MOBILE-EDGE COMPUTING FOR VEHICULAR NETWORKS

A Promising Network Paradigm with Predictive Off-Loading

Ke Zhang, Yuming Mao, Supeng Leng,
Yejun He, and Yan Zhang

Cloud-based vehicular networks are a promising paradigm to improve vehicular services through distributing computation tasks between remote clouds and local vehicular terminals. To further reduce the latency and the transmission cost of the computation off-loading, we propose a cloud-based mobile-edge computing (MEC) off-loading framework in vehicular networks. In this framework, we study the effectiveness of the computation transfer strategies with vehicle-to-infrastructure (V2I) and vehicle-to-vehicle (V2V) communication modes. Considering the time consumption of the computation task execution and the mobility of the vehicles, we present an efficient predictive combination-mode relegation scheme, where the tasks are adaptively off-loaded to the MEC servers through direct uploading or predictive relay transmissions. Illustrative results indicate that our proposed scheme greatly reduces the cost of computation and improves task transmission efficiency.

Motivations for Predictive Offloading

With the ever-increasing number of vehicles on the roads and the development of the Internet of Vehicles (IoV), vehicles constitute a considerable portion of the things connecting to the Internet. In the IoV paradigm, smart automobiles can provide intelligent vehicle control, traffic management, and interactive applications with their equipped computation units and communication technologies. These services and applications may require significant computation resources and constrained time delays [1]. However, in general, the computational capabilities of the automobile terminals are

Digital Object Identifier 10.1109/MVT.2017.2668838
Date of publication: 21 April 2017

limited. The computation resource-hungry applications pose a significant challenge to the resource-limited vehicular terminals.

To cope with the explosive computation demands of automobile terminals, cloud-based vehicular networking is widely considered as a new paradigm to improve service performance. By integrating communication and computing technologies, cloud-enabled networks allow applications to either run locally on the automobile terminals or be off-loaded to the remote computation cloud.

Mobile cloud computing greatly improves resource utilization and computation performance. However, considering the capacity limitation and delay fluctuation of the transmission on the backhaul and backbone networks, placement of the cloud servers far away from the mobile vehicles may cause serious degradation of the off-loading efficiency. MEC is proposed as a promising solution that pushes the cloud service to the edge of the radio access network and provides cloud-based computation off-loading in close proximity to the mobile vehicular terminals.


©ISTOCKPHOTO.COM/MIN6939

In MEC networks, various computation tasks have different resource requirements, including the computation resources for task execution and the communication resources for task transmission. Because MEC servers operate at the edge of radio access networks and transmit tasks with the aid of connected roadside units (RSUs), their service areas may be limited by the radio coverage of the RSUs. Due to their high mobility, vehicles in transit may pass through several RSUs and MEC servers during the task-off-loading process, and they can off load their computation task to any MEC servers that they can access. The selection of the target MEC servers affects the off-loading efficiency.

Furthermore, in vehicular networks, there are several ways for vehicles to access RSUs that are connected with MEC servers, such as V2I mode and V2V mode. The dynamic topology changes caused by the mobility of vehicles make off-loading transmission complex. To improve task accomplishment efficiency, it is imperative that an optimal task-off-loading scheme with MEC server selection and communication management be available in the MEC cloud-based vehicular networks.

In this article, we propose a cloud-based MEC off-loading framework in vehicular networks, where both the heterogeneous requirements of the computation tasks and the mobility of the vehicles are considered. We focus on MEC server selection and task transmission management. Based on the analysis of the characteristics of various off-loading strategies, we propose a predictive-mode transmission scheme for task-file uploading. This scheme greatly improves the transmission efficiency while meeting the delay constraint of the computation tasks. Moreover, we design an optimal predictive combination-mode MEC off-loading mechanism for various types of computation tasks in accordance with their different requirements. The mechanism minimizes the total cost of the off-loading process under the delay tolerance of each computation task. Illustrative results demonstrate that our proposed mechanism significantly reduces the off-loading cost by effectively utilizing multi-off-loading modes.

Related Work

Recently, there has been a large amount of work on cloud-enabled vehicular networks. A coalition game model to manage and share the resources among different cloud service providers is proposed in [2]. In [3], the authors

THE COMPUTATION RESOURCE-HUNGRY APPLICATIONS POSE A SIGNIFICANT CHALLENGE TO THE RESOURCE-LIMITED VEHICULAR TERMINALS.

exploited the cognitive radio and soft data fusion in vehicular access networks and designed a distributed traffic off-loading scheme for cognitive cloud vehicular networks. By combining the vehicular cloud with the fixed central cloud, [4] proposed a flexible off-loading strategy to discover unutilized resources and carry out task mitigation. In [5], the authors designed a vehicular fog computing architecture that utilizes a collaboration of vehicles and near-user edge devices to carry out communication and computation.

