LATENCY AND RELIABILITY-AWARE JOINT OPTIMIZATION FOR COOPERATIVE VEHICULAR COMPUTING WITH COOPERATIVE COMMUNICATION[1]

Xu Han, Beihang University
Daxin Tian, Beihang University
Xuting Duan, Beihang University
Zhengguo Sheng, University of Sussex Richmond
He Liu, Academy of Military Sciences PLA China
Yinsheng Gong, CHINA FAW GROUP CO., LTD.
Victor C.M. Leung, University of British Columbia

1. Introduction

With the innovation and development of related technologies, intelligent networked vehicles that can use vehicle-to-infrastructure (V2I) or vehicle-to-vehicle (V2V) communications to form a vehicle self-organizing network (VANET) have become the current development trend in the automotive industry. Large-scale mobile applications in vehicle networking systems based on traditional computing/design paradigms typically complete services with the assistance of an on-board unit (OBU). Due to the limited computational power of the OBU, studies have shown that using wireless access to offload computing tasks to the cloud is considered an effective way to solve this problem [1]. However, the capacity limitation and delay fluctuation of backbone network and backhaul network caused by long-distance deployment of mobile cloud computing (MCC) make the vehicle networking system unable to meet the ultra-low latency service demand, resulting in a serious decline in vehicle quality of service (QoS).

In order to provide vehicles with more reliable services, Mobile Edge Computing (MEC) has been proposed [2], [3], which enables IT tasks to be offloaded to servers deployed at the edge of the network by providing IT service environments and cloud computing capabilities at the edge of the Internet of Vehicles. MEC is considered to an effective method to meet the expansion requirements of the computing power of the vehicle terminal equipment, while compensates for the shortcomings of the core cloud computing delay.

The dynamic optimization theory can well describe the time-varying of the system and reflect the influence of the current decision of the system on the time-accumulated objective function.Previous studies usually expect to improve the calculation and communication performance by optimizing the existing protocol parameters.However, a more appropriate approach should be to use joint optimization as one of the main considerations for system design in the algorithm and protocol design phase.Low-latency mobile computing and communication collaboration are necessary goals for the design of the Internet of Vehicles, but there are certain contradictions when the two are designed and operated in isolation.The introduction of computation offloading will alleviate the pressure of local computing on data analysis [4], [5], but the accuracy of data filtering needs to be guaranteed, which will definitely affect the effective operation of the communication protocol, thus affecting the communication and cooperation capabilities of the network.In addition, in communication cooperation, continuous adjustment of power, continuous change of bit rate, continuous update of window, etc. all affect the stable transmission of perceived messages, affecting the real-time and correctness of situational awareness, and ultimately affecting the computing resources of vehicles.

The various algorithms and protocols designed to improve mobile computing and communication collaboration cannot simply sacrifice their own performance at the expense of the other party, otherwise the smart car network thus built will be difficult to apply. Both mobile computing and communication

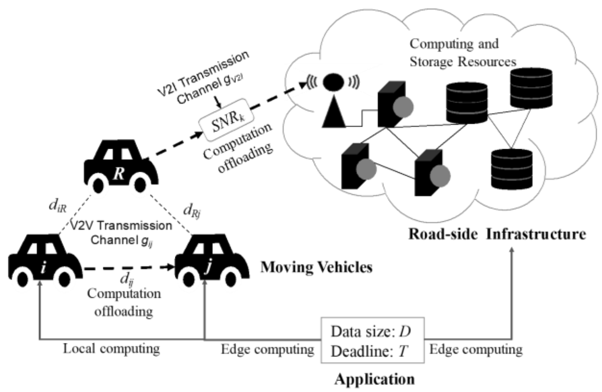
[1] 154th Annual Meetings of the *Canadian Transportation Research Forum*, May 26 - 29, 2019 at Vancouver, British Columbia

collaboration are one of the cutting-edge scientific issues that determine the successful deployment of smart car networks. It is necessary to establish a joint optimization balance between computing offload and communication performance that adapts to changing scenarios and resources. Therefore, mobile computing and communication collaboration can be optimized objects in dynamic optimization problems. In this paper, we will focus on the joint optimization of computational offloading and communication collaboration in Internet of Vehicles which in rapidly changing and resource-constrained driving behaviors and perception scenarios. The destination is to improve the quality of service of the Internet of Vehicles system.

