1

车辆边缘计算中解决任务卸载问题的遗传算法

段晓辉1,2

1清华大学,北京 100084

(sunrise_duan@126.com)

A Genetic Algorithm for Task Offloading problem in Vehicular Edge Computing

Xiaohui Duan^{1,2}

¹ (Tsinghua University, Beijing 100084)

Abstract In this paper, we consider task offloading in vehicular edge computing systems. The limited battery limits the mileage of electric vehicles. For a longer driving distance, the task of the vehicle can not only be performed locally, but also save energy by offloading the task. In this way, the energy utilization rate of the whole system is improved and the energy consumption of the system is reduced. The task offloading involves optimization variables which are where to offload tasks and the frequency, And The optimization variables are coupled. So it hard to decide how to allocate tasks. We propose that we can solve this problem by using genetic algorithm. Specifically, our experiment shows between 20% and 25% energy savings compared to the baseline which the tasks executes workload locally.

Key words Vehicular Edge Computing; Task Offloading; Genetic Algorithm

摘要 在本文中,我们考虑了车辆边缘计算系统中的任务卸载。电动汽车的行驶里程被有限的电池限制。为了实现更长的行驶距离,车辆的任务不仅可以在本地执行,还可以通过卸载任务来节省能源。这样可以提高整个系统的能量利用率,降低系统的能耗。任务卸载涉及优化变量有:卸载任务的位置和频率,并且这两个优化变量是耦合的。因此很难决定如何分配任务。我们提出可以使用遗传算法来解决这个问题。具体来说,我们的实验表明,与任务在本地执行工作负载的基线相比,可以节省 20% 到 25% 的能量。

关键词 车联网;任务卸载;基因算法

中图法分类号 TP391

1. INTRODUCTION

Clean energy is an important trend in the future. Electric energy is the representative of clean energy, so more and more people choose electric vehicles. Electric intelligent vehicles will be an integral part of urban transportation in the future. However, before major changes in battery technology, battery capacity is the shortcoming of electric intelligent vehicles[1]. Limited battery capacity will cause "range anxiety". Many methods have been proposed to solve the mileage problem, such as considering complete vehicle energy management[2]. The vehic-

ular coasts is part of the driving mode, thereby reducing energy consumption in[3].

Because of the large amount of data processed in the process of driving, the central processing unit(CPU) will consume considerable energy, which will have a great impact on the driving mileage. Some researchers proposed that send the task to mobile edge computing servers to save energy. But if all data is uploaded to cloud server for processing, the response time might be too long, and the requirements for low delay will not be met[4]. In paper[5], Liu et al.propose a system to guarantee task low

² (National Supercomputing Center in Wuxi, Wuxi 214072)

delay performance. In addition, the current bandwidth and storage simply cannot satisfy the requirements for transmitting all data.

To solve this problem, the recently popular edge computing technology is introduced. We can not only assign tasks to the nearby edge server, but also to resource rich users nearby. Yu et al.[6] focus on virtual machine resource allocation in vehicular edge clouds. Liu et al.[7] consider dynamic requirements and resource constraints, and classify all tasks into four types of pending lists. Then the tasks in each list will be offloaded to different nodes according to their features.

We propose to assign tasks to nearby vehicles to minimize energy consumption of the system. This offloading process forms a three-tier architecture, including vehicular cloud, edge servers which is roadside units, and central cloud processors. The edge servers are the controller of vehicular cloud. And the central cloud is responsible for global scheduling.

The problem is a non-linear integer programming. We propose a genetic algorithm to solve the task offloading problem. The experimental results show that when the number of vehicles is less than 30, the energy saving is obvious.

The remainder of this paper is organized as follows. Section II introduces the system model and problem formulation. Section III describes the genetic algorithm. Section IV describe our experiment and analysis the result. Section V is conclusion and future work.

2. SYSTEM MODEL AND PROBLEM

FORMULATION

We first describe the our system, including network artechiture and computation models in this section

2.1. Network Artechiture

Similar to[6], as shown in Fg.1, the hierarchical architecture consisting vehicular cloud, edge cloud, and central cloud.

The central cloud is responsible for global scheduling, and the tasks it performs include complicated computation and global decision.

