Energy-Efficient Task Offloading for Vehicular Edge Computing: Joint Optimization of Offloading and Bit Allocation

Youngsu Jang, Jinyeop Na, Seongah Jeong† and Joonhyuk Kang
Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea
Email: {jangyoung30, wlsduq37}@kaist.ac.kr, jhkang@ee.kaist.ac.kr
†Kyungpook National University, Daegu, Korea
Email: seongah@knu.ac.kr

Abstract—With the rapid development of vehicular networks, various applications that require high computation resources have emerged. To efficiently execute these applications, vehicular edge computing (VEC) can be employed. VEC offloads the computation tasks to the VEC node, i.e., the road side unit (RSU), which improves vehicular service and reduces energy consumption of the vehicle. However, communication environment is time-varying due to the movement of the vehicle, so that finding the optimal offloading parameters is still an open problem. Therefore, it is necessary to investigate an optimal offloading strategy for effective energy savings in energy-limited vehicles. In this paper, we consider the changes of communication environment due to various speeds of vehicles, which are not considered in previous studies. Then, we jointly optimize the offloading proportion and uplink/computation/downlink bit allocation of multiple vehicles, for the purpose of minimizing the total energy consumption of the vehicles under the delay constraint. Numerical results demonstrate that the proposed energy-efficient offloading strategy significantly reduces the total energy consumption.

Index Terms—Vehicular edge computing, energy efficiency, task offloading, bit allocation, vehicular networks.

I. INTRODUCTION

The emergence of the Internet of things (IoT) revolutionizes the computing era. IoT means a large network of interconnected devices, and IoT can be useful for various applications such as health care, aviation, social networking, transportation, and traffic control [1]. With the ever-increasing number of vehicles, the Internet of vehicles (IoV), which is typical application of IoT in the transportation area, has received much attention lately [2]. Furthermore, various vehicular applications such as autonomous driving, video streaming, speech recognition, and in-vehicle entertainment are expected to be implemented in vehicles [3]. However, these applications require not only strict delay constraints, but also enormous computation resources to process large volumes of workload data. Since existing cloud servers are located far from the vehicle, offloading tasks to the cloud server causes latency problems due to the capacity-constrained backhaul links [4]. To tackle this problem, vehicular edge computing (VEC), which provides the service at the network edge near the vehicle, has been proposed as a solution [5].

VEC is expected to support computation sensitive and delay sensitive applications effectively by offloading workloads to the VEC nodes that are close to the vehicle. Traditionally, all applications need to be processed locally on the vehicle, which causes enormous energy consumption, not desirable for energy-limited vehicles such as electric cars. In VEC, the vehicle can offload the tasks to adjacent VEC node, which has higher computational capability and abundant energy supply [2]. Thus, VEC can reduce the energy consumption of vehicle and increases the battery endurance time. However, task offloading does not always guarantee low energy consumption due to the energy consumption for communication. Therefore, it is necessary to study the energy-efficient offloading strategies.

Recently, there are some works that investigate offloading strategies in VEC. In [2], an offloading scheme is designed to enhance the transmission efficiency under the considerations of the time consumption of the task execution and the mobility of the vehicles. The authors in [4] develop a mobile edge computing (MEC) offloading framework to support various applications in smart city scenarios. Also, a VEC framework named autonomous vehicular edge (AVE) is proposed in [6] to increase the computational capabilities of vehicles in a decentralized manner. However, these works mainly solve the workload offloading problem from the perspective of minimizing latency, and have not considered the energy saving problems. Although, there are a few studies that consider the energy-efficient task offloading. An energy-efficient VEC framework is proposed in [5] to support in-vehicle user equipments (UEs) which have limited battery capacity. In [7], an energy-efficient workload offloading problem is studied, and a low-complexity distributed solution based on consensus alternating direction method of multipliers (ADMM) is proposed. The authors in [8] investigate the average transmission power minimization problem under the tasks quality of service (QoS) requirement. Nevertheless, [5] and [7] are mainly target on in-vehicle UEs, and [8] have not considered the multiple vehicle case.

