Joint Wireless Resource Allocation and Bitrate Adaptation for QoE Improvement in IRS-Aided RSMA-Enabled IoMT Streaming Systems

Nam-Phuong Tran[a], Thanh Phung Truong[a], Quang Tuan Do[a], Nhu-Ngoc Dao[b], Sungrae Cho[a,*]

[a] *School of Computer Science and Engineering, Chung-Ang University, Seoul 06974, South Korea*
[b] *Department of Computer Science and Engineering, Sejong University, Seoul 05006, South Korea*

Abstract

In the context of the increasing demand for Internet of Multimedia Things (IoMT) services, rate splitting multiple access (RSMA) and intelligent reflecting surface (IRS) technologies have been considered potential networking enablers to provide ultra-throughput wireless access. However, challenges arise due to the heterogeneity of IoMT devices and arbitrary network quality changes, resulting in unwanted service quality fluctuations and downgradation. This study addresses this problem by jointly optimizing wireless resource allocation and bitrate adaptation with Deep Reinforcement Learning (DRL)-based QoE management for IRS-aided RSMA-enabled IoMT streaming systems. We formulated the problem as a Markov decision process (MDP) and apply Proximal Policy Optimization (PPO) method to flexibly adjust IoMT bitrate, transmission beamforming, IRS phase shift, and RSMA parameters. As a result, our algorithm mitigates overestimation of client-side bandwidth, leading to smoother playback and reduced quality fluctuations. Simulations show that our approach outperforms baseline methods in terms of video resolution (up to 2.5 times) and achievable sum-rate (up to 50%), contributing to a superior streaming experience in IoMT systems.

Keywords: Quality of Experience, Internet of Multimedia Things, Bitrate Adaptation, IRS-aided RSMA, Proximal Policy Optimization

1. Introduction

The Internet of Things (IoT) has become a cornerstone of our era, connecting billions of intelligent devices across diverse infrastructure domains like healthcare, transportation, and smart homes [1]. This interconnected environment facilitates seamless data exchange and paves the way for novel applications. However, the integration of multimedia capabilities into the IoT framework, known as the Internet of Multimedia Things (IoMT), presents unique challenges.

IoMT involves utilizing IoT devices to capture, process, transmit, and display multimedia content such as audio, video, and images. This enables a new generation of multimedia applications and services, but necessitates specific considerations due to the limitations of constrained IoT networks. Extensive research has explored the architecture, protocols, and applications of IoMT, highlighting the critical role of Quality of Experience (QoE) and Quality of Service (QoS) in multimedia transmission over IoT networks [2].

QoS refers to the measurable and objective characteristics of a network service, such as bandwidth, delay, and, achievable rate, that influence its performance and reliability. These parameters significantly impact the user's QoE, such as video quality, smoothness, and buffering. Since QoS metrics are not inherently and directly correlated with a user's satisfaction with a service, recent years have seen the integration of user-centric QoE metrics. These metrics are now utilized to evaluate the quality of multimedia services in conjunction with QoS metrics, which consider the user's subjective perception of a specific service. Gaining insights into users' expectations and their actual experiences with a service is paramount for ensuring successful service delivery [3].

QoE management for multimedia streaming services focuses on ensuring a satisfactory user experience when consuming multimedia content over the Internet. It encompasses various techniques and tools for modeling, monitoring, optimizing, and controlling the users' experience of streaming services [3]. In the IoMT systems, QoE management needs to address the challenges posed by the surging demand for high-quality streaming content and the dynamic nature of IoT networks. Users expect smooth, high-resolution video playback, putting enormous strain on resource-constrained IoMT networks. Frequent changes in network conditions (bandwidth, latency, etc.) due to device heterogeneity and diverse network environments significantly impact streaming quality and lead to video rebuffering or disruptions.

To address these challenges, bitrate adaptation has emerged as a prominent strategy [3–5]. Adaptive bitrate streaming (ABS) dynamically adjusts the bitrate of multimedia content in real-time, tailoring it to the viewer's network capabilities and device limitations. This approach involves dividing the content into small segments encoded at different bitrates. The client de-

[*] Corresponding author
Email addresses: tnphuong@uclab.re.kr (Nam-Phuong Tran), tptruong@uclab.re.kr (Thanh Phung Truong), dqtuan@uclab.re.kr (Quang Tuan Do), nndao@sejong.ac.kr (Nhu-Ngoc Dao), srcho@cau.ac.kr (Sungrae Cho)

vices consistently observe the network conditions, encompassing aspects such as bandwidth, latency, and packet loss, and subsequently determine the suitable bitrate for playback. This ensures smooth playback and minimizes interruptions even under varying network conditions. It delivers consistent and high-quality multimedia content, minimizes interruptions, and ultimately enhances the overall QoE of multimedia services in the dynamic and resource-constrained environment of the IoMT. While users can choose their desired video bitrate, the heterogeneity of devices and the dynamic nature of network conditions may lead to overestimating the available bandwidth. This inaccurate estimation presents a challenge for ABS services, leading to sudden transitions between different video bitrate levels.

The advancement of wireless networks, exemplified by the sixth-generation (6G), is instrumental in accommodating the proliferation of connected devices within IoT networks [6]. This evolution is particularly fueled by the escalating demand for high data rate applications. The resultant surge in the number of devices within the IoMT network contributes to an enhanced heterogeneity of devices. This, in turn, presents a notable challenge for the network controller, necessitating dynamic management of user association, spectrum access, transmit power, and the efficient distribution of multimedia content to a vast array of IoMT devices within large-scale networks. Addressing this challenge demands the continuous real-time monitoring of network conditions, facilitating swift adaptation to changes within the IoMT network.

Orthogonal Frequency Division Multiple Access (OFDMA), a multiple access technique, dynamically allocates subcarriers within a channel to multiple devices, optimizing spectrum utilization [7, 8]. Nevertheless, the limited network capacity in OFDMA networks poses a significant obstacle to enhancing QoE [9, 10]. Given the data-intensive nature of IoMT networks and their susceptibility to latency [2], the inclusion of device heterogeneity introduces an extra layer of complexity, particularly in terms of interference management. The limited network capacity continues to be a primary obstacle to achieving substantial enhancements in QoE. Overcoming this challenge necessitates an efficient transmission paradigm capable of mitigating interference and achieving a higher level of spectral efficiency for the transmission of multimedia content over the wireless network.

In response to these challenges, we have proposed a server-side bitrate adaptation approach, where the server, situated at the Base Station (BS), determines the bitrate. We have constructed a QoE management model that continuously monitors, evaluates and enhances the overall QoE, encompassing real-time aspects including video quality, quality fluctuations, and rebuffering. A Deep Reinforcement Learning (DRL) model, trained at the server, continually monitors network conditions and user capabilities to optimize and allocate wireless resources in real-time, delivering suitable video quality to enhance the system's QoE. The video content is transmitted over an Intelligent Reflecting Surface (IRS) aided Rate Splitting Multiple Access (RSMA) downlink network.

The objectives of server-side adaptation are twofold: stabilize player experience by minimizing video quality fluctuations and mitigate the negative impact of bandwidth variations on streaming performance [4].

In DRL, Deep Neural Networks (DNNs) approximate the agent's optimal strategy. The generalization power of DNNs facilitates the solution of high-dimensional problems in IoMT networks [11, 12]. DRL proves instrumental in obtaining solutions for sophisticated network optimizations, enabling the system to tackle complex challenges related to joint wireless resource optimization and bitrate adaptation in real-time.

We deploy an RSMA network, emerging as a promising solution to mitigate interference, achieve a greater level of spectral efficiency, and provide increased degrees of freedom compared to Nonorthogonal Multiple Access (NOMA), or OFDMA networks [13–15]. RSMA, by allowing users to transmit multiple data streams concurrently, efficiently exploits the wireless channel, resulting in improved data rates and enhanced QoE. In scenarios where signal strength weakens due to distance or interference, we implement an IRS. This reconfigurable surface, with adjustable elements, controls how radio waves reflect, optimizing signal distribution and reducing interference [16, 17]. This benefits users near and far from the BS, ensuring a stronger and more reliable signal for all [18].

