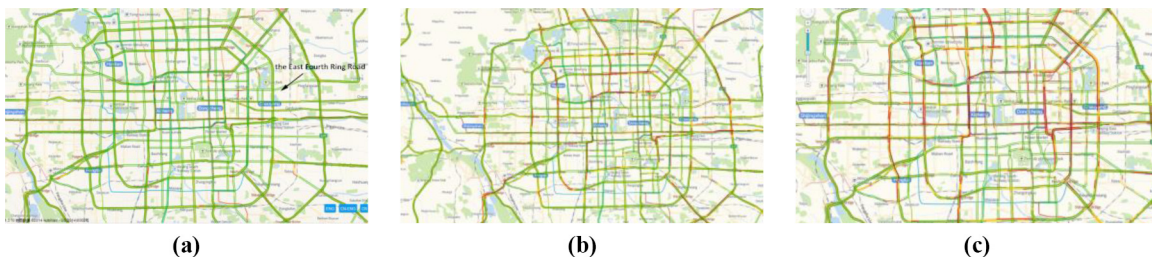
Simulation is the predominant tool used in research related to VANET (Liu et al., 2009). We will discuss a simulation that was based on real-world road traffic conditions on the East Fourth Ring Road in Beijing (Figure 6). We performed a field survey

Figure 5 Epidemic-based message dissemination on the road

and collected historical data from the East Fourth Ring Road to obtain reliable simulation parameters such as speed, density and congestion distribution. In Figure 6(a), the real-time traffic is colored in green, indicating that free-flow traffic conditions existed, where vehicles could travel at high speeds and at low densities. Under these conditions, drivers could arrive at their destinations within the prediction time. In Figure 6(b), the selected segment was mainly yellow, which indicated that traffic was moving at a relatively slower speed (approximately 45 km/h) and at densities on the middle level. In Figure 6(c), traffic congestion conditions were apparent, where both yellow and red sections were present on the selected segment. The red and yellow colors were distributed in two different directions on the road, where one direction was mainly yellow and the other direction was mainly red, indicating that vehicles were at low speeds and high densities.

3.2 Simulation setting

We simulated the data packages broadcasting in three scenarios (i.e. the free-flow traffic, heavy traffic and congestion scenarios), according to the traffic state of a road section of the East Fourth Ring Road of Beijing. To simplify the simulation, we simulated the traffic in a single lane for both directions. The length of the section was 5 kilometers. Two devices were located at each end of the section: the source node, which could generate data packages and the destination node, which would receive the message. In each scenario, the simulation interval was 1 s and the time was 150 s. The source node generated data packages every 5 s. Two vehicles could communicate only

Figure 6 Distribution of traffic conditions in Beijing

Notes: (a) Free-flow; (b) heavy traffic; (c) congestion

when their distance was less than 150 meters. When two vehicles were within communication range, one vehicle copied its particular data packages and sent them to the other. If two vehicles had the same data packages, there were no packages transmitted between them. The lifetime of each copy was 60 s. In the free-flow, heavy traffic and congestion scenarios, we set the average speed at 70, 50 and 25 kilometers per hour, respectively, and the number of vehicles in a lane was set to be 50, 250 and 400, respectively.

3.3 Results and analysis

We used the aforementioned simulation setting to simulate the process of data package broadcasting. In each scenario, we performed the simulation ten times within the same simulation setting. Figure 7 shows the results of the number of copies versus the time variation in the three scenarios. Based on the results, the destination node in the congestion traffic condition could receive more copies than in the free-flow condition. This result was as expected because the vehicles density is high in the congestion condition; thus, the copies can be more easily relayed than in the non-congestion condition.

We also observed the number of relays for a copy before it arrived at the destination node. Figure 8 shows the results of the number of copies versus the number of relays in the three scenarios. Intuitively, more relays were required to send the message when the vehicle density was higher. The average number of relays in the three scenarios (free-flow, heavy and congestion) were 39.8, 47.5 and 49.7, respectively, which was considered to be reasonable. In addition, we observed the time taken by a copy to arrive at the destination node, as shown in Figure 6. The average times obtained from the three scenarios (free-flow, heavy and congestion) were 9.3, 11.3 and 9.5 s, respectively. The reasons why the average times in the free-flow and congestion scenarios were less than that in the heavy traffic scenarios were attributed to the high vehicle speed in the free-flow scenario and the high vehicle density in the congestion scenario, both of which benefited the message transmission. In addition, we observed that many copies could arrive at the destination once they were generated. In the free-flow and