In this paper, we consider the VEC system that offloads workload to nearby road side unit (RSU) with computation server, when multiple vehicles arrive at different times on a unidirectional road with multiple lanes. The problem that min-
imizes the total energy consumption of vehicles is formulated as a joint bit allocation and offloading proportion problem. The main contributions of this paper are as follows:

- We formulate the energy-efficient task offloading problem to minimize the total energy consumption of all vehicles, under the consideration of the task deadline, computation and communication energy consumption, multiple access, and time-varying channel state. In addition, we further consider the offloading across multiple RSUs, various velocities of vehicles, and multiple lanes, that have been ignored in previous works [3], [7], and [8].
- We deal with the optimization problems for two different cases in the offloading process. One is the complete offloading case where the vehicles offload all their tasks. The other one is the partial offloading case where the vehicles offload a fraction of their tasks. Also, we provide the optimal solution for both cases. Simulation results demonstrate the effectiveness of the proposed offloading strategy.

The rest of the paper is organized as follows. First, the system model including the computation energy model and the communication energy model are introduced in Section II. Section III describes the formulation of optimization problem. The simulation results and analysis are depicted in Section IV. Finally, we conclude the paper in Section V.

II. SYSTEM MODEL

In this paper, we consider a VEC system that includes $K$ vehicles and $M$ RSUs as shown in Fig. 1. The RSUs are deployed along the unidirectional road with $J$ lanes, the distance between neighboring RSUs is $d$, and the coverage radius of each RSU is $r_{RSU}$. Define $\mathcal{M} = \{1, \ldots, M\}$ as the set of RSUs. The location of RSU $m$ is calculated as

$$p_m^r = (r_{RSU} + (m - 1)d, 0), \quad m \in \mathcal{M}. \quad (1)$$

We assume that $K$ vehicles arrive at the first RSU’s coverage edge in different time $t_k \in \{1, \ldots, T\}$. The set of vehicles is defined as $K = \{1, \ldots, K\}$. By further assuming that the vehicles in the same lane have the same velocity, the velocity of the vehicles in each lane is $v_j \in \{v_1, \ldots, v_J\}$.

The offloading process is described in the following steps. First, vehicle $k$ transmits the input data to the closest RSU, i.e., uplink transmission. Next, the RSU computes the received data, and finally, the RSU transmits the output data to vehicle $k$, i.e., downlink transmission. We assume frequency division duplex (FDD) that equal bandwidth $B$ is allocated for uplink and downlink. Thus, there is no interference between uplink and downlink communication. The time horizon $T$ is equally divided into $N$ frames, and each frame duration is $\Delta$, i.e., $T = N\Delta$. The frame duration $\Delta$ is sufficiently small so that the vehicle’s location is approximately constant within each frame. Thus, the position of the vehicle in the $j$th lane at the $n$th frame is expressed as

$$p_n^v = (n\Delta v_j, (j - 1)d_{lane}), \quad j = 1, \ldots, J \quad \text{and} \quad n \in \mathcal{N} = \{1, \ldots, N\}, \quad (2)$$

where $d_{lane}$ denotes the lane width. For the access scheme, we consider an orthogonal access that each vehicle has one time slot with duration $\delta = \Delta/K$ and communicates within its time slot.

For the sake of simplicity, we consider the communication channels between vehicles and RSUs which are dominated by line-of-sight (LoS) link [10]. Doppler effect due to the vehicle’s movement is assumed to be perfectly compensated by the receivers. In each frame, the vehicle communicates with the nearest RSU. Thus, the channel gain between the vehicle $k$ and the adjacent RSU at the $n$th frame is given by

$$h_{k,n}(p_n^v) = \frac{h_0}{\|p_n^v - p_{m_{\min}}^r\|^2 + H^2}, \quad (3)$$

where $\|\cdot\|$ means the norm 2 function, $m_{\min} \in M$ represents the index of the closest RSU, $H$ denotes the height of the RSU, and $h_0$ is the received power at the reference distance $d = 1m$ for a transmission power of 1W. We assume that the channel noise is additive white Gaussian with zero mean and power spectral density $N_0$ [dBm/Hz].

The task of the vehicle $k \in K$ is described by the number $L_k$ of input bits, the number $C_k$ of CPU cycles per input bit for computation, and the number $\kappa_k$ of output bits produced by computation per input bit. In most cases, the size of input data is much larger than the output data, i.e., $\kappa_k < 1$. All tasks must be computed within the deadline $T$.