As one of the most promising technical approaches to improve the effectiveness of cloud computing, MEC has attracted considerable attention recently. To investigate the MEC mechanisms, [6] studied a multiuser MEC computation off-loading problem in a multichannel wireless environment and designed a distributed game theoretic off-loading scheme. The authors in [7] made a case study where fifth-generation and MEC collaborate to provide real-time context-aware communication. The performance of an infrastructure cloud is compared with that of a mobile-edge cloud in terms of the application running time in [8]. In [9], the authors designed a mechanism for MEC networks to exploit cloud-based packet core infrastructures with virtualization.

Although many studies focus on cloud-based vehicular networking and MEC technologies, in the recent literature, very few articles feature studies on MEC cloud-enabled vehicular networks. In addition, the impact of both vehicle mobility and the various vehicular network communication schemes on computation off-loading performance has been ignored. Unlike these works, in this article, we study the computation off-loading mechanism in mobile-edge cloud-based vehicular networks and propose optimal strategies for various types of computation tasks for MEC server selection and task transmission management.

Computation and Communication in Vehicular Networks

Driven by the development of the IoV, more smart vehicles are on the road. These vehicles are always equipped with computation units, multicommunication technologies, a comprehensive sensor platform, and human–computer interaction devices. With the aid of this equipment and technology, smart vehicles can provide many intelligent traffic applications, such as active driving safety assistance, smart parking, road traffic monitoring, and automatic management. In addition, there are some on-board multimedia and convenience applications for drivers and passengers. Some of these applications require intensive computation and have tight delay constraints, especially the applications with dynamic video processing and real-time interaction, for example, the application of image-aided navigation, natural language processing, and interactive gaming.

The applications with high computational requirements pose great challenges to the existing vehicular terminals, especially in terms of their computational resources. However, in general, vehicles have weak computation capability and limited storage space. To meet the ever-increasing computation demand of these applications, off-loading the computation tasks to MEC servers through cloud-based vehicular networks is an appealing idea. The operation of cloud-based vehicular networks calls for effective communication between fast-moving vehicles and MEC servers, which transmit task-input files and computation output between them.

In vehicular communication networks, dedicated short-range communication (DSRC) is one of the most reliable wireless technologies, providing two-way, short-range communication between vehicles and infrastructures. In general, DSRC refers to a suite of standards that includes IEEE 802.11p [14], the IEEE 1609 protocol family, and a Society of Automotive Engineers message set [10]. In the architecture of DSRC, there are two types of communication entities. One is the RSU, which is located along the roadside, providing data access service to vehicles. The other is the on-board unit on vehicles, which enables them to communicate with adjacent RSUs and neighboring vehicles. DSRC provides two modes of wireless communications, i.e., V2I and V2V. When a vehicle is out of the radio coverage area of an RSU, it may access the RSU through multihop V2V relay transmissions [11].

By utilizing vehicular communication networks, the computation tasks of vehicular applications can be off-loaded to MEC servers. However, due to time-variant vehicular wireless communication channels and contention-based media access schemes as well as the limited transmission range of DSRC technologies, obtaining timely and reliable data transmission between vehicles and MEC servers is still a challenge.

Mobile-Edge Computing in Cloud-Enabled Vehicular Networks

Figure 1 shows the architecture of the mobile-edge cloud-enabled vehicular networks. We consider a unidirectional road, which has uninterrupted traffic in a free-flow state. Along the road, there are RSUs. The distance between every two adjacent RSUs is L. Each RSU provides wireless access service for the vehicles within its transmission range. We consider the transmission range to be $L/2$. The road can be divided into several segments with length L. Through V2I communication mode, the vehicles running within a given segment can only access the RSU located in the corresponding segment.

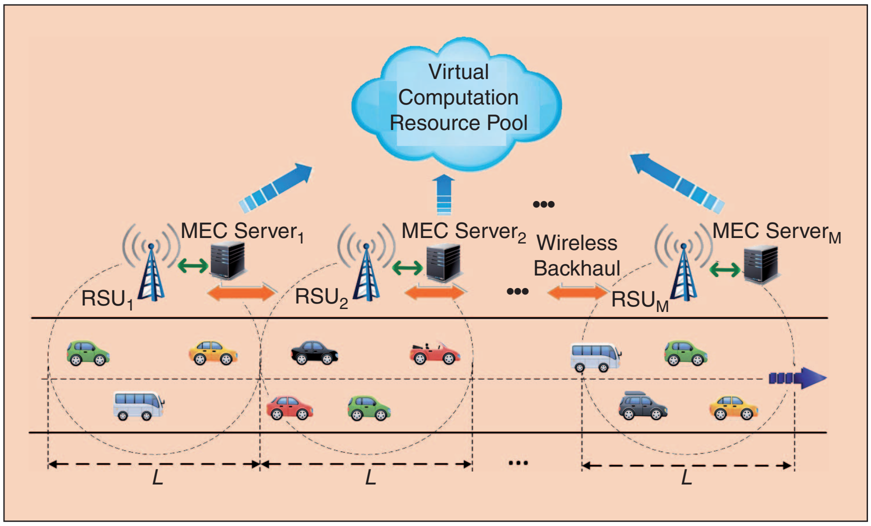

FIGURE 1 The architecture of the mobile-edge cloud-enabled vehicular networks.