Briefly, the principal contributions of this paper are elucidated in the subsequent points:

- **Joint optimization of wireless resources and video bitrate bitrate**: We leverage the unique capabilities of a potential advanced IRS-aided RSMA network to dynamically adjust both radio resources and video bitrates in real-time, maximizing overall QoE for IoMT devices.

- **DRL-powered QoE management**: We formulate the QoE optimization problem as a Markov Decision Process (MDP) and employ Proximal Policy Optimization (PPO), a powerful DRL technique, to learn an optimal policy for dynamic adaptation. This allows us to overcome the limitations of traditional methods and achieve better QoE performance under diverse network conditions.

- **Real-time QoE monitoring and evaluation**: Our approach continuously monitors and evaluates QoE metrics through the feedback of Channel State Information (CSI). This enables the DRL agent to continuously refine its adaptation policy and ensures smooth, high-quality streaming experiences for users.

- **Performance Assessment**: A public video streaming dataset [19] is used for the evaluation. Through meticulous simulations, the effectiveness of the proposed PPO-based algorithm is thoroughly assessed. The approach outperforms various baseline methods in terms of quality and latency, ultimately enhancing the overall streaming experience for users.

The subsequent sections of this paper are structured as follows to comprehensively address the outlined research objectives. Section 2 introduces relevant prior work. Section 3 details the explanation of the system model, providing a comprehensive understanding of the conceptual framework. Next, Section 4 describes the careful formulation of the problem, elaborating on the intricacies of IRS-aided RSMA IoMT streaming systems. The PPO algorithm development is explained in Section 5, clarifying the technicalities involved in its creation. Section 6 presents the empirical evaluation of the proposed approach. Finally, Section 7 culminates with conclusive insight from the research, elucidating the critical findings and implications of the study.

2. Related Work

In the contemporary multimedia environment, ensuring high-quality video streaming experiences and efficient resource utilization stands as a significant challenge. Achieving this objective requires optimal video service management and proactive system resource management. This section conducts a comprehensive review of the current research works within these domains, shedding light on key themes and solutions proposed by previous studies. Furthermore, we highlight limitations that motivate the development of our novel approach, presented in subsequent sections.

2.1. Video Service Management

Mao et al. [20] introduced a client-based ABS model, enhancing QoE through historical network throughput data. This algorithm utilizes past and upcoming video segment details to inform future decisions. Despite notable QoE improvements, their approach relies on linear functions for video segment quality assessment, potentially lacking precision in representing user visual experience. In [21], the authors designed an algorithm for joint bitrate adaptation and video quality enhancement at the client-side to maximize QoE in dynamic wireless networks with limited computation capacity. However, the use of Peak Signal-to-Noise Ratio (PSNR) for video assessment proved less effective than Video Multimethod Assessment Fusion (VMAF).

Ma et al. [22] proposed a QoE-aware ABS solution using DRL to dynamically adjust video stream bitrates based on client states and network conditions. Additionally, a study by Ma et al. [23] addressed bandwidth competition challenges in video streaming services with a server-based ABS model, considering historical data on network dynamics and client behaviors. Liu et al. [24] presented an ABS system with Edge-Client collaborative Super-resolution to enhance users' QoE, considering limited network bandwidth and computing resources. The common use of VMAF for video quality evaluation in these three works indicates its effectiveness in assessing system QoE.

However, it's crucial to note that the studies did not specify the multiple access technologies used in simulations, relying solely on available public network tracing datasets. Additionally, they utilized historical network information, such as throughput and bandwidth, which may offer limited value in rapidly changing network conditions requiring frequent updates.

2.2. System Resource Management

In the context of wireless video streaming systems, the primary focus of research centers on the adaptation of wireless radio resources, encompassing aspects like transmission power [10, 25–27, 29] or traffic consumption [30] between the BS and users. The overarching aim is to optimize both the wireless radio resources and the QoE of streaming services.

Researchers have delved into solutions for enhancing wireless video streaming by capitalizing on frequent updates of Channel State Information (CSI) to augment QoE for multimedia services and optimize wireless resources. With the CSI being updated at a faster timescale in milliseconds, it offers a more immediate and responsive insight into network conditions, making it highly effective for real-time streaming services. In [25], a method involving joint optimization of streaming rate control and power transmission is proposed, with the primary goal of minimizing power consumption and addressing challenges like playback overflow and rebuffering. On the other hand, Li et al. [26] addressed sustainable playback buffer stability through the optimization of power and subcarrier allocation. Similarly, Ye et al. [27] introduce a network-assisted ABS model based on CSI to minimize power transmission while ensuring uninterrupted video playback. However, these approaches are specifically designed for OFDMA networks, known to have lower performance compared to NOMA or RSMA networks. Moreover, these studies solely focus on events like buffer underflow or overflow, overlooking the significance of video bitrate or quality in evaluating the QoE.

In [10], a network-assisted ABS model is proposed for a NOMA network, entailing joint optimization of power allocation and bitrate adaptation. Additionally, Dao et al. [28] put forth ABS services within multi-user downlink NOMA edge caching systems, incorporating imperfect successive interference cancellation (SIC). The primary objective is to optimize the video bitrate for online streams with a focus on maximizing bitrate while ensuring uninterrupted playback smoothness. Nevertheless, these two works rely on the NOMA mechanism, known for its lower spectral efficiency and inferior interference management compared to RSMA. Furthermore, their evaluation is based solely on video bitrate, lacking the use of metrics such as PSNR or VMAF, which are essential for accurately reflecting the visual experience on the client side.

It is noteworthy that the previously mentioned studies neglect the crucial role of video quality in evaluating user QoE. These works fall short of comprehensively considering the three main aspects of QoE in streaming services, namely video quality, quality switching, and rebuffering, as they predominantly focus on one or two of these dimensions. Furthermore, there is a predominant emphasis on optimizing network power, neglecting the significance of optimizing other wireless resources, such as beamforming vectors. The optimization of beamforming vectors is instrumental in enhancing the Signal-to-Noise Ratio (SNR) at receivers, resulting in elevated data rates, improved interference management, and overall enhanced transmission efficiency. This optimization directly influences multimedia content quality and transmission latency—both critical factors in augmenting QoE [31–33].

Table 1. Related works

Network factors	QoE metrics	Multiple Access	Video Assessment	Reference
Channel state information	Video bitrate, Bitrate Switching	NOMA	No consideration	[10]
Network throughput	Video quality, Quality Switching, Rebuffering	Not specified	Linear function	[20]
Network throughput	Video quality, Quality Switching, Rebuffering	Not specified	PSNR	[21]
Network throughput	Video quality, Quality Switching, Rebuffering	Not specified	VMAF	[22]
Probed bottleneck delay	Video quality, Quality Switching, Rebuffering	Not specified	VMAF	[23]
Network bandwidth	Video quality, Quality Switching, Rebuffering	Not specified	VMAF	[24]
Channel state information	Rebuffering, Playback overflow	Not specified	No consideration	[25]
Channel state information	Rebuffering	OFDMA	No consideration	[26]
Channel state information	Rebuffering, Playback overflow	OFDMA	No consideration	[27]
Channel state information	Video bitrate, Rebuffering	NOMA	No consideration	[28]
Channel state information	Video quality, Quality Switching, Rebuffering	RSMA	VMAF	Our work

In table 1, we have identified crucial research gaps observed in previous studies, with a specific focus on limitations in network factors, QoE metrics, multiple access technology, and video assessment methods. Our research objectives are designed to effectively bridge these identified gaps. To overcome the limitations identified in prior research, we leverage real-time CSI feedback to achieve accuracy and responsiveness in QoE monitoring and adaptation. This allows us to dynamically adjust bitrates and resource allocation based on the actual network conditions, unlike traditional methods that rely on estimations or outdated information, leading to significantly reduced rebuffering and smoother playback. Our approach takes into account the three key dimensions of QoE assessment in multimedia services, namely video quality, quality switching, and rebuffering. This provides a more nuanced understanding of user experience and enables our DRL agent to optimize for overall QoE satisfaction. We also employ VMAF as a video assessment method for more accuracy. We utilize the unique capabilities of an IRS-aided RSMA network to overcome the limitations of traditional OFDMA and NOMA systems. By dynamically controlling IRS phase shifts, we achieve efficiency in resource allocation, ensuring consistent QoE for all users even in resource-constrained IoMT environments.

