A Q-learning Driven Energy-aware Multipath Transmission Solution for 5G Media Services

Lujie Zhong, Xiang Ji, Zhaoxue Wang, Jiuren Qin, Gabriel-Miro Muntean*, *Senior Member, IEEE*

***Abstract*—Supported by the latest evolution of the 5G technologies, Augmented Reality (AR) & Virtual Reality (VR) video streaming services are experiencing an unprecedented growth. However, the transmission issues caused by heterogeneous access and dynamic traffic are still challenging 5G communications. The Internet Engineering Task Force (IETF)'s Multipath Transmission Control Protocol (MPTCP) can aggregate bandwidth and balance traffic across multiple subflows in a heterogeneous network environment. However, in order to support delivery of high quality 5G media services, researchers should also address MPTCP's inefficient data scheduling to heterogenous sub-paths, consideration of multiple criteria, including energy consumption and its inconsistent behaviour when employed along with the Dynamic Adaptive Streaming over HTTP (DASH) adaptive application layer protocol. To address these issues, we propose a Q-Learning driven Energy-aware Data Scheduling (QLE-DS) mechanism for MPTCP-based media services. QLE-DS models the multipath scheduling as a Q-learning process and employs a novel quantum clustering approach to discretize the high dimensional continuous Q-table. An asynchronous framework is designed to improve the learning efficiency of QLE-DS. The simulation results show that QLE-DS performs better than other MPTCP scheduling algorithms in terms of flow completion time (FCT), retransmission rate, and energy consumption.**

***Index Terms*—Q-learning, MPTCP, 5G media services, energy-aware, data scheduling**

I. Introduction

The latest developments of communication technologies and intelligent devices, have fueled a rapid evolution of 5G services including high quality video streaming. According to a Cisco white paper, video will account for 82% of the total Internet traffic in 2022 [1]. This includes high quality 5G media services such as augmented reality (AR), virtual reality (VR) and other types of video content. These services require innovative network solutions for their distribution on one hand [2][3] and increased efficiency for the battery-powered user devices on the other hand [4][5] in order to maintain a high and smooth viewer experience.

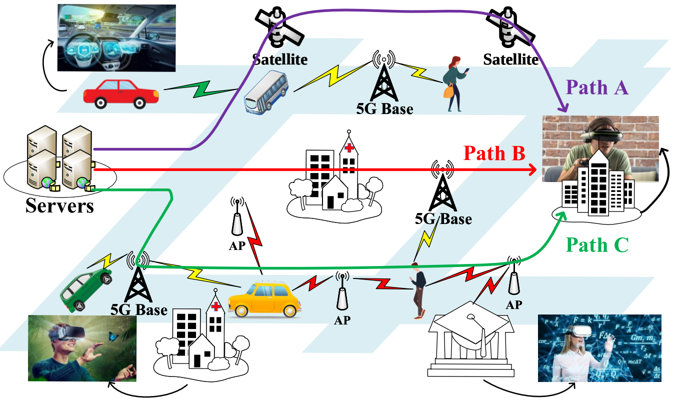


Fig. 1: Application scenario of bandwidth aggregation for real-time video transmission (e.g., AR&VR video delivery) to multihomed devices in a 5G heterogeneous wireless network environment.

Nowadays, a communication equipment often has multiple network interfaces [6]. In order to aggregate bandwidth and balance traffic in a multi-channel heterogeneous network environment, the Multipath Transmission Control Protocol (MPTCP), proposed by IETF in RFC 8684 [7], enables creation of a connection with multiple data subflows via different network interfaces (mmWave, Sub-6GHz, WiFi, etc.). By transmitting data concurrently, the idle bandwidth of different networks is aggregated which can improve the overall throughput. Additionally, MPTCP also achieves increased robustness as the sender can trade off the traffic among various subflows.

It is natural to see how MPTCP is increasingly being employed for 5G media services in heterogeneous wireless network environments. For instance, Fig. 1 illustrates how data reaches a user from the server via a 5G heterogeneous network environment over multiple paths. In such an environment, aggregating the video traffic transmitted over multiple network interfaces can improve the overall user experiences. However, the MPTCP scheduling algorithm has not been improved [8] to support the needs of the high speed 5G video application services. The standard MPTCP scheduling algorithm FastRTT [9] transmits segments on subflows with the smallest Round Trip Time (RTT) preferentially, which improves the utilization of the high-quality links, but also makes the flow allocation unbalanced. From an energy consumption perspective, the existence of oversaturated paths may result in energy waste [10]. Additionally, there is also increased use of the Dynamic Adaptive Streaming over HTTP (DASH) protocol at application layer. The DASH video traffic is adaptive, increasing and decreasing fast, while the traditional MPTCP scheduling

This work was partially supported by the National Natural Science Foundation of China (NSFC) under Grant No. 61872253, by the Science Foundation Ireland (SFI) Research Centres Programme under Grant No. 12/RC/2289_P2 (Insight SFI Centre for Data Analytics) and by the 111 Project (B18008).

L. Zhong and Z. Wang are with Information Engineering College, Capital Normal University, Beijing, China. E-mail: zhonglj@cnu.edu.cn, wang_zx@cnu.edu.cn.

X. Ji is with State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China. e-mail: jxiang_2019@bupt.edu.cn.

J. Qin is with National Key Laboratory of Science and Technology on Information System Security, Beijing, China. E-mail: qinjiuren@126.com.

G.-M. Muntean is with Performance Eng. Lab, School of Electronic Engineering, Dublin City University, Ireland. E-mail: gabriel.muntean@dcu.ie.

algorithm lacks responsiveness, making the scheduling behavior inefficient. The problems faced by the traditional MPTCP scheduling algorithm in the context of 5G media services can be summarized as follows:

1) **Static scheduling.** Traditional MPTCP scheduling methods (e.g., FastRTT, Roundrobin) obey static scheduling strategies and can not adjust scheduling to the dynamic networks. Especially important is the DASH-based adaptivity, affected by the heterogeneous network environment and mobility of users.
2) **Single criterion in decision making.** MPTCP allocates the segments depending on a single goal. For example, FastRTT's scheduling decision depends on RTT only, while other transmission factors such as congestion window size and packet loss rate are not considered.
3) **Energy unawareness.** Energy consumption is very important, especially to the mobile users in a 5G heterogeneous network environment, but this is ignored by the MPTCP's scheduling algorithm. High energy consumption will eventually result in battery depletion, affecting the quality of user experience.

A key component of MPTCP is the scheduling control algorithm, which decides the number of data packets to be transmitted over each particular subflow. Researchers have put much effort to address the problems associated with MPTCP scheduling for distribution of 5G media services. For instance, the authors of [11] proposed an energy-aware load balancing method for MPTCP scheduling and those of [12] introduced a low latency-focused scheduler. However, these researchers did not focus on the performance of 5G media services, ignoring the interaction between transport layer scheduling and application layer scheduling. Thanks to the latest development of artificial intelligence, some smart scheduling algorithms have been proposed. In order to make efficient use of bandwidth, Feng *et al* [13] proposes a novel hybrid transmission mode selection based on online reinforcement learning. At the same time, some researchers also start trying to apply intelligent algorithms to MPTCP scheduling [14] However, modifying the scheduling algorithm to consider optimizing the energy consumption was not attempted. This remains a challenge for MPTCP, because energy consumption can not be easily measured and estimated. Actually, some scholars have focuses on energy efficiency. In their view, frequent network selection is associated with energy waste[15]. Unfortunately, the same problem also exists in relation to MPTCP scheduling. The negative impact includes packet loss, which due to the retransmissions results in an increase of the energy consumption of video transmission.

In this paper, we propose a **Q-Learning driven Energy-aware Multipath Data Scheduling mechanism (QLE-DS)** for 5G media services. QLE-DS applies a Q-learning model to multipath scheduling and can adaptively find the optimal scheduling strategy based on comprehensive network-perception and energy-saving. The contributions of this paper can be summarized as follows:

- Model MPTCP data scheduling as a Q-learning process based on accurate evaluation of multiple important transmission factors for 5G media services: packet loss rate, throughput and energy consumption.
- Design a state reduction algorithm based on quantum clustering to discretize the high dimensional continuous state space.
- Propose an asynchronous learning framework to reduce the training costs and improve the solution training efficiency.
- Test the proposed QLE-DS in different scenarios and show that QLE-DS significantly outperforms alternative scheduling algorithms.

The rest of this paper is organized as follows. Section II discusses related works. The overall framework of the proposed mechanism is introduced in section III. Section IV gives a detailed description of the proposed QLEC model and the overall QLE-DS algorithm is presented in section V. Section VI illustrates and analyses the results of our simulation-based evaluation. Section VII presents the conclusions of our work.

II. Related Works

This section presents diverse MPTCP-based solutions for 5G media transmissions and scheduling in multipath environments. Additionally, several effective optimization methods for 5G media services are also discussed.