Figure 7 Distribution of the number of copies over the simulation times for different traffic conditions

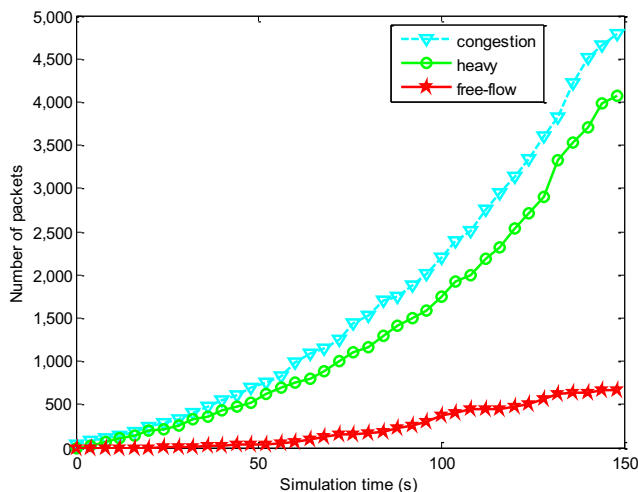
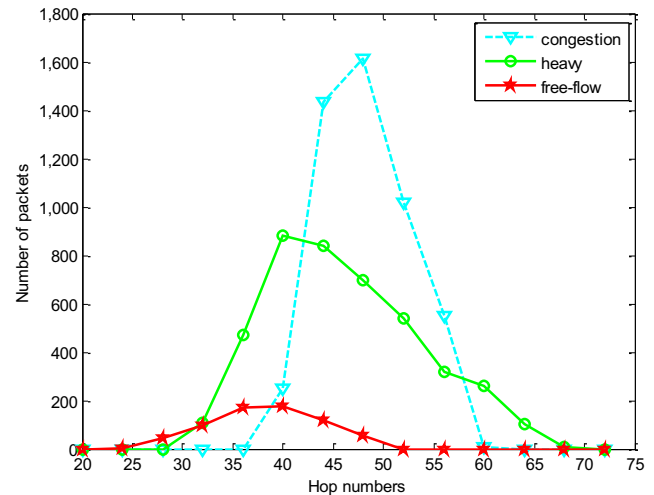


Figure 8 Relationship between the number of copies and the number of relays for different traffic conditions



heavy traffic scenarios, approximately 70 per cent of the copies could arrive at the destination node within 10 s, whereas approximately 80 per cent could arrived in the congestion scenario. Nevertheless, some copied packets were unable to reach the destinations when the density of vehicles in the free-flow scenario was too low.

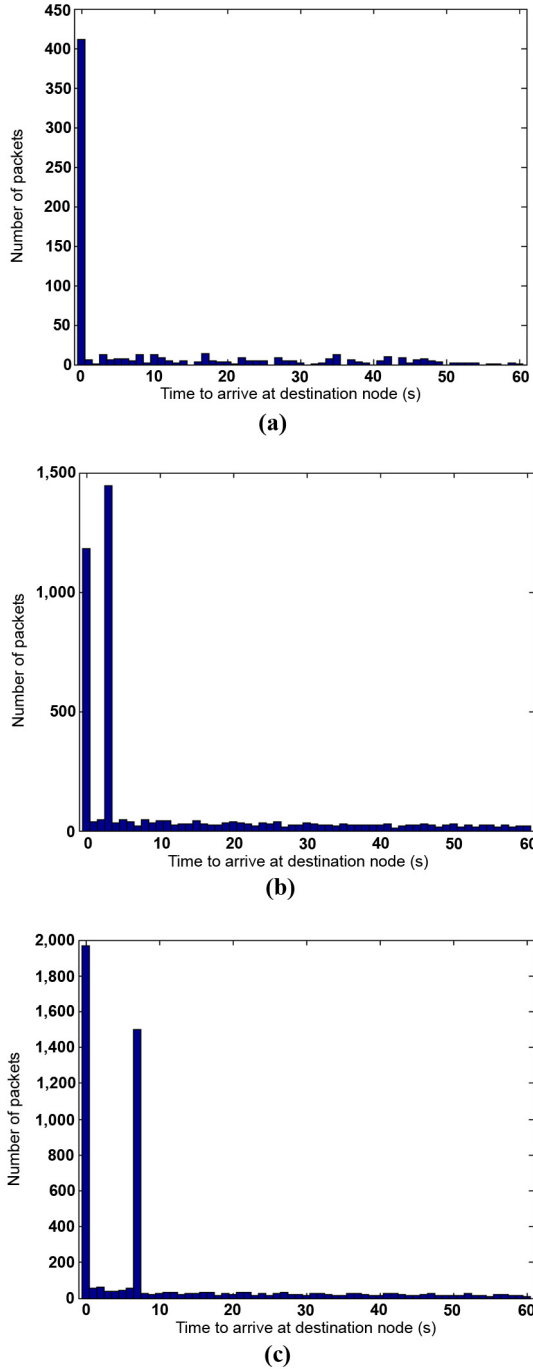
In the case study, we performed a broadcasting simulation based on real-road traffic conditions from the East Fourth Ring Road in Beijing, and observed that different vehicle densities can dramatically affect message transmission. Moreover, experiment results showed that when traffic conditions varied from free-flow to congestion, the number of message copies increased dramatically and the reachability was also improved. Nevertheless, the high network density may introduce other issues, such as the broadcast storm problem.