3. System Model

Figure 1 illustrates the overarching QoE Management paradigm for the IoMT streaming system. In our architecture, the BS incorporates both the Adaptive Bitrate Controller (ABC) module and the QoE Management module. At the heart of our system lies the server-side ABC module, powered by DRL. The BS receives live video streams and transmits them to IoMT devices through a wireless link. The QoE management continuously monitors the network conditions (CSI) user buffer levels, and video bitrate information through the QoE Monitoring module. This information is then fed into the QoE Evaluation module, which utilizes a multi-faceted QoE model to estimate the overall QoE for each user.

Based on the QoE evaluation, the DRL model within the ABC module dynamically adjusts the video bitrate for each user and optimizes parameters such as beamforming vectors, IRS phase shifts, and RSMA parameters. This optimization aims to maximize network efficiency, therefore improving video quality, and reducing transmission latency, ultimately leading to a superior QoE for all users. The DRL model continuously learns and refines its policy through a feedback loop from the QoE Evaluation module, ensuring adaptation to changing network conditions and user preferences.

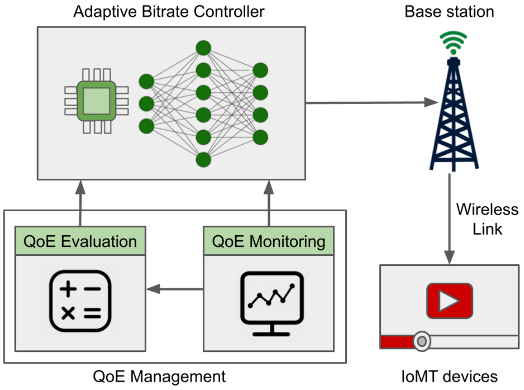

Figure 1. Overview of QoE Management for IoMT streaming systems.

In subsequent sections, we delve deeper into the details of the network model, multimedia model, and QoE model, providing a comprehensive analysis of the system's functionality and performance.

3.1. Channel Model

In this work, we consider an IRS-assisted multiple-input single-output downlink channel, as illustrated in Fig. 2. This configuration involves several key components: one BS, a set $\mathcal{K}$ of K users, and an IRS. Each user has a single antenna, whereas the BS has M transmit antennae. The IRS has a collection of passive reflecting elements, each indexed by $\mathcal{N} = \{1, 2, \dots, N\}$. For each user k, we let s_k represent the transmitted signal, and w_k denotes the corresponding beamforming vector. Additionally, in the context of RSMA, the BS transmits a common message signal s_0 with the corresponding beamforming vector w_0.

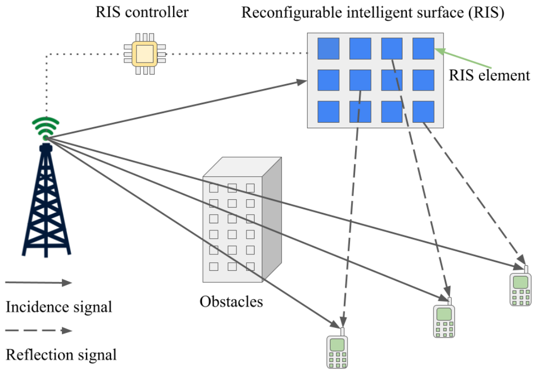

Figure 2. IoMT streaming system over IRS-aided RSMA network.

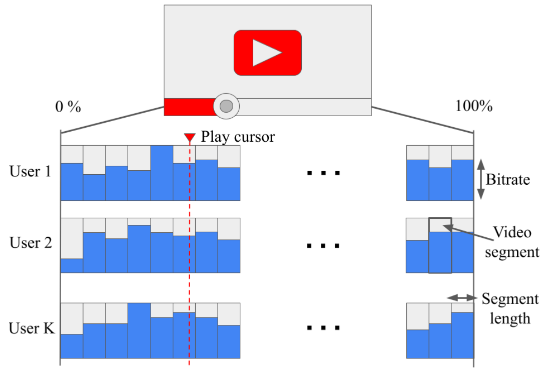

Figure 3. Video segment model.

Each signal s_k is assumed to have a zero mean and unit variance, represented by the expression $\mathbb{E}[s_k s_k^H] = 1, \forall k \in \mathcal{K} \cup 0$. The signal transmitted from the BS is mathematically represented as follows:

$$x = \sum_{k=0}^{K} w_k s_k. \tag{1}$$

The signal captured by user k is expressed as follows:

$$y_k = \left(g_k^H + h_k^H \Theta G\right) \sum_{i=0}^{K} w_i x_i + n_k, \tag{2}$$

where $g_k \in \mathbb{C}^M$, $G \in \mathbb{C}^{N\times M}$, and $h_k \in \mathbb{C}^N$ represent the channel responses from the BS to user k, BS to the IRS, and from the IRS to user k, respectively, and $n_k \sim \mathcal{CN}(0;\sigma^2)$ signifies additive white Gaussian noise. The phase-shift matrix Θ for the IRS is a diagonal matrix $\text{diag}(e^{j\theta_1}, e^{j\theta_1}, \ldots, e^{j\theta_N}) \in \mathbb{C}^{N\times N}$, where $\theta_n \in [0, 2\pi]$ represents the phase shift introduced by the nth element of the IRS to the incoming signal. Consequently, the achievable rate for the common message s_0 at the user k, is formulated as follows:

$$c_k = \log_2\left(1 + \frac{|(g_k^H + h_k^H \Theta G) w_0|^2}{\sum_{i=1}^{K} |(g_k^H + h_k^H \Theta G) w_i|^2 + \sigma^2}\right), \tag{3}$$

All users must decode the common message before deciphering their private messages as a prerequisite for subsequent actions [34]. This two-step process ensures all users understand the shared information and remove it from the received signals to further decode the respective private messages. The appropriate rate for transmitting the common message must be selected as the minimum among the rate for all users, denoted as $\min_{k\in K} c_k$, to guarantee the effective decoding of the common message across all users. The total data rates for all users receiving the common message should not surpass the rate assigned to the common message considering the chosen common message rate $\min_{k\in K} c_k$ and individual rate allocation a_k for user k, which is determined as follows:

$$\sum_{k=1}^{K} a_k \leq \min\{c_k, \forall k \in K\}, \tag{4}$$

Once each user has successfully decoded the common message s_0, the subsequent step involves decoding their respective private messages. The achievable rate for user k to decode its private message s_k can be determined as follows:

$$r_k = \log_2\left(1 + \frac{|(g_k^H + h_k^H \Theta G) w_k|^2}{\sum_{i=1, i\neq k}^{K} |(g_k^H + h_k^H \Theta G) w_i|^2 + \sigma^2}\right). \tag{5}$$

The overall transmission rate for user k can be expressed as follows, considering the common message rate a_k and the achievable rate for the private message r_k:

$$r_k^{sum} = a_k + r_k. \tag{6}$$

3.2. *Multimedia*

We established a set $\mathcal{S}$ containing elements $\{1, 2, 3, \ldots, S\}$ to represent the individual segments constituting a video stream, as illustrated in Fig. 3. Each segment spans τ s and is encoded at various bitrate levels. The size in bytes of the sth segment for user k is denoted as $d_k(s)$. As illustrated, the video stream is divided into segments with varying bitrates, allowing for dynamic adaptation based on network conditions and user buffer levels. The QoE monitoring module continuously gathers information about these factors, feeding it to the ABC module. This information is used by the DRL component within the ABC to determine the optimal bitrate for each segment, ensuring smooth playback and high QoE for users even in dynamic IoMT environments.