A. MPTCP in 5G Heterogeneous Networks

Recently, 5G solutions are widely deployed for supporting communications involving high mobility nodes including smartphones, vehicles and Unmanned Air Vehicles (UAVs). In order to enable them to have good services, it is necessary to employ heterogeneous network interfaces and protocols (i.e. cellular, Wi-Fi, 802.11p, etc.) to meet access requirements. Various researchers have performed much research on aggregating bandwidth across multiple interfaces. MPTCP is gaining significant attention in academia as an useful approach. Unfortunately, MPTCP can not alleviate the adverse effects of using multiple heterogeneous transmission technologies at the same time [12]. The problem of high packet loss rate and high delay caused by network heterogeneity has not been solved effectively yet. The impact of digital terrestrial television (DTT) promotes the digital transformation of television technology. Authors of [16] proposed a new method of calculating DTT coverage prediction based on clustering and a machine learning algorithm, which can achieve higher prediction accuracy in heterogeneous networks.

The authors of [17] improved the MPTCP protocol for URLLC scenario in 5G, and proposed a new URLLC-MPTCP protocol, which integrates a stream coding strategy in the congestion control algorithm to improve the delay unfriendliness of traditional MPTCP. The authors of [18], introduced a new MPTCP congestion control algorithm (CCA) named DEFT which achieves high throughput and low end-to-end latency in 5G networks. They found that reducing the length of the buffer queue at the receiver can make the delay more stable. Shiva *et al.* [19] have developed a novel MPTCP algorithm for the 5G Internet of vehicles, which employs FEC and makes use of path diversity to enable reliable communication

TABLE I: Comparisons Between the Proposed QLE-DS and the Existing Works

Algorithm	Overall Goal	Input Metrics	Characteristic	Processing mode
FastRTT [7]	Using less delay path to transmit data	RTT	Priority filling RTT smaller path(s)	Reactive
LRF [24]	Maximize throughput	CWND, RTT	The path with minimum SRTT and CWnd is preferred, and all paths are used	Reactive
ECF [25]	Minimise idle periods of the fast subflow(s) waiting for slow subflow(s)	CWND, RTT and buffered data to be sent	Takes lowest SRTT first, eventually skipping some subflow(s)	Proactive
QLE-DS	Minimize flow completion time, energy consumption	CWND, RTT, completion rate and energy	Separate video services from other services, using reinforcement learning training scheduler to optimize energy consumption	Proactive

over multiple parallel heterogeneous networks. Tang *et al* [20] explored employing deep reinforcement learning to adaptively allocate path resources in the high mobility 5G heterogeneous networks. Authors of [21] designed a transmission scheme to provide reliable transmitting resource for the purpose of resolving the problem of bandwidth consumption in transmission. For a traditional multimedia scenario, researchers of [22] proposed a video transmission framework, which can avoid the bandwidth starvation during the process of data transmission. With the emphasis on fault-tolerant data transmission and data-load-reduction processing, the authors of [23] proposed the REDPF mechanism to enhance reliability of data transmission and processing speed in heterogeneous networks.

The purpose of the above research is to improve the performance of MPTCP in heterogeneous network environments. However, the researchers considered their work in isolation, as they have not considered employment in conjunction with the latest applications. In this paper, we focus on solving the multi-path transmission scheduling problem for AR and VR applications, while considering multiple factors which affect the mobile network delivery systems.

B. Scheduling in Multipath Transmissions

The scheduling algorithm is an important component of the MPTCP protocol. It determines on which path the application layer data is sent during its delivery. However, as the traditional scheduling method is simple and uses a single performance factor in its decision making process (i.e. RTT), it can not meet the performance requirements of complex 5G media services. It is necessary to design an enhanced scheduling mechanism to minimize the influence caused by path heterogeneity.

Many researchers have developed new MPTCP scheduling schemes. Authors of [24] introduced a modular scheduler called the Lowest RTT First (LRF) selection framework, which allows MPTCP to select different paths for multiplexing data. Authors of [25] identified the cause of scheduling inefficiency: faster paths are under-utilized due to the idle periods and proposed the Earliest Completion First (ECF), a novel MPTCP path scheduler to improve the utilization of the fastest paths. The work in [26] makes scheduling decisions according to the time-varying network statistics over the millimeter wave cloud radio access networks. In [27], a new MPTCP scheduling method, the Dynamic Packet Scheduling and Adjusting with Feedback (DPSAF) was proposed. DPSAF takes packet loss rate and time offset into account when optimizing the scheduling strategy. Researchers of [28] developed a novel Deep Q-Learning (DQL)-based multipath packet scheduling scheme using policy gradients and prioritized replay buffer for the wireless networks. They provided a stability analysis of the scheduling algorithm and related practical design insights. We previously proposed a new game enhanced compensation handoff scheme for high quality transmission service based on multipath TCP [6], to solve transmission scheduling in the Internet of vehicles. Authors of [12] found that the MPTCP scheduler tends to overuse slower paths, resulting in waste of time as the data transmitted on faster paths needs to wait. The authors evaluate three most advanced schedulers for MPTCP: Delay-Aware Packet Scheduler (DAPS), Out-of-Order Transmission for In-Order Arrival Scheduler (OTIAS) and ECF. Additionally, they summarize the scheduling problems caused by traffic asymmetry.

The multipath schemes described in the literature mainly improve the media streaming transmission by performing more efficient allocation of existing traffic on the available paths. However, in AR, VR and other high quality video stream transmission scenarios, the application layer changes dynamically the stream bitrate in order to improve the user experience (i.e. by using DASH). Unfortunately, this increasingly common feature is ignored by the transport layer protocol and puts pressure on the transmission from both application layer and network environment at the same time. As a result of this, existing scheduling methods are not efficient in real world deployments and there is a need for proposal of improved ones, as is the solution proposed in this paper. Some classic scheduling algorithms relevant to this work and their characteristics are summarized in Table I.

C. Scheduling Optimization for Multimedia Services

Achieving both high speed and low energy consumption when delivering multimedia services represents a challenge for the design of an optimum MPTCP scheduling algorithm. Current RTT-based scheduling algorithms (e.g., FastRTT) do not meet the high quality latest adaptive application requirements [29]. In heterogeneous network environments, sudden bitrate adjustments determine packet loss, which increases the energy consumption due to retransmissions. Several works have tried to improve the energy transmission efficiency for multimedia services. Feng *et al* [30] analyzed the high energy consumption

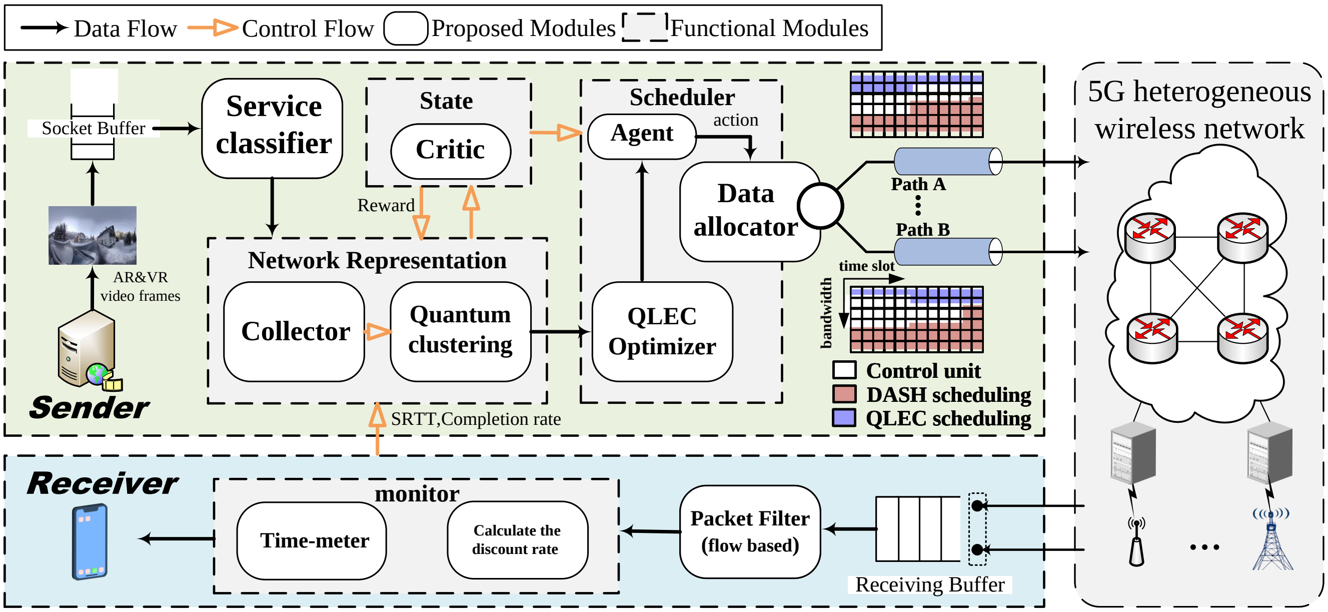


Fig. 2: QLE-DS in an AR/VR multipath transmission

problem in a multi-hop wireless network. They noted that there was a waste of energy in the decoding process at the base station. Following their observation, the researchers designed a new packet receiving algorithm based on energy accumulation according to the energy model of IEEE 802.11. The algorithm reduces the time required for the base station to decode data, thus reducing the energy consumption.