The RSUs communicate with each other through wireless backhauls. Each RSU is equipped with an MEC server with limited computation resource. For many applications, such as speech recognition, the size of the computation task-input data is much larger than that of the outcome [6]. To improve the transmission efficiency of the wireless backhauls, the task-input file cannot be transmitted between the RSUs. In other words, each MEC server only executes the computation task received from the RSU with which it connects. However, because the output data is small, the computation output can be transmitted between RSUs through wireless backhauls.

All the vehicles move at a constant speed along the road. The distribution of the vehicles on the road follows a Poisson distribution [11]. The traffic density in terms of vehicles per unit distance is λ. Each vehicle has a computation task that can be either accomplished locally by the vehicular terminal or be computed remotely on the MEC servers. The computation task is denoted as $T = \{c, d, t_{\max}\}$. Here, c is the amount of the computation resource required to accomplish task T. For example, c can be quantified as the required number of central processing unit cycles. The variable d denotes the size of the computation input file describing some information of the computation task, such as the program codes or the recorded video, and $t_{\max}$ is the delay tolerance of the task.

To study the effects of the computation task characteristics on the design of off-loading schemes, we classify the tasks into S types and present the tasks as $T_i = \{c_i, d_i, t_{i,\max}\}, i \in S$. According to their computation task types, the vehicles can be correspondingly classified into S types. The proportion of the vehicles with type-i tasks in all the vehicles on the road is given as ρ_i, where $i \in S$ and $\sum_{i=1}^{S} \rho_i = 1$ (see Table 1).

Off-Loading with Predictive Transmission

Due to proximity, MEC servers are able to reduce transmission cost and provide fast interactive response in the computation off-loading service. However, compared with the traditional cloud server location at the backbone network, which always has an enormous computation resource, the MEC server still suffers from the resource limitation. There is a certain time overhead for the MEC servers to accomplish each computation task, especially for the MEC servers located in the road segments with high density of vehicles and a large number of computation tasks.

In vehicular communication networks, both the vehicles and the RSUs are DSRC devices, which have short communication range. Under this range limitation, if a vehicle chooses to off-load its computation task to the MEC servers through the V2I transmission mode, it can only access the RSU whose communication area covers it. As a consequence of this data access, the task of this vehicle can only be executed on the MEC server connected to the RSU.

Figure 2 shows the scenario where the vehicles off-load their tasks to the MEC servers through direct V2I mode. Considering that a vehicle runs on an expressway with high speed, if the accomplishment of its computation task takes a relatively long time, the vehicle may pass through several RSUs during the task execution period. In this case, the sending back of the computation output to the vehicle needs to be transmitted

TABLE 1 *The main notations.*

Symbol	*Description*
S	Total number of vehicle types
ρ_i	Proportion of type-i vehicles
$J_{i,\max}$	Maximum hop of available MEC servers for type-i vehicles
$P_{i,j}$	Probability of type-i vehicles choosing MEC server j
$f_{i,j}$	Cost of type-i vehicles choosing MEC server j

> **OBTAINING TIMELY AND RELIABLE DATA TRANSMISSION BETWEEN VEHICLES AND MOBILE-EDGE COMPUTING SERVERS IS STILL A CHALLENGE.**

from the MEC server$_1$ that has accomplished the task to the RSU$_n$ that the vehicle newly accesses. Due to its easy deployment and low cost, wireless backhaul is widely used in the communication between adjacent RSUs. The computation output feedback takes a multihop wireless relay between several RSUs. The interference between the wireless links makes the wireless backhaul transmit at a low rate and with unpredictable delay [12]. The time overhead and transmission cost of the multihop relay seriously degrade the task transmission effectiveness improvement obtained through MEC technologies.

Given the large number of vehicles on the road, vehicles can communicate with each other through multihop V2V connections. Putting these underutilized V2V communication resources into use offers great help for load balancing and delay reduction in vehicular communication networks. Furthermore, unlike the RSU access service that is provided by some operators, the V2V communication is always self-organized by running vehicles and costs much less than V2I communication.

Motivated by the reasons just mentioned, we propose a computation off-loading scheme with predictive-mode transmission. Figure 3 illustrates the computation off-loading through the predictive-mode scheme. In this scheme, the vehicles send their task-input files to the MEC servers ahead of their running direction. Based on the accurate prediction of the file transmission time and the task execution time as well as the time spent for the vehicle on the road, vehicle k can arrive at the communication area of RSU$_n$ at the exact time when its task has just been accomplished. The computation output can be transmitted directly from RSU$_n$ to the vehicle through V2I transmission without multihop backhaul relay. Therefore, transmission cost for task off-loading can be saved.

We note that different types of computation tasks take various execution times and have different delay tolerances. In addition, the transmission delay of both V2I and V2V modes is affected by the wireless channel states, the vehicle density on the road, and the size of the task-input files. To make better utilization of the communication and computational resources of cloud-based vehicular networks, an optimal computation off-loading scheme is imperative.

Optimal Combination-Mode Off-Loading

With the aid of the MEC cloud servers, each vehicle can choose to accomplish its computation task either locally on its own vehicular terminal or remotely on an MEC server. The selection of the task-completion mode not only depends on the processing time requirement but also on the corresponding cost. This cost may include energy consumption, the fee charged by computing and communication service providers, etc.

