RIS-assisted MEC Computation Offloading for IoVT Networks

1st Daniar Estu Widiyanti
Dept. of IT Convergence Engineering
Kumoh National Institute of Technology
Gumi, South Korea
daniarestuw@kumoh.ac.kr

2nd Soo Young Shin
Dept. of IT Convergence Engineering
Kumoh National Institute of Technology
Gumi, South Korea
wdragon@kumoh.ac.kr

***Abstract*—The Internet of Things (IoT) encompasses a wide range of applications; despite that, some challenges persist and require further research and development. One such challenge is encountered in sensing devices, particularly in Internet of Video Things (IoVT) applications that demand high-capacity wireless transmission. These applications face limitations in computational capability and energy consumption of users, posing significant obstacles to overcome. Mobile edge computing (MEC) has gained attention as it offers potential solutions by offloading complex tasks to assist users. However, the line-of-sight (LoS) issue between users and base stations (BS) caused by blocking objects remains a challenge in MEC. To address latency challenges in IoVT communication affected by obstacles, a collaborative framework between reconfigurable intelligence surfaces (RIS) and task offloading techniques in MEC, specifically in IoVT networks, is proposed. This collaboration aims to optimize the performance and efficiency of IoVT systems by harnessing the combined advantages of RIS and task offloading strategies.**

***Index Terms*—internet of video things (IoVT), mobile edge computing (MEC), reconfigurable intelligence surfaces (RIS), task offloading**

I. INTRODUCTION

The Internet of Things (IoT) involves interconnection and communication between devices, sensors, and objects (Things). As IoT covers a wide range of applications, some difficulties require research efforts and further development. One of the difficulties occurs in sensing devices such as multimedia sensors [1], audio-visual cameras [2], and 3D cameras in the Internet of Video Things (IoVT) applications. In video applications, video encoding and processing are complex and computationally intense processes that are often unstructured and need additional processing [3]. Moreover, the size of this multimedia data requires extremely high networking capacity [4], which might cause high latency.

Additional challenges happen when there is an obstacle between the sender and receiver in a wireless communication system. High-frequency transmissions are particularly susceptible to obstacles [5], leading to decreased signal quality and lower transmission rates. This obstacle-induced degradation affects not only the overall performance of the communication link but also impacts users that are positioned in non-line-of-sight (nLoS) locations relative to the base station (BS). Consequently, the presence of obstacles introduces increased latency and can have a detrimental effect on the quality and reliability of the wireless communication system.

To address the latency challenges in wireless communication scenarios with obstacles, reconfigurable intelligence surfaces (RIS) are employed alongside base stations (BS). As RIS has the capability to enhance the signal-to-noise ratio (SNR), thereby impacting the overall capacity of the system. The contribution of this paper can be described as follows:

- Task Offloading scheme in a RIS-assisted edge computing system in IoVT networks. Some end-users may experience poor signal conditions due to nLoS conditions. In contrast to other studies, this work takes a more realistic approach by considering downlink (DL) transmission. In this work, RIS are deployed to improve both uplink (UL) and DL rates for minimizing the latency that occurs when user need to offload their tasks.
- Optimizing the task placement for the end-user based on the end-user task completion time and energy consumption, which depends on the upload time, computation latency, and result transmission.

The main focus of this study is to establish a collaborative framework between RIS and task offloading techniques in mobile edge computing (MEC), with a specific emphasis on Internet of Video Things (IoVT) networks. This collaboration aims to optimize the performance and efficiency of IoVT systems by leveraging the combined benefits of RIS and task offloading strategies.

II. PROPOSED SYSTEM

A. System Model

The proposed RIS-assisted edge computing system can be seen in Fig. 1. The set of user in user clusters is denoted as $k \in \{1, \cdots, K\}$, where K is the total number of users. Due to each users CPU limited computational capacity, the user could offload the data to the computational node in the base station (BS).

Consider the index of a task from total task L is denoted as $i \in [1, \cdots, L]$. The task offloading placement decision variable for user as $x_{i,k}$ which can be divided as $x_{i,k}$ = 1, *task offloaded to BS*; $x_{i,k}$ = 0, *task processed locally*.

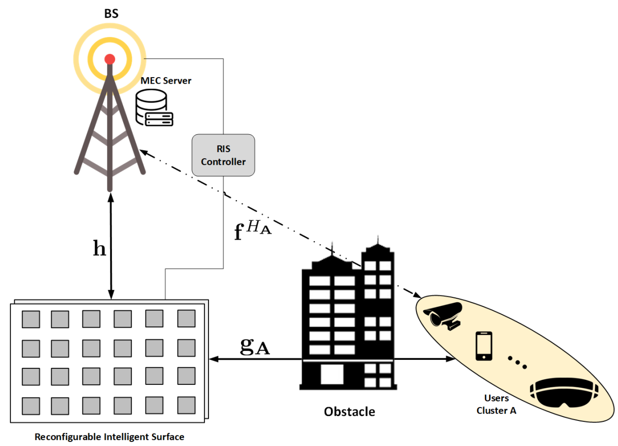

Fig. 1. Proposed system model.

