Reinforcement learning based energy-efficient internet-of-things video transmission

Yilin Xiao, Guohang Niu, Liang Xiao*, Yuzhen Ding, Sicong Liu, and Yexian Fan

Abstract: The video transmission in the Internet-of-Things (IoT) system must guarantee the video quality and reduce the packet loss rate and the delay with limited resources to satisfy the requirement of multimedia services. In this paper, we propose a reinforcement learning based energy-efficient IoT video transmission scheme that protects against interference, in which the base station controls the transmission action of the IoT device including the encoding rate, the modulation and coding scheme, and the transmit power. A reinforcement learning algorithm state-action-reward-state-action is applied to choose the transmission action based on the observed state (the queue length of the buffer, the channel gain, the previous bit error rate, and the previous packet loss rate) without knowledge of the transmission channel model at the transmitter and the receiver. We also propose a deep reinforcement learning based energy-efficient IoT video transmission scheme that uses a deep neural network to approximate Q value to further accelerate the learning process involved in choosing the optimal transmission action and improve the video transmission performance. Moreover, both the performance bounds of the proposed schemes and the computational complexity are theoretically derived. Simulation results show that the proposed schemes can increase the peak signal-to-noise ratio and decrease the packet loss rate, the delay, and the energy consumption relative to the benchmark scheme.

Key words: Internet-of-Things (IoT); video transmission; reinforcement learning; energy efficiency

1 Introduction

With the increasing demand of multimedia applications, such as virtual reality and video surveillance supported by Internet-of-Things (IoT) systems, the performances of video quality, packet loss rate, and delay of the video transmission must be guaranteed to satisfy the ever stringent quality-of-service requirements, especially with limited communication bandwidth and energy of IoT devices available[1]. However, because the number of the devices in an IoT system is huge, the used spectrum tends to be overlapped and in conflict[2], incurring interference that degrades the performance of the video transmission. Hence, an energy-constrained IoT device should apply the techniques of video encoding, modulation and coding, and power control in video transmission to guarantee the quality of service and save energy in the presence of the interference caused by other ambient devices.

A video encoder that applies the H.265/HEVC[3] or H.264/AVC[4] coding standard is employed by the IoT device to compress the captured video data before transmission to accommodate the source bit rate to the available bandwidth, thus increasing the spectral efficiency. The encoding rate controls the trade-off between the video compression distortion

● Yilin Xiao, Guohang Niu, Liang Xiao, Yuzhen Ding, and Sicong Liu are with the Department of Information and Communication Engineering, Xiamen University, Xiamen 361005, China. E-mail: ylxiao@stu.xmu.edu.cn; ghniu@stu.xmu.edu.cn; lxiao@xmu.edu.cn; dingding@stu.xmu.edu.cn; liusc@xmu.edu.cn.

● Yexian Fan is with the College of Information and Mechanical and Electrical Engineering, Ningde Normal University, Ningde 352100, China. E-mail: yfan@ndnu.edu.cn.

∗ To whom correspondence should be addressed.

and the number of bits which is used to encode the source video[5], which is usually optimized by the Lagrangian multipliers based on a predicted rate-distortion model[6-8]. Besides the compression distortion, the video transmission suffers from the transmission distortion due to packet errors at the receiver end under time-varying and error-prone channels, especially the channel with interference.

Forward Error Correction (FEC) codes (such as Reed-Solomon codes and rate-compatible punctured convolutional codes[9]) are adopted to reduce bit errors of video data by adding redundant coding bits to the video streams. The Adaptive Modulation and Coding (AMC) technique can effectively reduce the Bit Error Rate (BER) and improve the throughput by adaptively adjusting the modulation and coding scheme based on the channel state. For example, high-order modulation and coding schemes are applied to improve the throughput for good channel conditions.

Power control is another method to mitigate interference and thus decrease the transmission error probability[10]. The base station controls the transmit power properly to increase the Signal-to-Interference-plus-Noise Ratio (SINR), thereby improving the video transmission performance. However, in most existing studies, the modulation and coding scheme and the transmit power are fixed or selected according to a known channel model, which might not adapt to a dynamic and complex IoT environment in the presence of many devices and interferences.

In this paper, we propose a Reinforcement Learning (RL) based Energy-efficient IoT Video Transmission (REIVT) scheme to guarantee the multimedia service quality against interference and save energy. This scheme enables the base station to determine the transmission action of the IoT device (including the encoding rate, the modulation and coding scheme, and the transmit power) without being aware of the transmission channel model. More specifically, the base station observes the state that consists of the transmission channel gain, the queue length of the buffer, the previous Packet Loss Rate (PLR), and the previous BER. State-Action-Reward-State-Action (SARSA)[11], an RL algorithm that avoids the Q value overestimations compared with Q-learning, is applied to evaluate the long-term expected utility for each feasible transmission action under the state; this approach allows the base station to determine the transmission action of the IoT device. We also propose a Deep RL based Energy-efficient IoT Video Transmission (DREIVT) scheme that utilizes a Deep Neural Network (DNN) to compress the state space and further accelerate the learning process involved in choosing the optimal encoding rate, modulation and coding scheme, and transmit power, thereby improving the video transmission performance.

Moreover, the performance bounds of the proposed schemes in terms of the compression distortion, the PLR, the delay, and the energy consumption are theoretically derived, and the effect of the channel condition on the video transmission performance of the proposed scheme is analyzed. We also compare the computational complexities of REIVT and DREIVT. Simulation results of the proposed schemes in a video transmission system with good channel condition show that REIVT and DREIVT can achieve the performance bounds. We use Universal Software Radio Peripheral (USRP) to collect the raw data of the video transmission (such as PLR, BER, and channel gain) in an indoor WiFi video transmission system that involves a smart phone transmitting context sensing video data encoded by H.264/AVC to a base station 2 m away. The proposed schemes are evaluated by simulations in a scenario which is constructed according to the collected field data.

The rest of this paper is organized as follows. We review the related work in Section 2 and present the system model in Section 3. The REIVT and DREIVT schemes are introduced in Sections 4 and 5, respectively. The performance bounds and computational complexity are analyzed theoretically in Section 6. Simulation results and discussion are given in Section 7. Finally, Section 8 concludes this paper.

2 Related work

Encoding rate control methods can provide a trade-off between the bit rate and the video quality to generate a compressed video stream with little distortion at a given encoding rate. For example, a Lagrange multiplier based rate control framework for the high-efficiency video

coding, which was proposed in Ref. [12], allocates the encoding rate at the coding tree unit level to minimize the video distortion variation across video frames and provide more accurate rate regulations, more stable buffer fullness, and lower video quality fluctuation. The end-to-end delay also affects the multimedia service quality in real-time wireless video communication systems. In Ref. [13], a delay-rate-distortion based rate control method involves an algorithm based on Lagrange multiplier, Karush-Kuhn-Tucker conditions, and sequential quadratic programming methods. Both the source coding and the channel coding parameters are jointly selected to minimize the average total end-to-end distortion while being subject to the constraints of the end-to-end delay and the transmission rate. In Ref. [14], an energy-efficient adaptive source-FEC coding scheme for video surveillance systems manages the rate-distortion-power trade-off between the video encoding and the transmission and jointly controls the encoding rate and the FEC coding to generate energy-efficient and high-quality video streams.