Additionally, other works have tried to improve the quality of the diverse media services, focusing mostly on the vidio Yao *et al.* [31] proposed a novel multiview video stream adaptation strategy for interactive video playback which allows the client to navigate among multiple viewpoints. Authors of [32] proposed a price-based bandwidth allocation strategy for DASH, which can effectively reduce the impact of rebuffering on Quality of Experience (QoE). Spiteri *et al.* [33] formulated a video bitrate adaptation process for ABR streaming as a utility maximization problem and derived the Buffer Occupancy based Lyapunov Algorithm (BOLA), an online control algorithm that is provably near-optimal. Zhang *et al.* [34] analyzed theoretically the influence of varying video chunk size on streaming performance in wireless networks and the effect of block size on video stream performance. Praveen *et al.* [35] designed a rate adaptive algorithm to optimize the rate allocation of VR video. The authors selected tiles requiring high rate according to the viewer's field of view (FOV). Yaqoob et al. [36] proposed a novel solution for adaptive omnidirectional video delivery based on an improved accuracy user FOV prediction method. The authors of [37] formulated a DASH enabling service optimization problem for wireless networks, which was transformed into a queue stability problem that is highly tractable, scalable and general. There is a promising 5G-xcast project in Europe [38], which responded to the network environment dynamically and timely.

The researchers who made significant contributions to scheduling optimization tried to improve the user experience by controlling the flow rate or data block size at the level of application. Unfortunately, at application layer the quality of the transmission path can not be obtained directly, nor it can be interfered with the transmission resource allocation. Therefore more intelligent scheduling solutions should be considered and this paper introduces a reinforcement learning algorithm for scheduling optimization.

III. System Design And Analytical Framework

A. System Overview

Fig. 2 illustrates the system overview of the proposed solution, the Q-Learning driven Energy-aware Multipath Data Scheduling scheme (QLE-DS). It shows an AR/VR video to be transmitted, which is processed into frames and transmitted between a *Sender* and *Receiver* via a *5G Heterogeneous wireless network* in which multiple paths exist. QLE-DS is deployed at the transport layer.

1) Major Sender Components: The key components at the sender side include the *Service classifier*, *Critic* and *Data allocator* units. The *service classifier* labels the data to distinguish whether it is DASH data or not. *Critic* contains a neural network, which is used to evaluate the transmission quality of video rather than the quality of all services. The *data allocator* is used to receive the feedback from the agent and make scheduling decisions via the Q-table. The above three key components creatively process the services separately.

Following the DASH approach, the sender server encodes videos at different bitrates, chops them in different fragments and generates a media presentation description (MPD) file. The MPD file records the video fragment information and for each resource URLs. The service classifier module labels the data according to the types of services and divides it into video and other services, respectively. The data is used by the *Network Representation* module, which collects a variety of network-related information. This information informs the decision of the Q-Learning with Energy Consumption (QLEC) optimizer, which indicates how the data is allocated to different paths by the scheduler. The scheduler gets online suggestions from the *Critic* to modify its own algorithm, which is iterative and based on interaction with the environment. In the case of network environment changes (i.e., distinguish congestion losses from wireless losses), it can achieve good performance. The scheduling algorithm which uses the reinforcement learning method will be presented in section IV.

2) Major Receiver Components: At the receiver side, the data is aggregated from multiple paths into the receiving buffer. A *Packet filter* is introduced to distinguish packets from different paths according to the information in their MPTCP packet header. The *Packet filter* allows the receiver to feed back the energy consumption factor and RTT of each path to the sender via packet header option in the ACK packets. The newly designed *Monitor* calculates the changes in each path's status. Specifically, the monitor records for each path SRTT (as defined in Eq. (1)) based on the packets' time stamp and completion rate according to the missing packets based on their sequence numbers. This information will be provided to the sender's *Network representation* module. The end-to-end feedback mechanism allows the sender server to learn about the transmission efficiency (SRTT) and transmission quality (completion rate) of each path. This information is used as training data by the reinforcement learning algorithm.

The proposed solution takes energy minimization scheduling decisions based on a 5G Heterogeneous Network Transmission model, a Quantum clustering algorithm and an Energy consumption model. For the sake of completeness and integration within the system framework, these models are briefly introduced in the following subsection. Table II summarizes the basic notations used.

B. Model Description

1) 5G Heterogeneous Network Transmission model: We consider that the heterogeneous network is composed of a series $L = \{1, \ldots, l\}$ of links with c_l capacity. From the data source server s to the user r, link is instantiated as path p, which is used both to denote a path and the set of links l in the path. Let $P := \{p \big| p \in L\}$ be the collection of all paths. Let matrix $H \in \{0,1\}^{|L|\times|P|}$ denote whether link l exists in path p. $H(l,p) = 1$ if link l is in path p (denoted by $l \in p$), and 0 otherwise.

For each path, $\tau_p(t)$ represents its SRTT and t represents the time slot. $w_p(t)$ represents the *congestion control window (cwnd)* of path p, which defines the maximum number of packets that the path can support in a transmission round. Each path has a packet loss rate $pl(t)$ at time slot t. Let $q_p(t) = \sum_{l \in p} pl(t)$ be the approximate congestion level on path p. In this article, $q_p(t)$ is the packet loss rate of path p. We associate three state variables $(\tau_p(t), w_p(t), q_p(t))$ to each path p.

SRTT represents the transmission delay level of path p and is defined as in Eq. (1):

$$SRTT(i+1) = \alpha \cdot SRTT(i) + (1-\alpha) \cdot RTT(i) \qquad (1)$$

TABLE II: The basic notations used in this paper.

Symbol	Explanation
P	Collection of paths.
L	Collection of links.
e_p	the energy consumption of path p.
T_p	the flow completion time of path p.
C_p	Maximum traffic capacity of path p.
ω_p	the tail energy of data transferring of path p.
Seg_{totoal}	Total transmitted data.
q_p	the packet loss rate of path p.
$x_p(t)$	the transmission speed of path p.
I_p	the congestion window growth factor.
D_p	the congestion window descent factor.
MSS	maximum message segment length.
$SRTT$	the Smoothed RTT.
bu	the size of receiver buffer in use.
a	the action of the agent.
$\mathbb{S}$	the state set.
β	the utilization of path p.
r	the reward of Q-learning model.
α	the coefficient of $SRTT$.
g	the return factor.
λ, μ	the discount rate, the learning rate.
η	normalized representation of congestion level.
E_p	energy consumption factor of path p.
T_p	flow completion time factor of path p.

where i represents the transmission round and α is a coefficient greater than 0 and less than 1. This SRTT formula is used to solve the problem of outliers in the transmission process. At the same time, we define the completion rate $\frac{\beta \cdot w_p(t)}{SRTT(i)}$, which is used to represent the efficiency of path service flow. β is the utilization of path p's congestion window.

2) Quantum clustering Model: The proposed QLE-DS scheme includes a particular clustering algorithm needed for preprocessing of data packets. The input of this clustering algorithm is a large amount of continuous information related to data packets (e.g., *RTT*, *cwnd* and *energy consumption*) which is transformed into discrete data by the proposed Quantum clustering algorithm. The proposed algorithm makes use of the principle behind the dynamic quantum clustering (DQC) [39] to make the continuous environment inputs into discrete entries, without the need to indicate the cluster′ centers before the start of the clustering process. It takes three state variables as major characteristics of packets, and each packet can be described via a three component vector and represents a point $\vec{x}_i$ in the three-dimensional Euclidean space. In time, each packet corresponds to a Gaussian wave function $\psi_i(\vec{x}) = e^{-(\vec{x}-\vec{x}_i)^2/2\sigma^2}$. Next we construct the sum of all these Gaussian functions.

$$\psi(\vec{x}) = \sum_i e^{-(\vec{x}-\vec{x}_i)^2/2\sigma^2} \qquad (2)$$

Quantum clustering abstracts the information carried by data packets into the vector $\vec{x}$. The formula from Eq. (2) is the Gaussian wave function of the data packet. $\vec{x}_i$ represents the vector for the i-th packet. σ is the standard deviation (sometimes called Gaussian RMS width), which controls the width of the "wave" (set by ourselves). Quantum clustering has a unique advantage. It does not need to specify the number of clusters in advance. Different from the traditional clustering method and inspired by the Schrodinger equation, Quantum clustering regard the high-dimensional data as a wave. The Schrodinger equation contains a potential function, which can be analytically obtained from Eq. (3). We associate the minimum of the potential function with the cluster center. Through this method, we can determine the center of the cluster and data can be associated to each cluster. The method has only one variable parameter, that is, the scale σ of Gaussian kernel. We apply this method to perform the dimensionality reduction of packet high-dimensional data.

$$H\psi \equiv \left[-\frac{\sigma^2}{2}\nabla^2 + V\left(\vec{x}\right)\right]\psi = E_0\psi \tag{3}$$

The potential function $V\left(\vec{x}\right)$ is inextricably linked with the data point system. Eq. (3) is an extension of the wave function, which is used to solve the potential function. This can be regarded as an alternative model to the traditional clustering method, including the attraction to the cluster center and the generation of noise, both of which are inferred or implemented based on the given data packet info from *Collector*. The output of clustering algorithm can be inputted into the reinforcement learning model to act as the discrete state selection space for the agent.