Due to vehicular mobility, the edge cloud is composed of two parts: local server and communication site. And the edge cloud only serve nearby vehicles. The edge cloud is the controller of vehicular cloud, and is respon-

表 1 Notations in This Paper

Parameter	Interpretation
P	The provider
S_{it}	The provider <i>i</i> 's requester
V	The set of vehicles
T	The resource sharing time
C_{it}	The CPU capacity
r_i	The workload of vehicle i
r'_{jt}	The resource requester j need
$ec{M}_{it}$	Millions Instructions Per Second(MIPS)
θ_i	Estimated parameters
E_{it}^{blnc}	Energy consumed by the vehicle <i>i</i>
E_i^{rec}, E_i^{send}	Energy received or send by the vehicle i

sible for creating, maintaining, and deleting the cloud.

The vehicular cloud consists of vehicles which share their computation resources. Each vehicle can access the edge cloud and convey the information to it. The vehicles can save the energy through offloading task to other vehicles. It can improve the overall resource utilization.

First, the real-time traffic information is transmitted to the cloud server for analysis. The real-time information includes destination, current location, time, vehicle speed, and so on. After summarizing the data, the cloud server performs big data analysis to obtain the road congestion, and then infers the location information of the vehicle in a certain time period T in the future, and returns the information to the edge cloud which is near the vehicle. The information includes the set of vehicles which are closed to the vehicle within the range of time T, these vehicles will participate in resource sharing. Then, the edge cloud assigns tasks according to these information and divides the vehicles into providers and requesters, and the vehicles share resources at time slot T.

2.2. Constraints

To ensure the Quality of Service(QoS) between the requester and provider, the following distance constraints need to be met:

$$\min_{j \in S_{it}} l_{ij} < \delta, i \in P \tag{1}$$

Where l_{ij} is the distance between vehicle i and j. If the constraint is not satisfied, the vehicle will exit resource sharing.

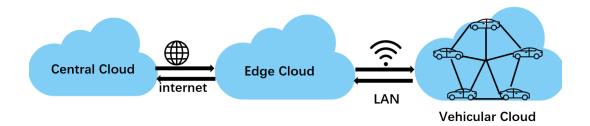


图 1 Fg1.Vehicular networks architecture

2.3. Computation Model

We define resource sharing time as $\mathcal{T} = \{1, \dots, T\}$. The number of vehicles is defined as N. We use a tuple $J_{it} = \{r_{it}, r'_{it}, d_{it}\}$ to represent the task of vehicle i at time slot t, where r_{it} is the number of instructions required by the vehicle i to complete the local task at time slot t, and r'_{it} is the number of instructions which the vehicle i task can offload to others at time slot t, and d_{it} is the task's data size.

To facilitate the description of the model, we denote $\mathbf{X}_t = \{x_{ij}\} \in \{0,1\}^{N \times N}$ allocation matrix, and $\mathbf{X}_t = [\mathbf{x}_{1t}, \mathbf{x}_{2t}, \dots, \mathbf{x}_{3t}, \dots, \mathbf{x}_{Nt}]$, where $\mathbf{x}_k = [x_{1k}, x_{2k}, \dots, x_{Nk}]^T$.

We characterize the computing capacity of vehicle i at time t by C_{it} . When a vehicle is selected as a provider (P) at time t, the computing resources of all requester should be less than its computing capacity.

$$\sum_{x_{ij}=1} r'_{jt} < C_{it}, i \in P \tag{2}$$

where r'_{jt} is the computing resource that the requester needs at time slot t, s_{it} is the set of customers of the provider at time slot t. For the convenience of description, the vector form is defined as $C_t = (C_{1t}, \ldots, C_{Nt})^T$, $R'_t = (r'_{1t}, \ldots, r'_{Nt})^T$.

A crucial step is how to quantify the computing resources. We regard the number of instructions that can be executed per unit time as the computing resources of a vehicle. The calculation capacity is defined based on the number of instructions executed per unit time (MIPS, M_{it}).

$$C_{it} = M_{it} \cdot T_s - r_{it} \tag{3}$$

For a single core CPU, the number of instructions executed per unit time (m_{it}) and CPU frequency (f_{it}) have

the following relationship:

$$M_{it} = v_i \cdot f_{it} + \theta_i \tag{4}$$

Where v_i and θ_i are parameters to be estimated. As a result, the CPU capacity C_{it} in Equation (3) can be calculated following:

$$C_{it} = (v_i \cdot f_{it} + \theta_i) \times T_s - r_{it} \tag{5}$$

2.4. Energy Consumption Model

When we consider the energy consumtion, we divide it into two parts: (1) the energy required for calculation, (2) the energy required for send or receive. In this paper[8], the energy consumption of processor contain dynamic energy consumption $E^{dynamic}$ and static energy consumption E^{static} . According to [9], the dynamic energy consumption is computed as follows:

$$E^{dynamic} = \alpha \cdot C \cdot V^2 \cdot f \cdot t \tag{6}$$

where α is a factor, C is the total capacitance load, V is the supply voltage, f is the processor frequency, t is a time period. E is proportional to $E^{dynamic}$, according to [10] [11].