AMC switches modulation and coding schemes adaptively to increase the data rate and satisfy the BER constraint. For example, a scalable video coding scheme combined with AMC, which was proposed in Ref. [15], chooses modulation and coding schemes for each video coding layer to provide high-quality videos for mobile devices that are close to the cellular towers and provide the required video quality for mobile devices in poor channel states. A multicast video delivery scheme in Ref. [16] jointly allocates resources and selects the modulation and coding scheme based on the average channel conditions of the users to improve the overall video quality in a multicast group. A Markov decision process based transmission scheme in Ref. [17] applies value iteration to determine the modulation and coding scheme that reduces the transmission cost and improves the video quality with higher Peak Signal-to-Noise Ratio (PSNR) for the reconstructed video corresponding to each channel state and transmission delay.

Power control methods are utilized to combat against interference and increase the energy efficiency. A feedback based transmission power control scheme, which was proposed in Ref. [18], builds a predictive model between the transmit power and the link quality by collecting the link quality history and adaptively adjusts the transmit power according to the observed link quality. In Ref. [19], the central controller of the cellular system controls the transmit power of the Device-to-Device (D2D) transmitter to limit the interference caused by undelying D2D users and applies convex programming approaches to maximize the SINR of the cellular link with the known global channel state information. The power control framework for wireless interference networks in Ref. [20] applies a branch-and-bound procedure to find the bounds for the energy-efficient maximization problem.

3 System model

3.1 Network model

As shown in Fig. 1, an energy-constrained IoT device
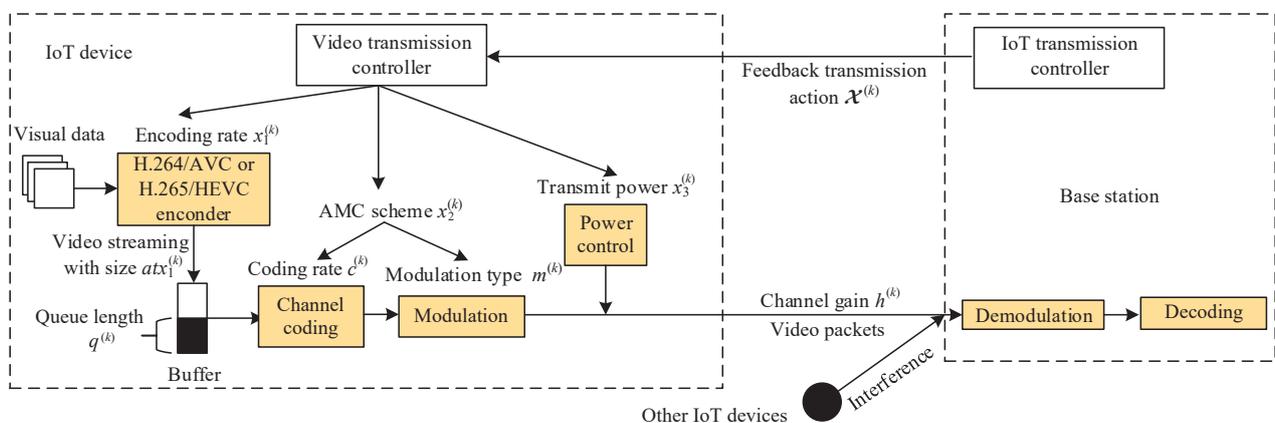
Fig. 1 **Illustration of an IoT video transmission system, in which an IoT device transmits the video stream to the base station in the presence of interference.**

captures the raw video data with a camera in the multimedia application such as video surveillance. At each time slot k, the IoT device employs an H.265/HEVC or H.264/AVC encoder to encode the captured video frames during the interval of one time slot denoted by t. The encoding rate $x_1^{(k)}$ is chosen from X feasible encoding rates, i.e., $x_1^{(k)} \in \{R_i | 1 \leqslant i \leqslant X\}$, where R_i is the i-th encoding rate level. The encoder encodes the video data with the group of pictures structure[4] and generates the video stream with size $atx_1^{(k)}$ comprised of frame sequences at each time slot, where a is the ratio of unit time slot that the IoT device generates the video stream. The encoded video frames are split into several data packets, each with size Z and a packet header that includes the information of the video frame number, the packet number, and the time stamp[14]. Without loss of generality, the IoT device has a buffer that stores the video packets to match the arrival rate and the service rate in a time-varying channel[21], where the queue length of the buffer is given by $q^{(k)}$. In order to protect video transmission against packet loss, the IoT device uses channel coding, such as the Reed-Solomon code, and applies the AMC that chooses the modulation and coding scheme, indexed by $x_2^{(k)}$, from M feasible schemes, i.e., $x_2^{(k)} \in \{1, \ldots, M\}$. As a consistent notation, the modulation and coding scheme indexed by $x_2^{(k)}$ encodes the video packets with coding rate $c^{(k)}$ and modulates the video packets with modulation type $m^{(k)}$. Afterwards, the IoT device transmits the video packets to the base station with transmit power $x_3^{(k)}$ that ranges from P_1 to P_N and is quantized into N levels with the i-th level denoted by P_i, i.e., $x_3^{(k)} \in \{P_i | 1 \leqslant i \leqslant N\}$. The video packets are transmitted in the channel at the central frequency F, bandwidth W, and channel gain $h^{(k)}$. The video transmission is interfered by other ambient IoT devices that transmit messages using the same channel, with the received interference power denoted by $y^{(k)}$ at the base station.

Once receiving the video packets, the base station demodulates and decodes the received video packets to reconstruct the video frames and evaluates the BER denoted by $\zeta^{(k)}$ and the PLR denoted by $\rho^{(k)}$ of the received video packets during this time slot. The energy consumption of the IoT device denoted by $E^{(k)}$ contains the contribution of both the encoding module and the signal transmitter. More energy-efficient video transmission saves the battery life of the IoT device. In order to guarantee the quality of service and increase the energy efficiency, the base station sends the transmission action denoted by $\mathcal{X}^{(k)}$, i.e., the parameters used for transmission (including the encoding rate $x_1^{(k)}$, the modulation and coding scheme $x_2^{(k)}$, and the transmit power $x_3^{(k)}$) to the IoT device through the control channel ($\mathcal{X}^{(k)} = [x_1^{(k)} \ x_2^{(k)} \ x_3^{(k)}]$).

3.2 Video traffic model

For simplicity, the buffer is assumed to have size large enough for the IoT transmission scenario. The size of the arrival data into the buffer is $atx_1^{(k)}$, and the size of the departure data from the buffer depending on the transmit rate $r^{(k)}$ is determined by the available channel bandwidth, the modulation type $m^{(k)}$, and the coding rate $c^{(k)}$. Accordingly, the dynamic variation of the buffer queue is given by

$$q^{(k+1)} = \max\{q^{(k)} + atx_1^{(k)} - tr^{(k)}, 0\} \quad (1)$$

The delay of video packets delivery denoted by $T^{(k)}$ affects the quality of service and is mainly composed of the queue delay and the transmission delay. For real-time video transmission, each packet has a deadline D and should be delivered before D; otherwise, this packet will be discarded[17]. In the rest of the paper, time slot k will be omitted for simplicity of notation without ambiguity. The important parameters are listed in Table 1.