2. System Model

A. System Analysis

Figure1. Cooperative V2X computing scenario.



In this part, we build a computing scenario (see Fig. 1) with multiple roadside infrastructures and Nrequesting vehicles, where these edge infrastructures and vehicles with limited wireless and computing resources are collectively referred to as service nodes (a total of L). Vehicle sets are represented as$V = \{v_1, v_2, \cdots, v_N\}$, and service nodes are represented as $G = \{g_1, g_2, \cdots, g_L\}$. Each service node g_i is equipped with limited computing resources C_i and radio resources B_i, where C_i and B_i in our model respectively indicate the number of CPU cycles and bandwidth required by a service. The computing task of the requesting vehicle can be executed locally by its own resources. Besides, the requesting vehicle within the communication range of the service node can offload the task to the service node, then obtaining the corresponding service. There is no core scheduler in the model, and the task offloading schedule is completed by the service node.

Following existing literature such as [6], [7], we can usetwo key parameters to characterize the profile of a mobileapplication: i) The input data size D that is the total numberof data bits as the application input. These D-bit data can bepartitioned and offloaded to a service node as thecloud edge for remote execution. ii) The application completiondeadline T that denotes the maximum number of successivetime slots before which the mobile application must becompleted. In addition, we use $k$to represent the index of a timeslot, and these D-bit data can also be partitioned into a series ofsmaller pieces$s_k \in [0, D]$, where s_k denotes the number of databits that should be transmitted to the service node in the k-thtime slot.

B. Communication Model

(1)Cooperative vehicle-vehicle communication

Suppose that each pair of V2V vehicles communicates through Rayleigh fading channels, it is possible to generate easily possible analysis results. We define a communication model based on cooperative transmission. The wireless link between service nodes$i$ and$j$ can be represented by the following model

$$g_{ij} = x_{ij}\frac{h_{ij}}{d_{ij}^{k/2}}(1)$$

where d_{ij} is the distance between service nodes$i$ and$j$, k is is the path loss coefficient, $d_{ij}^{k/2}$describes the large-scale behavior of channel gain, h_{ij}characterizes the fading characteristics of channels.$x_{ij}$is a connection variable, $x_{ij} = 1$when a connection is established between the requesting vehicle $i$and the service node$j$, otherwise it is 0.Consider a three-node scenario, thenode i wants to communicate with the node j with help of the relay node R. Thenthe data rate supported by V2V cooperative system can be shown as

$$I_{ij} = \begin{cases} \frac{1}{2}\log_2\left(1+2\rho\left|g_{ij}\right|^2\right) & \text{if } |g_{Ri}|^2 < q(\rho) \\ \frac{1}{2}\log_2\left(1+\rho[\left|g_{Rj}\right|^2+\left|g_{ij}\right|^2]\right) & \text{if } |g_{Ri}|^2 \geq q(\rho) \end{cases}(2)$$

where ρ denotes the signal-to-noise ratio, and$q(\rho) = \frac{2^{2\mathcal{R}}-1}{\rho}$ and $\mathcal{R}$ is the data rate in $bits/s/Hz$ defined by QoS requirement. Since h_{ij} is assumed complex Gaussian variables with zero mean and unit variance, $\left|\frac{h_{ij}}{d_{ij}^{k/2}}\right|^2$ is an exponentially distributed variable with parameter $\frac{1}{d_{ij}^2}$. According to the results in [8], the outage probability of cooperative transmissionbetween the nodes$i$ and$j$ is $P_{ij}^{DT} \approx d_{ij}^k\frac{(2^{\mathcal{R}}-1)}{\rho}$, and we have the outage probability of direct transmission between the nodes$i$ and$j$ as

