SATBAYEV UNIVERSITY

Computing & Engineering

Volume 1 (2023), Issue 2, 14-19

https://doi.org/10.51301/ce.2023.i2.03

Edge computing resource scheduling optimization method for Internet of Vehicles

Fang Ding*, Cao Yuning

Satbayev University, Institute of Automation and Information Technologies, Almaty, Kazakhstan

**Corresponding author: Ding.F@stud.satbayev.university*

Abstract. With the rapid development of the Internet of Things and 5G technology, the Internet of Vehicles has become one of the iconic application scenarios of 5G technology, promoting the emergence of a large number of emerging vehicle-mounted intelligent applications. However, vehicle terminals have limited computing resources, making it difficult to meet the latency and energy consumption requirements of emerging applications such as real-time traffic conditions and intelligent identification. The emergence of Internet of Vehicles technology based on edge computing solves the above problems. In-vehicle edge computing deploys edge servers to Road Side Units (RSU) close to the vehicle to provide computing and storage services to the vehicle, relieving the computing and storage pressure on the vehicle terminal. However, due to the limited resources of the vehicle edge network, unlimited task scheduling to the edge server will cause the server to be overloaded and affect the quality of service (QoS) of vehicle users. A reasonable resource scheduling strategy can ensure the comprehensive performance of the edge network and improve user QoS. Aiming at the problem of complex road traffic and dense vehicles in the Internet of Vehicles that generate too many computing tasks that need to be processed in real time, a partial resource scheduling strategy based on an improved simulated annealing algorithm is proposed for divisible tasks. With the goal of minimizing system overhead, a partial scheduling system model including the offloading ratio factor is constructed. Decompose part of the resource scheduling problem into two factors: offloading ratio and computing resource allocation. The optimal offloading ratio factor is solved through the improved simulated annealing algorithm. During the algorithm iteration process, the optimal resource allocation is obtained through the Lagrange multiplier method. Simulation verified the convergence of the proposed strategy and its efficiency in optimizing delay, energy consumption and overhead.

Keywords: *vehicular network, intelligent cockpit, edge computing, simulated annealing algorithm*.

1. Introduction

As a cutting-edge technology, the Internet of Things (IoT) is being fully integrated with the Internet [1]. It can be applied in a variety of fields, such as smart homes, smart transportation, and autonomous driving [2]. Due to its wide application in various fields, there is an explosion of data generated by intelligent transportation systems and autonomous driving applications. In order to handle data requests in intelligent transportation, autonomous driving, etc., and build smart cities, the Internet of Vehicles (Vehicle-To-Everything (V2X) came into being [3] and became one of the representative research scenarios in the IoT field. The Internet of Vehicles [4] consists of five elements, including people, vehicles, roads, communications, and service platforms. In the V2X scenario, a comprehensive Internet of Everything can be realized to interact with roadside infrastructure, vehicles and people [5-6]. However, as the intelligent applications of the Internet of Vehicles become more complex, people have stricter requirements for in-vehicle intelligent applications, which require low energy consumption, low latency, and high real-time performance. However, the computing resources of the vehicle itself are limited and cannot meet the delay and energy consumption requirements of computing-intensive on-vehicle intelligent applications, which will reduce the quality of service (QoS) of the vehicle [7].

With the emergence of cloud computing, the above problems have been gradually solved and the efficiency of task execution can be significantly improved. The cloud computing center consists of a large number of cloud servers and has powerful computing resources [8]. The vehicle can schedule locally generated tasks to the cloud computing center, which can effectively alleviate the problem of insufficient vehicle computing resources and improve the speed of processing tasks. However, cloud computing centers are often deployed in remote areas far away from vehicle terminals, which will cause high delays during data transmission and are prone to interference, which will also affect the QoS of vehicle applications [9].

In order to solve the shortcomings of cloud computing, edge computing has entered people's field of vision as an emerging technology. It deploys edge servers closer to users, reducing latency to a greater extent and avoiding long-distance transmission of cloud computing, enhances the real-time nature of task processing [10-11], alleviates the uncertainty of network transmission, and can bring better QoS to end users [12]. With the development of edge computing, its application in Internet of Vehicles scenarios can meet the requirements of low latency, high real-time performance, and low energy consumption for intelligent applications of Internet of Vehicles. Edge computing for the Internet of Vehicles deploys edge servers in RSUs, and provides computing resource scheduling services for vehicle terminals through reasonable deployment decisions and efficient service speeds brought by close-range deployment of servers, reducing the computing pressure of vehicles [13], saving vehicle energy consumption, reducing transmission delay, and bringing better QoS to vehicle users.