$$E = \beta \cdot E^{dynamic} \tag{7}$$

where β is a factor. Voltage is a linear function of the frequency[12].

$$f = \gamma \cdot V \tag{8}$$

As a result, the energy consumption is the cubic function of frequency.

$$E = \lambda_i \cdot f_{it}^3 \cdot T_s \tag{9}$$

4 3. ALGORITHMS

Where f_{it} is the frequency of the CPU in vehicle i at time slot t. If a vehicle is selected as the requester, its frequency will decrease to f'_{it} . Because its tasks are assigned, so the energy saved by the vehicle is:

$$E_i^{save} = \{\lambda_i \cdot f_{it}^3 - \lambda_i \cdot f_i^3\} \times T_s, \forall i \in V \setminus P \quad (10)$$

The transmission energy consumption is linear with the transmission time which depends on the ratio between the data size and the data transmission rate between the requester and the provider. Data transmission rate is given by Shannon theorem.

$$E_i^{rec} = \sum_{j \in S_i} P_0 \cdot \frac{d_{it}}{W \log(1 + \text{SNR})}, \quad \forall i \in P \quad (11)$$

where P_0 is the transmission power, and SNR is Signal-tonoise Ratio. When the vehicle i is selected as the provider, the energy required to receive the task is Similarly, the energy required by the requester for transmission is

$$E_i^{send} = \sum_{j|i \in S_j} P_0 \cdot \frac{d_{it}}{W \log(1 + SNR)}, \quad \forall i \in \mathcal{V} \backslash P$$
(12)

Define E_t^{blnc} is the energy consumption of the vehicle in the time slot t. The provider's energy consumption at this time is:

$$E_{it}^{blnc} = \sum_{j \in S_i} P_0 \cdot \frac{d_{it}}{W \log(1 + SNR)} + \lambda_i \cdot f_{it}^3 \cdot T_s, \quad \forall i \in P$$
(13)

In addition, when vehicle i is the requester at this time t, , the total energy consumption is

$$E_{it}^{blnc} = \sum_{j|i \in S_j} P_0 \cdot \frac{d_{it}}{W \log(1 + SNR)} + \lambda_i \cdot f_{it}^{\prime 3} \cdot T_s, \quad \forall i \in \mathcal{V} \setminus \mathbb{R}$$
(14)

The objective of the framework is to minimize the quadratic power of energy balance over all vehicles.

$$\min \sum_{t=1}^{T} \sum_{i} (E_{it}^{blnc})^2 \tag{15}$$

Algorithm 1: Greedy algorithm

```
Input: Vehicular real-time information
  while true do
1
      Update set of vehicles allocated to a
2
       vehicular network;
      for t = 1 to T do
3
         for each edge server do
4
             Establish vehicular network as
5
              controller;
             Run GA(Iterations, M, r', rloc,);
6
         end
7
     end
8
9 end
```

The problem can be formulated as:

$$\begin{aligned} & \min & & \sum_{t=1}^{T} \sum_{i} (E_{it}^{blnc})^{2} \\ & \text{s.t.} & & E_{it}^{blnc} = \sum_{j|i \in S_{j}} P_{0} \cdot \frac{d_{it}}{W \log(1+SNR)} + \\ & & \lambda_{i} \cdot f_{it}^{\prime 3} \cdot T_{s}, \quad \forall i \in \mathcal{V} \backslash P \\ & & E_{it}^{blnc} = \sum_{j \in S_{i}} P_{0} \cdot \frac{d_{it}}{W \log(1+SNR)} + \\ & & \lambda_{i} \cdot f_{it}^{3} \cdot T_{s}, \quad \forall i \in P \\ & & \mathbf{X}_{t} R_{t}^{\prime} \leq C \end{aligned}$$

3. ALGORITHMS

We design a greedy algorithm to minimize the energy consumption of all time slot. For each moment, the central cloud returns the vehicles contained in each edge cloud and the time of participating in resource sharing, and all the edge cloud forms a vehicular cloud, then run genetic algorithm to minimize the energy consumption of a time slot.