$$P_{ij}^{CT} \approx \frac{1}{2}d_{ij}^k(d_{iR}^k+d_{Rj}^k)\left[\frac{(2^{2\mathcal{R}}-1)}{\rho}\right]^2 = P_{ij}^{DT}\left[\frac{(d_{iR}^k+d_{Rj}^k)(2^{\mathcal{R}}-1)(2^{\mathcal{R}}+1)^2}{2\rho}\right](3)$$

The necessary condition for cooperative transmission to be superior to direct transmission is to ensure that$\frac{P_{ij}^{CT}}{P_{ij}^{DT}} \geq 1$, i.e.$\frac{(d_{iR}^k+d_{Rj}^k)(2^{\mathcal{R}}-1)(2^{\mathcal{R}}+1)^2}{2\rho} \geq 1$.

(2)Cooperative vehicle-infrastructure communication

We suppose that a direct transmission between the vehicle and the roadside infrastructure. Besides, we adopt a two-state Markov chain to model the V2Ichannel, which includes a high-SNR and a low-SNR states.The SNRs corresponding to these two states are denoted by$SNR_H$and SNR_L, respectively. Let SNR_k be the SNR levelof the transmission channel in time slot k. Thus, $SNR_k \in \{SNR_H, SNR_L\}$. We denote by pHH the state transition probabilitythat the next channel state has a high SNR given that itscurrent state has a high SNR and by pLL the probability thatthe next channel state has a low SNR given that its current statehas a low SNR. Additionally, the transition probability that thechannel changes from the current state with a high SNR to thatwith a low SNR can be calculated by $pHL = 1 - pHH$, and thetransition probability from a low-SNR state to a high-SNR statecan be $pLH = 1 - pLL$. Based on the two-stateMarkov model, the mean durations (represented by the numberof time slots) of being in the high-SNR state and in the low-SNRstate can be estimated by $T_H = 1 - \frac{1}{1-pHH}$ and $T_L = 1 - \frac{1}{1-pLL}$, respectively.Given the fading coefficient of the V2I channel as g_{V2I} and the corresponding fading variance as σ^2, we can express thechannel capacity between the vehicle and the wireless accesspoint of the infrastructure by

$$I_{V2I} = \log_2(1+SNR_k|g_{V2I}|^2)(4)$$

Rayleigh fading model is also used for V2I communications,in which $\left|g_{V2I}\right|^2$ is a random variable following an exponential distribution with parameter σ^{-2}. Accordingly, the successful probability of transmitting s_k-bit data in a time slot k through the V2I channel can be calculated by

$$p(s_k, SNR_k) = \text{Prob}\{I_{V2I} > s_k\} = \exp\left(-\frac{2^{s_k}-1}{\sigma^2 SNR_k}\right)(5)$$

C. Calculation Model

(1)Local execution

When the local application is executed on the mobile device, theenergy consumption is determined by the CPU workload. Theworkload is measured by the number of CPU cycles requiredby the application, denoted as $W = DX$, where Xtypically modeled as a random variableconsistent with empirical distribution.Local computing resources of vehicle node i are represented as C_i^L, and computing abilityisrepresented asW_i^L. Thus the calculation time required for local execution is expressed as

$$t_i^L = \frac{C_i^L}{W_i^L}(6)$$

According to the result in [6], the total computation energy for the mobile execution is $\sum_{w=1}^{W} \mathcal{E}_w(f_w) = k{f_w}^2$, where k is the effective switched capacitance depending onthe chip architecture, f is the clock frequency. Under the optimal CPU frequencyscheduling, the minimum amount of energy consumption forthe mobile execution is given by

$$\mathcal{E}_l^* = \min_{\psi\in\Psi}\{\mathcal{E}_l(D, T, \psi)\}(7)$$

where $\psi = \{f1, f2, \ldots fW\}$ is any clock-frequency vectorthat meets the delay deadline, Ψ is the set of all feasible clock-frequency vectors, and $\mathcal{E}_l(D, T, \psi)$ is the total energyconsumed by the mobile device.