1.1. Related Works

The offloading decision of vehicular edge computing, that is, the vehicle user makes an offloading strategy based on its own needs, includes whether to offload tasks and how much proportion to offload. Huy Hoang V [14] et al. proposed an edge computing infinite network for vehicular networks, which can be mobility-aware to perform task offloading. The value generated by the comparison of tasks performed by vehicles and edge servers is used to decide how to offload. For this, with the objective of minimizing the total overhead of the system, a system communication model and a computation model are established, and the simulation comparison shows that the strategy reduces costs by 17% compared with other benchmark strategies. Chen L [15] et al. proposed a new concept, service value, to generalize the total value of vehicle system delay and energy consumption. Xu [16] et al. proposed a joint multi-objective co-optimization task offloading strategy, which aims to reduce the pressure on the vehicle edge computing network, reduce the response time of the system and alleviate the consumption of energy. This strategy can directly schedule tasks to target edge servers through vehicle-to-base device communication according to the state of vehicle movement, or upload tasks to target edge servers through vehicle-to-vehicle communication. When all edge servers are insufficient in resources, it is necessary to consider scheduling tasks that cannot be processed by edge servers to cloud computing centers. The load balance of the system is achieved by using a heuristic algorithm such as non-dominated genetic algorithm. Zhang P [17] et al. made further optimizations to the delay-oriented edge computing based vehicular network framework, which comprehensively considers the mobility of vehicles and the demand of tasks, and proposes a comprehensive resource scheduling strategy, including the selection of edge servers and the management of tasks. A prediction mode transmission method is proposed for computing task uploading to improve transmission rate while satisfying delay constraints.

2. Materials and methods

In summary, partial scheduling strategies are proposed for the splittable task in the case of traffic congestion. And a partial scheduling model of vehicular network communication environment is established, and the system model diagram is shown in Figure 1.