4 REIVT scheme

We propose an REIVT scheme to guarantee the video quality and save the energy of the IoT device. In this scheme, the base station applies SARSA to select the transmission action $\mathcal{X}^{(k)}$ of the IoT device, which includes the encoding rate, the modulation and coding scheme, and the transmit power. The Q value of the current state-action pair is updated using the Q value of the next state-action pair instead of the maximum Q value of the next state to avoid the Q value overestimations suffered in Q-learning.

At time slot k, the base station receives the queue length q from the IoT device, estimates the channel gain

Table 1 Summary of the primary notations.

Symbol	Definition
$x_1^{(k)}$	Encoding rate at time slot k
$q^{(k)}$	Queue length of the buffer
$x_2^{(k)}$	Modulation and coding scheme index
$c^{(k)}$	Coding rate
$m^{(k)}$	Modulation type
$x_3^{(k)}$	Transmit power
$y^{(k)}$	Received interference power
$\zeta^{(k)}$	BER
$h^{(k)}$	Channel gain of the transmission channel
$r^{(k)}$	Transmit rate
$\rho^{(k)}$	PLR
$\tau^{(k)}$	Mean-Square Error (MSE)
$T^{(k)}$	Packet delay
t	Interval of the time slot
X	Number of the feasible encoding rates
M	Number of the feasible modulation and coding schemes
N	Transmit power quantization level
R_1, R_X	Minimum and maximum encoding rate
P_1, P_N	Minimum and maximum transmit power
D	Packet deadline

h, and formulates the state $s^{(k)}$ according to ζ, ρ, q, and h as

$$s^{(k)} = [\zeta \ \rho \ h \ q] \qquad (2)$$

which are all dynamic because of the time-varying channel and the dynamic chosen transmission action. More specifically, the mobility of the IoT device and the dynamic of the environment determine the dynamic of the channel gain. The BER and PLR are determined by the channel gain, the chosen modulation and coding scheme, and the transmit power. The queue mainly depends on the encoding rate and the modulation and coding scheme. Because the transmission action is chosen dynamically over the time, the resulting state is dynamic.

The base station selects the transmission action $\boldsymbol{\mathcal{X}}^{(k)}$ from the feasible transmission action space denoted by Ω consisting of all the feasible encoding rates, modulation and coding schemes, and transmit power levels via the ε-greedy exploration[22] according to the Q values of state-action pairs. The chosen $\boldsymbol{\mathcal{X}}^{(k)}$ is sent back to the IoT device.

Once receiving the transmission action, the encoder in the IoT device compresses the raw visual data captured by the camera with the encoding rate x_1. The IoT device applies the modulation and coding scheme x_2 to encode the video packets with channel coding rate c and modulate them with modulation type m, and then transmits the video packets in the buffer to the base station at transmit power x_3. Upon receiving the video packets, the base station estimates the MSE denoted by τ between the raw video frames and the encoded video frames that indicates the video distortion using a predicted rate-distortion model in Ref. [23]. The base station also evaluates the packet delay T according to the packet header information, the energy consumption E, ρ, and ζ. The aim of this scheme is to reduce the compression distortion and the transmission distortion, decrease the delay, and improve the energy efficiency. Hence, the utility of the base station that guides the system to achieve the aim is given by

$$u^{(k)} = \frac{\omega_0}{\omega_1 - x_1} - \omega_2 T - \omega_3 \rho - \omega_4 E \qquad (3)$$

where the first term is the estimated compression MSE and ω_0, ω_1, ω_2, ω_3, and ω_4 represent coefficients.

The next state $s^{(k+1)}$ is formulated in a manner similar to Eq. (2), and the next transmission action $\boldsymbol{\mathcal{X}}^{(k+1)}$ is selected via the ε-greedy exploration approach similar to the selection of $\boldsymbol{\mathcal{X}}^{(k)}$. The video transmission experience $[s^{(k)} \ \boldsymbol{\mathcal{X}}^{(k)} \ u^{(k)} \ s^{(k+1)} \ \boldsymbol{\mathcal{X}}^{(k+1)}]$ is stored in a first-in-first-out experience pool $\mathcal{D}$ that can store K experiences on a rolling basis.

As shown in Algorithm 1, this scheme estimates the Q value of the transmission action $\boldsymbol{\mathcal{X}}^{(k)}$ under state $s^{(k)}$ according to previous K experiences and the Q value of the next state-action pair and updates the Q value with

$$Q(s^{(k)}, \boldsymbol{\mathcal{X}}^{(k)}) \leftarrow \alpha \sum_{i=k-K+1}^{k} \gamma^{i-k+K-1} u^{(i)} +$$
$$(1-\alpha)Q(s^{(k)}, \boldsymbol{\mathcal{X}}^{(k)}) + \alpha \gamma^K Q(s^{(k+1)}, \boldsymbol{\mathcal{X}}^{(k+1)}) \quad (4)$$

where α is the learning rate and γ is the discounted factor.

5 DREIVT scheme

We propose a DREIVT scheme to further accelerate the learning process of the base station in choosing the encoding rate, the modulation and coding scheme index, and the transmit power. This scheme utilizes

---

Algorithm 1   REIVT

1: Initialize ζ, ρ, γ, α, K and $Q = 0$ for all state-action pairs
2: **for** $k = 1, 2, 3, \ldots$ **do**
3:   Receive the queue length q from the IoT device
4:   Estimate h
5:   Formulate $s^{(k)}$ via Eq. (2)
6:   **if** $k = 1$ **then**
7:     Select $\mathcal{X}^{(k)}$ via the ε-greedy exploration
8:   **else**
9:     $\mathcal{X}^{(k)} = \mathcal{X}^{(k-1)}$
10:   **end if**
11:   Send $\mathcal{X}^{(k)}$ to the IoT device
12:   Estimate MSE between the raw video frames and the encoded video frames
13:   Evaluate T, E, ρ, and ζ
14:   Calculate $u^{(k)}$ via Eq. (3)
15:   Formulate $s^{(k+1)}$ similar to Eq. (2)
16:   Select $\mathcal{X}^{(k+1)}$ via the ε-greedy exploration
17:   **if** $k > K$ **then**
18:     Remove $[s^{(k-K)} \, \mathcal{X}^{(k-K)} \, u^{(k-K)} \, s^{(k-K+1)} \, \mathcal{X}^{(k-K+1)}]$ from $\mathcal{D}$
19:   **end if**
20:   Store $[s^{(k)} \, \mathcal{X}^{(k)} \, u^{(k)} \, s^{(k+1)} \, \mathcal{X}^{(k+1)}]$ in $\mathcal{D}$
21:   **if** $k \geqslant K$ **then**
22:     Estimate and update Q value via Formula (4)
23:   **end if**
24: **end for**

Algorithm 2   DREIVT

1: Initialize γ, θ, $\mathcal{D} = \varnothing$, ζ, ρ, and B
2: **for** $k = 1, 2, 3, \ldots$ **do**
3:   Receive the queue length q from the IoT device
4:   Estimate h
5:   $s^{(k)} = [\zeta \, \rho \, h \, q]$
6:   Input $s^{(k)}$ to the DNN
7:   obtain $Q(s^{(k)}, \cdot, \theta)$
8:   Select $\mathcal{X}^{(k)}$ via ε-greedy exploration according to $Q(s^{(k)}, \cdot, \theta)$
9:   **Same as Lines 11–15 in Algorithm 1**
10:   $\mathcal{D} \leftarrow \mathcal{D} \cup [s^{(k)} \, \mathcal{X}^{(k)} \, u^{(k)} \, s^{(k+1)}]$
11:   **for** $i = 1, 2, 3, \ldots, B$ **do**
12:     Choose $[s^{(b(i))} \, \mathcal{X}^{(b(i))} \, u^{(b(i))} \, s^{(b(i)+1)}]$ from $\mathcal{D}$ randomly
13:   **end for**
14:   Update the DNN parameters via Formula (5)
15: **end for**

---

a deep neural network to compress the state space and approximate Q value more accurately, thereby obtaining better video transmission performance against interference than REIVT for base stations with sufficient computational resources.