(2)Edge-cloud execution

The calculation time required for edge-cloud execution is expressed as

$$t_i^E = \frac{D}{I_{ij}} + \frac{C_j^L}{W_j^L}, I_{ij} \in \{I_{ij}^{DT}, I_{ij}^{CT}\}(8)$$

ForapplicationA (D, T), $S \leq D$ bits data need to be transmitted toservicenodeintimeT, therefarethetotal computation energy for edge-cloud execution is$\sum_{t=1}^{T} \mathcal{E}_t(S_t, g_t, n)$, where S_t denotes the number of bits transmitted and g_t denotes the channel state inslot t.Under the optimaltransmission scheduling, the minimum amount of transmissionenergy for the edge-cloud execution is given by

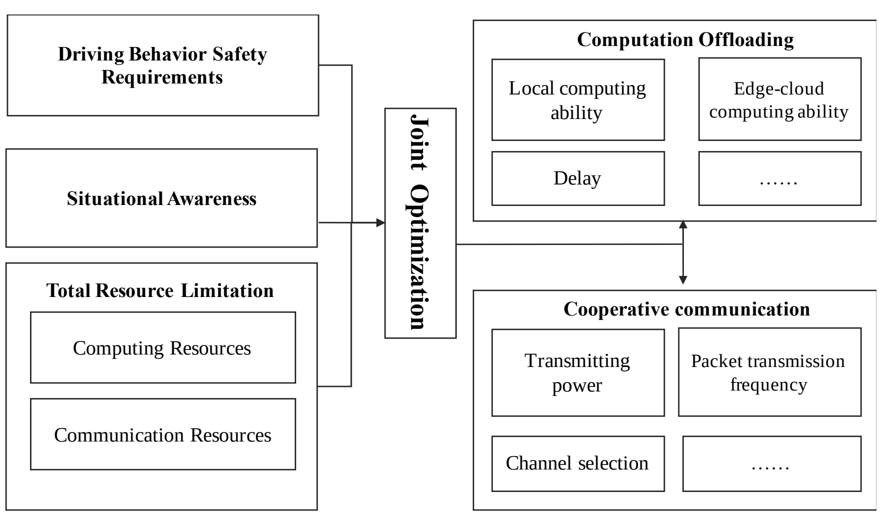
$$\mathcal{E}_E^* = \min_{\emptyset\in\Phi} \mathbb{E}\{\mathcal{E}_c(D, T, \emptyset)\}(9)$$

where $\varphi = \{s_1, s_2, \ldots s_T\}$denotes a data transmission schedule that meets the delay deadline (T time slots), Φ is the set of all feasible data scheduling vectors, and $\mathcal{E}_c(D, T, \emptyset)$ denotes the transmission energy.

3. Stochastic Optimization Model for Data Transmission Scheduling

The structure of the optimization system of computational uninstallation and cooperative communication control is shown in Fig.2.The system model takes the safety requirements of driving behavior and the total limit of network resources as the main line and optimization constraints, the control quantities of computation offloading and communication cooperation as the optimization objects.

Figure 2. The structure of the optimization system.



In V2I transmission schedulingsystem, we consider the optimization of transmission schedulingof D-bit application data in Ttime slots. For simplicity,we arrange the time slot index, k, in descending order, i.e.,$k = T, T-1, \ldots, 1$. That is, k is used as a countdown timer,and the initial time slot is indexed by$k = T+1$ .We also denoteby $s = [s_T, s_{T-1}, \ldots, s_1]^T$ a feasible scheduling solution andby $\mathbb{S}$ the corresponding feasible region.