A. Computation energy model

When the CPU is operated at frequency $f^i$, the computation energy consumption to execute the application of vehicle $k$ with $l$ input bits is obtained by [11], [12]

$$E_k^i(l, f^i) = \gamma_i C_k l (f^i)^2, \quad (4)$$

where $i = v$ for the vehicle and $i = r$ for the RSU, respectively. Here, $f^i$ [CPU cycles/s] is the operating frequency of each processor, and $\gamma_i$ denotes the effective switched capacitance of each processor which is related to the chip architecture.

B. Communication energy model

When vehicle $k$ transmits $L_{k,n}^v$ bits during the slot duration $\delta$, the following equation is obtained by Shannon theory.

$$B\delta \log_2 \left(1 + \frac{E_k^v l_{k,n}^v}{N_0 B\delta} h_{k,n}(p_n^v) \right) = L_{k,n}^v, \quad (5)$$
where \( q = u \) for uplink while \( q = d \) for downlink. From (5), the communication energy consumption of the vehicle \( k \) at time slot \( n \) is calculated as

\[
E_{k,n}^u(L_{k,n}^u, P_k^u) = \frac{N_A B \delta}{C_k} \left( \frac{L_{k,n}^u}{T} - 1 \right)^2.
\]

From (6), we can see that the communication energy consumption is related to the number of transmission bits and the channel condition affected by the communication distance.

### III. Minimize Total Energy Consumption of Vehicles

In this section, we study the problem to minimize the total energy consumption of \( K \) vehicles. First, the energy consumption when locally computing in the vehicle is obtained as a benchmark for comparison. Then, without considering the offloading proportion, the optimal bit allocation for uplink, computation, and downlink is found under the maximum delay constraint of \( T \). Finally, the joint optimization of bit allocation and offloading ratio is introduced.

#### A. Energy consumption for local execution

For reference, we consider the total energy consumption of overall vehicles when all applications are processed locally. To process \( L_k \) bits within \( T \) seconds, the CPU frequency of vehicle \( k \) needs to be selected as

\[
f_k = \frac{C_k L_k}{T}.
\]

From (4), the total energy consumption for local execution is obtained by

\[
\sum_{k=1}^{K} E_{k,n}^u(L_k) = \sum_{k=1}^{K} \frac{\gamma_k C_k^3}{C_k} L_k^2,
\]

where \( \gamma_k \) is the effective switched capacitance of the vehicle \( k \)'s processor.

#### B. Optimization for complete offloading

In this section, we study the optimal bit allocation that minimize the total energy consumption of all vehicles. It is assumed that all input data for each vehicle is offloaded to the RSU so that the data is not processed locally in the vehicle. At the \( n \)th frame, we denote \( L_{k,n}^u \) as the number of uplink bits transmitted from the vehicle \( k \) to the RSU, \( L_{k,n}^c \) as the number of bits computed for the task of the vehicle \( k \) at the RSU, and \( L_{k,n}^d \) as the number of downlink bits transmitted from the RSU to the vehicle \( k \). Also, we define \( \hat{N} = \{1, \ldots, N-2\} \).

We analyze the offloading process in frame-by-frame manner. This method is commonly used to handle the offloading process [13]. For example, vehicle \( k \) transmits \( L_{k,1}^u \) bits to the RSU at the first frame and there is no computation and no downlink transmission, i.e., \( L_{k,1}^c = 0 \). At the second frame, vehicle \( k \) transmits \( L_{k,2}^u \) bits to the RSU, the RSU computes \( L_{k,2}^c \) bits, and there is no downlink transmission, i.e., \( L_{k,2}^d = 0 \). At the third frame, vehicle \( k \) transmits \( L_{k,3}^u \) bits to the RSU, the RSU computes \( L_{k,3}^c \leq L_{k,1}^u + L_{k,2}^u - L_{k,2}^c \) bits, and the RSU transmits \( L_{k,3}^d \leq \kappa_k L_{k,2}^c \) bits to vehicle \( k \). This process is continued until the \( N \)th frame. The optimization problem is formulated as follows:

\[
\begin{align*}
\text{minimize} & \quad \sum_{k=1}^{K} \sum_{n=1}^{N-2} E_{k,n}^u(L_{k,n}^u) \\
\text{s.t.} & \quad \sum_{i=1}^{n} L_{k,i+1}^c \leq \sum_{i=1}^{n} L_{k,i}^c, \quad \text{for } k \in K \text{ and } n \in \hat{N}, \quad (9a) \\
& \quad \sum_{i=1}^{n} L_{k,i+2}^c \leq \kappa_k \sum_{i=1}^{n} L_{k,i+1}^c, \quad \text{for } k \in K \text{ and } n \in \hat{N}, \quad (9b) \\
& \quad L_{k,n}^u \leq L_{k,n}^u, \quad \text{for } k \in K \text{ and } n \in \hat{N}, \quad (9c) \\
& \quad L_{k,n}^d \leq L_{k,n}^d, \quad \text{for } k \in K \text{ and } n \in \hat{N}, \quad (9d) \\
& \quad \sum_{k \in A} L_{k,n}^u \leq L_{k,n}^u, \quad \text{for } k \in A \text{ and } n \in \hat{N}, \quad (9e) \\
& \quad L_{k,n}^u, L_{k,n}^c, L_{k,n}^d \geq 0, \quad \text{for } k \in K \text{ and } n \in \hat{N}, \quad (9f) \\
& \quad \sum_{n=1}^{N-2} L_{k,n}^u = L_k, \quad \text{for } k \in K, \quad (9g) \\
& \quad \sum_{n=1}^{N-2} L_{k,n}^c = L_k, \quad \text{for } k \in K, \quad (9h) \\
& \quad \sum_{n=1}^{N-2} L_{k,n}^d = \kappa_k L_k, \quad \text{for } k \in K, \quad (9i) \\
& \quad \sum_{n=1}^{N-2} L_{k,n}^d \leq \kappa_k L_k, \quad \text{for } k \in K, \quad (9j)
\end{align*}
\]

where \( L_{\text{max}}^u \) and \( L_{\text{max}}^d \) represent the maximum number of uplink and downlink bits that can be transmitted within slot duration, \( A \) is the set of vehicles that communicate with the same RSU at \( n \)th frame, and \( L_{\text{max}} \) is the limitation of total uplink bits transmitted to the same RSU at each frame.

In problem (9), the constraint (9b) guarantees that the number of bits processed at the \( (n+1) \)th frame by the RSU is no larger than the number of uplink bits transmitted from the vehicle in the previous \( n \) frames. In a similar manner, the constraint (9c) guarantees that the number of downlink bits transmitted from the RSU at the \( (n+2) \)th frame is no larger than the number of output bits processed in the previous \( (n+1) \) frames. Next, (9d) and (9e) are the constraints of uplink and downlink transmission bits at each frame. Also, (9f) is the constraint of the number of uplink bits received by one RSU at each frame, which is mainly due to the computation capability of the RSU. The constraint (9g) enforces non-negative bit allocations. Finally, the equality constraints (9b)–(9j) ensure that the given input bits are completely processed via offloading within the deadline \( T \).

The problem (9) is convex, since the objective function (9a) is the sum of convex exponential functions and the constraints (9b)–(9j) are linear. Thus, we can solve this problem using standard convex optimization solver such as CVX MOSEK [14].

#### C. Optimization for partial offloading

In order to obtain the additional energy savings, we consider not only the bit allocation, but also the offloading proportion.
Assume that vehicle $k$ offloads the ratio $\rho_k$ of the input bits to the VEC node and locally computes the remaining portion of the input bits. In this case, the energy consumption of the vehicle is the sum of the communication energy consumed in offloading process and computation energy consumed in local execution. Thus, the joint optimization problem of bit allocation and offloading proportion is as follows:

$$\text{minimize} \quad \sum_{k=1}^{K} \sum_{n=1}^{N-2} E_k^n (\rho_k L_k)$$

subject to

$$\sum_{n=1}^{N-2} L_k^n = \rho_k L_k, \quad \text{for } k \in K,$$

$$\sum_{n=1}^{N-2} L_k^n = \rho_k L_k, \quad \text{for } k \in K,$$

$$0 \leq \rho_k \leq 1, \quad \text{for } k \in K,$$

where the equality constraints (10b)-(10d) ensure the completion of total offloaded data, and the inequality constraint (10e) denotes the offloading ratio constraint.