This study focuses on a synchronous IoMT streaming framework. Initially, all users attempt to download segments at the lowest bitrate level. Once this stage concludes, ABS is initiated. Users download the sth segment when the $(s-1)$th segment begins playback. The buffer undergoes depletion during video playback, and upon a successful download, the buffer experiences a τ-s increase.

3.3. *Quality of Experience Model*

3.3.1. *Perceptual Quality*

The bitrate has a significant influence on the QoE in multimedia applications. Higher bitrates generally result in better video and audio quality. When the bit rate increases, more data are allocated to each frame or sample, improving the resolution, detail, clarity, and fidelity. This outcome results in a more immersive and enjoyable viewing or listening experience for users.

Video multimethod assessment fusion (VMAF) [35] is a metric for video quality that combines multiple objective quality assessment methods to provide a comprehensive and perceptually accurate evaluation of video quality. It is designed to mimic human perception and is widely used to assess the visual quality of videos. The fusion of multiple quality assessment methods in VMAF captures various aspects of video quality, making it a more reliable and versatile metric than individual assessment methods. It has gained popularity and is widely used in video encoding, streaming, and other multimedia applications to evaluate and optimize the quality of video content. This work employs VMAF to determine the users' perceptual quality. We define $q(\cdot)$ mapping between VMAF and bitrates of the segments. The perceived quality of segment s at user k can be described as follows:

$$Q_k(s) = q(b_k[s]), \tag{7}$$

where $b_k[s]$ is the video bitrate of the sth segment at user k.

3.3.2. Temporal Quality Oscillations

Video quality switching adversely affects the QoE in multimedia services, causing frustration and interrupting the user experience. When a video quality switch suddenly occurs, a visible change often exists in the perceived video quality. If the switch leads to a significant drop in quality, such as lower resolution or increased compression artifacts, it can create a jarring experience and break immersion. Users may perceive the video as less enjoyable or even unacceptable, leading to a negative effect on the QoE. Therefore, smooth transitions during quality switches of successive segments, such as fading or intensifying between quality levels, can minimize the visual effects and reduce the perceived quality degradation. This type of temporal quality fluctuation of successive segments at user k can be calculated as follows:

$$\Delta Q_k(s) = \frac{1}{s}\sum_{i=1}^{s} |q(b_k[i]) - q(b_k[i-1])|. \tag{8}$$

3.3.3. Re-buffering

Rebuffering, stalling, or starvation in multimedia services refers to the interruption or pause in the playback due to loading problems. Stalling disrupts the immersive experience and engagement with the content. Users expect a seamless and uninterrupted streaming experience. When playback stalls, it breaks the narrative or visual experience flow, creating a disjointed viewing process. Therefore, mitigating the effects of stalling on the QoE is crucial.

We assume the latency associated with other tasks, such as data storage orf message division for RSMA, is insignificant. Given this assumption, we can calculate the transmission latency of the sth segment for user k using the following equation:

$$l_k(s) = \frac{d_k(s)}{B \times r_k^{sum}}, \tag{9}$$

where $d_k(s)$ signifies the size in bytes of the video segment indexed as s transmitted to user k, and B represents the available bandwidth.

We let $T_k^{buf}(s)$ (in seconds) be the amount of buffered video for user k when the BS starts broadcasting the sth segment. Hence, the starvation time of segment s at client k can be expressed as follows:

$$T_k^L(s) = \max\{l_k(s) - T_k^{buf}(s), 0\}. \tag{10}$$

The starvation time over the displaying time is a critical factor influencing the QoE in IoMT streaming. The starvation time over the displaying time of segment s at client k can be expressed as follows:

$$L_k(s) = \frac{T_k^L(s)}{T_k^L(s) + \tau}. \tag{11}$$

4. Problem Formulation

The QoE within multimedia services can be assessed through a comprehensive perspective encompassing several factors. This assessment involves evaluating the quality of individual video segments, accounting for the temporal variation in quality across consecutive segments, and addressing the adverse effects of the rebuffering time. A higher value of Q_k typically corresponds to a more favorable user experience, reflecting better video quality. Conversely, a higher ΔQ_k often indicates an unsatisfactory user experience, as it signifies significant fluctuations in quality over time. The concept of rebuffering, characterized by interruptions or delays in video playback due to buffering or network problems, can substantially influence the QoE. The rebuffering time denotes the duration during which users await the resumption of video playback, causing frustration and interrupting the seamless viewing experience. Therefore, in evaluating the multimedia service QoE, the inherent quality of the content must be considered while accounting for the detrimental influences of quality fluctuations and rebuffering instances.

$$QoE_k(s) = Q_k(s) - \alpha\Delta Q_k(s) - \beta L_k(s), \tag{12}$$

where α and β are non negative weighting parameters corresponding to the variation in quality and rebuffering of content.

The optimization of the QoE for the sth segment is formulated as follows:

$$\max_{\boldsymbol{\theta}[s],\boldsymbol{w}[s],\boldsymbol{a}[s],\boldsymbol{b}[s]} \sum_{k=1}^{K} QoE_k \tag{13}$$

$$\text{subject to } \sum_{k=1}^{K} a_k \leq \min\{c_k, \forall k \in K\} \tag{14a}$$
$$a_k \geq 0, \forall k \in \mathcal{K} \tag{14b}$$
$$B \times r_k^{sum} \geq R_{min}, \forall k \in \mathcal{K} \tag{14c}$$
$$||\boldsymbol{w}[\boldsymbol{s}]||_2^2 \leq P_{max} \tag{14d}$$
$$\theta_n \in [0; 2\pi], \forall n \in \mathcal{N} \tag{14e}$$
$$b_{min} \leq b_k[s] \leq b_{max} \tag{14f}$$
$$\alpha, \beta > 0, \tag{14g}$$

where $\boldsymbol{\theta}[\boldsymbol{s}] = [\theta_1, \ldots, \theta_N]^T$, $\mathbf{w}[\boldsymbol{s}] = [w_0, \ldots, w_K]^T$ and $\boldsymbol{a}[\boldsymbol{s}] = [a_1, \ldots, a_K]^T$, $\boldsymbol{b}[\boldsymbol{s}] = [b_1[s], \ldots, b_K[s]]$, b_{min}, b_{max} are the minimum and maximum bitrates of the video segments, respectively. In addition, R_{min} denotes the minimum data rate to ensure the QoS. Constraints (14a) to (14b) ensure the decodability of the common message across all users. Constraint (14c) reflects the QoS constraint. Constraint (14d) limits the BS transmit power. Constraint (14e) addresses the IRS phase-shift limits. Last, Constraint (14f) defines the range of admissible bitrates for the video segment.

5. Proposed Solution

The initial strategy to swiftly enhance the QoE within the IoMT streaming system entails converting the system into an MDP. Subsequently, we employed the PPO-based approach to effectively resolve this transformed scenario. In this context, the BS is positioned as the principal agent. As each time step, denoted as t, advances, the involvement of the agent evolves by observing the state s_t, executing the action a_t, and subsequently receiving the reward r_t. The resolution of the optimal approach for addressing the problem (13) manifests through the accumulation of multiple time steps.

5.1. Markov Decision Process Formulation

A MDP formally represents a sequential decision-making problem under uncertainty. It consists of a set of states, representing possible situations, and a set of actions available in each state. Taking an action in a state leads to a new state with a certain probability, and each transition comes with a reward, quantifying the desirability of the outcome. The objective of an MDP is to choose a policy, mapping states to actions, that maximizes the expected cumulative reward over time. MDPs offer a powerful framework for modeling and solving problems in various domains, such as reinforcement learning, resource management, and economic analysis.