4. Vehicular positioning enhancement with connected vehicle

The availability of high-accuracy location-awareness is essential for a diverse set of vehicular applications including intelligent transportation systems, location-based services, navigation and a couple of emerging cooperative vehicle-infrastructure systems (Alam and Dempster, 2013). However, it still faces a big challenge in the areas with inconsistent availability of satellite networks, especially in dense urban areas where the standalone global navigation satellite systems (e.g. GPS).

To tackle the aforementioned problems, a new class of cooperative positioning (CP) methods that relies on V2V communications and data fusion filtering (Sayed et al., 2005; Win et al., 2011; Alam et al., 2012) has been presented in recent years, which can further improve the accuracy of positioning. Actually, such concern raised in CP is the reliability of the localization approaches in heavy multipath and NLOS scenarios, which is similar to that in indoor environment (Sayed et al., 2005; Gu et al., 2009). An exploration into CP from theoretical foundations to specific applications, including advanced technologies and cooperative algorithms was presented in Win et al. (2011). Several CP techniques have

Figure 9 Relationship between the number of copies and the time required for the copy to arrive at the destination node in different traffic condition scenarios



Notes: (a) Free-flow; (b) heavy traffic; (c) congestion

been proposed to achieve lane-level vehicular positioning using DSRC signals broadcasted by onboard both units (OBUs) and RSUs (Alam *et al.*, 2012; Dao *et al.*, 2007).

Because of the low speed of vehicles, the DFS is too difficult to be extracted from noise, and thus for DFS vehicular positioning methods, the STD of positioning error increases as

the relative speed between the target vehicle (TV) and the other vehicles decreases. So we will investigate the method to overcome this problem. In this section, we will focus on the scenario that the neighbor vehicles travel in the opposite direction of the TV, for this case can provide obviously detectable Doppler effect.

4.1 Problem statement

The problem to be solved is to estimate the position of a TV moving on a 2D road, where there are many other neighbors around the TV. All of the vehicles are able to know their own state information, including position and velocity, provided by coarse GPS receiver, and they can know the neighbors' state information via vehicular communications as well. Consequently, this case can be treated as a simple but practical CV scenario. A TV is considered as a research object for positioning enhancement based on CV, and a neighbor is considered as the vehicle who is within a certain communication range to the TV and travel in the opposite direction of the TV. Each vehicle is assumed to be with an offshore banking unit providing both the DSRC and the DFS measurements (Duan *et al.*, 2015).

Considering the i th moving vehicle at time instant k with a state vector $\theta_k^i = [p_{x,k}^i, p_{y,k}^i, v_{x,k}^i, v_{y,k}^i]^T$, $i = 1, \dots, n_p$, where the $(p_{x,k}^i, p_{y,k}^i)$ and $(v_{x,k}^i, v_{y,k}^i)$ denote the i th vehicle's position and velocity, respectively, n_p is the total number of the vehicles driving on the road and T is a transpose operator. The dynamic state can be modeled by the following system:

$$\theta_k^i = \Phi_{k-1}^i \theta_{k-1}^i + G_{k-1}^i (u_{k-1}^i + w_{k-1}^i) \quad (6)$$

where Φ_{k-1}^i is the state transition matrix, and G_{k-1}^i is the noise distribution matrix. u_{k-1}^i is the control vector and w_{k-1}^i is zero-mean white Gaussian noise with covariance matrix C_{k-1}^i .