B. Communication Model

The RIS with finite elements N is deployed to assist the users with UL and DL. Let $\mathbf{s} = [s_1, \cdots s_K]^T \in \mathbb{C}^K$ is the modulated symbol of offloaded task of users k. Therefore, the transmission signal can be written as $s_{UL} = \sum_{j=1}^{K} \boldsymbol{q}_j s_j$ and $s_{DL} = \sum_{j=1}^{K} \boldsymbol{w}_j s_j$.

The offloaded task to BS is denoted as s_{UL}. $\mathbf{q}_j \in \mathbb{C}^{1\times Nt}$ and $\mathbf{w}_j \in \mathbb{C}^{Nt\times 1}$ denote UL and DL precoding, respectively. Hence, the UL signal of the suggested system is indicated as: $y_k^{UL} = (\boldsymbol{f}_k^{\mathrm{H}} + \boldsymbol{h}_{s,k}^{\mathrm{H}} \boldsymbol{\Theta} \boldsymbol{h}_{d,k}) s_i + \boldsymbol{n}$, where the direct channel from BS to the user k can be written as follows: $\boldsymbol{f}_k^{\mathrm{H}} \in \mathbb{C}^{1\times Nt}$; The channel from RIS to the user as follows: $\boldsymbol{h}_{d,k}^{\mathrm{H}} \in \mathbb{C}^{1\times N}$; The optimized RIS phase shifter (in form of diagonal matrix) can be written as: $\boldsymbol{\Theta}^* \in \mathrm{diag}\{\theta \in \mathbb{C}^{1\times N}\}$; channel from BS to RIS can be denoted as: $\boldsymbol{h}_{s,k} \in \mathbb{C}^{N\times Nt}$; and additive white gaussian noise (AWGN) with $\sim \mathcal{CN}\ (0, \sigma_n^2)$ can be denoted as: $\boldsymbol{n} \in \mathbb{C}^{1\times K}$.

In contrast with other works, where only the UL scenario is considered, in this system, UL and DL are both considered to minimize latency. Therefore, the optimal signal-to-interference-plus-noise-ratio (SINR) for the UL transmission can be written as:

$$\gamma_{i,k}^{UL} = q_k \tilde{\boldsymbol{H}}_k^H \left[\boldsymbol{W}_k(\boldsymbol{\Theta}, \boldsymbol{q})\right]^{-1} \tilde{\boldsymbol{H}}_k, \tag{1}$$

where q_k is user power, and $\tilde{\boldsymbol{H}}_k = \tilde{\boldsymbol{f}}_k^{\mathrm{H}} + \tilde{\boldsymbol{g}}_k^{\mathrm{H}} \boldsymbol{P} \Theta \tilde{\boldsymbol{h}}$ is the estimated channel.

The inter-user interference is denoted as $\boldsymbol{W}_k(\boldsymbol{\Theta}, \boldsymbol{q})$ and written as:

$$\boldsymbol{W}_k(\boldsymbol{\Theta}, \boldsymbol{q}) = \sigma_n^2 \boldsymbol{I}_{Nt} + \sum_{j\neq k} q_j \tilde{\boldsymbol{H}}_j \tilde{\boldsymbol{H}}_j^H, \tag{2}$$

where I_{Nt} is identity square matrix with $Nt \times Nt$ size. Furthermore, the SINR for DL transmission can be expressed as:

$$\gamma_k^{DL} = \frac{|\tilde{\boldsymbol{H}}_k^H w_k|^2}{\sum_{j=1, j\neq k}^{K} |\tilde{\boldsymbol{H}}_j^H w_k|^2 + \sigma_n^2}. \tag{3}$$

Finally, data rates for both UL can be calculated as follows; and

$$r_{i,k}^{UL} = log_2(1 + \gamma_{i,k}^{UL}). \tag{4}$$

While the DL can be calculated as follows;

$$r_{i,k}^{DL} = \gamma_k^{DL}. \tag{5}$$

C. Computation Model

When the user compute the task locally, the computation latency for local processing is denoted as:

$$t_{i,k}^{L} = \frac{\omega_i}{f_k^l}. \tag{6}$$

where ω_i is the computation amount of the i^{th} task and f_k^l is the CPU computing capabilities of k^{th} user. If the task is offloaded to the edge server through BS, the computation latency is denoted as:

$$t_{i,k}^{E} = \underbrace{\frac{D_i}{r_{i,k}^{UL}}}_{SendingData} + \underbrace{\frac{\omega_i}{f_k^c}}_{Computation} + \underbrace{\frac{D_i'}{r_{i,k}^{DL}}}_{RecievingData}. \tag{7}$$

where D_i is the data size of each task, while D_i' is the data size of the task after computation, f_k^c is the CPU computing capabilities in edge server and $r_{i,k}^{UL}$, $r_{i,k}^{DL}$ is the data rate for UL and DL transmission.