As shown in Algorithm 2, the base station receives the queue length q from the IoT device, estimates the channel gain h, and formulates the state $s^{(k)}$ via Eq. (2) at time slot k. This scheme inputs $s^{(k)}$ into a three-layer fully connected DNN with parameters denoted by θ and consisting of an input layer, a hidden layer, and an output layer to obtain Q value $Q(s^{(k)}, \cdot, \theta)$ under the current state. More specifically, the input layer involves 4 nodes, the hidden layer contains f nodes with each activated by a rectified linear unit, and the output layer contains *XMN* nodes, as shown in Fig. 2. The transmission action $\mathcal{X}^{(k)}$ is selected according to the ε-greedy exploration and the output of the DNN, i.e., the Q value for each feasible transmission action, is sent to the IoT device.

The IoT device encodes the captured video data at

the encoding rate x_1, uses the modulation and coding scheme x_2 to modulate and encode the video packets, and transmits the video packets to the base station at transmit power x_3. Upon receiving the video packets, the base station estimates τ, T, E, ρ, and ζ. The utility and the next state are given by Eqs. (3) and (2), respectively. The base station stores the video transmission experience $e^{(k)} = [s^{(k)} \, \mathcal{X}^{(k)} \, u^{(k)} \, s^{(k+1)}]$ in a memory pool $\mathcal{D}$.

A minibatch is formulated by randomly and uniformly choosing B experiences from $\mathcal{D}$, i.e., $[e^{(b(i))}]_{1 \leqslant i \leqslant B}$, where $b(i) \sim U(1, k)$. Similar to the scheme of Ref. [22], this scheme uses the stochastic gradient descent algorithm to train the DNN. The DNN parameters θ are updated to minimize the error between the target Q value and the estimated Q value with the minibatch:

$$\theta \leftarrow \arg\min_{\theta'} \frac{1}{B} \sum_{i=1}^{B} (Q(s^{(b(i))}, \mathcal{X}^{(b(i))}, \theta) - u^{(b(i))} - \gamma \max_{\mathcal{X}'} Q(s^{(b(i)+1)}, \mathcal{X}', \theta))^2 \quad (5)$$

6   Performance evaluation

We evaluate the performance bounds of REIVT and DREIVT, including the compression distortion, the PLR, the energy consumption, and the utility of the base station. The computational complexity of each proposed scheme is analyzed.

At time slot k, the base station chooses the encoding rate $x_1 \in [R_1, R_X]$ and the transmit power $x_3 \in$
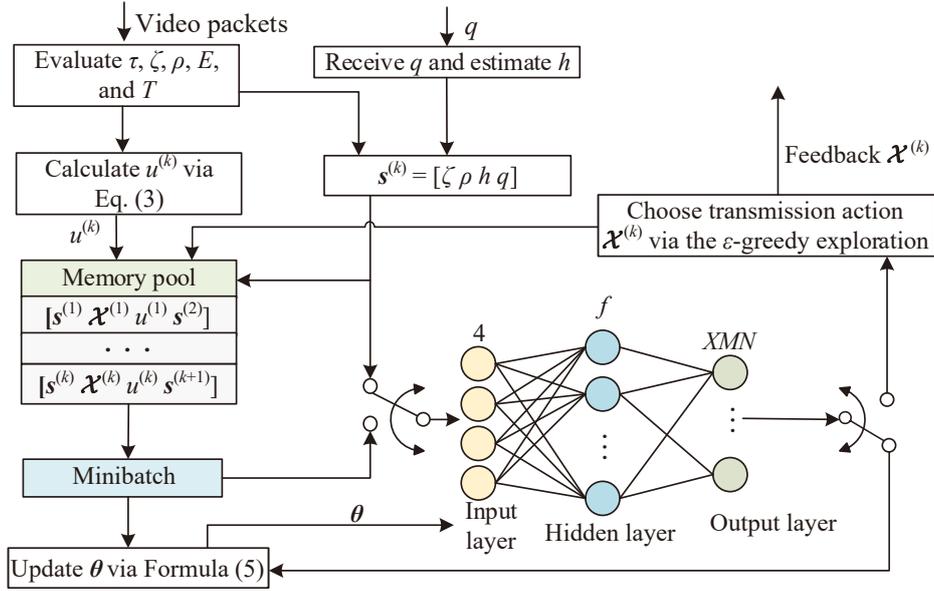
Fig. 2   Illustration of DREIVT scheme.

$[P_1, P_N]$ for the IoT device to compress the video and transmit the video packets under channel gain h with the received interference power denoted by Y. The IoT device uses the M-order modulation and coding scheme to encode and modulate the video packets to guarantee the throughput, resulting in the transmit rate r_M of the video packets. Assuming that r_M is higher than the maximum encoding rate R_X, i.e., $r_M > R_X$, and then the number of arrival video packets to the buffer is less than that of departure video packets from the buffer at each time slot. The queue delay is negligible and the delay of the video packets is equal to the transmission delay given by $T = Z/r_M$. The data rate is assumed to be high enough so that the maximum delay is less than the packet deadline D, i.e., $Z < Dr_M$. Thus, the packet loss is only caused by the packet error during transmission, which is approximated as Formula (6) according to Ref. [24],

$$\rho = \begin{cases} \beta_0 \exp\left(\dfrac{-\beta_1 W h x_3}{\sigma + Y}\right), & \text{if } W h x_3 \geqslant \beta_2(\sigma + Y); \\ 1, & \text{otherwise} \end{cases}$$

$$(6)$$

where β_0, β_1, and β_2 are positive fitting parameters for the M-th modulation and coding scheme, and σ is the noise.

The energy consumption of the IoT device mainly contains the consumption of encoding given by $\beta_3 x_1 t +$ $\beta_4 t$ and the consumption of transmission given by $x_3 t$, where β_3 and β_4 are fitting parameters.