3) Energy Consumption Model: This research employs the e-aware model developed in [41] to characterize the energy consumption of mobile terminals. This model considers the energy consumption of data temporary storage, data transmission and tailing. The variable e_p represents the energy consumption intensity of the transmission path, which is defined as the consumed energy for delivering the same amount of data traffic (*J/Kbps*). t_p represents the flow completion time, which refers to the time needed to complete the transmission of all the content in the video service - the main goal of the scheduler. ω_p represents tail energy and refers to the energy consumed by the network interface after a data transmission. ω_p depends on the tail energy, and represents the energy consumption of the system when processing packets and waiting in the queue. Seg_{totoal} represents the total transmission required in a video service. Assuming that the packet loss rate q_p tends to zero, we can roughly estimate the effective transmission speed of each path as $x_p\left(t\right)$ which can be calculated by multiplying the number of transmission rounds per unit time by the number of scheduled packets per round.

$$\begin{aligned} E &= \sum_{p\in P}\left(\lambda_p \cdot e_p + t_p \cdot \omega_p\right) \\ &= \sum_{p\in P}\left(\frac{\beta\cdot w_p(t)}{SRTT_p(i)}\cdot e_p + t_p\cdot\omega_p\right) \\ &= \sum_{p\in P}\left(\frac{\beta\cdot w_p(t)}{SRTT_p}\cdot e_p + SRTT_p\cdot\left\lceil\frac{Seg_{total}}{MSS_p\cdot x_p(t)}\right\rceil\cdot\omega_p\right) \end{aligned} \tag{4}$$

The energy in Eq. (4) is divided into the energy consumed during transmission and the tail energy consumed by network equipment during transmission. The data collection interval of the model is the time interval of two adjacent transmission rounds. We use $SRTT$ to represent this time interval, which is defined in Eq. (1).

However, it is not enough to indicate the calculation method of energy consumption for each round. We also define the global objective function as follows:

$$\begin{aligned} &\min \sum_i \left(\gamma E + (1-\eta)\max\{T_p\,|p\in P\} - \eta\sum_{p\in P} x_p\right) \\ &s.t.\ \omega_p > 0,\ e_p > 0,\ \eta\in(0,1)\,,\ \sum_{p\in P} x_p \le \sum_{p\in P} C_p \end{aligned} \tag{5}$$

Where C_p represents the maximum transmission capacity of each path p, and η is a coefficient we define, which represents the congestion of the whole transmission process. We add it to the objective function to make our algorithm more flexible and automatically select whether to pay more attention to flow completion time or throughput according to the degree of network congestion. η is described in more details in the next section. As shown in Eq. (5), the optimization objective of our algorithm considers three different dimensions at the same time, and changes the strategy according to the constraints of the environment to minimize the objective function.

IV. QLE-DS Model

The major components of the proposed QLE-DS model are: agent, network state, action, critic and reward function. Agent represents the scheduler, which acts according to the Q-table obtained by reinforcement learning. The action here refers to when and over which path the packets are scheduled to be transmitted. Scheduling results will affect the network's state. The feedback brought by the state will determine the agent to optimize the Q-table using a reward function. Critic is an independent component and adjusts the update function of the Q-table to match the scheduling of DASH video.

The goal of the proposed Q-learning method is to find a set of actions to maximize the cumulative reward. To avoid any additional delay triggered by the Q-learning model in the actual network, we propose an asynchronous framework for the purpose of separating the training and decision process of Q-learning. The framework includes two stages: *offline training* and *online decision*.

Offline training: This stage aims to obtain a basic rule table, which is the basis for the online decision. The rule table, named Q-table, represents the expected reward of each executable action for all states and it filled by conducting pre-training with the samples collected from the environment. This stage employs three modules: collector, cluster and trainer, as illustrated in Fig. 4. The complete description is provided in section V. B.

Online decision: In this stage, with the help of the Q-table acquired in the offline training stage, the system will execute the action and schedule the segments to be transmitted over the selected path. In addition, the system will continue to collect the data to optimize the Q-table. This state involves two modules: decision maker and optimizer. The process of this stage are presented in Algorithm 2.

Next, we introduce the proposed MPTCP scheduling method which extends the Q-learning with energy consumption (QLEC) model. QLEC is a model composed of a series of algorithms, which is a part of our proposed QLE-DS scheme. Q-learning is a reinforcement learning approach which learns a policy that maximizes the expected reward through training and feedback. In MPTCP, we will design a scheduler with a Q-learning model to select the scheduling path to improve network performance. We innovatively divide 5G media business into video services and other services (e.g.,position information, cookie and bullet screen comments). Our QLEC optimizer can schedule other services traffic according to the video traffic scheduling of application layer DASH protocol. Specifically, it can balance the performance of scheduling algorithm and reduce the energy consumption of transmission.

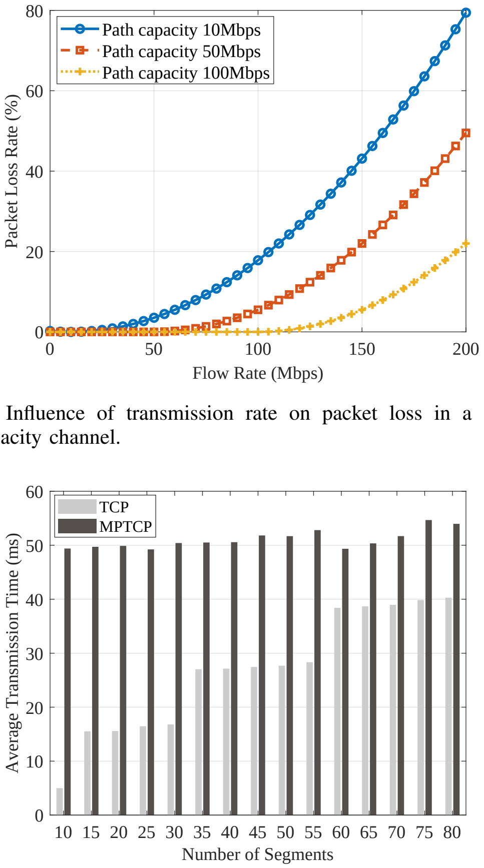
A. Learning Goals for Q-learning

The goal of the MPTCP data scheduling is to select the appropriate path for each packet delivery to improve the overall performance of the multipath transmission. The goals of Q-learning in MPTCP scheduling is to choose the path that improves overall media service user experience and reduce energy consumption.

In this paper, the energy consumption of data transmission is computed in two ways. First, the energy consumption E is computed with Eq. (4), which represents the relationship between the energy utilization of path p and the scheduling action. According to Eq. (4), the energy consumption of data transmission is mainly related to the transmission medium, transmission efficiency and flow completion time. Two aspects will be covered in the following paragraphs and two examples are used to explain our optimization ideas.

The packet loss model proposed in [42] regards the derivative of packet loss rate $\dot{L}_p$ of multiple paths as $\dot{L}_p = \gamma_p \left(x_p - c_p\right)^{+}_{L_p}$, where $\gamma_p > 0$, γ_p is a positive gain factor, and c_p represents the capacity of the path p. By importing the packet loss rate model and flow model formulas of [42] into Matlab, we get the curve between throughput and packet loss rate in Fig. 3 (a). The change trend of packet loss rate is shown in Fig. 3 (a), when the path capacity is 10Mbps, 50Mbps and 100Mbps, respectively (this packet loss occurs only because the rate exceeds the transport capacity of the path).

$$\dot{x}_p = \frac{x_p}{\tau_p}\left(I_p - D_r \cdot q_r\right) \tag{6}$$



(a) Influence of transmission rate on packet loss in a fixed capacity channel.

(b) Micro transmission time of TCP and MPTCP, with two available paths having asymmetric RTTs (10ms vs 100ms).

Fig. 3: Energy consumption bottleneck of MPTCP.

From this point of view, we can find that when bad scheduling occurs, the traffic size exceeds the capacity of the path, and retransmission probability of packets increases. This will increase the energy consumption according to Eq. (4).

Eq. (6) presents the throughput of path p and uses the parameter L_p as the number of packets transmitted since the last packet loss. The energy consumption is calculated using the model shown in Eq. (4), extending the e-aware model proposed in [43], which calculates the consumed energy following data transmission. It is assumed that a small number of retransmitted packets can be received, so we can get $x_p = (1 + L_p)\frac{w_p}{\tau_p}$. If the transmitted data Seg_{total} is small enough, the following formula is obtained:

$$T_p = \begin{cases} \frac{Seg_{total}}{MSS_p \cdot \bar{x}_p} + SRTT_P, & if \quad L_p = 0 \\ \frac{Seg_{total} \cdot (1+L_p)}{MSS_p \cdot \bar{x}_p} + 2 \cdot SRTT_P, & if \quad L_p \neq 0 \end{cases} \tag{7}$$

Using Eq. (7), we simulated in MATLAB data transmission using TCP and MPTCP over two paths set with delays of 10ms and 100ms, respectively. Fig. 3 (b) shows the results in terms of transmission time. Due to the different delays, the data arriving first needs to wait for the data arriving later, which affects the flexibility of MPTCP scheduling.