In order to maximize the expected successful probability of computation offloading in the optimization model for V2I transmission scheduling based on (5) is expressed as

$$\max_{s\in\mathbb{S}}: \mathbb{E}\left[\prod_{k=1}^{T} p(s_k, \mathrm{SNR}_k)\right] = \mathbb{E}\left[\sum_{k=1}^{T} \exp\left(-\frac{2^{sk}-1}{\sigma^2 \mathrm{SNR}_k}\right)\right]$$

$$\mathrm{s.t.}\begin{cases}\sum_{k=1}^{T} s_k = D \\ \forall s_k > 0\end{cases} \tag{10}$$

To maximize the objective function in (10) asmuch as possible, we can turn to solve the following modelrather than the original model (10).

$$\min_{s\in\mathbb{S}}: \mathbb{E}\left[\sum_{k=1}^{T} \exp\left(-\frac{2^{sk}-1}{\sigma^2 \mathrm{SNR}_k}\right)\right]$$

$$\mathrm{s.t.}\begin{cases}\sum_{k=1}^{T} s_k = D \\ \forall s_k > 0\end{cases} \tag{11}$$

Thenwe can provethatThe optimal values of the objective function in (11) under $SNR_{T+1} = SNR_H$and $SNR_{T+1} = SNR_L$ are

$$\emptyset(\mathrm{D},\mathrm{T}|\mathrm{SNR_H}) = \frac{\mathrm{T}2^{\frac{lT}{T}}\left(\prod_{t=1}^{T} H_t^{\frac{t}{T}}\right) - \mathrm{TH}_1}{\sigma^2} \tag{12}$$

$$\emptyset(\mathrm{D},\mathrm{T}|\mathrm{SNR_L}) = \frac{\mathrm{T}2^{\frac{lT}{T}}\left(\prod_{t=1}^{T} L_t^{\frac{t}{T}}\right) - \mathrm{TL}_1}{\sigma^2} \tag{13}$$

respectively, where $l_T = D$. The minimum expected objective function in (11), denoted by $\Phi\ (D, T)$, is

$$\Phi\ (D,T) = \frac{T_H}{T_H+T_L}\emptyset(\mathrm{D},\mathrm{T}|\mathrm{SNR_H}) + \frac{T_H}{T_H+T_L}\emptyset(\mathrm{D},\mathrm{T}|\mathrm{SNR_L}) \tag{14}$$

Refer to the above,we propose the joint optimization model for V2V and V2I transmission scheduling based on (2), (5) and (8) with the goal to maximize the expected successful probability of computation offloadingunder the constraints of system demand and delay. It should be noted that service nodes can provide limited services in each slot.Next, we give the utility function ofmaximizing benefits to request node i.

$$U_i = \mathbb{E}\left[\sum_{v_i\in V, g_i\in G}\sum_{k=1}^{T}\left\{\exp\left(-\frac{2^{sk}-1}{\sigma^2 SNR_k}\right) \cap d_{ij}^k \frac{(2^{\mathcal{R}}-1)}{\rho} \cap \left[d_{ij}^k \frac{(2^{\mathcal{R}}-1)}{\rho} + \frac{(d_{iR}^k + d_{Rj}^k)(2^{\mathcal{R}}-1)(2^{\mathcal{R}}+1)^2}{2\rho}\right]\right\} t_i^E\right] \tag{15}$$

The planning problem as

$$\max_{s\in\mathbb{S}}: U_i$$

$$s.t.\begin{cases}\sum_{k=1}^{T} s_k = D \\ \forall s_k > 0 \\ \Theta_{ij} \geq c_i + w_i \\ \sum_{k=1}^{T} x_{ij} c_i \leq C_j, v_i \in V, g_i \in G \\ \sum_{k=1}^{T} x_{ij} w_i \leq W_j, v_i \in V, g_i \in G\end{cases} \tag{16}$$