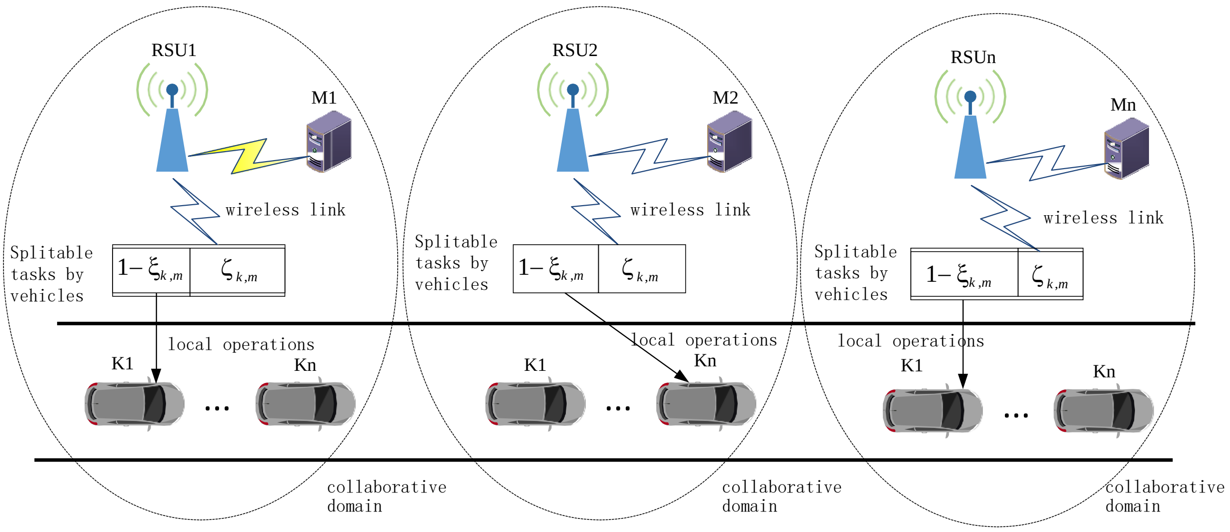

Figure 1. System model

Assume that there are M RSUs uniformly distributed on the one-way road, and an edge server is deployed in each RSU. The set of edge servers in partial scheduling can be expressed as $M = \{1, 2, 3,..., m\}$。 Suppose there are K vehicles, and the set is $K = \{1, 2, 3,..., k\}$, Each vehicle has a computation task, then there are K computation tasks. Suppose all vehicles obey Poisson distribution, then a vehicle can produce one divisible processing task at a time[7]. The computational task generated by the vehicle can be represented as $G_k = \{S_k, C_k, T^{max}_k, V^k\}$, S represents the size of data volume generated by vehicle-generated tasks, and C represents the CPU cycles required to process vehicle-generated tasks, T^{max} represents the maximum delay time that can be tolerated by the processing vehicle to produce a task, V_k represents the value generated by the vehicle corresponding to the task. The offloading ratio factor of the task scheduling to edge server for vehicle terminal can be defined as $\xi_{k,m}$ $(0 \le \xi_{k,m} \le 1)$. Therefore the offloading ratio factor at vehicle local is $1-\xi_{k,m}$.

This study employs the Improved Simulated Annealing Algorithm (ISAA) to determine the optimal offloading ratio factor. In contrast to the conventional simulated annealing algorithm, this research utilizes the particle swarm optimization algorithm to identify the adaptive temperature decay coefficient, thereby securing the optimal temperature decay coefficient in the current state [18]. The temperature decay coefficient within the simulated annealing algorithm process dictates the rate of temperature reduction. Building upon the traditional simulated annealing algorithm, this study adaptively modifies the temperature decay coefficient based on the rate of temperature reduction. This adjustment allows the rate of temperature decrease to align with the characteristics of searching for the optimal offloading ratio factor, thereby significantly enhancing both the performance and search efficiency of the simulated annealing algorithm.

A. Partially scheduled communication model

Each vehicle node can communicate with the RSU upload communication system data through an independent channel, according to Shannon theory, the upload rate $r_{k,m}$ can be expressed as:

$$r_{k,m} = W_{k,m} log_2 (1 + \frac{\rho_{k,m} G}{\sigma^2}) \tag{1}$$

$W_{k,m}$ denotes the communication link bandwidth, $\rho_{k,m}$ denotes the transmit power, G denotes the channel gain, and σ^2 denotes the average noise power.

B. Partially schedule local computation model

Tasks generated by the vehicle terminal are concurrently scheduled to both the edge server and the vehicle terminal, adhering to a predetermined offloading ratio.

If the calculation task of □$_{k,m}$ offloading ratio factor is scheduled to edge server for processing, then the remaining 1□ □$_{k,m}$ offloading ratio factor needs to be processed at vehicle terminal. The time delay tlocal that the calculation task of remaining 1□ □$_{k,m}$ ratio factor is processed at vehicle terminal can be defined as:

$$t_{k,m}^{local} = \frac{S_k (1 - \zeta_{k,m})}{f_{k,m}^{local}} \tag{2}$$

f denotes the vehicle terminal CPU computing power.

C. Partially Scheduled Edge Computing Model

The ratio factor of the scheduling to the edge server and the total delay t for computation can be expressed as:

$$t_{k,m}^{sum} = t_{k,m}^{trans} + t_{k,m}^{edge} = \zeta_{k,m} \left(\frac{S_k}{f_{k,m}^{edge}} + \frac{D_k}{r_{k,m}} \right) \tag{3}$$

2.1. Overall algorithm design

The optimization problem in this article is split into two sub-problems. First, the simulated annealing algorithm is used to find the optimal offloading ratio factor. After the solution is obtained, it is then substituted into the resource allocation problem as a known solution. It is mathematically proven that problem P3 is a convex optimization problem. To find the optimal resource allocation, the specific steps of the ISAA-COA algorithm (Improved Simulated Annealing Algorithm-Convex Optimization Algorithm, ISAA-COA) proposed in this article are as follows:

(1) Initialize simulated annealing parameters and start iteration.

(2) Generate a new unloading ratio factor through random perturbation, and use the particle swarm optimization algorithm to obtain the optimal temperature attenuation coefficient.

(3) The latest solution is obtained in each iteration, and the objective function at this time is updated to P2.

(4) Substitute the latest unloading ratio factor into P2 to obtain P3. It is proved that P3 is a convex problem. It is solved by the Lagrange multiplier method and the optimal resource allocation solution of P3 is obtained.