However, classic scheduler designers ignore the fact that poor quality paths may not be suitable for early allocation of data for transmission. This, despite the fact that differences in path delays are very common in heterogeneous networks. Fig. 3(b) shows how the performance of MPTCP in a heterogeneous network environment is weaker than that of TCP over the better path. The proposed reinforcement learning method can detect this situation and avoid it.

B. Q-learning Model with MPTCP

The proposed QLE-DS model based on reinforcement learning is illustrated in Fig. 2, as discussed in Section III. The MPTCP server agent (a part of the scheduler) interacts with the path environment and the receiving end user to collect the state set $\mathbb{S}(t) = \{s(t,p)\}$. It includes the delay, packet loss rate and energy consumption of path p. The scheduler balances two kinds of resources on multipath: 5G video service and other services. The critic module uses the Long Short-Term Memory (LSTM) network to monitor long-term changes in the transmission quality of video services. This innovatively prevents the double role of transport layer scheduling and application layer DASH scheduling from leading to a sharp reduction in the quality of service.

Network Representation: The network representation module constructs a vector space to represent the network environment of each path ($s(t,p)$ as inputs). However, the heterogeneity of 5G network and the time-varying characteristics of wireless channel quality make the network state of each time slot t change. Dealing with a lot of state information of multipath is a challenge. Therefore, we use quantum clustering algorithm as a network information filter to process the information from the collector. After the continuous network state is sampled by collector, it is classified into discrete state space. In this way, some slight network changes and the superposition of white noise, preventing over fitting of neural network.

The state after clustering is described as $\mathbb{S}_c(t,p)$, the state space is limited, and it is fed to the agent network for training. At the same time, the scheduling action and video transmission effect obtained from agent network training are sent to the critic network. Critic network has a unique state space. Part of its input is the output of agent network. Its training results are used to evaluate the impact of action on video transmission quality in agent network. The results of relevance feedback will affect the attenuation factor of agent reinforcement learning.

Agent-Critic: The agent in Q-learning refers to an entity that gives the environment a guide to achieve the desired goals. In our proposed QLE-DS model, the agent is in charge of deciding the appropriate subflow to schedule the data in multipath transmission. Critic networks are designed to guide agent to improve the service quality of video service. This dual network design is to make the overall traffic scheduling adapt to the network scheduling and make the video service obtain relatively better network resources. As we know, DASH protocol is a complete application layer protocol, which enables users to request videos with different bit rates according to the network environment. However, scheduling above the transport layer will make it difficult for the transport layer scheduling to cooperate. If the agent is used to schedule the overall traffic only according to the network environment, the decision result will lag behind the change of video traffic.

Algorithm 1: Offline Training Algorithm For Agent

```
Input: The replay buffer
Output: A basic $Q$-table $Q$
Acquire the currentCluster of state using quantum
    clustering algorithm;
repeat:
    $\bar{\omega} \leftarrow a$batch of samples from the replay buffer;
    Cluster the state using *currentCluster*;
for $each(s,a,r,s')$ in $\bar{\omega}$ do
    Calculate reward
      $r = \sum_{p \in P} \eta \cdot x_p - (1-\eta) \cdot T_p - \gamma \cdot E_p$ ;
    Update $Q$ value for each state and action using
      Eq.(7);
    if the replay buffer is empty then
        Update return factor $g$;
        break;
```

$$Q_t(s,a) = Q_{t-1}(s,a)$$

$$+\lambda \cdot \left[r + \mu \cdot \underset{a'}{\mathrm{maxQ_{t-1}}}(s',a') - Q_{t-1}(s,a) \right] \tag{8}$$

As shown in Eq. (8), Q-table is a table that includes the expected reward of all executable actions for each state. The main process of Q-learning is continuous optimizing of the Q-table through training. The update function of value in Q-table for each state and corresponding action is done according to Eq. (8). In this equation, $\mu \in [0,1)$ represents the learning rate while λ represents the discount (or attenuation) rate. Parameters s and a represent the current state and action and parameters s' and a' represent the next state and action. Updates at each stage are provided based on Eq. (9), which is based on the temporal difference:

$$\Delta_t = R_t + \mu \max_{a'} Q(s',a') - Q(s,a) \tag{9}$$

Motivated by the above analysis, based on the generalization of Eq. (7) to the approximation setting:

$$g' = g + \mu(\lambda \cdot Q_t(s,a)) \nabla_t Q_t(s,a) \tag{10}$$

where Q_g is a continuous function with return factor g. On the other hand, g' is the return factor of the next time slot, as shown in Eq. (10). A gradient descent strategy is used to solve the gain in the next moment.

In addition, the critic network is a fully connected LSTM with two hidden layers, which are composed of 128 neurons. Long Short-Term Memory (LSTM) is a special Recurrent neural network (RNN), which can well predict time series. In the dynamic network environment of data transmission, its prediction is better than RNN network [44]. We use the correction linear function in the hidden layer and the hyperbolic tangent function in the output layer. The critic and

Algorithm 2: Online Decision Algorithm

```
Input: The basic $Q$-table
Output: $(s, a, r, s')$
for time do
    if $n < \varepsilon$ then
        a=argmax($Q$ value);
    else
        Select action a randomly;
    Apply action a;
    Compute reward r and observe next state $s'$;
    Save record $(s, a, r, s')$ into the replay buffer;
    Classify the state $s$ using the currentCluster;
    Update $Q$-table;
    $s \leftarrow s'$;
    Adjust the attenuation factor $\lambda$ according to the
      critic network training results;
```

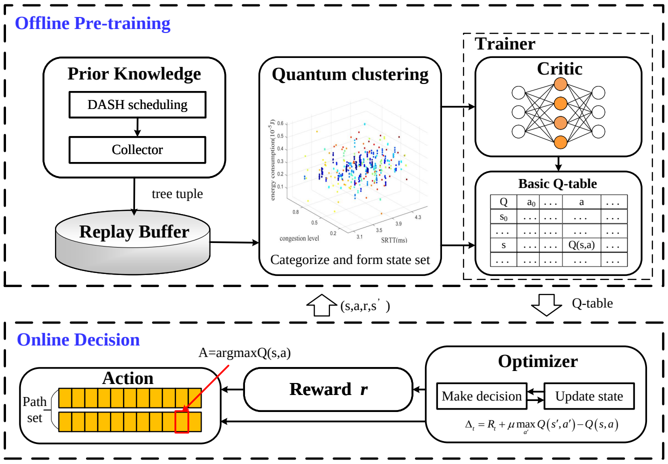


Fig. 4: The asynchronous framework of proposed mechanism.

participant networks used Adam optimizer for joint training, and we set the learning rates to 0.001 and 0.01, respectively.

Action: Action is the behaviour that the agent chooses in order to achieve the goal of maximizing the cumulative reward. In our proposed model, we defined the action as the scheduling result in MPTCP, which can be shown in Eq. (11). Parameter sf_i has only two possible values: 0 and 1, and represents the scheduling result of i-th subflow. When the value of sf$i$ equals to 1, the agent choose i-th subflow to transmit the data. The agent can only choose one subflow to schedule at any particular time, that is, only one element in action can have the value of 1, while the other should be equal to 0. We express the decision action of each round as a one-dimensional vector:

$$a = (sf_0, \dots, sf_m) \tag{11}$$

State: Under normal conditions, the state in Q-learning can be explained as the status of the environment at a specific time. In reality, we can represent the state with some network parameters to express the network condition. The state at time slot t in our proposed model is represented with SRTT, as defined in Eq. (1).

In order to improve the transmission efficiency and optimize the energy consumption of user, we define three state indexes to distinguish the advantages and disadvantages of states. As Eq. (12) cl represents the congestion level of the subflow. We used variable T to express the flow completion time of subflow. The SRTT can be measured based on timestamp.

$$cl = \frac{bu}{buffer} \tag{12}$$

The congestion level can be calculated according to Eq. (12), in which bu is the size of receiver $buffer$ in use and buffer is the maximum size of the receiver buffer.

Reward: In reinforcement learning, the design of reward impacts on the result of learning. Generally, the agent performs a set of actions that maximize the acumulated reward. In this model, we design the reward with throughput th, packet loss pl and energy consumption e, which prompt improving the throughput, reducing loss rate and declining the energy expend in MPTCP. From formula Eq. (4), we can get the reward function of agent network. Energy consumption is an expression of transmission efficiency. Bad scheduling results in both poor energy consumption efficiency and low user experience. The reward r can be computed with Eq. (13).

$$r = \sum_{p \in P} \eta \cdot x_p - (1 - \eta) \cdot T_p - \gamma \cdot E_p \tag{13}$$

where $\eta \in (0, 1)$ depends on the congestion level of the network and is obtained by normalizing cl. γ is a coefficient which ensures dimensional alignment in the formula.