Theorem 1   The bounds of the performance metrics of the proposed schemes are given by

$$\tau = \frac{\omega_0}{\omega_1 - R_X} \tag{7}$$

$$\rho = \beta_0 \exp\left(\frac{-\beta_1 P_1 W h}{\sigma + Y}\right) \tag{8}$$

$$E = \beta_3 R_X t + \beta_4 t + P_1 t \tag{9}$$

$$u = \frac{\omega_0}{\omega_1 - R_X} - \omega_2 \frac{Z}{r_M} - \omega_3 \beta_0 \exp\left(\frac{-\beta_1 P_1 W h}{\sigma + Y}\right) -$$
$$(\beta_3 R_X + \beta_4 + P_1)\omega_4 t \tag{10}$$

if

$$R_X - \sqrt{\frac{\omega_0 P_1 \exp\left(\frac{\beta_3 P_1 W h}{\sigma + Y}\right)}{\beta_0 \beta_1 \beta_2 \beta_3 \omega_3}} \leqslant \omega_1 < R_1 \tag{11}$$

Proof   Based on Eq. (3) and Formula (6), the utility of the base station is given by

$$u(x_1, M, x_3) = \frac{\omega_0}{\omega_1 - x_1} - \omega_2 \frac{Z}{r_M} - \omega_3 -$$
$$\omega_3\left(\beta_0 \exp\left(\frac{-\beta_1 W h x_3}{\sigma + Y}\right) - 1\right)\mathcal{I} -$$
$$(\beta_3 x_1 + \beta_4 + x_3)\omega_4 t \tag{12}$$

where $\mathcal{I}$ is an indicator function that is equal to 1 if $W h x_3/(\sigma + Y) \geqslant \beta_2$ is satisfied and is equal to 0 otherwise.

We denote $u_1(x_1)$, $u_2(M)$, and $u_3(x_3)$ as the functions of x_1, M, and x_3, respectively, given by

$$u_1(x_1) = \frac{\omega_0}{\omega_1 - x_1} - \beta_3 \omega_4 t x_1 \tag{13}$$

$$u_2(M) = -\frac{\omega_2 Z}{r_M} \tag{14}$$

$$u_3(x_3) = -\omega_3 - \omega_4 t x_3 -$$
$$\omega_3 \left(\beta_0 \exp\left(\frac{-\beta_1\, Whx_3}{\sigma + Y} \right) - 1 \right) \mathcal{I} \tag{15}$$

By Eq. (15), if Formula (11) holds, we have

$$\frac{\mathrm{d}u_1(x_1)}{\mathrm{d}x_1} = \frac{\omega_0}{(\omega_1 - x_1)^2} - \beta_3 \omega_4 t \geqslant 0 \tag{16}$$

Thus, $x_1^* = R_X$ is the maximum of $u_1(x_1)$. If Formula (11) holds, we have

$$\frac{\mathrm{d}u_3(x_3)}{\mathrm{d}x_3} = \frac{\omega_3\beta_0\beta_1 Wh}{\sigma + Y} \exp\left(\frac{-\beta_1 Whx_3}{\sigma + Y} \right) - \omega_4 t \leqslant 0 \tag{17}$$

Thus, $x_3^* = P_1$ is the maximum of $u_3(x_3)$.

Because $\mathrm{d}u_2(M)/\mathrm{d}x_2 = 0$ and $u(x_1, M, x_3) = u_1(x_1) + u_2(M) + u_3(x_3)$, $[R_X\ M\ P_1]$ is the maximum of $u(x_1, M, x_3)$.

Thus, $\forall x_1 \in [R_1, R_X]$ and $x_3 \in [P_1, P_N]$; as a result, we have

$$u(R_X, M, P_1) \geqslant u(x_1, M, x_3) \tag{18}$$

Using Formula (6) and Eq. (12), we have the performance bounds of the proposed schemes given by Eqs. (7)–(10). ∎

Remark 1   If the channel condition between the IoT device and the base station is good compared with a threshold determined by the channel bandwidth, the noise, and the minimum transmit power of the IoT device, as shown in Formula (11), the minimum encoding rate and the maximum encoding rate satisfy Formula (11), and then the IoT device chooses the maximum encoding rate and the minimum transmit power to achieve a lower PLR and energy consumption to guarantee the received video quality and save energy of the IoT device.

Figure 3 shows that the performances of MSE, PLR, energy consumption, and utility of both REIVT and DREIVT converge to the bounds provided in Eqs. (7)–(10) after approximately 2000 time slots.

According to Ref. [25], the computational complexity of REIVT mainly depends on the number of the feasible encoding rates X, the feasible modulation and coding schemes M, and the transmit power quantization level N, and is given by $O(XMN)$. The computational

(a) MSE

(b) PLR

(c) Energy consumption

(d) Utility
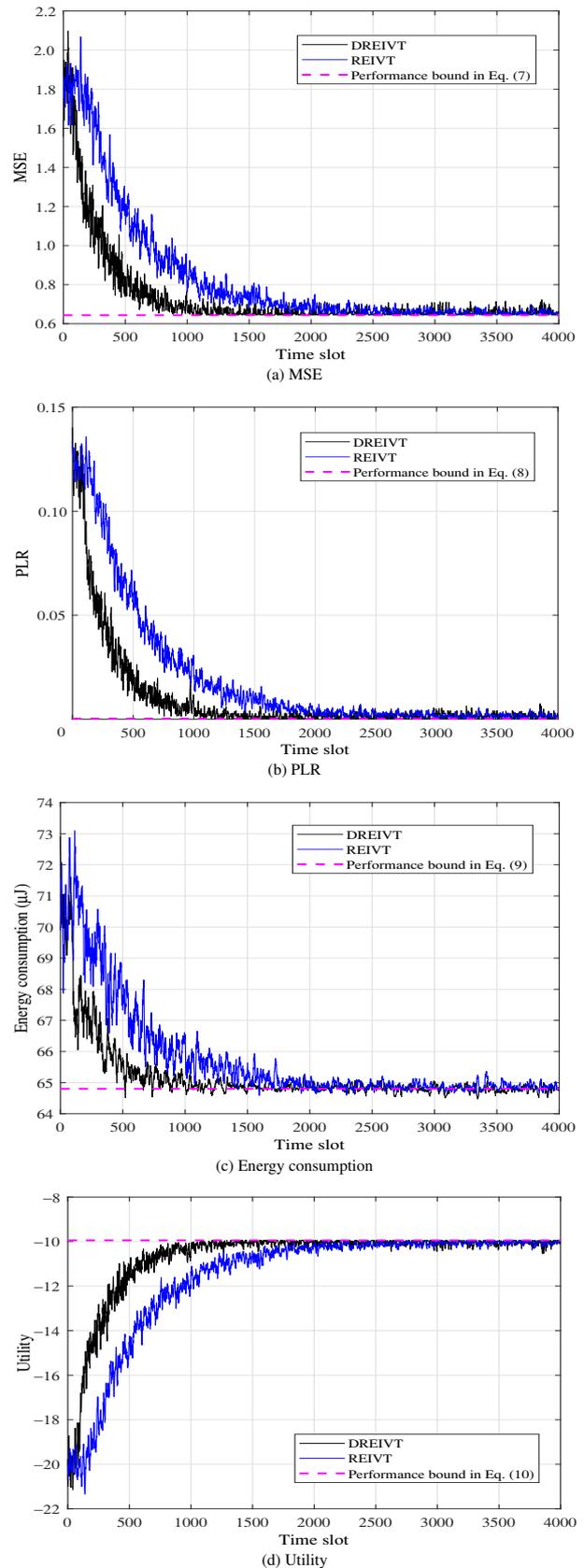
Fig. 3   Performance of DREIVT and REIVT encoded with H.264/AVC using 2 MHz bandwidth, Quadrature Phase Shift Keying (QPSK), and code rate 3/4 against interference.

complexity of DREIVT is determined by the number of multiplications in both the forward propagation and the back propagation of the DNN[26]. Hence, the computational complexity of DREIVT given by $O(BXMNf)$ linearly decreases with the number of experiences in a minibatch B, the number of feasible actions XMN, and the number of neural nodes f in the hidden layer of the DNN.