Genetic algorithm is a adaptive heuristic optimization method, which is proposed by John Holland in 1970s[13]. Based on Darwin's theory of natural selection, $\forall i \in \mathcal{V} \backslash P$ it simulates the process of natural selection and reproduction to solve optimization problems. Similarly to individuals with favorable variation surviving, the solutions in the algorithm imitate the behavior of chromosomes, such as the mutations of genes and the crossovers of chromosomes. In recently, the genetic algorithm has been used for real-time path planning[14], maximum coverage deployment in Wireless Sensor Networks[15], and intrusion detection[16].

In genetic algorithm, a population is composed of randomly generated solutions, and all individuals in the population are composed of encoded strings similar to

Algorithm 2: Genetic Algorithm

```
Input: input parameters T, M, r', rloc,
   Output: X, f, E
1 t = 0;
2 Initialize \mathbf{X}(0), f(0);
3 while t < T do
       Encoder \mathbf{X}(t), f(t);
       for i = 1 to M do
5
           Evaluate fitness P(ti) = -E_i;
 6
       end
 7
8
       Eliminate fitness smaller individuals from
        P(t);
       Replicate the fitness larger individuals from
9
        P(t);
       Divide the array composed of optimization
10
        variables into two.;
       for i = 1 to M/2 do
11
           Crossover operation to P(t);
12
13
       for i = 1 to M do
14
        Mutation operation;
15
       end
16
17
       Decoder \mathbf{X}(t), f(t);
       t \leftarrow t + 1;
18
19 end
20 for i = 1 to N do
      E \leftarrow E_i + E
22 end
23 X = X(t), f = f(t);
24 return \mathbf{X}, f, E
```

chromosomes[17]. Similar to natural selection, with the iteration, the diversity decreases, but the preserved individuals are excellent adaptive individuals. In other words, the preserved solutions are are locally optimal.

We develop a genetic algorithm called GA-Veh. Algorithm 1 is the pseudo-code of GA-Veh in our experiment. The algorithm has as input the vector of vehicles with their request size, r'_i , and their local workload rloc. The output consists of the allocation matrix X_t , the frequency f_t , and the energy consumption E_t in the current time. At the initial stage of the algorithm, it will randomly generate a set of feasible solutions which is the first generation of chromosomes(Line2). After that, $\mathbf{X}(t)$ and f(t) is encoded as a string like a chromosome. Then the fitness function is used to calculate the fitness degree for each chromosome(Line5-7), and the probability of each chromosome being selected in the next evolution is calculated according to the fitness degree(Line8). The chromosomes with lower fitness will be eliminated, and only the chromosomes with higher fitness will be retained(Line9). Divide the individual into two parts at random(Line10). A certain position of these two chromosomes is cut off and spliced together to generate a new chromosome. As a result, this new chromosome contains a certain number of genes of both father and mother(Line11-13).

Because the solution thus solved is closer to the local optimal solution, and there is no way to achieve the global optimal solution. In order to solve this problem, we need to introduce mutation(Line14-16). The next step is to decode the string to the original solution(Line17). Finally we calculate all the energy consumption(Line20-22).

4. EXPERIMENT

We set the parameters of the experiment and analyzed the result of the experiment in the section.

4.1. The Setup

As we all known, U[x,y] is the uniform distribution between x and y, and N(x,y) is the normal distribution which mean is x and variance is y. We assume that the time which vehicles share their resource is $T_s=10$ seconds, because of their limited space.

The Cortex-A57 processor is ARM's highest performing processor, designed to further mobile and enterprise computing applications[18]. The Cortex-A57 has the frequency from 700 MHz to 1900 MHz. The frequency set is $\{700, 800, 900, , , 1900\}$, as a result, there is only 13 frequency levels.