(5) Substitute the optimal offloading ratio solution and resource allocation solution for each update into problem P1 to calculate the current system average cost, and compare it with the system cost of the initial solution. If the average system cost at this time is less than the system cost of the initial solution. The cost indicates that this iteration is valid and the perturbation solution of this iteration is accepted. On the contrary, screen out the poor solutions with a certain probability and re-enter step (2).

(6) When the simulated annealing temperature decays to the termination temperature, the iteration ends, and the optimal unloading ratio factor and resource allocation are output.

2.2. Experimental simulation

The MATLAB simulation platform is used for simulation verification, assuming that there are 50 vehicles on the one-way road, and there are 5 RSUs on the road surface, each RUS configures an edge server, a vehicle terminal generates a computing task at a certain moment, so the number of computing tasks and the number of vehicles is the same, and there are 50 tasks. The specific parameters are shown in Table 1.

Table 1 Experimental parameter table

Parameter	Parameter value
Number of vehicles	40
Curbside Unit RSU Quantity	5
Task data volume S_k	[900,1600] KB
CPU cycles required for the task C_k	[0.3,1] GHz
Vehicle terminal computing power focal	1GHz
Energy coefficient k	$1x10^{-27}$
Network bandwidth B	20MHz
Uplink channel gain G	140+36.71og d
The maximum computing resource Fmax of the edge server	20GHz
Delay weight factor	0.2
Initial temperature T	1000
Initial temperature attenuation coefficient	0.95

In order to verify the ISAA-COA algorithm proposed in this article, the algorithm in this article is compared with local computing (LC), adaptive genetic algorithm (AGA), and random scheduling algorithm (RA).

3. Results and discussion

According to the simulation results in Figure 2, it can be observed that there is a positive correlation between the number of CPU cycles of the computing task and the system delay. However, compared with the other three benchmark algorithms (local computing (LC), adaptive genetic algorithm (AGA), and random scheduling algorithm (RA)), the ISAA-COA algorithm proposed in this article has a higher average delay when processing tasks. short. Specifically, the average delays of the ISAA-COA algorithm, AGA algorithm, Random algorithm and LC algorithm are 0.751s, 0.906s, 1.04s and 1.13s respectively. It can be calculated that the ISAA-COA algorithm reduces the optimization delay by 20.6%, 38.5%, and 50.5% respectively compared with the above three solutions. It shows that the ISAA-COA algorithm proposed in this article has certain advantages in optimizing delay.

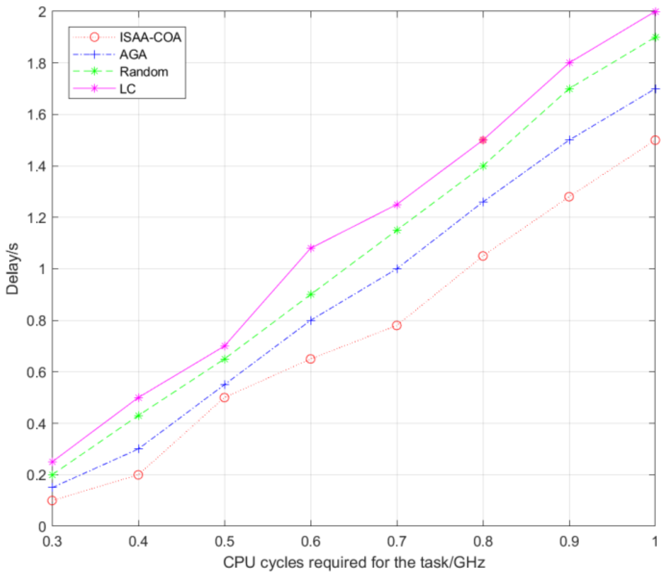

Figure 2. Simulation relationship diagram of computing task CPU cycles and latency

According to the simulation results in Figure 3, it can be observed that there is a positive correlation between the number of CPU cycles of the computing task and the system energy consumption. However, the ISAA-COA algorithm proposed in this article produces lower average energy consumption when processing tasks than the other three benchmark algorithms (AGA algorithm, Random algorithm and LC algorithm). Specifically, the average energy consumption of the ISAA-COA algorithm, AGA algorithm, Random algorithm and LC algorithm is 6.506J, 8.638J, 8.888J and 10.15J respectively. It can be concluded that compared with the above three benchmark algorithms, the energy consumption of the ISAA-COA algorithm is reduced by 32.8%, 36.6%, and 56.0% respectively. It shows that the ISAA-COA algorithm is efficient in reducing energy consumption.