- **State space** During each discrete time step, the BS must acquire pertinent details concerning the transmission conditions and the specifics of the video segment. These inputs are vital for facilitating informed decision-making processes. We define d_s as the file size and q_s as the VMAF score of segment s at different bitrates. The state is expressed as follows:

$$\mathcal{S} = \left[G, \mathbf{g}, \mathbf{h}, \mathbf{b}[s-1], \mathbf{T}^{buf}(s), d_s, q_s\right], \tag{15}$$

where $\mathbf{g} = \{g_k, \forall k\}$, $\mathbf{h} = \{h_k, \forall k\}$, $\mathbf{b}[s-1] = \{b_k[s-1], \forall k\}$, and $\mathbf{T}^{buf}(s) = \{T_k^{buf}(s), \forall k\}$.

- **Action space**

Under the present conditions of each user, the agent chooses and implements actions guided by policy π [36]. As alterations occur in the network state for users and the forthcoming quality of the video segment, the agent undertakes modifications to the phase-shift matrix, beamforming vector, common rate, and video bitrate for each user to improve the system QoE. A collection of optimization variables of the problem (13) is encompassed within the action vector denoted by A.

$$\mathcal{A} = [\boldsymbol{\theta}[s], \boldsymbol{w}[s], \boldsymbol{a}[s], \boldsymbol{b}[s]] \tag{16}$$

- **Reward** The reward function was formulated to maximize the system's QoE, as delineated in equation (13). In the pursuit of promoting QoE optimization while diligently considering the QoS requisites of users, the BS is subject to penalties for any QoS violations among users. At given time step t, the BS undertakes observation of the present state $s_t \in \mathcal{S}$, enacts an action $a_t \in \mathcal{A}$, and subsequently acquires an instantaneous reward $r_t(s_t, a_t)$. The immediate reward is established through the subsequent definition:

$$r_t(s_t, a_t) = \sum_{k=1}^{K} QoE_k(1 - \kappa_t), \tag{17}$$

where κ_t signifies the punitive consequence incurred by the BS for an action a_t that fails to meet the QoS criteria delineated in Constraint (14c). Further, the penalty κ_t corresponding to the specific time step t is

$$\kappa_t = \frac{1}{K} \sum_{k=1}^{K} \text{sgn}(R_k - R_{min}), \tag{18}$$

$$\text{sgn}(x) = \begin{cases} 1 & \text{if } R_k > R_{min} \\ 0 & \text{if } R_k < R_{min}. \end{cases} \tag{19}$$

In contrast to the QoS constraint, the common rate and power constraints must remain unbroken during any individual time step due to the inherent constraints from the total BS transmission power limitations and the essential requirement for the common message to be decodable across all users. Consequently, we chose not to incorporate penalties into the immediate reward for violations of the common rate and power constraints. Instead, we integrated these constraints into the algorithm framework, as explained in the next section.

5.2. Proximal Policy Optimization-based Approach

By formulating the problem as an MDP, we utilized the state-of-the-art PPO algorithm to learn an optimal policy that dynamically selects actions in each state, maximizing the expected cumulative QoE for users over time. This DRL algorithm was developed by OpenAI organization [37, 38], and adheres to the actor-critic framework. In this framework, the actor network generates precise actions based on a given state, and the critic network provides a value function to assess the performance of the actor network, effectively adapting to the dynamic nature of IoMT environments. In our context, the actor network, informed by the current network state and user state, generates optimal bitrate and resource allocation decisions. Simultaneously, the critic network evaluates the long-term impact of these choices on overall QoE, guiding the actor network toward achieving superior performance. A noteworthy advantage of PPO is its adaptability to continuous and discrete action spaces within various environments. Moreover, PPO shares certain attributes with the trust region policy optimization (TRPO) method regarding dependability and stability. However, due to its first-order optimization nature, it surpasses TRPO in terms of generality and ease of implementation. As explicitly detailed in [37], PPO optimizes a clipped surrogate objective function, defined as

$$J(\theta) = \hat{\mathbb{E}}\left[\min(u_t(\theta)\hat{A}_{\theta_{old}}(s_t, a_t), \text{clip}(u_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_{\theta_{old}}(s_t, a_t)\right], \tag{20}$$

where $\hat{A}_{\theta_{old}}(s_t, a_t)$ corresponds to an advantage function derived through a generalized advantage estimation. The parameter ϵ denotes a predefined clipping threshold. Finally, $u_t(\theta)$ represents the probability ratio between the policies of the old and new iterations, formulated as follows:

$$u_t(\theta) = \left[\frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}\right]. \tag{21}$$

Particularly, the clip function serves to prevent excessive policy updates, effectively addressing the problem of a drastic performance decline. In other words, its role is to curtail any incentives for the new policy to significantly diverge from the old policy. Consequently, if $\hat{A}_{\theta_{old}}(s_t, a_t)$ has a positive value, the value of $u_t(\theta)$ is bounded by $1+\epsilon$. Conversely, when $\hat{A}_{\theta_{old}}(s_t, a_t)$ is negative, the value of $u_t(\theta)$ is constrained by $1-\epsilon$. The calculation of $\hat{A}_{\theta_{old}}(s_t, a_t)$ is

$$\hat{A}_{\theta_{old}}(s_t, a_t) = \delta_t + (\gamma\lambda)\delta_{t+1} + \ldots + (\gamma\lambda)^{L-t}\delta_U, \tag{22}$$

where L signifies the size of the mini-batch, and the parameter γ denotes the discount factor. Additionally, λ represents a parameter associated with the generalized advantage estimation technique to reduce the variance and facilitate more stable training. Subsequently, δ_t denotes the temporal difference error, computed as follows:

$$\delta_t = r_t + \gamma V_\phi(s_{t+1})V_\phi(s_t), \tag{23}$$

where $V_\phi(s_t)$ represents the value-function approximation. By minimizing the subsequent loss function, the critic network can undergo enhancements:

$$J(\phi) = \hat{\mathbb{E}}\left[(V_\phi(s_t) - \hat{V}_t)^2\right]. \tag{24}$$

The computation of $\hat{V}_t$ is expressed as follows:

$$\hat{V}_t = \sum_{j=t}^{U} \gamma^{j-t} r_j. \tag{25}$$

We assigned ξ_c to represent the learning rate applied to the critic network

$$\hat{V}_t = \sum_{j=t}^{U} \gamma^{j-t} r_j \tag{26}$$

We assigned ξ_c to represent the learning rate applied to the critic network, and ξ_a signifies the learning rate applied to the actor network. Finally, the update process involves adjusting the parameters of the actor network θ using mini-batch stochastic gradient ascent and the critic parameter ϕ using mini-batch stochastic gradient descent, as indicated below:

$$\theta \leftarrow \theta + \xi_a \nabla_\theta J(\theta), \tag{27}$$

$$\phi \leftarrow \phi - \xi_c \nabla_\phi J\phi). \tag{28}$$

Given a state, the agent selects an action based on the output of the actor network at each time step t, receives a reward, and transitions to the next state. This trajectory, represented as (s_t, a_t, r_t, s_{t+1}), is stored in the replay buffer $\mathcal{B}$. The agent collects L samples from this buffer to form a mini-batch and subsequently iterates through V updates of its network parameters using the Adam optimizer. Additionally, the architecture of the actor and critic networks involves two fully connected hidden layers, each comprising 256 neurons. The rectified linear unit activation function is applied within these hidden layers. Moreover, the actor-network output layer is followed by a tanh layer.

To ensure adherence to the common rate vector constraint specified in (14a) to (14b), we employed the softmax activation function for actions related to the common rate, following the network generation of an action as the output. By integrating Constraints (14e) and (14f), we adjust the network output actions to a suitable range that aligns with the conditions set by these constraints. Algorithm 1 presents the pseudo-code outlining the proposed PPO-based QoE management strategy to comprehensively understand this approach.