For the dynamic model presented by equation (6), the following observation model can be defined:

$$\psi_k = h(\theta_k) + \gamma_k(r_e) \quad (7)$$

where $h = [p_{x,k}^j, p_{y,k}^j, v_{x,k}^j, v_{y,k}^j, \omega_k^1, \dots, \omega_k^j]^T$ is a nonlinear observation vector in terms of θ_k . ω_k^j is the DFS of the received signal from the j th neighbor, $j = 1, \dots, n_k$, $n_k < n_p$, and n_k is the total number of the neighbors on the road. $\gamma_k(r_e)$ is the observation noise that can be used to describe the M types of observation errors by assuming a set of another covariance matrixes. The transition among M types of the errors is generally modeled as a first-order M -state homogeneous Markov chain r_e , $e = 1, 2, \dots, M$.

Specifically, assuming that the DFS measurements from the OBU can be modeled in a derivative form of the DSRC carrier frequency, f , as follows:

$$\begin{aligned} \omega_k^j &= -\frac{f}{c} \nabla_t (d_k^j + \vartheta_k^j r_e) \\ d_k^j &= \sqrt{(p_{x,k} - p_{x,k}^j)^2 + (p_{y,k} - p_{y,k}^j)^2} \end{aligned} \quad (8)$$

where c is the speed of light, d_k^j is the relative distance between the TV and its neighbor j and ϑ_k^j is the DFS observation noise of neighbor j . where $(p_{x,k}^j, p_{y,k}^j)$ and $(v_{x,k}^j, v_{y,k}^j)$ is the position and velocity vector of the neighbor j . To solve this nonlinear observation function, with the first-order Taylor expansion of (8) around an arbitrary state vector, h can be transformed to a fixed form of matrix, in which all of the components are supposed to obtain from both the GPS and OBU. As a result, the observation model of the TV can be reformulated as a linear one:

$$Z_k = H_k \theta_k + \gamma_k(r_e) \quad (9)$$

with the observation transition matrix H_k .

Based on the aforementioned models from equation (6) and (9), it is reasonable to assume that Φ_{k-1}^i and G_{k-1}^i in the system model are invariable at both each time instant and vehicle. Therefore, the position estimation of the TV can be formulated as the problem of linear filtering for M-state jump Markov systems and the model can be simplified as:

$$\begin{aligned} \theta_k^i &= \phi \theta_{k-1}^i + G(u_{k-1}^i + w_{k-1}^i) \\ Z_k &= H_k \theta_k + \gamma_k(r_e) \end{aligned} \quad (10)$$

Because H_k can be estimated by data fusion from both the GPS and OBU at each time instant k , a CV-enhanced interacting multiple model extended Kalman filter (IMM-EKF) can be deployed.

4.2 Connected vehicles-enhanced interacting multiple model extended Kalman filtering for vehicular positioning

4.2.1 Mixing probabilities and state estimates:

$$\mu_{k+1,s|t} = \pi_{st} \mu_{k,s} / c_t \quad (11)$$

where $\mu_{k+1,s|t}$ is known as the mixing probability in the IMM estimator, $\mu_{k,s}$ is the probability of the event that the s th motion model is in effect at time step k , $s, t = 1, 2, \dots, M$, correspond to the s, t th mode of the Markov chain r_e and c_t is a normalization constant:

$$\begin{aligned} \theta_{k|k,t}^0 &= \sum_{s=1}^M \mu_{k+1,s|t} \theta_{k|k,t}^{0,0} \\ P_{k|k,t}^0 &= \sum_{s=1}^M \mu_{k+1,s|t} \times \{P_{k|k,t}^{0,0} + [\theta_{k|k,t}^{0,0} - \theta_{k|k,t}^0] \\ &\quad \times [\theta_{k|k,t}^{0,0} - \theta_{k|k,t}^0]^T\} \end{aligned} \quad (12)$$

4.2.2 Mode update and prediction steps:

The CV-IMM-EKF advanced prediction is given by:

$$\theta_{k|k-1,t} \approx \phi_k \theta_{k-1|k-1,t} = \phi \theta_{k-1|k-1,t} \quad (13)$$