Local Computing or Cloud Computing

In vehicular networks, we consider that each vehicle has a homogeneous computation resource, which is denoted as c_0. For easy presentation, the vehicles are named with type-i tasks as *type-i* vehicles. If a type-i vehicle chooses to accomplish its task T_i through local computing, the task execution time can be given as $t_{i,\text{local}} = c_i/c_0$, and the cost of the local computing is $f_{i,\text{local}}$. If $t_{i,\text{local}} > t_{i,\max}$, type-$i$ vehicles should off-load their tasks to the MEC servers to accomplish the tasks under the delay constraints.

When portions of the vehicles choose cloud computing mode, their tasks are off-loaded to the MEC servers. Let $P_{i,j}$ denote the probability of type-i vehicles that choose to off-load their tasks to the MEC servers connected to the RSUs j road segments away from their current position. Here, we define that $j = 0$ means the vehicles choose to execute the computation tasks locally on their own vehicular terminals. In addition, $j = 1$

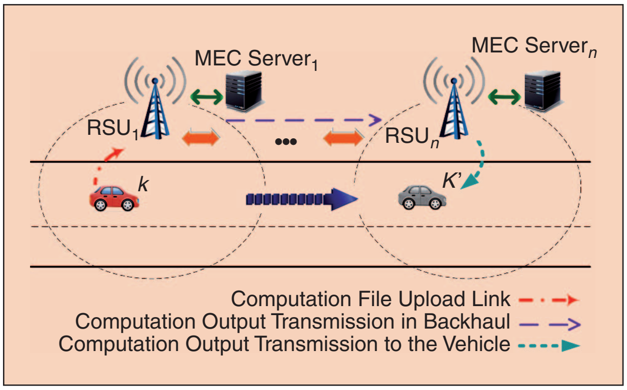

FIGURE 2 The off-loading through direct V2I mode transmission.

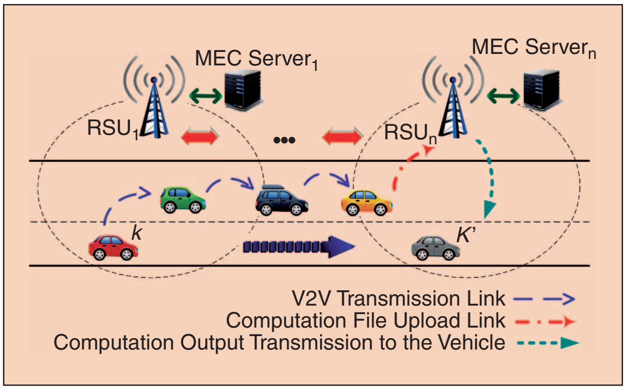

FIGURE 3 The off-loading through the predictive-mode transmission.

indicates that the vehicles off-load the tasks to the MEC servers through direct V2I communication, i.e., each destination server and the task-off-loading vehicles are located in the same road segment.

Task-file transmission from the vehicles to the MEC servers takes some time consumption, which acts as a main part in the delay of the off-loading process. Furthermore, data transmission to the farther MEC server means more wireless transmission hops and longer time delay. As there is a latency tolerance for each type of task, we define $J_{i,\max}$ as the maximum hop of the MEC servers that type-i vehicles can offload their tasks to under the tasks' delay constraints.

Considering that the distribution of the vehicles on the road follows Poisson distribution, the arrival of the off-loading tasks to an MEC server is a Poisson process. We model the arrivals and the fulfilments of the computing tasks on an MEC server as an $M/M/1$ queue. In the queuing model of each MEC server, the computation service time has an exponential distribution with mean service time $1/\mu$. The average time spent on accomplishing a task is $t_{\text{remote}} = 1/(\mu - \bar{\lambda})$, which includes the waiting time and the computing execution time. Here, $\bar{\lambda}$ is the average rate of the tasks arriving at an MEC server and can be given as

$$\bar{\lambda} = \sum\nolimits_{i=1}^{S} \sum\nolimits_{j=1}^{J_{i,\max}} P_{i,j}\rho_i\lambda. \qquad (1)$$

Because the establishment and the maintenance of the MEC servers are always provided by operators, there are some fees charged for providing the off-loading computing service. We denote the cost for computing a type-i task on the MEC servers as $f_{i,\text{remote}}$.

Off-Loading Through Direct Transmission or Through Predictive Transmission

When a vehicle chooses to accomplish its computation task remotely on an MEC server, it needs to deliver its input file to the corresponding RSU connected with the MEC server. There are two methods of file transmission from the vehicle to the RSU. The first is direct V2I mode and the other is predictive-mode transmission. In the following text, we discuss the details of these two methods and compare them in terms of time consumption and cost. To facilitate modeling and analysis of the off-loading process, the wireless interference and the wireless transmission capacity limitation have not been considered.