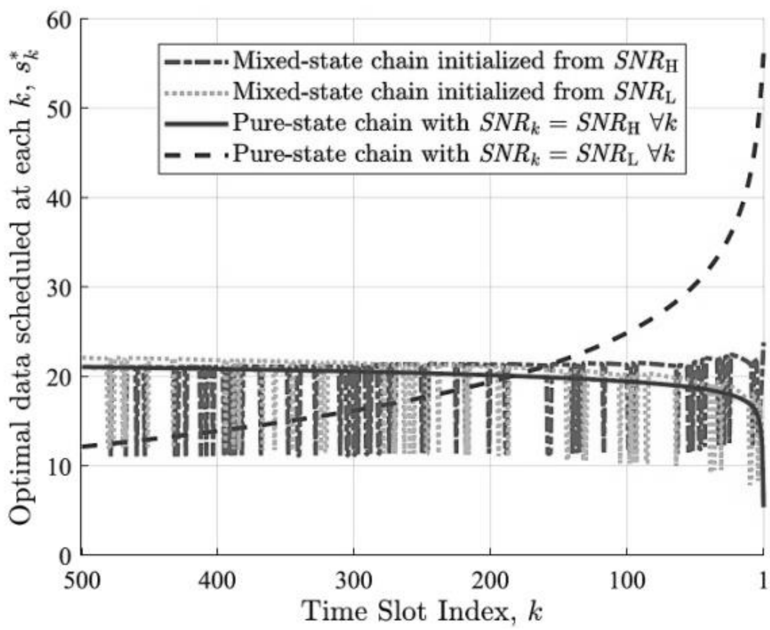
where $\Theta_{ij} = c_j + b_j$ represents the sum of resources obtained by vehicle i at service node j, c_i and w_i represent computing power and computing resources, respectively. Thensolving the above problems and get the result $\mathbb{U}\ (D, T)$.

The decision for latency and reliability-aware joint optimization application execution, is to choose where to execute the application, with an objective to minimize the total energy and computing recourses consumed on the mobile device while minimize the latency. Specifically, the optimal policy under a given threshold r is determined by the following decision rule

$$\begin{cases} Local\ Excution, & \Phi\ (D,T) \le r \cap\ \mathbb{U}\ (D,T) \le r \\ Roadside\ Excution, & \Phi\ (D,T) > r \cap\ \Phi\ (D,T) > U\ (D,T) \\ \text{Service vehicle Excution,} & \mathbb{U}\ (D,T) > r \cap \Phi\ (D,T) \le \mathbb{U}\ (D,T) \end{cases} \quad (17)$$

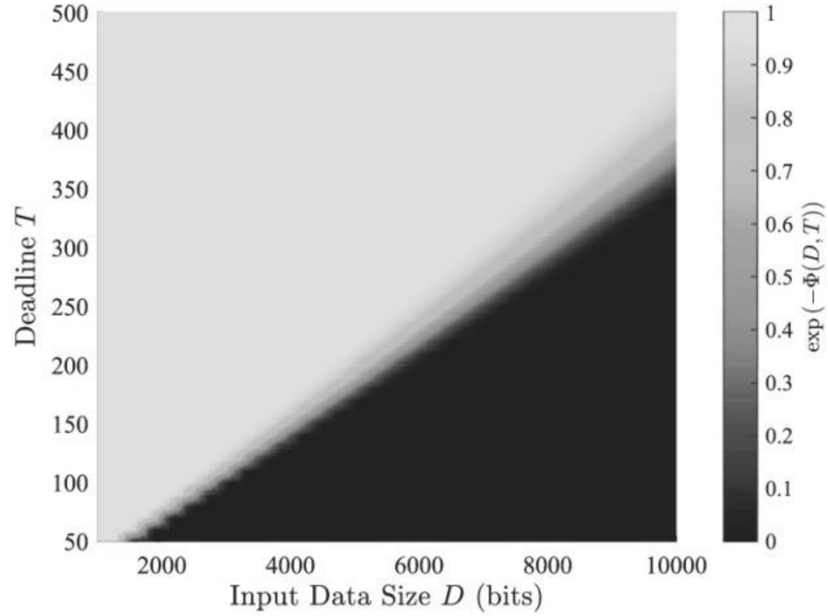
4. Numerical Results

Figure 3. Optimal V2I transmission scheduling with $D = 104$ bits, $T = 500$, $p_{HH} = 0.9$ and $p_{LL} = 0.2$.