7 Simulation result and discussion

Simulations were performed to evaluate the proposed REIVT and DREIVT in an indoor WiFi-based video transmission system created by field data, such as channel gain, PLR, and BER, collected on a USRP platform, in which a smart phone executes a context sensing task and transmits video packets to the base station 2 m away. The smart phone captures the video with a resolution of 352×288 at 15 frames/s and encodes them with H.264/AVC. The encoding rate is chosen from {300, 500, 700, 1000, 1200} kbps in each time slot with the duration of 400 ms. The smart phone chooses a modulation and coding scheme from 2 feasible modulation and coding schemes including Binary Phase Shift Keying (BPSK) and QPSK with the code rate of 3/4, and transmits video packets with the power of 9, 36, or 100 mW in the channel with the center frequency 2.412 GHz and the bandwidth 2 MHz.

Let the discounted factor $\gamma = 0.5$, the learning rate of REIVT $\alpha = 0.4$, and $K = 10$. The initial exploration rate is $\varepsilon = 0.9$, and the rate linearly decreases to 0.1 after 2000 time steps. In DREIVT, the number of nodes in the hidden layer $f = 256$, and the size of memory pool $\mathcal{D}$ is 5000. Similar to the approach of Ref. [23], we use PSNR denoted by χ to measure the video quality given by

$$\chi = 10 \lg \frac{255^2}{\tau} \qquad (19)$$

As shown in Fig. 4, REIVT outperforms the benchmark scheme λ Domain Rate Control (LDRC)[8] in PSNR, PLR, delay, energy consumption, and utility because LDRC only controls the video encoding rate to optimize the rate-distortion trade-off without controlling the video transmission scheme. For instance, this scheme improves the video quality after compression by 5.8%,

reduces the PLR of the video packets by 86.7%, reduces the delay by 73.9%, reduces the energy consumption of the IoT device by 8.5%, and increases the utility by 50.1% after 3000 time slots.

DREIVT further accelerates the learning process of choosing the optimal action and improves the performance; at time slot 2000, it improves the video quality after compression by 1.43% and 6.67%, reduces the PLR of the video packets by 88.6% and 97.8%, reduces the delay by 35.1% and 78.6%, saves energy of the IoT device by 1.94% and 10.1%, and increases the utility by 19.2% and 56.4%, compared to the REIVT and the LDRC scheme, respectively.

Considering that the estimated channel gain in the proposed REIVT is not perfectly accurate in practical IoT systems, we provide the performance of REIVT and DREIVT under different levels of channel estimation error in Fig. 5, where the channel estimation error is calculated with the estimated channel gain and the actual channel gain and normalized to the range between 0 and 1.

The performances of REIVT and DREIVT (including the PSNR, PLR, delay, energy consumption, and utility) get worse as the channel estimation error increases. For example, for REIVT, the PSNR decreases from 48.2 to 46.8 dB, the PLR increases by 1.8 times from 0.042 to 0.119, the delay increases by 81.1% from 9.5 to 17.2 ms, the energy consumption increases from 64.2 to 67.7 μJ, and the utility decreases from −15.3 to −19.5 as the channel estimation error changes from 0 to 1. In addition, REIVT and DREIVT outperform LDRC even if the level of channel estimation error is high. For example, the PLR of REIVT is 15.1% lower than that of LDRC and the delay of REIVT is 13.5% lower than that of LDRC when the channel estimation error is 1. DREIVT is more robust than REIVT against the channel estimation error. For example, the PLR of DREIVT is nearly 200% lower than that of REIVT and the delay of DREIVT is approximately 100% lower than that of REIVT when the channel estimation error is 1.

8 Conclusion

In this paper, we proposed an REIVT scheme for IoT

(a) PSNR

(b) PLR

(c) Delay

(d) Energy consumption

(d) Utility
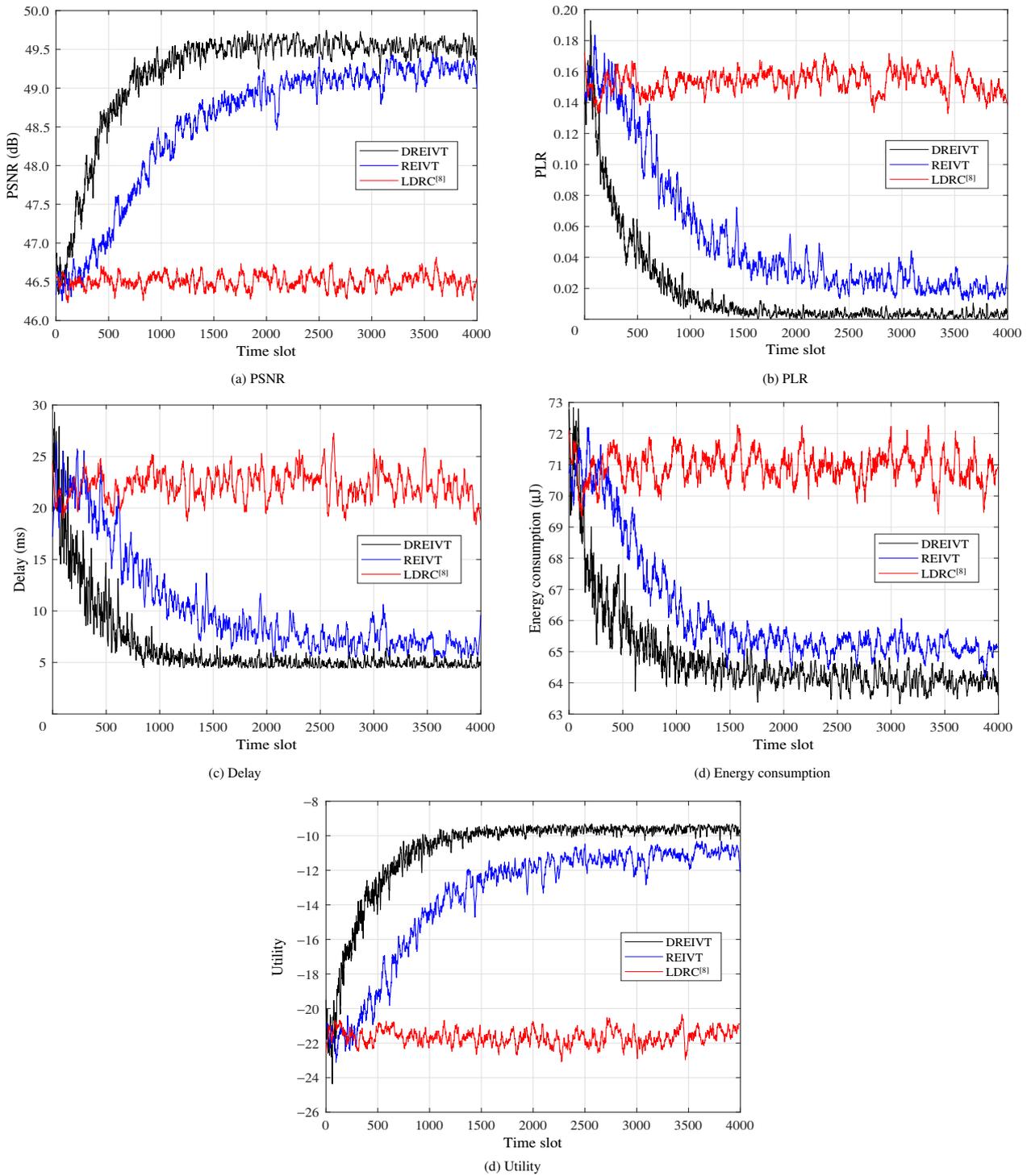
Fig. 4   Performance of REIVT and DREIVT for IoT system in which the IoT device transmits the encoded video packets to the base station 2 m away using the bandwidth of 2 MHz with the center frequency 2.412 GHz.