We consider a scenario where a type-i vehicle runs on road segment l. In this scenario, this vehicle can directly access the RSU located in this road segment. We denote this RSU as RSU l. If the vehicle adopts the direct V2I mode to off-load its task to an MEC server, it transmits the file to RSU l directly. In this mode, the process of the file upload is simple and efficient. Both the file upload time and cost are saved. However, the execution of the computation task on the MEC server takes time $t_{i,\text{remote}}$. As the vehicle runs at a high speed, it may be outside the transmission range of RSU l at the end of time $t_{i,\text{remote}}$. The output data of the task needs to be transmitted from RSU l to the RSU located on the road segment where the vehicle arrives. The transmission between these RSUs goes through the wireless backhauls.

MOBILE-EDGE COMPUTING SERVERS ARE ABLE TO REDUCE TRANSMISSION COST AND PROVIDE FAST INTERACTIVE RESPONSE IN THE COMPUTATION OFF-LOADING SERVICE.

Let $t_{i,\text{backhaul}}$ and $f_{i,\text{backhaul}}$ be the time delay and the cost for transmitting the output data of a type-i task across one road segment, respectively. We can get the total time consumption of the task accomplishment through V2I mode as $t_{i,1} = t_{i,\text{upload}} + t_{i,\text{remote}} + x_i \cdot t_{i,\text{backhaul}} + t_{i,\text{download}}$, where x_i is the number of the road segments that the vehicle has passed through during $t_{i,\text{remote}}$. $t_{i,\text{upload}}$ and $t_{i,\text{download}}$ are the time spent on uploading the task file and downloading the computation output, respectively. Similarly, we can compute the total cost of the task accomplishment in V2I mode as $f_{i,1} = f_{i,\text{upload}} + f_{i,\text{remote}} + x_i \cdot f_{i,\text{backhaul}} + f_{i,\text{download}}$. It is easy to see that, with the higher vehicle speed and the longer task execution time, the vehicle runs farther from the MEC server that computes the task. Longer delay and more cost are induced in the output transmission process.

In the case that a vehicle chooses to off-load its task in the predictive-mode transmission, the computation file is first delivered through the multihop V2V relay transmission. Then, the file is transmitted to an RSU through V2I by the vehicle at the final hop of the transmission relay. In this way, the computation task is off-loaded to the MEC server located at the road segment ahead of the vehicle's current position in advance. After getting the computation output, the MEC server stores the results on the RSU connected to it. When the task-off-loading vehicle arrives in the transmission range of the RSU, it obtains the results directly from the RSU. By adopting this approach, the transmission capability of the vehicular network is fully utilized, while the resource of the wireless backhauls is greatly saved.

Let $t_{i,v2v}$ denote the average time delay for the transmission of type-i task's input file through a one-hop V2V relay. The total time consumption of the task accomplishment in this predictive mode is $t_{i,j} = y_j \cdot t_{i,v2v} + t_{i,\text{upload}} + t_{i,\text{remote}} + t_{i,\text{download}}$. Recall that j is the hops of the upload destination RSU away from the vehicle's current position, so $j > 1$ means the vehicles adopt the predictive-mode transmission. Here, we define y_j as the V2V relay hops that are required in transmitting the input file to the j-hop-away RSU. Furthermore, the total cost of this type of task off-loading can be similarly

GIVEN THE LARGE NUMBER OF VEHICLES ON THE ROAD, VEHICLES CAN COMMUNICATE WITH EACH OTHER THROUGH MULTIHOP VEHICLE-TO-VEHICLE CONNECTIONS.

given as $f_{i,j} = y_j \cdot f_{i,v2v} + f_{i,\text{upload}} + f_{i,\text{remote}} + f_{i,\text{download}}$, where $1 < j \leq J_{i,\max}$.

Optimal Off-Loading Schemes for Computation Tasks

To minimize the off-loading cost of both data transmission and task execution while satisfying the latency constraints, the objective function of the optimal off-loading schemes is shown as

$$\begin{aligned} &\min_{\{P_{i,j}\}} \sum_{i=1}^{S} \sum_{j=0}^{J_{i,\max}} \rho_i P_{i,j} f_{i,j} \\ &\text{s.t.} \quad t_{i,j} \leq t_{i,\max}, \quad i \in \{1, S\}, j \in \{0, J_{i,\max}\}. \end{aligned} \tag{2}$$

The objective function in (2) gives the average off-loading costs of all types of vehicles when they choose off-loading strategies $\{P_{i,j}\}$. To obtain the solution of (2), we take a game approach to find the optimal off-loading strategies of each type of vehicle. In this game, there are S players. Each player is a set of vehicles with the same type tasks. We denote the vehicle set with type-i tasks as set i. The strategies of vehicle set i $(i = \{1, 2, \ldots, S\})$ are $\{P_{i,j}\}$. In other words, the strategies are the probabilities of vehicle set i's off-loading choices. Vehicle set i can choose either to execute tasks locally or to offload them to j-hop-away MEC servers. The payoff for set i is the sum of their off-loading costs.