The objective function (10a) is the sum of convex exponential functions and convex power functions, and the constraints (10b)-(10f) are linear. Hence, the problem (10) is convex, which can be solved by CVX MOSEK.

### IV. Numerical Results

In this section, we provide simulation results to evaluate the performance of our proposed offloading design. In the simulations, we consider a system with the number of vehicles $K = 10$ and the number of lanes $J = 3$. The number of RSUs $M$ is determined by $M = \lceil (v_{\text{max}} T - r_{\text{RSU}})/d + 1/2 \rceil$, where $v_{\text{max}}$ is the velocity of the fastest vehicle and $\lceil \rceil$ means the ceiling function. We assume that all vehicles have the same deadline $T$ and the vehicles are arrived at the first RSU’s coverage edge in randomly. The remaining parameters are summarized in Table I. We consider the following two schemes for reference. First, local execution scheme that all applications are processed locally. Second, we consider the proposed scheme.

Fig. 2 represents the optimal uplink bit allocation at each frame. The red, blue, and black lines mean the vehicles in lane 1, 2, and 3, respectively. We set the deadline $T = 20s$, but the proposed scheme can be applied even if the deadline is shorter. As shown in Fig. 2, we can observe that the vehicles need to offload more data when they are closer to the RSU, to reduce the communication energy consumption. Additionally, when two or more vehicles communicate with the same RSU, some vehicles cannot offload the maximum number of bits due to the limited computation capability of the RSU, as described in constraint (9f). In this case, the vehicle in the better channel condition transmits more bits. If the channel conditions of the vehicles are similar, the vehicle which has more total input bits, i.e., $\rho_k L_k$, offloads more bits.

In Fig. 3, we compare the total energy consumption of each scheme as a function of the deadline $T$. From Fig. 3, we can see that the energy savings of proposed offloading scheme becomes more effective compared to the benchmarks if the deadline $T$ becomes stricter. Also, we can observe that the total energy consumption of the equal bit allocation scheme is larger than the local execution scheme. This means that task offloading does not always guarantee low energy consumption due to the communication energy. In other words,

| Parameter | Definition | Value |
|-----------|------------|-------|
| $r_{\text{RSU}}$ | coverage radius of RSU | 100m |
| $d$ | distance between neighboring RSUs | 200m |
| $\delta_{\text{lane}}$ | lane width | 3.5m |
| $H$ | height of RSU | 10m |
| $B$ | bandwidth | 40MHz |
| $L_k$ | number of input bits | 10 - 25Mbits |
| $L_{\text{up max}}$ | uplink constraint | 180Kbits |
| $L_{\text{down max}}$ | downlink constraint | 140Kbits |
| $L_{\text{access max}}$ | multiple access constraint | 250Kbits |
| $C_k$ | number of CPU cycles per bit | 1550.7 |
| $\gamma_k^r$, $\gamma_k^d$ | switch capacitance constant | $10^{-28}$ |
| $N_0$ | noise spectrum density | 0.5 |
| $h_0/N_0 B$ | reference SNR | 20dB |
| $v_j$ | velocity of vehicle | 20 - 30m/s |
| $\Delta$ | frame duration | 4ms |
| $\tau_k$ | arrival time of vehicle | 0 - 20s |
the vehicles continue to transmit the input bits to the RSU when they are located in far from the RSU, so it causes a case where they consume more energy than when they process locally. Furthermore, it is seen that the proposed optimal bit allocation significantly reduces the total energy consumption, and the joint optimization of bit allocation and offloading ratio achieves additional energy savings.

Fig. 4 shows the total energy consumption of all vehicles as a function of offloading proportion. In this simulation, we assume that all vehicles offload their task with equal offloading proportion $\rho$. In Fig. 4, we can observe that the local execution, i.e., $\rho = 0$, causes enormous energy consumption and offloading about 70% of the task is more effective than offloading all the task, i.e., $\rho = 1$. The reason for this is that if the vehicle moves beyond a certain distance away from the RSU, local execution consumes less energy than when data offloading is applied. Thus, in order to minimize the energy consumption, some portion of the task needs to be processed locally.