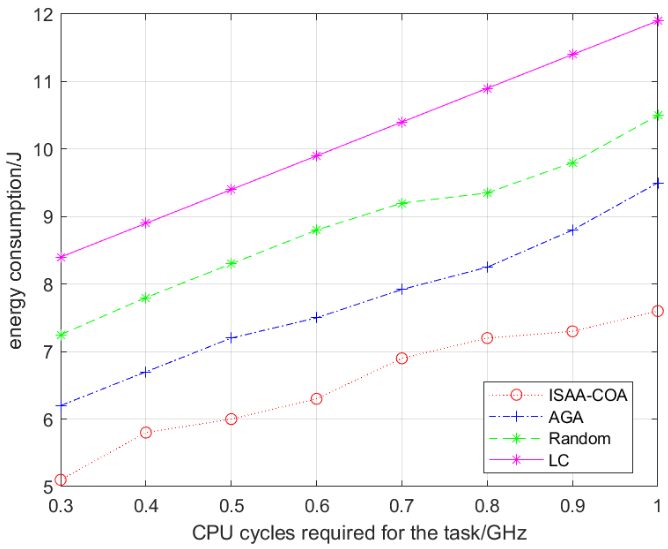

Figure 3. Computing task CPU cycles versus energy consumption graph

According to the simulation results in Figure 4, a positive correlation between the number of CPU cycles of a computing task and the system overhead can be observed. However, compared with the other three benchmark algorithms (AGA algorithm, Random algorithm and LC algorithm), the ISAA-COA algorithm proposed in this paper produces lower average overhead when processing tasks. Specifically, the average overheads of the ISAA-COA algorithm, AGA algorithm, Random algorithm, and LC algorithm are 5.355, 6.391, 7.318, and 8.346 respectively. It can be concluded that the average overhead of the ISAA-COA algorithm is reduced by 19.3%, 36.7%, and 55.9% respectively compared with the above three benchmark algorithms. Note that the ISAA-COA algorithm has certain advantages in reducing the average system overhead.

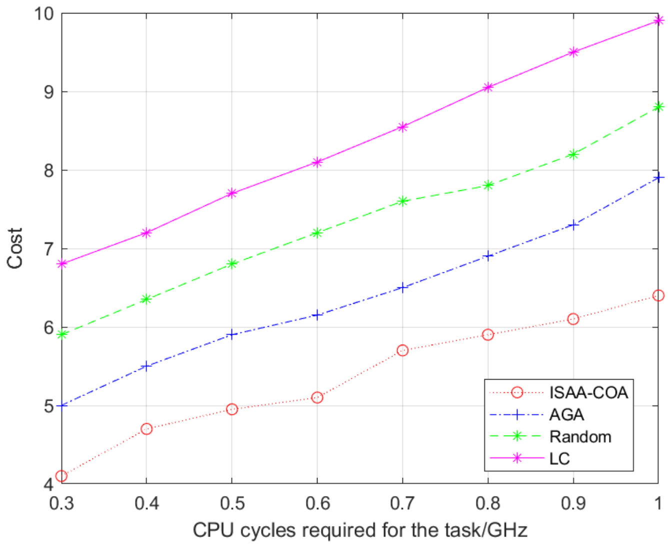

Figure 4. Compute task CPU cycles versus system overhead graph

In summary, the proposed ISAA-COA algorithm in this paper can effectively reduce the time delay, energy consumption and average system overhead of vehicular ad hoc networks. The algorithm can obtain the optimal offloading ratio factor under the optimization of the ISAA algorithm, which further improves the performance and efficiency of the system. Its advantages are mainly reflected in the accurate calculation and adjustment of task allocation and offloading ratio factor, so that the system can better adapt to different task requirements and network environments, improving the reliability and stability of the system.