D. Energy Consumption Model

The energy consumption for offloading cases can be expressed as follows:

$$\varepsilon_{i,k}^{E} = \frac{p_i s_i}{r_{i,k}^{UL}}. \tag{8}$$

Where ϱ_i is the power coefficient of energy consumed for each CPU cycle, $r_{i,k}$ is the uplink rate from the user to offload to the MISO BS. Furthermore, energy consumption for local task processing can be expressed as follows:

$$\varepsilon_{i,k}^{L} = \varrho_i \omega_i. \tag{9}$$

III. Problem Formulation

To minimize the time latency of the proposed system, the placement of the task can be formulated as follows:

$$\min_{\boldsymbol{x}} \sum_{k=1}^{K} \left[x_{i,k} t_{i,k}^{L} + (1 - x_{i,k}) t_{i,k}^{E}\right] \tag{10a}$$

$$\text{s.t: } \mathrm{C}_1 : \ x_{i,k} \in \{0, 1\}. \tag{10b}$$

$$\mathrm{C}_2 : \ r_{i,k} \geq r_{min}, \tag{10c}$$

$$\mathrm{C}_3 : \ \varepsilon_{i,k} \geq \varepsilon_{min}, \tag{10d}$$

$$\mathrm{C}_4 : \ x_{i,k} \varepsilon_{i,k}^{L} \leq \alpha_{i,k} E_{max}^{i,k}, \tag{10e}$$

$$\mathrm{C}_5 : \ (1 - x_{i,k}) \varepsilon_{i,k}^{E} \leq \alpha_{i,k} E_{max}^{i,k},. \tag{10f}$$

$$\mathrm{C}_6 : \ \sum_{i=1}^{I} \omega_i s_i \leq \omega_{MEC}^{Max}, \tag{10g}$$

$$\mathrm{C}_7 : \ |\boldsymbol{\Theta}| \leq 1. \tag{10h}$$

The main objective of the formulation described in 10a is to minimize the total duration of the task by optimizing the placement of the tasks. This placement is determined by the offloading decision, as indicated in 10b. The offloading process is subject to a minimum rate and energy requirement, as expressed in 10c and 10d. Additionally, there are energy constraints for local computation 10e and offloading 10f, as well as a computation constraint for the base station 10g, and RIS phase shifter optimization in 10h.

This formulated problem requires an appropriate offloading design x_i to minimize system latency. In addition, precoding matrix UL and DL can not exceed P_{user} and P_{BS}^{max}, respectively. However, the problem in Eq. 10a is considered NP-hard and non-convex. Therefore, first, the problem can be decompose as:

$$\min_{\mathbf{\Theta}} \sum_{k=1}^{K} \frac{s_{i,k}}{r_{i,k}^{UL}(\mathbf{\Theta})} \tag{11}$$
$$\text{s.t. } r_{i,k}^{UL}(\mathbf{\Theta}) \geq r_k, \forall k \in \mathcal{K} \tag{12}$$
$$|\mathbf{\Theta}| \leq 1, \forall k \in \mathcal{N}. \tag{13}$$

Where r_k is the required data rate. However, the problem is still non-convex and hard to solve. Therefore, a Block Coordinate Descent (BCD) method is proposed to alternate the problem. The previous problem can be equivalently solved as:

$$\min_{\boldsymbol{\theta},\boldsymbol{\alpha}} \alpha \tag{14a}$$
$$\text{s.t. } \theta^H \mathbf{Q}_k \theta + 2Re\left\{\mathbf{q}_k^H \theta\right\} + d_k \leq \alpha, \tag{14b}$$
$$|\theta| \leq 1. \tag{14c}$$

Where α denotes an auxiliary variable to alternate problem Eq. 14. Notice that Q_k is semi-definite positive, then all the constraints are guaranteed convex. Furthermore, it can be easily solved with solver programming, i.e., CVX program.