The State prediction error covariance matrix is as follows:

$$\begin{aligned} P_{k|k-1,t} &\approx \phi_k P_{k-1|k-1,t} \phi_k^T + C_{k-1} \\ &= \phi P_{k-1|k-1,t} \phi^T + C_{k-1} \end{aligned} \quad (14)$$

From the previous data, the CV-IMM-EKF gain is given by:

$$\begin{aligned} T_k &= P_{k|k,t} H_k^T (\varphi_k(N)) \times \{H_k(\varphi_k(N)) P_{k|k-1,t}(\varphi_k(N)) \\ &\quad + R_k(r_e(N))\}^{-1} \end{aligned} \quad (15)$$

where $\varphi_k(N)$ and $r_e(N)$ are functions of N and can change the dimension of the observation transition matrix H_k and the covariance matrix R_k , respectively.

The CV-IMM-EKF update steps are given by:

$$\begin{aligned} \theta_{k|k,t}^0 &= \theta_{k|k-1,t}^0 + T_k \{Z_k - H_k(\varphi_k(N)) \theta_{k|k-1,t}^0\} \\ P_{k|k,t}^0 &= P_{k|k-1,t}^0 - T_k \{Z_k - H_k(\varphi_k(N)) \theta_{k|k-1,t}^0 \\ &\quad + R_k(r_e(N))\} T_k^T \end{aligned} \quad (16)$$

The CV-IMM-EKF prediction steps are given by:

$$\begin{aligned} \theta_{k+1|k,t}^0 &= \phi \theta_{k|k,t}^0 + G \mu_{k,t} \\ P_{k+1|k,t}^0 &= \phi \theta_{k|k,t}^0 \phi^T + GCG^T \end{aligned} \quad (17)$$

4.2.3 Estimates combination:

$$\begin{aligned} \theta_{k|k} &= \sum_{t=1}^M \mu_{k,t} \theta_{k|k,t} \\ P_{k|k} &= \sum_{t=1}^M \mu_{k,t} \times \{P_{k|k,t} + [\theta_{k|k,t} - \theta_{k|k}] \\ &\quad \times [\theta_{k|k,t} - \theta_{k|k}]^T\} \end{aligned} \quad (18)$$

4.3 Numerical study

A basic set with N neighbors for the TV can be formed through Algorithm CV-IMM-EKF. Considering a section of urban roads, which is with a width of four lanes (each one is 3.5m wide) and a length of 1 kilometer. It is assumed that the traffic density of the road section is 20 vehicles/km and the average speed of traffic is generated stochastically in duration from 50km/h to 60km/h following a uniform distribution.

The noise vector $w_{k-1}^i = [\sigma_{ax,k-1}, \sigma_{ay,k-1}]^T \sim \mathcal{N}(0, C)$, with covariance matrix $C = \text{diag}[\sigma_{ax}^2, \sigma_{ay}^2]$, where the elements $\sigma_{ax,k-1} = \sqrt{0.99/2}$ and $\sigma_{ay,k-1} = \sqrt{0.01/2}$ are the acceleration noises along the x and y -axis, respectively, with STD in m/s^2 . The covariance matrix $R(r_e)$ of observation noise $Y_k(r_e) \sim \mathcal{N}(0, R(r_e))$ is described as a first-order Markov chain

switching between two models $R(r_1) = \text{diag}[\sigma_{px}^2, \sigma_{py}^2, \sigma_{vx}^2, \sigma_{vy}^2, \sigma_{\omega 1}^2(r_1), \dots, \sigma_{\omega N}^2(r_1)]$ and $R(r_2) = \text{diag}[\sigma_{px}^2, \sigma_{py}^2, \sigma_{vx}^2, \sigma_{vy}^2, \sigma_{\omega 1}^2(r_2), \dots, \sigma_{\omega N}^2(r_2)]$, of which the elements are with STDs in units of $m, m/s^2$ and Hz . The transition probability for this Markov chain is $\pi_{r_1 r_2} = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}$, and their initial probability is $\mu_0 = [0.5 \ 0.5]$. According to the achievable performance discussed in [Alam and Dempster \(2013\)](#) and [Tian et al. \(2015\)](#), as the number of the neighbors is increasing, the performance enhancement can be less obvious and lead to more additionally computational burden.

Therefore, we set $N = 4$, which is a relatively conservative number of the neighbors for the basic set, which is mentioned in algorithm CV-IMM-EKF.