C. *Online Training*

In this subsection, our proposed QLEC optimizer is introduced. Fig. 4 illustrates the two major stages of the algorithm: *offline training* and *online decision* making. From a global perspective, they operate on the same *Q-table*. The *offline training* stage aims at building a *basic Q-table* to provide support for the *online decision* stage. The offline training process is completed in advance and its results are compiled into the kernel file. The *online decision* stage focuses on both making the decision on what action will be performed and updating the *Q-table* to help improve future decisions. This second stage is performed during the transmission. Next, the *QLEC* optimizer will be introduced.

Offline training in Algorithm 1 employs three modules: *Replay Buffer*, *cluster* and *trainer*. First, the *collector* collects data from the environment and stores them in the *replay buffer* with the form (s, a, r, s'). The parameters represents state, action, reward and next state respectively. Then the cluster sort the continuous state space into different categories. Finally training the samples and calculate a *basic Q-table* for next stage. The details of each modules are described as follows.

In trainer, the agent collect data from the environment and store in the replay buffer in the form of (s, a, r, s'). To keep the agent away from making bad decisions because of the reason that there is no prior knowledge at the beginning, QLEC

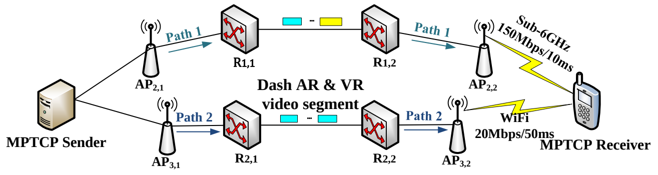


Fig. 5: The simulation topology.

optimizer selects the existing $FastRTT$ scheduling algorithm as the initial scheme.

In our proposed QLEC model, we defined state s as a formula represented with $SRTT$ and congestion level, which makes the state space $S = \{s_1, ..., s_n\}$ high dimensional continuous. In order to process high-dimensional data, we use quantum clustering method, which can reduce the data to low-dimensional and processable. In online decision, the agent first determines the action, and then performs the action in the network. Afterwards, update the *Q-table* according to the feedback of the environment. This stage includes two modules: decision making module and optimizer module, which will be introduced below.

1) Decision making:
At the beginning, the agent will initialize the *Q-table* the same as the basic *Q-table* that was calculated in the offline training stage. In decision making, we can observe the current state s and take action with the ε-policy, which is a strategy to weaken the contradiction in reinforcement learning between exploration and exploitation. The model explores the system with probability ε to take random action and with probability ε to exploit the *Q-table*. After deciding the action, the system schedules the data to the appropriate subflow.
2) Optimizer:
The goal is to collect data from the environment and optimize the *Q-table*. After the steps described above, we compute the reward r according to Eq. (6) and next state s' after performing the action. The records of (s, a, r, s') are saved and ready to be delivered to the buffer. The *Q-table* will be updated using Eq. (11). The algorithm of online decision is described in Algorithm 2.

V. Performance Evaluation

This section describes the performance evaluation of the proposed QLE-DS scheme. The evaluation involves simulations in a two path heterogeneous network environment. The assessment is performed in terms of energy consumption, flow completion time and congestion level and the proposed QLE-DS results are compared with those of other two state-of-art algorithms.

A. Simulation Setup

Our performance testing uses models implemented in the Network Simulator (NS-3) version 3.29, enhanced with MPTCP implementation published by [46]. A two path topology is considered between a server and a client, as illustrated in Fig. 5. This is a typical scenario widely used in practice when a mobile device has access to a cellular and a wireless broadband network, respectively. In order to simulate a 5G heterogeneous network environment and reflect the performance of different scheduling algorithms in such an environment, we set *sub-6ghz* and *WiFi* as the two access paths, which have great differences in bandwidth and delay. The details of the simulation parameters are described in Table III. This is a typical evaluation method, often used by other MPTCP-based research works [45].

TABLE III: Parameters Setting of Simulation

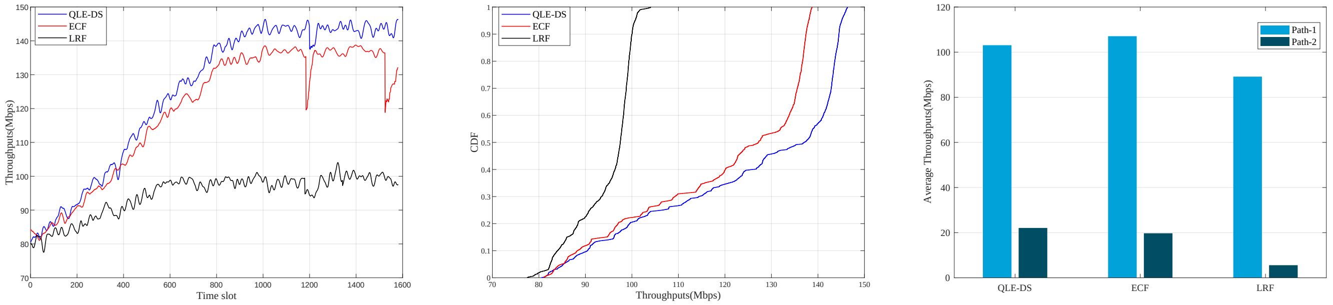
Parameters	Value	
Wireless technology	Sub-6GHz communication	WiFi communication
Access bandwidth	150Mbps	20Mbps
Access link delay	10ms	50ms
Core network link delay	10ms	15ms
Loss rate (%)	0.2	0.1

The performance of the proposed QLE-DS is compared with that achieved when ECF and LRF are employed. These three algorithms are used in turn for MPTCP transmission scheduling. ECF assesses path state based on congestion window and RTT, LRF employs a RTT-based scheduling strategy and our proposed QLE-DS considers the path transmission completion rate and energy consumption in scheduling decision making.

The major parameters of the QLE-DS model are set as follows. The learning rate μ in Eq. (7) is set to 1, and the initial value of the discount rate λ is set to 0.5. The values of η and γ in Eq. (12) which relates throughput, packet loss rate and energy consumption are set to 0.5 and 2, respectively. LRF and ECF scheduling algorithms do not have specific variable settings. Their models follow the algorithmic descriptions from [24] and [25]. In addition to the scheduling algorithms, the congestion control in the test environment employs the Linked Increases Convergence Control Algorithm (LIA), which is the default congestion control algorithm of MPTCP in NS3 [46].

B. Performance Analysis

The performance of QLE-DS is compared with that of the other two state-of-art algorithms in the two path heterogeneous network environment in terms of three aspects. First we assess throughput performance in a setup in which the video bit rate varies based on DASH to test the adaptability of the different scheduling algorithms. Second, the performance is evaluated in terms of SRTT and data flow completion time. The results reflect the effect of the trade off between the level of consideration of energy efficiency and performance. Finally,

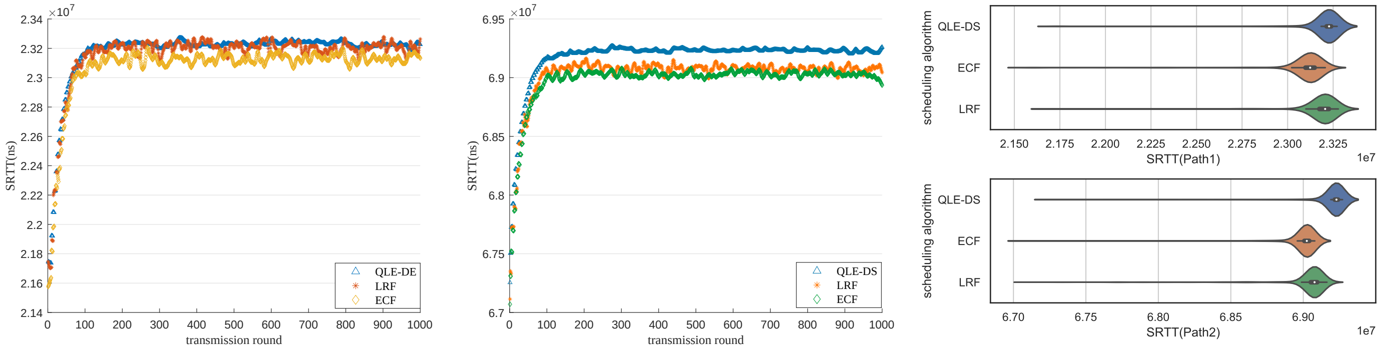


(a) The time-varying throughput

(b) The CDF of throughput

(c) The average throughput

Fig. 6: Throughput comparison of DASH services.



(a) SRTT comparison for path 1

(b) SRTT comparison for path 1

(c) Violin diagram of SRTT distribution on each path

Fig. 7: The SRTT distribution comparison when using QLE-DS, ECF and LRF.

TABLE IV: Flow completion time (FCT) of the three methods

Algorithms	Data Size	FCT(s)
QLE-DS	5GB	40.62
ECF	5GB	43.02
LRF	5GB	54.12

the energy consumption when employing the three algorithms is assessed, respectively.