Firstly, we analyze the V2I scenario, and conduct different simulation experiments where we set $SNRH = 50$ dB and $SNRL = 20$ dB to simulate the good andthe bad SNR conditions of the V2I channel, respectively, and let $\sigma = 10^3$. In Fig. 3, we illustrate the optimal data transmissionscheduling s_k^* in different specified cases of the channel statevariation. From two extreme cases (See the solid and theforth dashed lines), we can see that the number of data bits to betransmitted in each k decreases as time proceeds when the SNRis always high, while it will increase when the SNR always stayslow. The main reason is that since the expected SNR in the nexttime slot is lower given that the current SNR is high, the data bitsto be transmitted in the next time slot should be reduced in orderto guarantee the offloading reliability. In contrast, given thecurrent channel is at a low-SNR level, the expected SNR in thenext time slot is higher, such that the data bits to be transmittedin the next time slot should be larger. This proposition can alsobe confirmed in other two specific cases where the transitionchain consists of mixed states. It is worth pointing out that sincean optimal scheduling solution for the stochastic optimizationmodel (11) is obtained from the optimal expectation perspective,it may not be the best for a specific and deterministic case.For instance, in Fig. 3, in the extreme case where the SNR isalways low, the corresponding scheduling solution is not thebest solution for that case, and may even perform worse than asimple solution that schedules equal data bits in each time slot.

Figure 4. The lower bound of the expected reliability $\exp(-\Phi(D,T))$, $p_{HH} = 0.9$ and $p_{LL} = 0.9$.



Nonetheless, in actual implementation, we can decide to performancelocal execution instead of computation offloading whenthe lower bound of the optimal expected successful probabilityof computation offloading associated with an optimal schedulingsolution, $\exp(-\Phi(D,T))$, is lower than a given threshold.Moreover, we show the profile of $\exp(-\Phi(D,T))$ under differentapplication profiles in Fig. 4. A finite region with highreliability of computation offloading (where $\exp(-\Phi(D,T))$ ismore than 0.9) does exist.