In the simulations, the sampling period and length are taken to be 0.2 and 100, respectively, and the communication range of the DSRC is 300 m. As the DFS measurements presented in [Tian et al. \(2015\)](#) and [Schmidl and Cox \(1997\)](#), the probability density function (PDF) of the DFS is approximately zero-mean asymmetric Gaussian with the left and right STDs of 100 and 120 Hz, when the vehicles travel at the speed of 60 km/h, broadcasting the DSRC packets with a frequency of 5.89 GHz and a rate of 100 packets. It is worth noting that the PDF of the DFS remains a fairly consistent estimation from LOS to NLOS. Considering the noise of the DFS measurements as zero-mean Gaussian with two states of STDs: $\sigma_{wN}(r_1) = 100Hz$ and $\sigma_{wN}(r_2) = 120Hz$. Specifically, the state of the observation noise remains unchanged in r_1 between 0 and 6 s, and changes in the following 10 s to r_2 . Finally, the state changes back to r_1 for another 4 s. The position and velocity measured by GPS are assumed to be added noise with the variance $(\sigma_{px} = \sqrt{200/2m}, \sigma_{py} = \sqrt{200/2m})$ and $(\sigma_{vx} = \sqrt{15/2m}, \sigma_{vy} = \sqrt{15/2m})$, respectively.

To quantify the performance of the proposed approach, the root mean square error (RMSE) of vehicular positioning is calculated to assess the closeness of the estimated trajectory $(\hat{p}_{x,k}, \hat{p}_{y,k})$ to the true trajectory $(p_{x,k}, p_{y,k})$ at each time instant over $N_m = 500$ Monte Carlo simulations. In equation (28),

$(\hat{p}_{x,k}(m), \hat{p}_{y,k}(m))$ denotes the estimated position vector in the m th Monte Carlo run at the k th step:

$$RMSE = \sqrt{\frac{1}{N_m} \sum_{m=1}^{N_m} [(\hat{p}_{x,k}(m) - p_{x,k})^2 + (\hat{p}_{y,k}(m) - p_{y,k})^2]} \quad (19)$$

The performance comparison between the proposed CV-IMM-EKF and the GPS-only approach is shown in [Figure 10](#) with respect to the RMSE in distance. It is obvious that the proposed CV-IMM-EKF method outperforms the GPS alone localization. To indicate the enhancement of vehicular positioning of the proposed approach, the enhancement indicator μ is calculated as follows:

$$\mu = 1 - \frac{A_RMSE}{B_RMSE} \times 100\% \quad (20)$$

The enhancement of vehicular positioning is shown in [Table I](#). Compared to the GPS-based localization, the proposed CV-IMM-EKF approach achieves the enhancement of $\mu = 35.59$ per cent.

If A_RMSE is better than B_RMSE , μ will be greater than 0. The increase of μ is link to the good performance of A_RMSE . By describing the transition of the measurement noise as a first-order M -state jump Markov chain, the proposed CV-IMM-EKF approach has been proved to achieve better performance in a scenario that is similar to a practical one.

5. Conclusions

In this paper, we discuss three applications of intelligent computing in vehicular networks and their performance and effect is better. In vehicular networks, based on the intelligent computing, multidimensional analysis of vehicular network testbed data, the number of message copies increased dramatically, and the reachability and vehicular positioning enhancement have been improved. So the application of intelligent computing in the vehicle network system will be more extensive, and the computing intelligence will be an

Figure 10 CV-IMM-EKF and GPS performance in positioning error

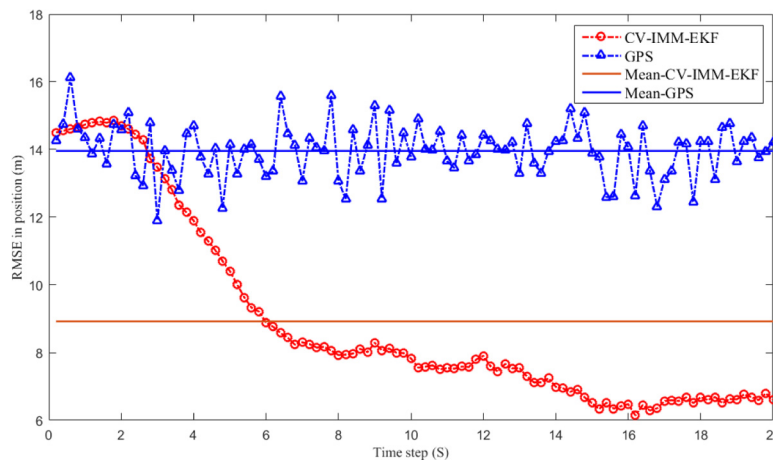


Table I CV-IMM-EKF and GPS error comparison

Method	RMSE	Enhancement
GPS	13.9593	N/A
CV-IMM-EKF	8.9920	35.59%

important technical means of car networking and improve the development of the car networking system.