1) Throughput Performance: Fig. 6(a) shows a comparison of the time-varing throughput achieved by QLE-DS, ECF and LRF over a 1600 second simulation time. The three curves are the same at the starting point, but the curves of QLE-DS and ECF rise faster. Finally, the convergence position of QLE-DS is higher than those of ECF and LRF. We calculate the average throughput of this experiment from the perspective of subflow and draw Fig. 6(c). The average throughput values of QLE-DS, ECF and LRF are 125.16 Mbps, 122.79 Mbps and 94.71 Mbps respectively. Compared with ECF and LRF, QLE-DS records improvements of 1.9% and 32.2%. Fig 6(b) shows the CDF function, which also validates the superiority of QLE-DS. According to the curve, QLE-DS has the best throughput performance and can reach the highest convergence point. For ECF, the curve in Fig. 6 (a) drops sharply twice, which is due to the reset of congestion window caused by timeout, which is not desirable. As shown in Fig 6(c), QLE-DS has more balanced scheduling on the two paths, while LRF prefers the better path, and ECF performance is in between the ones of the other two solutions.

2) SRTT Performance and Flow Completion Time Analysis: In the context of a large path delay difference in the heterogeneous network environment considered, Fig. 7 (a) and Fig. 7 (b) show the SRTT change trend of the three algorithms in the two paths, respectively. The average SRTT values of QLE-DS, ECF and LRF in path 1 are 23.17ms, 23.07ms and 23.15ms, and the average SRTT values in path 2 are 69.18ms, 68.98ms and 69.03ms. According to these results, ECF is the best in terms of SRTT performance. Fig. 7 (c) shows the distribution of packet SRTT of different scheduling algorithms. QLE-DS does not achieve the lowest round trip delay because our proposed algorithm mainly focuses on improving the energy consumption for MPTCP and the SRTT performance was sacrificed. This is due to the relatively large energy consumption weight factor in Eq. (12), which makes energy the most important evaluation index.

Next, based on the model proposed in Section III, we calculate the flow completion time of the three scheduling algorithms when transmitting 5GB data, which simulates a AR & VR video transmission in a 5G network. As shown in Table IV, the flow completion time of QLE-DS is the shortest, which means that our scheduling algorithm makes fewer errors and achieves the highest transmission rate in the same network conditions, in comparison with the other solutions.

3) Energy consumption performance: Fig. 8 illustrates the user battery energy when using QLE-DS, ECF and LRF

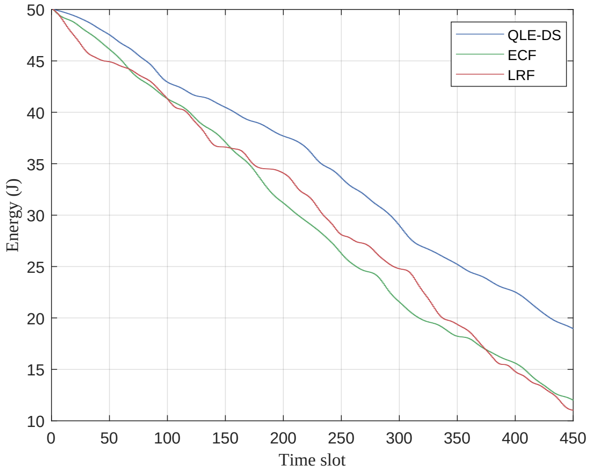


Fig. 8: Comparison of energy depletion.

scheduling algorithms, respectively. As expected, the curves show a decreasing trend, but they differ according to algorithm energy efficiency performance. They show how QLE-DS is superior to the other two algorithms in terms of energy consumption. It can be seen how, when compared with ECF and LRF, the energy decline rate of QLE-DS is slower and more stable. In the energy consumption model presented in Eq. (4), lower energy consumption indicates higher efficiency and better scheduling of the algorithm. Therefore, although the energy consumption is related to the inherent properties and congestion level of each path, QLE-DS successfully gives priority to allocating packets to paths with less energy consumption in a changing environment. In contrast, ECF and LRF algorithms achieve higher energy consumption in this multipath scenario.

VI. Conclusions

This paper proposes a new MPTCP scheduling solution based on reinforcement learning method to optimize energy consumption. QLE-DS relies on the new designed cooperative scheduling architecture and QLEC algorithm. QLE-DS analyzes the impact of path capacity and data retransmission on energy consumption, and then estimates the real situation of each path to reduce the flow completion time. In addition, QLE-DS introduces a new quantum clustering method which helps transform multidimensional discrete parameters into finite state ones. Moreover, we use the critic network to cooperate with the scheduling results of DASH. Based on the energy consumption evaluation and definition model, QLE-DS adopts the solution of offline training and online decision-making, which is a compromise solution, sacrificing the SRTT performance of data packets to meet the energy-aware design objectives. Simulation results show that the proposed QLE-DS achieves improved throughput, flow completion time and energy consumption performance in comparison with two alternative scheduling solutions. Future work will consider deploying our proposed solution in an actual system and test it in real network situations.

References

[1] Visual Networking Index: Forecast and Methodology, 2017–2022, Cisco Inc., San Jose, CA, USA, 2017. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/serviceprovider/visual-networking-index-vni/white-paper-c11–741490.html

[2] A. Zhang et al., ”Video Super-Resolution and Caching–An Edge-Assisted Adaptive Video Streaming Solution,” in IEEE Transactions on Broadcasting, Early Access, 2021.

[3] I. -S. Comşa, G. -M. Muntean and R. Trestian, ”An Innovative Machine-Learning-Based Scheduling Solution for Improving Live UHD Video Streaming Quality in Highly Dynamic Network Environments,” in IEEE Transactions on Broadcasting, vol. 67, no. 1, pp. 212-224, March 2021.

[4] L. Zhong, C. Xu, J. Chen, W. Yan, S. Yang and G. -M. Muntean, ”Joint Optimal Multicast Scheduling and Caching for Improved Performance and Energy Saving in Wireless Heterogeneous Networks,” in IEEE Transactions on Broadcasting, vol. 67, no. 1, pp. 119-130, March 2021.

[5] P. Scopelliti, A. Tropeano, G.-M. Muntean and G. Araniti, ”An Energy-Quality Utility-Based Adaptive Scheduling Solution for Mobile Users in Dense Networks,” in IEEE Transactions on Broadcasting, vol. 66, no. 1, pp. 47-55, March 2020.

[6] K. Gao, C. Xu, P. Zhang, J. Qin, L. Zhong and G. -M. Muntean, ”GCH-MV: Game-Enhanced Compensation Handover Scheme for Multipath TCP in 6G Software Defined Vehicular Networks,” in IEEE Transactions on Vehicular Technology, vol. 69, no. 12, pp. 16142-16154, Dec. 2020.

[7] A. Ford, C. Raiciu, M. J. Handley, O. Bonaventure, and C. Paasch, ”TCP Extensions for Multipath Operation with Multiple Addresses,” RFC 8684, March 2020.

[8] A. Brighente, M. Cerutti, M. Nicoli, S. Tomasin and U. Spagnolini, ”Estimation of Wideband Dynamic mmWave and THz Channels for 5G Systems and Beyond,” in IEEE Journal on Selected Areas in Communications, vol. 38, no. 9, pp. 2026-2040, Sept. 2020.

[9] M. Casoni, C. A. Grazia and M. Klapez, ”SDN-Based Resource Pooling to Provide Transparent Multi-Path Communications,” in IEEE Communications Magazine, vol. 55, no. 12, pp. 172-178, Dec. 2017.

[10] S. R. Pokhrel, Y. Qu, and L. Gao, “Qos-aware personalized privacy with multipath tcp for industrial iot: Analysis and design,” IEEE Internet of Things Journal, vol. 7, no. 6, 2020.

[11] J. Wu, R. Tan and M. Wang, ”Energy-Efficient Multipath TCP for Quality-Guaranteed Video Over Heterogeneous Wireless Networks,” in IEEE Transactions on Multimedia, vol. 21, no. 6, pp. 1593-1608, June 2019.

[12] P. Hurtig, K. Grinnemo, A. Brunstrom, S. Ferlin, Ö. Alay and N. Kuhn, ”Low-Latency Scheduling in MPTCP,” in IEEE/ACM Transactions on Networking, vol. 27, no. 1, pp. 302-315, Feb. 2019.

[13] L. Feng, Z. Yang, Y. Yang, X. Que and K. Zhang, ”Smart Mode Selection Using Online Reinforcement Learning for VR Broadband Broadcasting in D2D Assisted 5G HetNets,” in IEEE Transactions on Broadcasting, vol. 66, no. 2, pp. 600-611, June 2020.

[14] J. Luo, X. Su and B. Liu, ”A Reinforcement Learning Approach for Multipath TCP Data Scheduling,” 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), 2019.

[15] J. Montalban, G.-M Muntean and P. Angueira, ”A Utility-Based Framework for Performance and Energy-Aware Convergence in 5G Heterogeneous Network Environments,” in IEEE Transactions on Broadcasting, vol. 66, no. 2, pp. 589-599, June 2020.