6. Performance Evaluation

This section extensively examines the effectiveness of the PPO algorithm. This assessment involves a comprehensive comparison with the alternative benchmark methodologies. The evaluation is achieved through a series of simulation experiments conducted in a programming environment using Python 3.10.12 and PyTorch 2.0.1, executed on a personal computer equipped with the following hardware specifications: a 12th-generation Intel Core i7-12700 CPU operating at 4.90 GHz with 32.0 GB of RAM and an NVIDIA GeForce GTX 3070 Lite Hash Rate graphics card featuring CUDA 12.2.

Algorithm 1: Proximal policy optimization-based quality of experience management algorithm

```
Initialize parameters of the actor network $\theta$
Initialize parameters of the critic network $\phi$
Initialize the old actor parameter $\theta_{old} \leftarrow \theta$
for $ep \leftarrow 1$ to $max\ episode$ do
    Set the initial state $s_t$
    for $step \leftarrow 1$ to $max\ step$ do
        Obtain the current state $s_t$
        Generate an action $a_t$ based on the old actor $\pi_{\theta_{old}}$
        Get the reward $r_t$ and the next state $s_{t+1}$
        Store $\{s_t, a_t, r_t, s_{t+1}\}$ in the replay buffer $\mathcal{B}$
        if $size\ of\ \mathcal{B}\ equals\ U$ then
            for $m \leftarrow 1$ to $V$ do
                Using the sampled data from $\mathcal{B}$
                Compute $J(\theta)$ by (20) then update the
                  parameter of actor network $\theta$ by (27)
                Compute $J(\phi)$ by (24) then update the
                  parameter of the critic $\phi$ by (28)
            end
            Update the old actor parameter $\theta_{old} \leftarrow \theta$
        end
    end
end
```

Table 2. Simulation parameter configurations.

Parameter	Value
Duration of each video segment τ	4 sec
Initial buffer size	24 sec
QoE weighting parameter α	1.0
QoE weighting parameter β	200.0
Mini-batch size L	300
Generalized advantage estimation parameter λ	0.95
Clipping rate ϵ	0.2
Discount factor γ	0.99
Number of steps per episode	300
Number of episodes	10000
Critic network learning rate ξ_c	0.0003
Actor-network learning rate ξ_a	0.0003

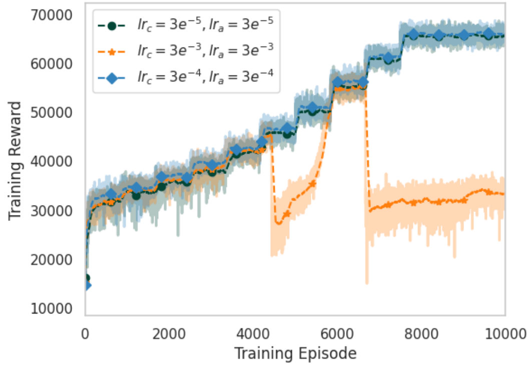

Figure 4. Comparison of the reward proximal policy optimization by learning rate (K = 3).

6.1. Data Description and Experiment Setup

This study employs the comprehensive video quality dataset Comyco [19], which was designed explicitly for dynamic adaptive streaming over the Hypertext Transfer Protocol (DASH) scenarios and offers a wide variety of content types, including movies, games, sports, news, television shows, and music videos. The dataset involves breaking down these video clips into smaller units called segments to facilitate effective analysis and evaluation. These segments adhere to a predefined encoding ladder, ensuring consistency and comparability across various video sequences.

The encoding ladder corresponds to a hierarchy of bitrate levels, specifically {235, 375, 560, 750, 1050, 1750, 2350, 3000, 4300} kbps. Each segment (or "chunk") is encoded to last for 4 seconds. This segmentation strategy facilitates adaptive streaming by allowing devices to dynamically adjust the video stream's quality in response to the available network conditions. Each video undergoes assessment using the VMAF metric as part of the quality evaluation process. This quality metric offers insight into the perceptual quality of the videos by considering factors that influence the user experience. The assessments are conducted using a reference resolution of 1920×1080, ensuring consistency in the evaluation process. We assumed that initially, the buffer size for each user was set at 24 seconds.

In the context of the network configuration, the setup involves a BS equipped with six antennas (M = 6), and the IRS comprises six reflecting elements (N = 6). The transmit power of the BS is standardized at 1 W. When considering the communication channel characteristics, we set the path loss exponent to 3.8 for direct channels that connect the BS to users. Additionally, for channels that link the IRS to users, the path loss exponent is established at 2.2. The noise power spectral density, represented as σ^2, is valued at -174 dBm/Hz. Another critical consideration is ensuring the user experience regarding receiving video segments at the minimum required bitrate before the subsequent segment download initiates. A minimal rate threshold, R_{min}, is set to 235,000 bits/s to fulfill this requirement.

The weighting parameters linked to the quality variation (α) and content rebuffering (β) were established at 1.0 and 200.0, respectively. The PPO algorithm undergoes training for 10,000 episodes, with each episode consisting of 300 steps. The simulation's system parameters are outlined in Table 2.

6.2. Convergence Performance of Proximal Policy Optimization

We analyzed the convergence behavior exhibited by the proposed PPO algorithm. We monitored the acquired reward over multiple training iterations to evaluate its performance. Additionally, we assessed the evolution of the achievable sum-rate throughout the training process, as this metric significantly influences content quality and latency during user delivery. In addition, Figs. 4 and 5 present a graphical depiction of these metrics. Furthermore, we conducted a comparative evaluation

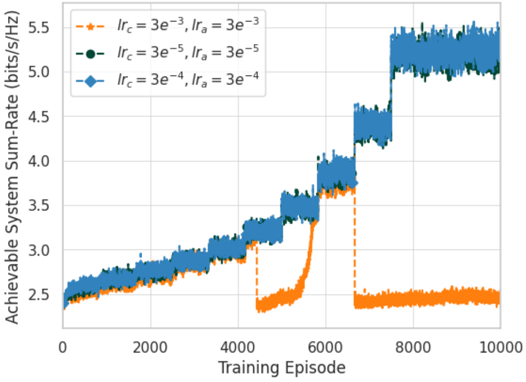

Figure 5. Comparison of the achievable system sum rate proximal policy optimization by learning rate ($K = 3$).

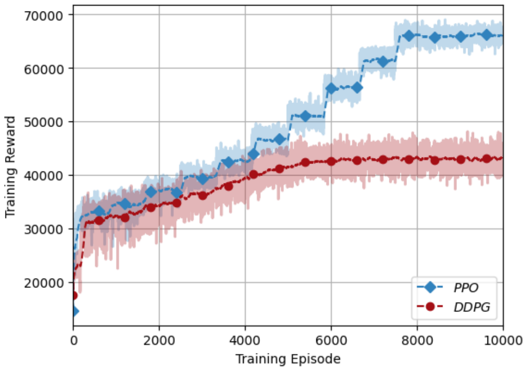

Figure 6. Comparison of the rewards for PPO and DDPG algorithm ($K = 3$).

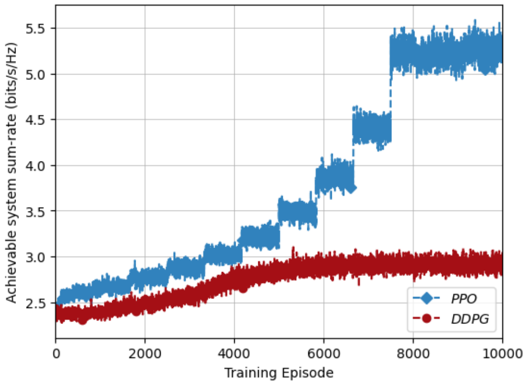

Figure 7. Comparison of the achievable rate between PPO and DDPG algorithm ($K = 3$).