For the DL suppose the set of precoding vector of K UEs is denoted as $\boldsymbol{W} = [w_1^T, \cdots, w_K^T]^T$ and the RIS phase is $\mathbf{\Theta} = \text{diag}\{e^{j\theta_1} \cdots e^{j\theta_N}\}$. The problem can be formulated as:

$$\max_{\boldsymbol{W}, \mathbf{\Theta}} \quad \eta = \frac{R}{P} \tag{15a}$$
$$\text{s.t. } \mathrm{C}_1: \ \xi \sum_{k=1}^{K} \|\boldsymbol{w}_k\|^2 + W_{\mathrm{BS}} \leq P_{\mathrm{BS}}^{\max}, \tag{15b}$$
$$\mathrm{C}_2: \ |\theta_n| \leq 1, \ \forall n \in [N], \tag{15c}$$

Where C_1 and C_2 are the maximum power consumption of BS and passive RIS. C_3 denotes the feasible sets of N RIS phase shifter and no amplification factor. However, solving the problem (15a) is very hard due to its non-convexity. Therefore, the problem is reformulated by utilizing fractional programming [6].

TABLE I
SIMULATION PARAMETER

Symbol	Value
Area	500 x 500 m
No. of User	5
No. of BS	1
Data size of the visual task	[1,7] Mbits
Each user computing capacity	[1.5, 2.5] Mb/s
Edge server capacity	[7, 10] Gb/s
Bandwidth	200 KHz
BS Power	40 dBm
User Power	0.25 dBm

IV. RESULTS

Simulation setups are divided into full offloading with RIS and without RIS, optimized task placement with RIS and without RIS, and local computation. Each parameter can be seen in Table I. The coordinate location (x, y) of BS and RIS are $(0, 20)$ and $(100, 5)$, respectively (in meters). The users location follows uniformly distributed pseudo-random integers ranging from 0 to 500.

The simulation result can be seen in Fig. 2. According to the results, local computation achieves a lower computation time result due to no required communication transmission. Full offloading with RIS and optimized placement with RIS has a lower duration compared to full offloading without RIS and optimized placement without RIS, with a difference of almost 0.5 seconds. Despite the overall task duration result, the local computation requires far higher energy consumption, whereas the full offloading requires less energy for computation but more energy for data transmission and receiving the data. In addition, the optimized task placement can balance the task duration latency.

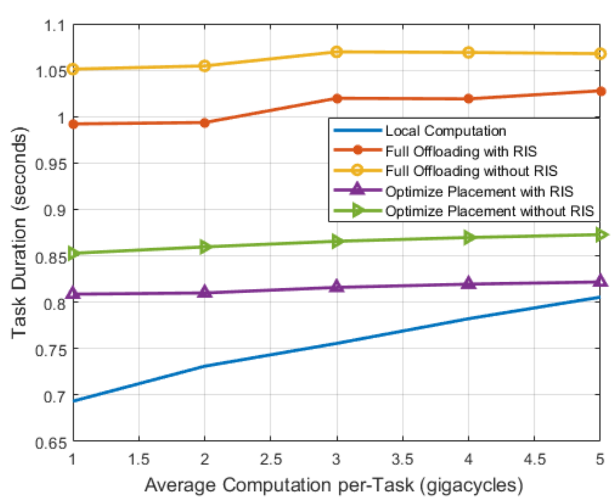

Fig. 2. Task duration with joint optimization of RIS-BS precoding and Task Placement

V. CONCLUSION

The result has demonstrated that RIS-assisted MEC computation offloading can assist users who are in non-line-of-sight (nLoS) positions within IoVT networks. This method has been shown to minimize latency when users offload their data. Future research around the topics of dynamic user, UAV communication, and multiple user clusters can be considered.

ACKNOWLEDGMENT

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the Innovative Human Resource Development for Local Intellectualization support program (IITP-2023-2020-0-01612) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation) and This work was supported by Priority Research Centers Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology (2018R1A6A1A03024003).

REFERENCES

[1] K. Bajaj, B. Sharma, and R. Singh, "Implementation analysis of iot-based offloading frameworks on cloud/edge computing for sensor generated big data," *Complex & Intelligent Systems*, vol. 8, no. 5, pp. 3641–3658, 2022.

[2] M. F. Alotaibi, M. Omri, S. Abdel-Khalek, E. Khalil, and R. F. Mansour, "Computational intelligence-based harmony search algorithm for real-time object detection and tracking in video surveillance systems," *Mathematics*, vol. 10, no. 5, p. 733, 2022.

[3] A. Sammoud, A. Kumar, M. Bayoumi, and T. Elarabi, "Real-time streaming challenges in internet of video things (iovt)," in *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2017, pp. 1–4.

[4] X. Yang, H. Lu, F. Guo, Y. Wang, C. Kong, and Q. Lu, "Joint grouping and offloading in noma-assisted multi-mec iovt systems," in *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*, 2022, pp. 5699–5704.

[5] M. Alsenwi, M. Abolhasan, and J. Lipman, "Intelligent and reliable millimeter wave communications for ris-aided vehicular networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 11, pp. 21 582–21 592, 2022.

[6] K. Shen and W. Yu, "Fractional programming for communication systems—part i: Power control and beamforming," *IEEE Transactions on Signal Processing*, vol. 66, no. 10, pp. 2616–2630, 2018.