V. Conclusion

In this paper, we studied the energy-efficient task offloading strategy with the purpose of minimizing the total energy consumption of overall vehicles. First, we investigated the problem of finding the optimal uplink/computation/downlink bit allocation of multiple vehicles, when the vehicles offload all the input bits without considering the offloading proportion. We then formulated the joint optimization problem involving the offloading proportion of multiple vehicles. Since both optimization problems are convex, optimal bit allocation and offloading proportion are obtained by using CVX MOSEK. Numerical results demonstrate that the proposed joint optimization of bit allocation and offloading ratio achieves significant energy savings compared to the benchmarks.

REFERENCES

[1] M. Sookhak et al., Fog vehicular computing: Augmentation of fog computing using vehicular cloud computing, IEEE Veh. Technol. Mag., vol. 12, no. 3, pp. 5564, Sep. 2017.

[2] K. Zhang, Y. Mao, S. Leng, Y. He, and Y. Zhang, Mobile-edge computing for vehicular networks: A promising network paradigm with predictive off-loading, IEEE Veh. Technol. Mag., vol. 12, no. 2, pp. 3644, Jun. 2017.

[3] Z. Zhou, P. Liu, Z. Chang, C. Xu, and Y. Zhang, Energy-efficient workload offloading and power control in vehicular edge computing, in Proc. IEEE Wireless Commun. Netw. Conf. Workshops, Barcelona, Spain, Apr. 2018, pp. 191196.

[4] T. Taleb, S. Dutta, A. Ksentini, M. Iqbal, and H. Flinck, Mobile edge computing potential in making cities smarter, IEEE Commun. Mag, vol. 55, no. 3, pp. 3843, Mar. 2017.

[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. 38603873, Jun. 2016.

[6] J. Feng, Z. Liu, C. Wu, and Y. Ji, AVE: Autonomous vehicular edge computing framework with ACO-based scheduling, IEEE Trans. Veh. Technol., vol. 66, no. 12, pp. 1066010675, Dec. 2017.

[7] Z. Zhou, J. Feng, Z. Chang, and X. Shen, "Energy-Efficient Edge Computing Service Provisioning for Vehicular Networks: A Consensus ADMM Approach," IEEE Trans. Veh. Technol., vol. 68, no. 5, pp. 5087-5099, May. 2019.

[8] S. Li, G. Zhu, and S. Lin, “Joint Radio and Computation Resource Allocation with Predictable Channel in Vehicular Edge Computing,” in Proc. IEEE Int. Conf. Intell. Transp. Syst., Maui, HI, USA, Nov. 2018, pp. 3736-3741.

[9] L. Alvarez, R. Horowitz, and P. Li, Traffic flow control in automated highway systems, Control Eng. Practice, vol. 7, no. 9, pp. 10711078, Sep. 1999.

[10] H. Wang, X. Li, H. Ji, and H. Zhang, Federated offloading scheme to minimize latency in MEC-enabled vehicular networks, in Proc. IEEE GLOBECOM Workshops, Abu Dhabi, UAE, Dec. 2018, pp. 1-6.

[11] W. H. Yuan and K. Nahrstedt, Energy-efficient soft real-time CPU scheduling for mobile multimedia systems, ACM SIGOPS Oper. Syst. Rev., vol. 37, no. 5, pp. 149163, Dec. 2003.

[12] W. H. Yuan and K. Nahrstedt, Energy-efficient CPU scheduling for multimedia applications, ACM Trans. Comput. Syst., vol. 24, no. 3, pp. 20331, Aug. 2006.

[13] S. Jeong, O. Simeone, and J. Kang, Mobile edge computing via a UAV-mounted cloudlet: Optimization of bit allocation and path planning, IEEE Trans. Veh. Technol., vol. 67, no. 3, pp. 20492063, Mar. 2018.

[14] M. Grant and S. Boyd, "Cvx: Matlab software for disciplined convex programming, version 2.1," Mar. 2014, Available on-line at http://cvxr.com/cvx

[15] M. Hua et al., Optimal resource partitioning and bit allocation for UAV-enabled mobile edge computing, in Proc. IEEE Veh. Technol. Conf., Chicago, USA, Aug. 2018, pp. 1-6.