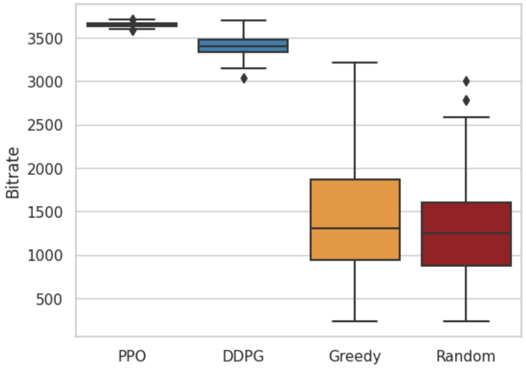

Figure 8. Comparison of the average bitrate ($K = 3$).

by integrating the deep deterministic policy gradient (DDPG) algorithm [39] into the experimental setup.

The PPO algorithm effectively enhances the learned policy by selecting an appropriate learning rate, such as 3×10^{-4}. The outcomes illustrated in Fig. 4 emphasize that employing an excessively high learning rate (i.e., 3×10^{-5}) can result in inefficient policy learning. Specifically, with a learning rate of 3×10^{-5}, a slightly lower reward is observed. A similar trend is noticeable regarding the achievable sum rate, as illustrated in Fig. 5.

We thoroughly examined Fig. 6, observing that during the initial 6,000 episodes, the performance of PPO exhibited a slight advantage over the DDPG algorithm. However, an intriguing shift occurred between episodes 6,000 and 10,000, where PPO notably improved over the DDPG algorithm. Between episodes 6000 and 10000, as the DDPG algorithm maintains coverage, its reward remains stable. Conversely, the PPO algorithm exhibits a consistent increase in reward, stabilizing after the 9000th episode. The comparative stability of PPO is apparent in Fig. 6, illustrating that the PPO algorithm achieved higher rewards post-training and displayed reduced variability compared to DDPG.

Furthermore, an examination of the achievable system sum rate reinforced the superiority of the PPO over the DDPG algorithm. Additionally, Fig. 7 indicates that PPO exhibited superior performance. Specifically, the achievable system sum rate attained by PPO was 185% higher than that achieved by the DDPG. This observation underscores the substantial enhancement the PPO algorithm offers regarding system efficiency and overall performance.

6.3. *Performance Comparison*

Comparing the average video segment bitrates reveals that the application of the PPO algorithm yields higher bitrates and enhanced stability than the DDPG, greedy, and random algorithms, demonstrated in Fig. 8.

Likewise, the pattern observed in Fig. 8 is consistent with the trend in VMAF scores depicted in Fig. 9. Although the average bitrate achieved by the PPO algorithm surpasses that of the greedy algorithm by a factor of 2.5, the corresponding VMAF score improvement is modest at a 35% increase over the score for the greedy algorithm. This outcome is due to the nonlinear correlation between the VMAF score and video bitrates.

While the average bitrate improvement provided by the PPO algorithm over the DDPG algorithm is only marginal, it is crucial to consider the average rebuffering per segment (4 s). In this aspect, the PPO algorithm outperforms DDPG. Notably, in scenarios involving three, five, or eight users, the PPO algorithm experiences nearly no interruptions. However, with DDPG, the rebuffering time increases with a growing number of users, as presented in Table 3. This method leads to a rebuffering time

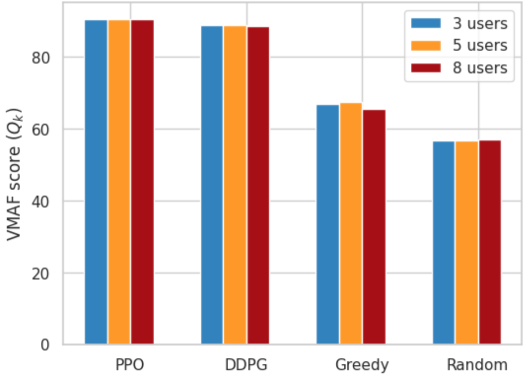

Figure 9. Comparison of the video multimethod assessment function (VMAF) score.

Table 3. Rebuffering time per video segment (in seconds)

	PPO	DDPG
3 users	0.0	0.020015
5 users	0.0	0.416756
8 users	$5.5 \times 10^{-}6$	0.531961

that constitutes approximately 12% of the video segment duration when the user count reaches five or eight. The primary cause lies in DDPG's lower system achievable rate, leading to playback interruptions for users.

The trend is evident from Fig. 10, where the achievable system sum rate declines as the number of users rises. However, the PPO algorithm demonstrates its capability to enhance the achievable system sum rate. Specifically, when dealing with eight users, the PPO algorithm achieves a sum rate that is 50% higher than that achieved by the DDPG algorithm. This comparison highlights the efficacy of the PPO algorithm in consistently maintaining higher and stabler bitrates for video segments and smoothness in video playback.

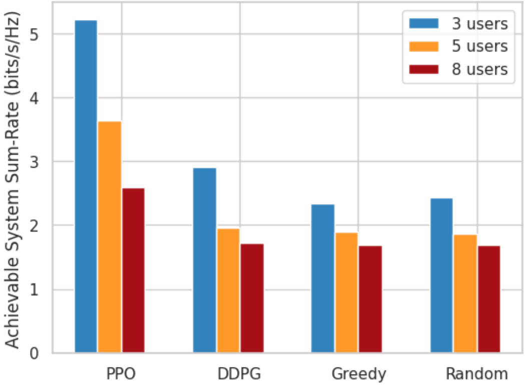

Figure 10. Comparison of the achievable system sum rate.

7. Conclusion

This work tackled the multifaceted challenge of delivering high-fidelity multimedia in resource-constrained IoMT networks, plagued by dynamism, heterogeneous devices, and limited capacity. We proposed a novel server-side bitrate adaptation approach, empowered by DRL-based QoE management, residing on an advanced IRS-aided RSMA network. This solution combats overestimated bandwidth, ensuring smoother playback with minimized quality fluctuations. By formulating the QoE optimization problem as an MDP and employing powerful PPO-DRL, we achieve real-time optimization of crucial aspects like bitrates, beamforming, IRS phase shifts, and RSMA parameters. Extensive simulations leveraging real-world datasets showcase that our PPO-based approach excels compared to baseline methods in both video quality and latency, ultimately leading to an enhanced streaming experience for IoMT users. While demonstrating effectiveness, future research directions include exploring large-scale deployments, dynamic content adaptation, and alternative DRL algorithms, contributing to the advancement of QoE in IoMT multimedia streaming.

Acknowledgments

This work was supported in part by the Chung-Ang University, South Korea, Young Scientist Scholarship in 2022, in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2023-00209125) and in part by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2024-RS-2022-00156353) supervised by the IITP (Institute for Information and Communications Technology Planning and Evaluation).

References

[1] M. Stoyanova, Y. Nikoloudakis, S. Panagiotakis, E. Pallis, E. K. Markakis, A survey on the internet of things (iot) forensics: challenges, approaches, and open issues, IEEE Communications Surveys & Tutorials 22 (2) (2020) 1191–1221.

[2] A. Nauman, Y. A. Qadri, M. Amjad, Y. B. Zikria, M. K. Afzal, S. W. Kim, Multimedia internet of things: A comprehensive survey, Ieee Access 8 (2020) 8202–8250.

[3] A. A. Barakabitze, N. Barman, A. Ahmad, S. Zadtootaghaj, L. Sun, M. G. Martini, L. Atzori, Qoe management of multimedia streaming services in future networks: A tutorial and survey, IEEE Communications Surveys & Tutorials 22 (1) (2019) 526–565.

[4] J. Kua, G. Armitage, P. Branch, A survey of rate adaptation techniques for dynamic adaptive streaming over http, IEEE Communications Surveys & Tutorials 19 (3) (2017) 1842–1866.