As is shown in Equation (4), the number of instructions is $M_{it} = v_i \cdot f_{it} + \theta_i$. And the CPU capacity in Equation (5) is $C_{it} = (v_i \cdot f_{it} + \theta_i) \times T_s - r_{it}$,

We estimate the CPU capacity parameters v_i , θ_i is based on the analysis provided in [19], and the value is 7.683 and -4558.52. After calculation, the maximum and minimum M values are respectively 819, 10039. And the r'_{it} , For all tasks the required computing resource r_{it} to execute workloads is uniformly drawn from [820, 10000], and we assume that the task computing resource value which vehicle i can offload to others is $U[200, r_{it}]$, because some subtask must be execute locally. In our experiment the size of data varies from 1MB to 10 MB. And we set the value of bandwith d is 27Mb/s We have selected seven kinds of vehicle number to participate in resource sharing, which are $\{10, 15, 20, 25, 30, 35, 40\}$.

Our experiment are implemented in python and executed on an Intel Core i7 with 8 GB RAM.

表 2 The values of parameters

Parameters	Values / Distribution
f_i	[700, 1900]
v_i	7.683
θ	-4558.52
λ	0.00125
T_s	1
r_{it}	U[820, 10000]
r'_{it}	$U[200, r_{it}]$
P_0	0.2
$W \log(1 + \text{SNR})$	27

4.2. Performance Metrics

The performance of our experiment is the percentage of energy saving, which is define as follows:

$$P = 100 \cdot (1 - \frac{\sum_{t=1}^{T} \sum_{i} E_{it}^{blnc}}{E_{loc}})$$
 (16)

where E_{loc} is the energy consumed by local task execution.

In order to reflect the fairness of our framework, we define the fairness coefficient (FC) based on the standard deviation As with the standard deviation, the smaller the value the greater the fairness. FC is define as follows.

$$FC = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(E_{it}^{blnc} - \bar{E} \right)^2}}{\bar{E}}$$
 (17)

4.3. Experimental Results

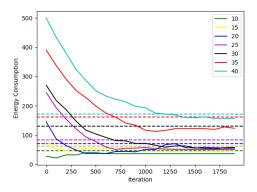


图 2 Energy Consumption

In Fg.2, the dotted line is the energy consumed by local execution. We can see that when the number of vehicles is 10, the algorithm can get good results by running

a few times. While the number of vehicles is 20, it takes 350 iterations to get a better result. And when the number of vehicles is 30, the reduction rate of energy consumption is faster before 750 iterations, and will be significantly slower after 750 iterations. Almost no energy saving when the number of vehicles is 40. We can conclude that with the increase of the number of vehicles, the number of iterations is also increasing. Therefore, when we run the algorithm on the edge server, we need to adjust the number of iterations reasonably as the number of vehicles changes.

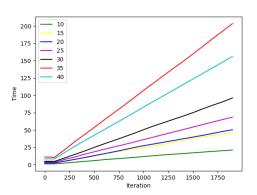


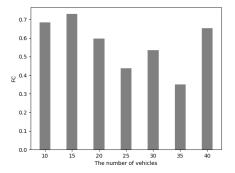
图 3 Experimental running time

In Fg.3, after 100 iterations, the running time and the number of iterations show a linear relationship. However, when the number of iterations is fixed, the relationship between time and vehicle is not linear. As Fg.4(a) shows, we can see that with the increase of vehicles, the balance of vehicle energy consumption increases first and then decreases. As Fg.4(b) shows, when the number of vehicles is 10 and 20, our system can save half of the energy consumption, but when the number of vehicles is 25, only 25% can be saved, and when the number of vehicles is 40, almost no savings.

Considering the performance measurement, energy consumption and time, it is appropriate to choose 25 vehicles to participate in resource sharing.

5. CONCLUSIONS AND FUTURE

In this paper, we proposed task offloading based on genetic algorithm for electric smart vehicle. We evaluate the performance of the algorithm through simulation experiments. Experimental results show that our algorithm can achieve the goal of energy conservation. In the future research, we plan to establish the relationship between the data size and the number of instructions and consider the



(a) The values of FC

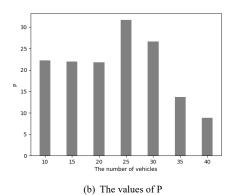


图 4 Performance metrics

deadline constraints.