References

[1] Wei, F, Chen, S, Zou, W. (2018). A greedy algorithm for task offloading in mobile edge computing system[J]. *China Communications, 15(11), 149-157*

[2] Shen Qing. (2016). Service quality optimization solution for Internet of Vehicles[D]. *Southeast University*

[3] Han Yi. (2019). Research on task offloading based on joint scheduling of Internet of Vehicles communications and computing resources[D]. *Xiamen: Xiamen University*

[4] Ji, B, Zhang, X, Mumtaz, S. (2020). Survey on the internet of vehicles: Network architectures and applications[J]. *IEEE Communications Standards Magazine, 4(1), 34-41*

[5] Mao, Y, You, C, Zhang, J. (2017). A survey on mobile edge computing: The communication perspective[J]. *IEEE Communications Surveys & Tutorials, 19(4), 2322-2358*

[6] Shi, W, Cao, J, Zhang, Q. (2016). Edge computing: Vision and challenges[J]. *IEEE Internet of Things Journal, 3(5), 637-646*
[7] Abbas, N, Zhang, Y, Taherkordi, A. (2017). Mobile edge computing: A survey[J]. *IEEE Internet of Things Journal, 5(1), 450-465*
[8] Li Zixuan. (2021). Research on mobile edge computing resource deployment in the Internet of Vehicles environment [D]. *Beijing: Beijing University of Posts and Telecommunications*
[9] Sun Jianan. (2020). Research on vehicle networking task offloading and data distribution technology in edge computing environment [D]. *Beijing: Beijing Jiaotong Universit*
[10] Huy, Hoang V, Ho, T.M. & Le, L.B. (2020). Mobility-Aware Computation Offloading in MEC- Based Vehicular Wireless Networks[J]. *IEEE Communications Letters, 24(2), 466- 469*
[11] Chen, R, Wang, X. (2021). Maximization of value of service for mobile collaborative computing through situation aware task offloading[J]. *IEEE Transactions on Mobile Computing*
[12] Li, J, Cao, X, Guo, D. (2020). Task scheduling with UAV-assisted vehicular cloud for roaddetection in highway scenario[J]. *IEEE Internet of Things Journal, 7(8), 7702- 7713*
[13] Santoyo-González, A., Cervelló-Pastor, C. (2020) Network-aware placement optimization for edge computing infrastructure under 5G[J]. *IEEE Access, (8), 56015-56028*
[14] Yousefpour, A, Patil, A, Ishigaki, G. (2019). FogPlan: a lightweight QoS-aware dynamic fog service provisioning framework[J]. *IEEE Internet of Things Journal, 6(3), 5080- 5096*
[15] Chen, L, Liu, S, Li, B. (2018). Scheduling jobs across geo-distributed datacenters with max- min fairness[J]. *IEEE Transactions on Network Science and Engineering, 6(3), 488- 500*
[16] Wang, Z, Zhao, Z, Min, G. (2018). User mobility aware task assignment for mobile edge computing[J]. *Future Generation Computer Systems, (85), 1-8*
[17] Tao, Y, You, C, Zhang, P. (2019). Stochastic control of computation offloading to a helper with a dynamically loaded CPU[J]. *IEEE Transactions on Wireless Communications*
[18] Tao, M, Nakano, K, Ito, Y. (2020). A work-time optimal parallel exhaustive search algorithm for the QUBO and the Ising model, with GPU implementation[C]. *2020 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), 557-566*

Көлік құралдарының интернеті үшін шеткі есептеу ресурстарын жоспарлауды оңтайландыру әдісі

Динг Фанг*, Цао Юнин
Satbayev University, Автоматтандыру және ақпараттық технологиялар институты, Алматы, Қазақстан
**Корреспонденция үшін автор: Ding.F@stud.satbayev.university*