Figure 5. Total utility of different amount of request nodes

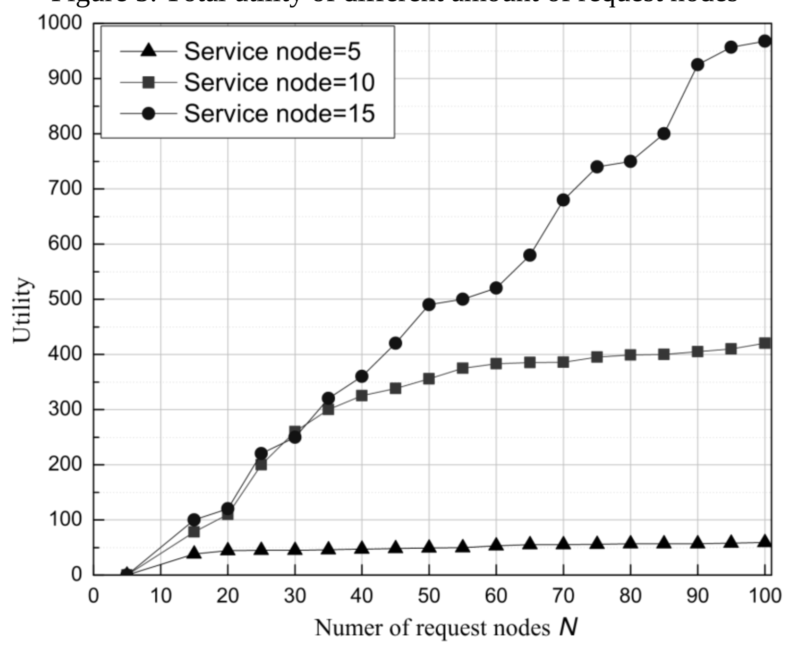


Fig.5 provides the total utility of different amount of request nodeswhen the number of service nodes is 5, 10 and 15, respectively.When there are only 5 service nodes, it is difficult to increase the system utility due to the limitation of resource capacity.However, when the amount of service nodes increases to 15, the system utility is increased to more than 950. Moreover, when the number of requestvehicles reach to 100, the utility can be improved by using the proposed framework.

5. Conclusions

Low latency and high reliability brought by calculation offloading and cooperative communication are the necessary goals for the design of the Internet of Vehicles, but there are certain contradictions when the two are designed and operated in isolation. Various optimization algorithms and protocols cannot simply maximize their performance at the expense of the other party. Otherwise, the intelligent mobile vehicle network thus established will be difficult to apply. Matching problem between requesting nodes and servicing nodes is studied when a vehicle wants to offload tasks, a roadsideinfrastructure-based offloading framework in vehicular networks is proposed, Vehicle can either offload task to roadsideinfrastructure sever as V2I link or neighboring vehicle as V2V link. Through theoretical modeling and dynamic road network simulation analysis, future work will focus on the joint optimization of computing offload and communication cooperative for the Internet of Vehicles, under rapidly changing and resource-constrained driving behaviors and perception scenarios. The proposed research can provide theoretical basis and key technologies for the design and joint control of new computing offloading strategies and new communication cooperation protocols, and promote the rapid development and comprehensive promotion of smart car networking systems.

Acknowledgement

This research was supported by the National Key Research and DevelopmentProgram of China (2017YFB0102502).

References

[1] XU Lida, HE Wu, and LI Shancang. Internet of things in industries: a survey[J]. IEEE Transactions on Industrial Informatics, 2014, 10(4): 2233–2243. doi: 10.1109/TII.2014.2300753.
[2] ABUELELA M and OARIUL S. Taking VANET to the clouds[C]. International Conference on Advances in Mobile Computing and Multimedia, Paris, France, 2010: 6–13.
[3] HAN Guangjie, LIU Li, CHAN S, et al. HySense: A hybrid mobile crowd Sensing framework for sensing opportunities compensation under dynamic coverage constraint[J]. IEEE Communications Magazine, 2017, 55(3): 93–99. doi: 10.1109/MCOM.2017.1600658CM.
[4] CHEN Xu, JIAO Lei, LI Wenzhong, et al. Efficient multiuser computation offloading for mobile-edge cloud computing[J]. IEEE/ACM Transactions on Networking, 2016, 24(5): 2795–2808. doi: 10.1109/TNET.2015.2487344.
[5] ZHANG Ke, MAO Yuming, LENG Supeng, et al. Predictive offloading in cloud-driven vehicles: using mobile-edge computing for a promising network paradigm[J]. IEEE Vehicular Technology Magazine, 2017, 12(2): 36–44. doi: 10.1109/MVT.2017.2668838.
[6] Zhang W , Wen Y , Guan K , et al. Energy-Optimal Mobile Cloud Computing under Stochastic Wireless Channel[J]. IEEE Transactions on Wireless Communications, 2013, 12(9):4569-4581.
[7] You C , Huang K , Chae H , et al. Energy-Efficient Resource Allocation for Mobile-Edge Computation Offloading[J]. IEEE Transactions on Wireless Communications, 2017, 16(3):1397-1411.
[8] Laneman J N ,Tse D N C , Wornell G W . Cooperative diversity in wireless networks: Efficient protocols and outage behavior[J]. IEEE Transactions on Information Theory, 2004, 50(12):3062-3080.