参考文献

- [1] CANO Z P, BANHAM D, YE S, et al. Batteries and fuel cells for emerging electric vehicle markets[J]. Nature Energy, 2018, 3(4): 279-289.
- [2] ROMIJN T C J, DONKERS M C F, KESSELS J T B A, et al. A distributed optimization approach for complete vehicle energy management [J/OL]. IEEE Transactions on Control Systems Technology, 2019, 27(3): 964-980. DOI: 10.1109/TCST.2018.2789464.
- [3] BARTA D, MRUZEK M, LABUDA R, et al. Possibility of increasing vehicle energy balance using coasting[J]. Advances in Science and Technology. Research Journal, 2018, 12(1).
- [4] PASSAS V, MAKRIS N, NANIS C, et al. V2mec: Low-latency mec for vehicular networks in 5g disaggregated architectures[C]//2021 IEEE Wireless Communications and Networking Conference (WCNC). IEEE, 2021: 1-6.
- [5] LIU C F, BENNIS M, DEBBAH M, et al. Dynamic Task Offloading and Resource Allocation for Ultra-Reliable Low-Latency Edge Computing [J/OL]. IEEE Transactions on Communications, 2019, 67(6): 4132-4150. DOI: 10.1109/TCOMM.2019.2898573.
- [6] YU R, ZHANG Y, GJESSING S, et al. Toward cloud-based vehicular networks with efficient resource management[J]. Ieee Network, 2013, 27 (5): 48-55.

- [7] LIU C, LIU K, GUO S, et al. Adaptive offloading for time-critical tasks in heterogeneous internet of vehicles[J/OL]. IEEE Internet of Things Journal, 2020, 7(9): 7999-8011. DOI: 10.1109/JIOT.2020.2997720.
- [8] KIM K H, BUYYA R, KIM J. Power aware scheduling of bag-of-tasks applications with deadline constraints on dvs-enabled clusters[C/OL]// Seventh IEEE International Symposium on Cluster Computing and the Grid (CCGrid '07). 2007: 541-548. DOI: 10.1109/CCGRID.2007.85.
- [9] GE R, FENG X, CAMERON K. Performance-constrained distributed dvs scheduling for scientific applications on power-aware clusters[C/OL]//SC '05: Proceedings of the 2005 ACM/IEEE Conference on Supercomputing. 2005: 34-34. DOI: 10.1109/SC.2005.57.
- [10] LI J, MARTINEZ J. Dynamic power-performance adaptation of parallel computation on chip multiprocessors[C/OL]//The Twelfth International Symposium on High-Performance Computer Architecture, 2006. 2006: 77-87. DOI: 10.1109/HPCA.2006.1598114.
- [11] PIGUET C, SCHUSTER C, NAGEL J L. Optimizing architecture activity and logic depth for static and dynamic power reduction[C/OL]//The 2nd Annual IEEE Northeast Workshop on Circuits and Systems, 2004. NEW-CAS 2004. 2004: 41-44. DOI: 10.1109/NEWCAS.2004.1359011.
- [12] ELNOZAHY E N M, KISTLER M, RAJAMONY R. Energy-efficient server clusters[C]//FALSAFI B, VIJAYKUMAR T N. Power-Aware Computer Systems. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003: 179-197.
- [13] SAMPSON J R. Adaptation in natural and artificial systems (john h. hol-land)[J/OL]. SIAM Review, 1976, 18(3): 529-530. https://doi.org/10.1137/1018105.
- [14] ROBERGE V, TARBOUCHI M, LABONTÉ G. Comparison of parallel genetic algorithm and particle swarm optimization for real-time uav path planning[J]. IEEE Transactions on industrial informatics, 2012, 9(1): 132-141.
- [15] YOON Y, KIM Y H. An efficient genetic algorithm for maximum coverage deployment in wireless sensor networks[J/OL]. IEEE Transactions on Cybernetics, 2013, 43(5): 1473-1483. DOI: 10.1109/TCYB.2013.225095
 5.
- [16] KANNAN A, MAGUIRE JR G Q, SHARMA A, et al. Genetic algorithm based feature selection algorithm for effective intrusion detection in cloud networks[C]//2012 IEEE 12th International Conference on Data Mining Workshops. IEEE, 2012: 416-423.
- [17] SIVARAJ R, RAVICHANDRAN T. A review of selection methods in genetic algorithm[J]. International journal of engineering science and technology, 2011, 3(5): 3792-3797.
- [18] Cortex-A57[EB/OL]. [2022-07-20]. https://developer.arm.com/Processors/Cortex-A57.
- [19] BAHREINI T, BROCANELLI M, GROSU D. Vecman: A framework for energy-aware resource management in vehicular edge computing systems [J/OL]. IEEE Transactions on Mobile Computing, 2021: 1-1. DOI: 10.1 109/TMC.2021.3089338.