[5] N.-N. Dao, A.-T. Tran, N. H. Tu, T. T. Thanh, V. N. Q. Bao, S. Cho, A contemporary survey on live video streaming from a computation-driven perspective, ACM Computing Surveys 54 (10s) (2022) 1–38.

[6] N.-N. Dao, Internet of wearable things: Advancements and benefits from 6g technologies, Future Generation Computer Systems 138 (2023) 172–184.

[7] E. Yaacoub, Z. Dawy, Background on downlink resource allocation in ofdma wireless networks.

[8] R. O. Afolabi, A. Dadlani, K. Kim, Multicast scheduling and resource allocation algorithms for ofdma-based systems: A survey, IEEE Communications Surveys & Tutorials 15 (1) (2012) 240–254.

[9] J. Cui, Y. Liu, Z. Ding, P. Fan, A. Nallanathan, Qoe-based resource allocation for multi-cell noma networks, IEEE Transactions on Wireless Communications 17 (9) (2018) 6160–6176.

[10] J. Zhang, H. Wu, X. Tao, X. Zhang, Adaptive bitrate video streaming in non-orthogonal multiple access networks, IEEE Transactions on Vehicular Technology 69 (4) (2020) 3980–3993.

[11] N. C. Luong, D. T. Hoang, S. Gong, D. Niyato, P. Wang, Y.-C. Liang, D. I. Kim, Applications of deep reinforcement learning in communications and networking: A survey, IEEE Communications Surveys & Tutorials 21 (4) (2019) 3133–3174.

[12] A. Feriani, E. Hossain, Single and multi-agent deep reinforcement learning for ai-enabled wireless networks: A tutorial, IEEE Communications Surveys & Tutorials 23 (2) (2021) 1226–1252.

[13] Y. Mao, B. Clerckx, V. O. Li, Rate-splitting multiple access for downlink communication systems: Bridging, generalizing, and outperforming sdma and noma, EURASIP journal on wireless communications and networking 2018 (2018) 1–54.

[14] B. Clerckx, Y. Mao, R. Schober, E. A. Jorswieck, D. J. Love, J. Yuan, L. Hanzo, G. Y. Li, E. G. Larsson, G. Caire, Is noma efficient in multi-antenna networks? a critical look at next generation multiple access techniques, IEEE Open Journal of the Communications Society 2 (2021) 1310–1343.

[15] Y. Mao, O. Dizdar, B. Clerckx, R. Schober, P. Popovski, H. V. Poor, Rate-splitting multiple access: Fundamentals, survey, and future research trends, IEEE Communications Surveys & Tutorials.

[16] A. Bansal, K. Singh, B. Clerckx, C.-P. Li, M.-S. Alouini, Rate-splitting multiple access for intelligent reflecting surface aided multi-user communications, IEEE Transactions on Vehicular Technology 70 (9) (2021) 9217–9229.

[17] H. Li, Y. Mao, O. Dizdar, B. Clerckx, Rate-splitting multiple access for 6g—part iii: Interplay with reconfigurable intelligent surfaces, IEEE Communications Letters 26 (10) (2022) 2242–2246.

[18] Y. Liu, X. Liu, X. Mu, T. Hou, J. Xu, M. Di Renzo, N. Al-Dhahir, Reconfigurable intelligent surfaces: Principles and opportunities, IEEE communications surveys & tutorials 23 (3) (2021) 1546–1577.

[19] T. Huang, C. Zhou, R.-X. Zhang, C. Wu, X. Yao, L. Sun, Comyco: Quality-aware adaptive video streaming via imitation learning, in: Proceedings of the 27th ACM international conference on multimedia, 2019, pp. 429–437.

[20] H. Mao, R. Netravali, M. Alizadeh, Neural adaptive video streaming with pensieve, in: Proceedings of the conference of the ACM special interest group on data communication, 2017, pp. 197–210.

[21] J. Yang, Y. Jiang, S. Wang, Enhancement or super-resolution: Learning-based adaptive video streaming with client-side video processing, in: ICC 2022-IEEE International Conference on Communications, IEEE, 2022, pp. 739–744.

[22] X. Ma, Q. Li, L. Zou, J. Peng, J. Zhou, J. Chai, Y. Jiang, G.-M. Muntean, Qava: Qoe-aware adaptive video bitrate aggregation for http live streaming based on smart edge computing, IEEE Transactions on Broadcasting 68 (3) (2022) 661–676.

[23] X. Ma, Q. Li, Y. Jiang, G.-M. Muntean, L. Zou, Learning-based joint qoe optimization for adaptive video streaming based on smart edge, IEEE Transactions on Network and Service Management 19 (2) (2022) 1789–1806.

[24] X. Liu, Z. Ke, X. Zhou, T. Qiu, K. Li, Qoe-oriented adaptive video streaming with edge-client collaborative super-resolution, in: GLOBECOM 2022-2022 IEEE Global Communications Conference, IEEE, 2022, pp. 6158–6163.

[25] S. Chai, V. K. Lau, Joint rate and power optimization for multimedia streaming in wireless fading channels via parametric policy gradient, IEEE Transactions on Signal Processing 67 (17) (2019) 4570–4581.

[26] J. Li, Q. Yang, J. Yang, M. Qin, K. S. Kwak, User perceived qos provisioning for video streaming in wireless ofdma systems: Admission control and resource allocation, IEEE Access 6 (2018) 44747–44762.

[27] C. Ye, M. C. Gursoy, S. Velipasalar, Power control for wireless vbr video streaming: From optimization to reinforcement learning, IEEE Transactions on Communications 67 (8) (2019) 5629–5644.

[28] N.-N. Dao, D.-N. Vu, W. Na, T.-M. Hoang, D.-T. Do, S. Cho, Adaptive bitrate streaming in multi-user downlink noma edge caching systems with imperfect sic, Computer Networks 212 (2022) 109064.

[29] F. Raphel, S. Sameer, A speed adaptive joint subcarrier and power allocation technique for downlink ofdma video transmission over doubly selective channels, IEEE Transactions on Vehicular Technology 69 (2) (2019) 1879–1887.

[30] A. Xiao, X. Huang, S. Wu, H. Chen, L. Ma, Traffic-aware rate adaptation for improving time-varying qoe factors in mobile video streaming, IEEE Transactions on Network Science and Engineering 7 (4) (2020) 2392–2405.

[31] A. H. Sodhro, N. Zahid, S. Pirbhulal, N. M. Garcia, L. Wang, Towards qoe optimization in medical multimedia services for decentralized iot-based applications, in: 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring), IEEE, 2020, pp. 1–5.

[32] N. Zahid, A. Alkhayyat, M. Ismail, A. H. Sodhro, An effective traffic management approach for decentralized bsns, in: 2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall), IEEE, 2022, pp. 1–5.

[33] N. Zahid, A. H. Sodhro, U. R. Kamboh, A. Alkhayyat, L. Wang, Ai-driven adaptive reliable and sustainable approach for internet of things enabled healthcare system, Math. Biosci. Eng 19 (2022) 3953–3971.

[34] A. Mishra, Y. Mao, O. Dizdar, B. Clerckx, Rate-splitting multiple access for 6g—part i: Principles, applications and future works, IEEE Communications Letters 26 (10) (2022) 2232–2236.

[35] R. Rassool, Vmaf reproducibility: Validating a perceptual practical video quality metric, in: 2017 IEEE international symposium on broadband multimedia systems and broadcasting (BMSB), IEEE, 2017, pp. 1–2.

[36] R. S. Sutton, A. G. Barto, Reinforcement learning: An introduction, MIT press, 2018.

[37] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov, Proximal policy optimization algorithms, arXiv preprint arXiv:1707.06347.

[38] Proximal Policy Optimization — Spinning Up documentation, https://spinningup.openai.com/en/latest/algorithms/ppo.html, [Accessed 16-02-2024].

[39] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, D. Wierstra, Continuous control with deep reinforcement learning, arXiv preprint arXiv:1509.02971.