References

- Alam, N. and Dempster, A.G. (2013), "Cooperative positioning for vehicular networks: facts and future [J]", *IEEE Transactions on Intelligent Transportation Systems*, Vol. 14 No. 4, pp. 1708-1717.
- Alam, N., Balaei, A.T. and Dempster, A.G. (2012), "An instantaneous Lane-Level positioning using DSRC carrier frequency offset [J]", *IEEE Transactions on Intelligent Transportation Systems*, Vol. 13 No. 4, pp. 1566-1575.
- Bishop, C.M. (2006), "Pattern Recognition and Machine Learning (Information Science and Statistics) [M]", Springer-Verlag, New York, NY.
- Biswas, A.R. and Giaffreda, R. (2014), "IoT and Cloud Convergence: Opportunities and Challenges[C]//Internet of Things", IEEE, 375-376.
- Çatay, B., Chiong, R. and Siarry, P. (2013), "Computational intelligence in production and logistics systems [J]", *International Journal of Production Economics*, Vol. 145 No. 1, pp. 1-3.
- Chen, M., Hao, Y., Hwang, K., Wang, L. and Wang, L. (2017), "Disease prediction by machine learning over big data from healthcare communities [J]", *IEEE Access*, Vol. 5 No. 99, pp. 8869-8879.
- Chen, S. and Zhao, J. (2014), "The requirements, challenges, and technologies for 5G of terrestrial mobile telecommunication [J]", *Communications Magazine IEEE*, Vol. 52 No. 5, pp. 36-43.
- Chetty, M., Ngom, A. and Marchiori, E. (2010), "Computational intelligence in bioinformatics [J]", *Journal of Proteomics & Bioinformatics*, Vol. 73 Nos 13/15, pp. 2291-2292.
- Chiong, R., Weise, T. and Michalewicz, Z. (2012), "Variants of evolutionary algorithms for Real-World applications [J]".
- Chiong, R. and Weise, T. (2011), "Special issue on modern search heuristics and applications [J]", *Evolutionary Intelligence*, Vol. 4 No. 1, pp. 1-2.
- Dao, T.S., Leung, K.Y.K., Clark, C.M. and Huissoon, J.P. (2007), "Markov-Based lane positioning using intervehicle communication [J]", *IEEE Transactions on Intelligent Transportation Systems*, Vol. 8 No. 4, pp. 641-650.
- Du, H., Li, W., Zhang, N., Zhao, K., Wang, J. and Shao, Z. (2015), "A vehicle cooperative positioning enhancement for GNSS-dark areas based on GRW model[C]//International Conference on Wireless Communications & Signal Processing", IEEE, pp. 1-6.
- Duan, X., Wang, Y., Tian, D., Sun, L., Michelson, D.G. and Leung, V.C. (2015), "A Vehicular Positioning Enhancement with Connected Vehicle Assistance[C]//IEEE, Vehicular Technology Conference". IEEE, pp. 1-5.
- Golden, B.L., Magnanti, T.L. and Nguyen, H.Q. (2010), "Implementing vehicle routing algorithms[J]", *Networks*, Vol. 7 No. 2, pp. 113-148.
- Golestan, K., Sattar, F., Karray, F., Kamel, M. and Seifzadeh, S. (2015), "Localization in vehicular ad hoc networks using data fusion and V2V communication[J]", *Computer Communications*, Vol. 71, pp. 61-72.
- Gu, Y., Lo, A. and Niemegeers, I. (2009), "A survey of indoor positioning systems for wireless personal networks [J]", *IEEE Communications Surveys & Tutorials*, Vol. 11 No. 1, pp. 13-32.
- Hou, X., Li, Y., Chen, M., Wu, D., Jin, D. and Chen, S. (2016), "Vehicular fog computing: a viewpoint of vehicles as the infrastructures [J]", *IEEE Transactions on Vehicular Technology*, Vol. 65 No. 6, pp. 3860-3873.
- Jin, Y. and Hammer, B. (2014), "Computational intelligence in big data [guest editorial][J]", *IEEE Computational Intelligence Magazine*, Vol. 9 No. 3, pp. 12-13.
- Jin, Y. and Meng, V. (2010), "Guest Editorial: special Issue on Evolutionary and Developmental Robotics [M]", IEEE Press.
- Kargupta, H. (2010), "Onboard driver, vehicle and fleet data mining: uS", US 7715961 B1 [P].
- Kargupta, H., Bhargava, R., Liu, K., Powers, M., Blair, P., Bushra, S. and Handy, D. (2004), VEDAS: A Mobile and Distributed Data Stream Mining System for Real-Time Vehicle Monitoring[C]//Siam International Conference on Data Mining, April. DBLP, Lake Buena Vista, FL.
- Liu, B., Khorashadi, B., Du, H., Ghosal, D., Chuah, C.N. and Zhang, M. (2009), "VGSim: an integrated networking and microscopic vehicular mobility simulation platform [J]", *Communications Magazine IEEE*, Vol. 47 No. 5, pp. 134-141.
- Liu, B., Khorashadi, B., Ghosal, D., Chuah, C.N. and Zhang, M.H. (2010), "Assessing the VANET's local information storage capability under different traffic mobility[C]//IEEE INFOCOM", IEEE, pp. 1-5.
- Mumtaz, S., Huq, K.M.S., Ashraf, M.I., Rodriguez, J., Monteiro, V. and Politis, C. (2015), "Cognitive vehicular communication for 5G[J]", *IEEE Communications Magazine*, Vol. 53 No. 7, pp. 109-117.
- Panichpapiboon, S. and Pattara-Atikom, W. (2012), "A review of information dissemination protocols for vehicular ad hoc networks [J]", *IEEE Communications Surveys & Tutorials*, Vol. 14 No. 3, pp. 784-798.
- Ratnasamy, S., Stoica, I. and Shenker, S. (2002), *Routing Algorithms for DHTs: Some Open Questions [J]*, Lecture Notes in Computer Science, 2429:45-52.
- Rodgers, J.L. and Nicewander, W.A. (1988), "Thirteen ways to look at the correlation coefficient [J]", *American Statistician*, Vol. 42 No. 1, pp. 59-66.
- Saravanan, P. and Arunkumar, T. (2014), "Fuzzy enabled geographic routing protocol for vehicular ad hoc networks [J]", *International Review on Computers & Software*, Vol. 9 No. 6, pp. 1101-1107.
- Saravanan, P. and Arunkumar, T. (2016), "Identification of malicious vehicles in vehicular adhoc networks using prime product calculation [J]", *International Journal of Pharmacy & Technology*, Vol. 8 No. 4, pp. 21896-21904.