Аңдатпа. Заттар интернеті мен 5G технологиясының қарқынды дамуымен Көлік құралдарының интернеті 5G технологиясының таңғажайып қолданбалы сценарийлерінің біріне айналды, бұл көлікке орнатылған көптеген интеллектуалды қосымшалардың пайда болуына ықпал етті. Дегенмен, көлік терминалдарында шектеулі есептеу ресурстары бар, бұл нақты уақыттағы қозғалыс жағдайлары және интеллектуалды сәйкестендіру сияқты пайда болатын қолданбалардың кідіріс және энергия тұтыну талаптарын қанағаттандыруды қиындатады. Шеттік есептеулерге негізделген «Internet of Vehicles» технологиясының пайда болуы жоғарыда аталған мәселелерді шешеді. Көлік ішіндегі шеттік есептеулер көлік құралына есептеу және сақтау қызметтерін қамтамасыз ету үшін көлік терминалындағы есептеу және сақтау қысымын жеңілдету үшін көлікке жақын жол бойындағы блоктарға (RSU) шеткі серверлерді орналастырады. Дегенмен, көлік құралының шеткі желісінің шектеулі ресурстарына байланысты шеткі серверге шектеусіз тапсырмаларды жоспарлау сервердің шамадан тыс жүктелуіне әкеледі және көлік пайдаланушыларының қызмет көрсету сапасына (QoS) әсер етеді. Ақылға қонымды ресурстарды жоспарлау стратегиясы шеткі желінің жан-жақты өнімділігін қамтамасыз ете алады және пайдаланушының QoS деңгейін жақсартады. Нақты уақытта өңдеуді қажет ететін тым көп есептеу тапсырмаларын генерациялайтын Көлік құралдары интернетіндегі күрделі жол қозғалысы мен тығыз көліктер мәселесіне бағытталған, бөлінетін тапсырмалар үшін жақсартылған модельденген жасыту алгоритміне негізделген ішінара ресурстарды жоспарлау стратегиясы ұсынылады. Жүйенің үстеме шығындарын азайту мақсатында түсіру коэффициенті коэффициентін қамтитын ішінара жоспарлау жүйесінің моделі құрастырылады. Ресурстарды жоспарлау мәселесінің бір бөлігін екі факторға бөліңіз: түсіру коэффициенті және есептеу ресурстарын бөлу. Оңтайлы түсіру коэффициенті жақсартылған имитациялық жасыту алгоритмі арқылы шешіледі. Алгоритмді итерациялау процесінде ресурстарды оңтайлы бөлу Лагранж мультипликаторы әдісі арқылы алынады. Модельдеу ұсынылған стратегияның конвергенциясын және оның кешіктіруді, энергияны тұтынуды және үстеме шығындарды оңтайландырудағы тиімділігін тексерді.

Негізгі сөздер: *көлік желісі, интеллектуалды кабина, жиектерді есептеу, имитациялық жасыту алгоритмі.*

Метод оптимизации планирования ресурсов периферийных вычислений для Интернета транспортных средств

Динг Фанг*, Цао Юнин

Satbayev University, Институт автоматики и информационных технологий, Алматы, Казахстан
**Автор для корреспонденции: Ding.F@stud.satbayev.university*

Аннотация. Благодаря быстрому развитию Интернета вещей и технологии 5G Интернет транспортных средств стал одним из знаковых сценариев применения технологии 5G, способствуя появлению большого количества новых интеллектуальных приложений, устанавливаемых на транспортных средствах. Однако автомобильные терминалы имеют ограниченные вычислительные ресурсы, что затрудняет удовлетворение требований к задержке и энергопотреблению новых приложений, таких как условия дорожного движения в реальном времени и интеллектуальная идентификация. Появление технологии Интернета транспортных средств на основе периферийных вычислений решает вышеуказанные проблемы. Периферийные вычисления в автомобиле развертывают периферийные серверы в придорожных устройствах (RSU), расположенных рядом с транспортным средством, для предоставления транспортных услуг и услуг хранения данных, снижая нагрузку на вычислительные ресурсы и хранилище данных на автомобильном терминале. Однако из-за ограниченности ресурсов пограничной сети транспортного средства неограниченное планирование задач на пограничном сервере приведет к перегрузке сервера и ухудшит качество обслуживания (QoS) пользователей транспортных средств. Разумная стратегия планирования ресурсов может обеспечить комплексную производительность периферийной сети и улучшить качество обслуживания пользователей. Стремясь решить проблему сложного дорожного движения и большого количества транспортных средств в Интернете транспортных средств, которые генерируют слишком много вычислительных задач, которые необходимо обрабатывать в реальном времени, для делимых задач предлагается стратегия частичного планирования ресурсов, основанная на улучшенном алгоритме имитации отжига. С целью минимизации накладных расходов системы создается частичная модель системы планирования, включающая коэффициент коэффициента разгрузки. Разделите часть проблемы планирования ресурсов на два фактора: коэффициент разгрузки и распределение вычислительных ресурсов. Оптимальный коэффициент разгрузки определяется с помощью улучшенного алгоритма моделирования отжига. В процессе итерации алгоритма оптимальное распределение ресурсов достигается с помощью метода множителей Лагранжа. Моделирование подтвердило сходимость предложенной стратегии и ее эффективность в оптимизации задержки, энергопотребления и накладных расходов.

Ключевые слова: *автомобильная сеть, интеллектуальная кабина, периферийные вычисления, алгоритм моделирования отжига*.

Received: 13 February 2023
Accepted: 15 June 2023
Available online: 30 June 2023