systems, in which the base station chooses the encoding rate, the modulation and coding scheme, and the transmit power for the IoT device to compress the video data and protect the video transmission against interference. We also proposed a DREIVT scheme that uses a deep neural network to compress the state space, thereby improving the video transmission performance for base stations with sufficient computational resources. By analyzing the performance of the proposed schemes theoretically, we provided the performance bounds that

(a) PSNR

(b) PLR

(c) Delay

(d) Energy consumption
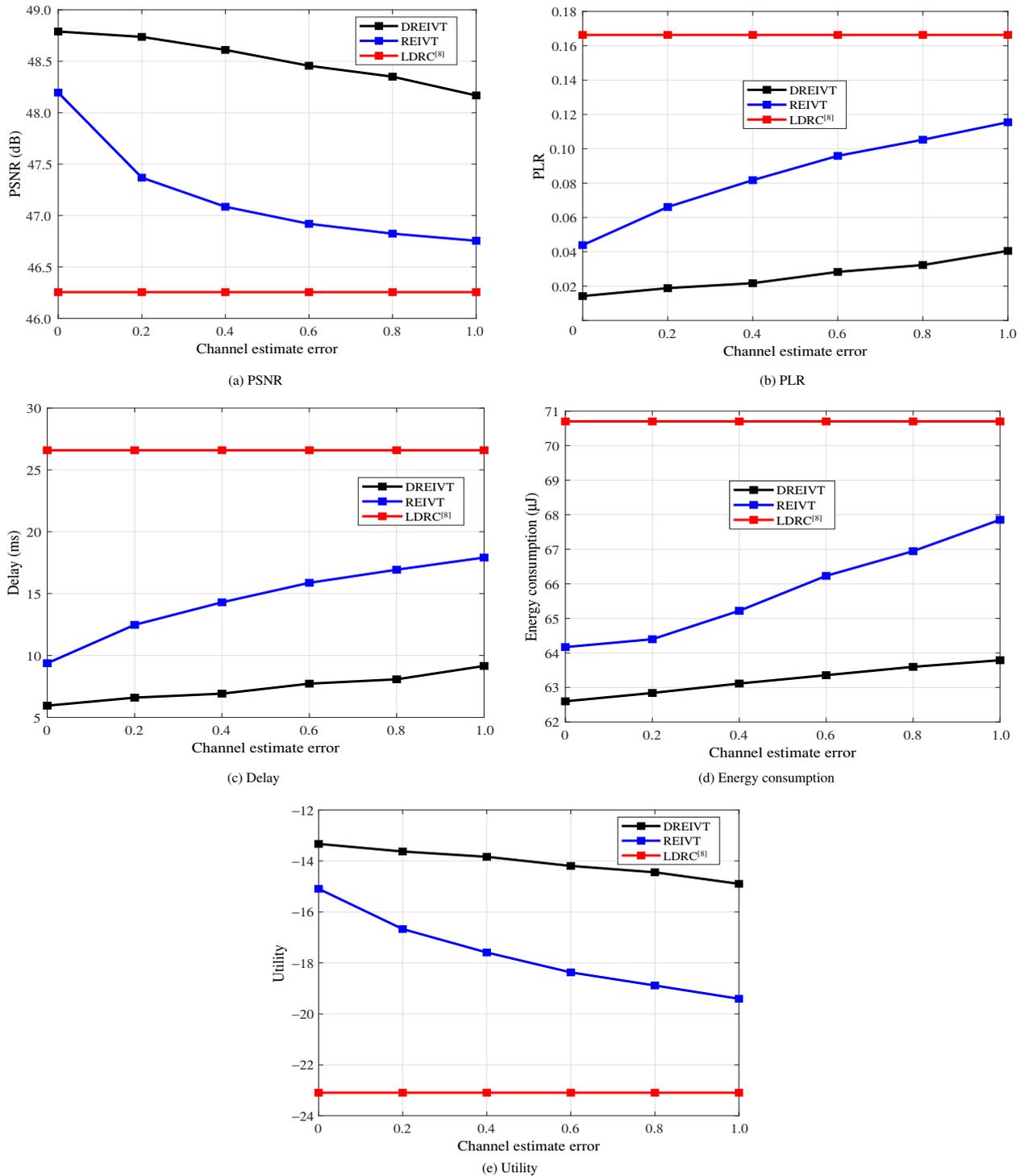
(e) Utility

Fig. 5   Averaged performance of REIVT and DREIVT for IoT systems against interference with respect to channel estimation error in range between 0 and 1.

contain the MSE, PLR, energy consumption, and utility. We also analyzed the computational complexity of the proposed schemes. Simulations on a video transmission scenario built by the data collected in an indoor WiFi-based video transmission system showed that our scheme outperforms the benchmark scheme. For example, the DREIVT scheme was found to improve the video quality by 6.67%, reduce the PLR by 97.8%, reduce the delay by 78.6%, and save energy by 10.1% compared with LDRC.

References

[1] W. Ji, J. C. Xu, H. X. Qiao, M. D. Zhou, and B. Liang, Visual IoT: Enabling internet of things visualization in smart cities, *IEEE Network*, vol. 33, no. 2, pp. 102–110, 2019.

[2] Z. Li, Y. H. Liu, K. G. Shin, J. Liu, and Z. Yan, Interference steering to manage interference in IoT, *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10 458–10 471, 2019.

[3] Z. Z. Chen and X. Pan, An optimized rate control for low-delay H.265/HEVC, *IEEE Trans. Image Process.*, vol. 28, no. 9, pp. 4541–4552, 2019.

[4] J. Chakareski, Uplink scheduling of visual sensors: When view popularity matters, *IEEE Trans. Commun.*, vol. 63, no. 2, pp. 510–519, 2015.

[5] P. Z. Wu, P. C. Cosman, and L. B. Milstein, Resource allocation for multicarrier device-to-device video transmission: Symbol error rate analysis and algorithm design, *IEEE Trans. Commun.*, vol. 65, no. 10, pp. 4446–4462, 2017.

[6] D. Liu, J. Wu, H. Cui, D. D. Zhang, C. Luo, and F. Wu, Cost-distortion optimization and resource control in pseudo-analog visual communications, *IEEE Transactions on Multimedia*, vol. 20, no. 11, pp. 3097–3110, 2018.