- Sayed, A.H., Tarighat, A. and Khajehnouri, N. (2005), "Network-based wireless location: challenges faced in developing techniques for accurate wireless location information [J]", *IEEE Signal Processing Magazine*, Vol. 22 No. 4, pp. 24-40.
- Schmidl, T.M. and Cox, D.C. (1997), "Robust frequency and timing synchronization for OFDM [J]", *IEEE Transactions on Communications*, Vol. 45 No. 12, pp. 1613-1621.
- Smith, D., Timms, G. and De Souza, P. (2011), "A Quality Control Framework for Marine Sensing Using Statistical, Causal Inference[C]//Oceans", IEEE, pp. 1-7.
- Tian, D., Zhou, J., Wang, Y., Xia, H., Yi, Z. and Liu, H. (2015), "Optimal epidemic broadcasting for vehicular ad

- hoc networks [J]", *International Journal of Communication Systems*, Vol. 27 No. 9, pp. 1220-1242.
- Tian, D., Zhu, K., Zhou, J., Wang, Y. and Liu, H. (2015), "Swarm model for cooperative multi-vehicle mobility with inter-vehicle communications [J]", *Iet Intelligent Transport Systems*, Vol. 9 No. 10, pp. 887-896.
- Win, M.Z., Conti, A., Mazuelas, S., Shen, Y., Gifford, W.M., Dardari, D. and Chiani, M. (2011), "Network localization and navigation via cooperation [J]", *IEEE Communications Magazine*, Vol. 49 No. 5, pp. 56-62.

Corresponding author

Xuting Duan can be contacted at: duanxuting@buaa.edu.cn