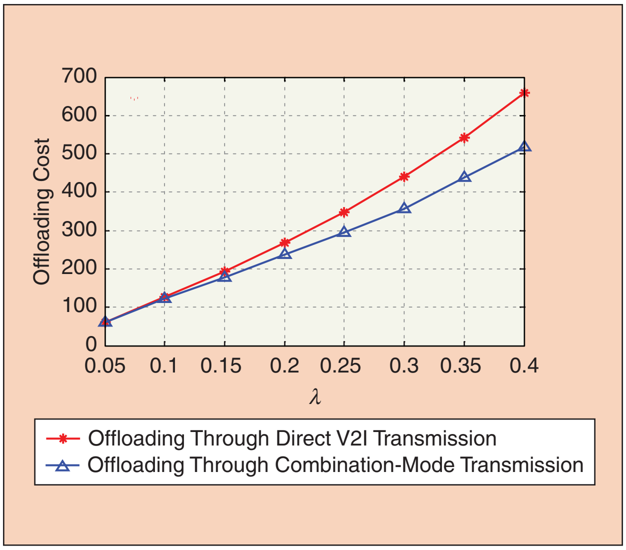

FIGURE 4 The computation off-loading costs in terms of the density of vehicles.

In a game with finite players, a Nash equilibrium (NE) is a solution in which no player can improve its pay off by changing the strategy unilaterally, given the strategies of the other players. In our proposed game, there are S players whose total number is finite. Furthermore, the off-loading strategies of each vehicle set i can be regarded as a finite set $\mathcal{J}_i = \{0, 1, 2, \ldots, J_{i,\max}\}$ with probability set $\mathcal{P}_i = \{P_{i,0}, P_{i,1}, P_{i,2}, \ldots, P_{i,\max}\}$, which satisfies the definition of mixed strategies. According to the Nash existence theorem, a finite-player game, in which each player can choose a pure strategy from a finite strategy set, has at least one mixed strategies NE [13]. By a heuristic way where each vehicle set takes its best response action given the strategies of other vehicle sets, we can get an NE, which is the solution of (2).

As MEC servers play a key role in the task-off-loading process, it is worth discussing the point of failure at an MEC server. In this case, although the available computing resources and the distance between two adjacent available MEC servers have been changed, the vehicles can still off-load tasks to other available MEC servers using similar off-loading mechanisms as discussed previously.

Performance Evaluation

In this section, we show illustrative results to demonstrate the performance of our proposed optimal off-loading schemes. We consider ten RSUs located along a four-lane one-way road and use Simulation of Urban Mobility to simulate the road traffic. The density of the vehicles on the road is set as $\lambda = 0.3$, and the vehicles are running at speed 120 km/h. The computation tasks of these vehicles are classified into five types with the probabilities $\{0.05, 0.15, 0.3, 0.4, 0.1\}$. As resource requirement is the most important factor affecting the off-loading performance, we set the computation resource requirement of each type of the tasks as $\{7, 13, 27, 33, 48\}$ units, respectively.

Figure 4 evaluates the total computation off-loading costs in terms of the density of the vehicles on the road. We compare the performance of our proposed predictive combination-mode scheme with the V2I direct transmission scheme. It can be seen that the combination-mode scheme greatly reduces the cost when the road has high vehicle density. However, the cost-saving effectiveness is weak when the density is low, because with low traffic density, the total number of vehicles on the road is small. The difference between the costs of the two schemes is small. Furthermore, when λ is small, the computation load of each MEC server is light. A large portion of the off-loaded tasks on the MEC servers can be accomplished within the period when the vehicles accessing RSUs have not been changed. As a result, there is no need to adopt predictive-mode transmission.

On the contrary, in the case of high traffic density, long task execution time on the MEC servers leads to more RSUs being passed through by the running vehicles. Due to the transmission cost of the wireless backhaul between RSUs, in Figure 4, the total costs of the direct V2I scheme rise rapidly with the increase of λ. However, in the predictive combination-mode scheme, part of the transmission is off-loaded to the V2V relay transmission, which has less cost compared with the wireless backhaul transmission. Therefore, the computation off-loading cost has been saved.

Figure 5 shows the percentage of various types of the tasks that are off-loaded through the predictive-mode transmission. We found that the percentage of the type-1 and the type-5 tasks stay as 0% and 100% in terms of all the λ values, respectively. With each λ value, there is a critical task type. The percentage of the tasks whose type index is less than the critical type index is always 0%, and the percentage of the tasks whose type index is larger than the critical index is 100%. This can be explained as follows. Given the MEC servers working at a certain computation load level, the execution of the tasks with a higher type index takes a longer time. As the vehicles are running at an identical speed, longer execution time means more RSUs may be passed through. More backhaul transmission cost may be induced if the vehicles adopt V2I direct transmission. To save the off-loading cost, computation tasks with higher type index are more likely to prefer predictive-mode transmission.

Figure 6 shows the total computation off-loading costs in terms of the vehicle speed. We can see that both the two transmission schemes have higher off-loading costs with faster vehicle speed. For the direct V2I transmission scheme, there is a sudden increase between the costs at vehicle speed 110 km/h and 120 km/h. In the case where the speed is 110 km/h, at the time when type-3 and type-4 vehicles' off-loaded tasks are completed, these vehicles are running near the coverage border of the currently accessing RSU. Increasing speed to 120 km/h gives a large portion of these vehicles access to the next RSU. As a result, more wireless backhaul transmission cost has been added to the total costs. However, in our proposed predictive combination-mode scheme, the transmission mode can be dynamically changed to cope with different vehicle speeds. Compared with direct V2I transmission scheme, increase of the total costs of the combination-mode scheme is smaller.