[7] M. L. Zhou, X. K. Wei, S. Q. Wang, S. Kwong, C. K. Fong, P. H. W. Wong, and W. Y. F. Yuen, Global rate-distortion optimization-based rate control for HEVC HDR coding, *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 12, pp. 4648–4662, 2020.

[8] B. Li, H. Q. Li, L. Li, and J. L. Zhang, λ domain rate control algorithm for high efficiency video coding, *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 3841–3854, 2014.

[9] S. Pudlewski and T. Melodia, A tutorial on encoding and wireless transmission of compressively sampled videos, *IEEE Commun. Surv. Tut.*, vol. 15, no. 2, pp. 754–767, 2013.

[10] L. Xiao, H. L. Zhang, Y. L. Xiao, X. Y. Wan, S. C. Liu, L. C. Wang, and H. V. Poor, Reinforcement learning-based downlink interference control for ultra-dense small cells, *IEEE Transactions on Wireless Communications*, vol. 19, no. 1, pp. 423–434, 2020.

[11] R. S. Sutton and A. G. Barto, *Reinforcement Learning*: *An Introduction*, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.

[12] M. H. Wang, K. N. Ngan, and H. L. Li, Low-delay rate control for consistent quality using distortion-based lagrange multiplier, *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 2943–2955, 2016.

[13] C. L. Li, H. K. Xiong, and D. P. Wu, Delay-rate-distortion optimized rate control for end-to-end video communication over wireless channels, *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 10, pp. 1665–1681, 2015.

[14] J. Y. Wu, B. Cheng, and M. Wang, Adaptive source-FEC coding for energy-efficient surveillance video over wireless networks, *IEEE Trans. Commun.*, vol. 66, no. 5, pp. 2153–2168, 2018.

[15] P. L. Li, H. H. Zhang, B. H. Zhao, and S. Rangarajan, Scalable video multicast with adaptive modulation and coding in broadband wireless data systems, *IEEE/ACM Trans. Netw.*, vol. 20, no. 1, pp. 57–68, 2012.

[16] J. Yoon, H. H. Zhang, S. Banerjee, and S. Rangarajan, Video multicast with joint resource allocation and adaptive modulation and coding in 4G networks, *IEEE/ACM Trans. Netw.*, vol. 22, no. 5, pp. 1531–1544, 2014.

[17] C. Gong and X. D. Wang, Adaptive transmission for delay-constrained wireless video, *IEEE Transactions on Wireless Communications*, vol. 13, no. 1, pp. 49–61, 2014.

[18] S. Lin, F. Miao, J. B. Zhang, G. Zhou, L. Gu, T. He, J. A. Stankovic, S. Son, and G. J. Pappas, ATPC: Adaptive transmission power control for wireless sensor networks, *ACM Trans. Sensor Netw.*, vol. 12, no. 1, p. 6, 2016.

[19] N. Lee, X. Q. Lin, J. G. Andrews, and R. W. Heath, Power control for D2D underlaid cellular networks: Modeling, algorithms, and analysis, *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 1, pp. 1–13, 2015.

[20] B. Matthiesen, A. Zappone, K. L. Besser, E. A. Jorswieck, and M. Debbah, A globally optimal energy-efficient power control framework and its efficient implementation in wireless interference networks, *IEEE Trans. Signal Process.*, vol. 68, pp. 3887–3902, 2020.

[21] A. A. Khalek, C. Caramanis, and R. W. Heath, Delay-constrained video transmission: Quality-driven resource allocation and scheduling, *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 1, pp. 60–75, 2015.

[22] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, et al., Human-level control through deep reinforcement learning, *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[23] D. W. Wang, L. Toni, P. C. Cosman, and L. B. Milstein, Resource allocation and performance analysis for multiuser video transmission over doubly selective channels, *IEEE Transactions on Wireless Communications*, vol. 14, no. 4, pp. 1954–1966, 2015.

[24] Q. W. Liu, S. L. Zhou, and G. B. Giannakis, Cross-layer combining of adaptive modulation and coding with

truncated ARQ over wireless links, *IEEE Transactions on Wireless Communications*, vol. 3, no. 5, pp. 1746–1755, 2004.

[25] C. Jin, Z. Allen-Zhu, S. Bubeck, and M. I. Jordan, Is *Q*-learning provably efficient? in *Proc. 32nd Int. Conf.* *Neural Information Processing Systems*, Red Hook, NY, USA, 2018, pp. 4868–4878.

[26] G. B. Ou and Y. L. Murphey, Multi-class pattern classification using neural networks, *Pattern Recognition*, vol. 40, no. 1, pp. 4–18, 2007.

Yilin Xiao received the BS degree in automation and the MS degree in pattern recognition and intelligent system from Hefei University of Technology, Hefei, China in 2015 and 2018, respectively. He is currently pursuing the PhD degree at the Department of Information and Communication Engineering, Xiamen University, Xiamen, China. His research interests include network security and wireless communications.

Guohang Niu received the BS degree in communication engineering from Xiamen University, Xiamen, China in 2020. He is currently pursuing the MS degree at the Department of Information and Communication Engineering, Xiamen University. His research interests include network security and wireless communications.
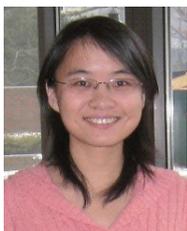
Liang Xiao is currently a professor at the Department of Information and Communication Engineering, Xiamen University, Fujian, China. She has served as an associate editor of *IEEE Trans. Information Forensics and Security* and the guest editor of *IEEE Journal of Selected Topics in Signal Processing*. She is the recipient of the best paper award for 2016 INFOCOM Big Security WS and 2017 ICC. She received the BS degree in communication engineering from Nanjing University of Posts and Telecommunications, China in 2000, the MS degree in electrical engineering from Tsinghua University, China in 2003, and the PhD degree in electrical engineering from Rutgers University, NJ, USA in 2009. She was a visiting professor at Princeton University, Virginia Tech, and University of Maryland, College Park.

Yuzhen Ding received the BS degree in communication engineering from Wuhan University, Wuhan, China in 2018. She is currently pursuing the MS degree at the Department of Information and Communication Engineering, Xiamen University, Xiamen, China. Her research interests include network security and wireless communications.

Sicong Liu received the BS and PhD degrees with honor both in electronic engineering from Tsinghua University, Beijing, China in 2012 and 2017, respectively. From 2010 to 2011, he was a visiting scholar at City University of Hong Kong. Currently, he is an assistant professor at Department of Information and Communication Engineering, Xiamen University, China. He has published over 50 research papers. He owns 8 Chinese invention patents. He has won the second prize in the Natural Science Award of Chinese Institute of Electronics, and the Best Doctoral Dissertation Award of Tsinghua University. He has served as an associate editor of *Frontiers in Communications and Networks*, the publication chair of IEEE SmartGridComm 2019, and TPC member of IEEE ICC, Globecom, and several international conferences. His research interests lie in wireless communications, sparse signal processing, machine learning, and visible light communications.
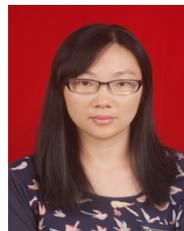
Yexian Fan received the MS degree from Soochow University, Suzhou, China in 2007. Currently, she is an associate professor at the College of Information and Mechanical and Electrical Engineering, Ningde Normal University. Her research interests include wireless network, network security, and machine learning.