[16] C. E. G. Moreta, M. R. C. Acosta and I. Koo, ”Prediction of Digital Terrestrial Television Coverage Using Machine Learning Regression,” in IEEE Transactions on Broadcasting, vol. 65, no. 4, pp. 702-712, Dec. 2019.

[17] S. R. Pokhrel, N. Kumar, A. Walid, ”Towards Ultra Reliable Low Latency Multipath TCP For Connected Autonomous Vehicles”, IEEE Transactions on Vehicular Technology, 2021.

[18] C. Lee, J. Jung, J. M. Chung, ”DEFT: Multipath TCP for high speed low latency communications in 5G networks”, IEEE Transactions on Mobile Computing, 2020.

[19] S. R. Pokhrel and J. Choi, ”Low-Delay Scheduling for Internet of Vehicles: Load-Balanced Multipath Communication With FEC,” in IEEE Transactions on Communications, vol. 67, no. 12, pp. 8489-8501, Dec. 2019.

[20] F. Tang, Y. Zhou and N. Kato, ”Deep Reinforcement Learning for Dynamic Uplink/Downlink Resource Allocation in High Mobility 5G HetNet,” in IEEE Journal on Selected Areas in Communications, vol. 38, no. 12, pp. 2773-2782, Dec. 2020.

[21] W. Zhao, J. Liu, H. Guo, and T. Hara, “Etc-iot: Edge-node-assisted transmitting for the cloud-centric internet of things,” IEEE Network, vol. 32, no. 3, 2018.

[22] C. Long, Y. Cao, T. Jiang, and Q. Zhang, "Edge computing framework for cooperative video processing in multimedia iot systems," IEEE Transactions on Multimedia, vol. 20, no. 5, 2018.

[23] K. Wang, Y. Shao, L. Xie, J. Wu, and S. Guo, "Adaptive and faulttolerant data processing in healthcare iot based on fog computing," IEEE Transactions on Network Science and Engineering, vol. 7, no. 1, 2020.

[24] Paasch C, Ferlin S, Alay O, et al, "Experimental evaluation of multipath TCP schedulers," ACM SIGCOMM workshop on Capacity sharing workshop, 2014.

[25] Lim, Yeon-sup, et al, "ECF: An MPTCP path scheduler to manage heterogeneous paths," The 13th CoNEXT: international conference on emerging networking experiments and technologies, 2017.

[26] X. Chen, P. Liu, H. Liu, C. Wu, and Y. Ji, "Multipath transmission scheduling in millimeter wave cloud radio access networks," in 2018 IEEE International Conference on Communications (ICC), 2018.

[27] K. Xue, J. Han, D. Ni, W. Wei, Y. Cai, Q. Xu, and P. Hong, "Dpsaf: Forward prediction based dynamic packet scheduling and adjusting with feedback for multipath tcp in lossy heterogeneous networks," IEEE Transactions on Vehicular Technology, vol. 67, no. 2, 2018.

[28] S. R. Pokhrel, A. Walid, "Learning to harness bandwidth with multipath congestion control and scheduling", IEEE Transactions on Mobile Computing, 2021.

[29] I. Alqerm and B. Shihada, "Sophisticated Online Learning Scheme for Green Resource Allocation in 5G Heterogeneous Cloud Radio Access Networks," in IEEE Transactions on Mobile Computing, vol. 17, no. 10, pp. 2423-2437, 1 Oct. 2018.

[30] M. Khabbazian and K. Gharouni Saffar, "The Gain of Energy Accumulation in Multi-Hop Wireless Network Broadcast," in IEEE/ACM Transactions on Networking, vol. 27, no. 5, pp. 1830-1844, Oct. 2019.

[31] C. Yao, J. Xiao, Y. Zhao and A. Ming, "Video Streaming Adaptation Strategy for Multiview Navigation Over DASH," in IEEE Transactions on Broadcasting, vol. 65, no. 3, pp. 521-533, Sept. 2019.

[32] X. Tao, Z. Chen, M. Xu and J. Lu, "Rebuffering Optimization for DASH via Pricing and EEG-Based QoE Modeling," in IEEE Journal on Selected Areas in Communications, vol. 37, no. 7, pp. 1549-1565, July 2019.

[33] K. Spiteri, R. Urgaonkar and R. K. Sitaraman, "BOLA: Near-Optimal Bitrate Adaptation for Online Videos," in IEEE/ACM Transactions on Networking, vol. 28, no. 4, pp. 1698-1711, Aug. 2020.

[34] T. Zhang, F. Ren, W. Cheng, X. Luo, R. Shu and X. Liu, "Towards Influence of Chunk Size Variation on Video Streaming in Wireless Networks," in IEEE Transactions on Mobile Computing, vol. 19, no. 7, pp. 1715-1730, 1 July 2020.

[35] Praveen Kumar Yadav and Wei Tsang Ooi, "Tile Rate Allocation for 360-Degree Tiled Adaptive Video Streaming," The 28th ACM International Conference on Multimedia, 2020.

[36] A. Yaqoob and G. -M. Muntean, "A Combined Field-of-View Prediction-Assisted Viewport Adaptive Delivery Scheme for 360° Videos," in IEEE Transactions on Broadcasting, vol. 67, no. 3, pp. 746-760, Sept. 2021.

[37] Z. Jiang, C. Xu, J. Guan, Y. Liu and G.-M. Muntean, "Stochastic Analysis of DASH-Based Video Service in High-Speed Railway Networks," in IEEE Transactions on Multimedia, vol. 21, no. 6, pp. 1577-1592, June 2019.

[38] D. Mi et al., "Demonstrating Immersive Media Delivery on 5G Broadcast and Multicast Testing Networks," in IEEE Transactions on Broadcasting, vol. 66, no. 2, pp. 555-570, June 2020.

[39] D. Horn, A. Gottlieb, "Algorithm for data clustering in pattern recognition problems based on quantum mechanics," in Physical review letters, 2001.

[40] D. Horn, A. Gottlieb, "The Method of Quantum Clustering," in Nips. 2001.

[41] J. Wu, B. Cheng, M. Wang, and J. Chen, "Energy-efficient bandwidth aggregation for delay-constrained video over heterogeneous wireless networks," IEEE Journal on Selected Areas in Communications, vol. 35, no. 1, 2017.

[42] Q. Peng, A. Walid, S H. Low, "Multipath TCP algorithms: theory and design", ACM SIGMETRICS Performance Evaluation Review, 2013.

[43] E. Harjula, O. Kassinen, and M. Ylianttila, "Energy consumption model for mobile devices in 3g and wlan networks," in 2012 IEEE Consumer Communications and Networking Conference (CCNC), 2012.

[44] S. Agethen and W. H. Hsu, "Deep Multi-Kernel Convolutional LSTM Networks and an Attention-Based Mechanism for Videos," in IEEE Transactions on Multimedia, vol. 22, no. 3, pp. 819-829, March 2020.

[45] Wischik D, Raiciu C, Greenhalgh A, et al. Design, Implementation and Evaluation of Congestion Control for Multipath TCP, NSDI. 2011.

[46] N. Kashif. (2018) mptcp implemention in ns3 : https://github.com/Kashif-Nadeem/ns-3-dev-git.

Lujie Zhong received the Ph.D. degree from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2013. She is currently an Associate Professor with the Information Engineering College, Capital Normal University, Beijing, China. She has published papers in prestigious international journals and conferences in the related area, including IEEE Communication Magazine, IEEE Transactions on Mobile Computing, IEEE Transactions on Multimedia, IEEE Internet Things Journal, IEEE INFOCOM and Acm Multimedia, etc. Her research interests include communication networks, computer system and architecture, and mobile Internet technology.

Xiang Ji received B.S. degree from the North China Electric Power University, BaoDing, China, in 2019. He is currently working toward the Ph.D. degree with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China, His research interests include multipath transmission protocol, pareto optimization, and deep learning.

Zhaoxue Wang received B.S. degree from the Capital Normal University, Beijing, China, in 2019. She is currently working toward the M.S. degree with the Information Engineering College, Capital Normal University, Beijing, China. Her research interests include multipath transmission control protocol and reinforcement learning.

Jiuren Qin received the Ph. D. degree from Beijing University of Posts and Telecommunications, China, in 2021. She is currently an engineer at the National Key Laboratory of Science and Technology on Information System Security. Her research interests include network security, multipath transmission protocol, multimedia communication, etc.

Gabriel-Miro Muntean is a Professor with the School of Electronic Engineering, Dublin City University (DCU), Ireland, and coDirector of DCU Performance Engineering Laboratory. He has published 4 books and over 450 papers in top international journals and conferences. His research interests include rich media delivery quality, performance, and energy-related issues, technology enhanced learning, and other data communications in heterogeneous networks. He is an Associate Editor of the IEEE Transactions on Broadcasting, the Multimedia Communications Area Editor of the IEEE Communications Surveys and Tutorials, and reviewer for important international journals, conferences, and funding agencies. He coordinated the EU project NEWTON and leads the DCU team in the EU project TRACTION.