In our proposed predictive combination-mode scheme, the transmission mode can be dynamically changed to cope with different vehicle speeds.

Conclusions

In this article, we proposed a new framework of mobile-edge cloud-based vehicular networks. Based on the framework, we discussed the time consumption and the off-loading cost of various transmission modes. We designed a task-file transmission strategy with predictive V2V relay and proposed an optimal predictive combination-mode off-loading scheme. The results demonstrated that our scheme greatly reduces the off-loading cost.

Acknowledgments

This research is partially supported by the National Natural Science Foundation of China under grants 61374189 and 61372077, the projects 240079/F20 funded by the Research Council of Norway, and the project Security in IoT for Smart Grids, with 248113/O70 part of the program funded by the Norwegian Research Council.

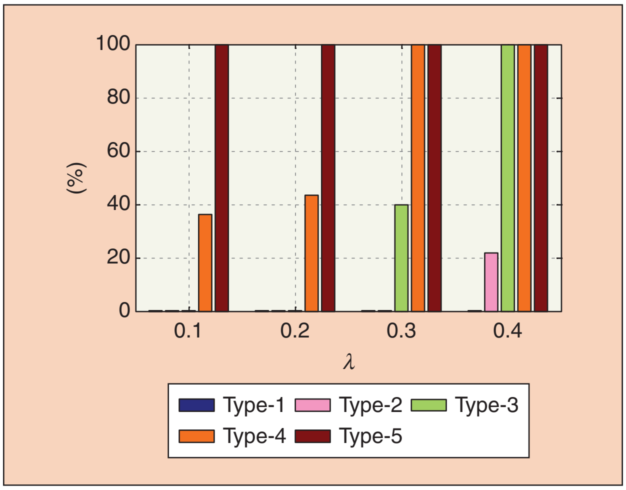

FIGURE 5 The percentage of various types of tasks off-loaded through predictive-mode transmission.

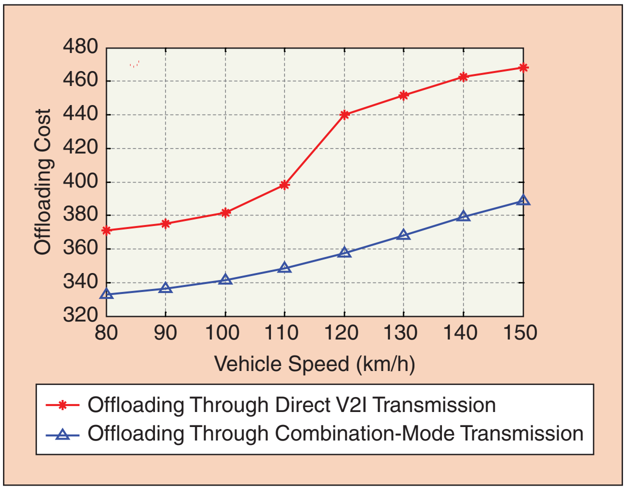

FIGURE 6 The computation off loading costs in terms of vehicle speed.

Author Information

Ke Zhang (zhangke@uestc.edu.cn) is a lecturer and Ph.D. degree candidate in the School of Communication and Information Engineering, University of Electronic Science and Technology of China, Chengdu. His research interests include energy management and communications in the smart grid, design and optimization of next-generation wireless networks, and the Internet of Things.

Yuming Mao (ymmao@uestc.edu.cn) is a professor with the School of Communication and Information Engineering, University of Electronic Science and Technology of China, Chengdu, where he is chair of the Department of Network Engineering. His main research interests include broadband communication networks, network organization and protocol analysis, transmission control protocol/Internet Protocol technology, network management and protocol, routing protocol, and network engineering. He received several awards, including the first-grade, second-grade, and third-grade awards of the Ministry of Electronic Industry for science and technology progress and the second-grade national award for science and technology progress.

Supeng Leng (spleng@uestc.edu.cn) is a professor with the School of Communication and Information Engineering, University of Electronic Science and Technology of China, Chengdu. He has been a research fellow with the Network Technology Research Center. He has published more than 100 research papers. His research interests include resources, spectrum, energy, routing, and networking in broadband wireless access networks, vehicular networks, the Internet of Things, next-generation mobile networks, and smart grids. He serves as an organizing committee chair and a technical program committee member for many international conferences as well as a reviewer for more than ten international research journals. He is a Member of the IEEE.

Yejun He (heyejun@126.com) is a full professor with the College of Information Engineering, Shenzhen University, China, where he is the drector of Guangdong Engineering Research Center of Base Station Antennas and Propagation and the director of Shenzhen Key Laboratory of Antennas and Propagation. He has authored or coauthored more than 100 research papers and books or book chapters and holds 13 patents. His research interests include wireless mobile communication, antennas, and radio frequency. He currently serves as associate editor of the *IEEE Access* journal and *Security and Communication Networks* journal. He is a fellow of IET and a Senior Member of the IEEE.

Yan Zhang (yanzhang@ieee.org) is a full professor in the Department of Informatics, University of Oslo, Norway. He is an associate technical editor of *IEEE Communications Magazine*, an editor of *IEEE Transactions on Green Communications and Networking*, an editor of *IEEE Communications Surveys & Tutorials*, and an associate editor of *IEEE Access*. He served as chair for a number of conferences, including the 2017 IEEE Global Telecommunications Conference, Wireless Communications and Signal Processing 2016, IEEE SmartGridComm 2015, and IEEE CloudCom 2015. He also served as a technical program committee member for numerous international conferences. His current research interests include next-generation wireless networks leading to fifth generation and green and secure cyberphysical systems (e.g., smart grid, health care, and transport). He is an IEEE Vehicular Technology Society Distinguished Lecturer. He is a fellow of the Institution of Engineering and Technology, and he is a Senior Member of the IEEE.

References

[1] Y. He, M. Chen, B. Ge, and M. Guizani, “On WiFi offloading in heterogeneous networks: Various incentives and trade-off strategies,” *Commun. Surveys Tuts.*, vol. 18, no. 4, pp. 2345–2385, Apr. 2016.

[2] R. Yu, X. Huang, J. Kang, J. Ding, S. Maharjan, S. Gjessing, and Y. Zhang, “Cooperative resource management in cloud-enabled vehicular networks,” *IEEE Trans. Ind. Electron.*, vol. 62, no. 12, pp. 7938–7951, Dec. 2015.

[3] N. Cordeschi, D. Amendola, and E. Baccarelli, “Reliable adaptive resource management for cognitive cloud vehicular networks,” *IEEE Trans. Veh. Technol.*, vol. 64, no. 6, pp. 2528–2537, June 2015.

[4] H. Zhang, Q. Zhang, and X. Du, “Toward vehicle-assisted cloud computing for smartphones,” *IEEE Trans. Veh. Technol.*, vol. 64, no. 12, pp. 5610–5618, Sept. 2015.

[5] X. Hou, Y. Li, M. Chen, D. Wu, D. Jin, and S. Chen, “Vehicular fog computing: A viewpoint of vehicles as the infrastructures,”” *IEEE Trans. Veh. Technol.*, vol. 65, no. 6, pp. 3860–3873, June 2016.

[6] X. Chen, L. Jiao, W. Li, and X. Fu, “Efficient multi-user computation offloading for mobile-edge cloud computing,” *IEEE/ACM Trans. Networking*, vol. 24, no. 5, pp. 2795–2808, Oct. 2016.

[7] S. Nunna, A. Kousaridas, M. Ibrahim, M. Dillinger, C. Thuemmler, H. Feussner, and A. Schneider, “Enabling real-time context-aware collaboration through 5G and mobile edge computing,” in *Proc. 12th International Conference on Information Technology – New Generations (ITNG)*, Apr. 2015, pp. 601–605.

[8] U. Drolia, R. Martins, J. Tan, A. Chheda, M. Sanghavi, R. Gandhi, and P. Narasimhan, “The case for mobile edge-clouds,” in *Proc. IEEE 10th International Conference on Ubiquitous Intelligence and Computing, and 10th International Conference on Autonomic and Trusted Computing (UIC/ATC)*, Dec. 2013, pp. 209–215.

[9] E. Cau, M. Corici, P. Bellavista, L. Foschini, G. Carella, A. Edmonds, and T. Bohnert, “Efficient exploitation of mobile edge computing for virtualized 5G in EPC architectures,” in *Proc. 4th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering (MobileCloud)*, Mar. 2016, pp. 100–109.

[10] K. Liu, J. K. Y. Ng, V. C. S. Lee, S. H. Son, and I. Stojmenovic, “Cooperative data scheduling in hybrid vehicular ad hoc networks: VANET as a software defined network,” *IEEE/ACM Trans. Networking*, vol. 24, no. 3, pp. 1759–1773, June 2016.

[11] C. Shao, S. Leng, Y. Zhang, A. Vinel, and M. Jonsson, “Performance analysis of connectivity probability and connectivity-aware MAC protocol design for platoon-based VANETs,” *IEEE Trans. Veh. Technol.*, vol. 64, no. 12, pp. 5596–5609, Sept. 2015.

[12] G. Zhang, T. Q. S. Quek, M. Kountouris, A. Huang, and H. Shan, “Fundamentals of heterogeneous backhaul design—analysis and optimization,” *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 876–889, Feb. 2016.

[13] J. F. Nash, “Equilibrium points in n-person games,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 36, no. 1, pp. 48–49, Jan. 1950.

[14] Q. Wang, S. Leng, H. Fu, and Y. Zhang, “An IEEE 802.11p-based multichannel MAC scheme with channel coordination for vehicular Ad Hoc networks,” *IEEE Trans. Intell. Transp. Syst.,* vol. 13, no. 2, pp. 449–458, June 2012.

VT