JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 38, 165-183 (2022)
DOI: 10.6688/JISE.202201_38(1).0009

Q-Learning Based and Energy-Aware Multipath Congestion Control in Mobile Wireless Network*

JIUREN QIN, KAI GAO, LUJIE ZHONG AND SHUJIE YANG
State Key Lab of Networking and Switching Technology
Beijing University of Posts and Telecommunications
Haidian District, Beijing, 100876 P.R. China
E-mail: {jrqin; gaokai; sjyang}@bupt.edu.cn; zhonglj@cnu.edu.cn

Along with the development of mobile wireless communication technologies, many devices are equipped with more than on network interfaces (4G/5G, Wi-Fi, Bluetooth, *etc*.). To aggregate the idle bandwidth of different network interfaces, Multipath Transmission Control Protocols (MPTCP) are standardized by the Internet Engineering Task Force (IETF). MPTCP can establish sub-flows through different network interface in one connection and improve the transmission efficiency by transmitting data concurrently. However, there are still two problem for MPTCP to work in the mobile wireless network: (1) Unawareness to the network changes; (2) No consideration of energy consumption. To address these two issues, we propose the Q-Learning based and Energy-aware Multipath Congestion Control (QE-MCC) scheme in this paper. Firstly, the stability and trend parameters are introduced to formulate the system state. Then, an energy-aware transmission utility model is presented to evaluate the effects of congestion control. Finally, the Q-learning based congestion control algorithms are designed to improve transmission efficiency. The simulation results shows that QE-MCC performs better on throughput, delay and, energy consumption compared with standard and similar solutions.

Keywords: Q-learning, energy, MPTCP, congestion control, mobile wireless networks

1. INTRODUCTION

Recently, the fast-developing wireless networks enables various mobile services (self-driving, AR/VR, online game, and video conference, *etc*.) [1]. These services not only facilitate our life, but also bring challenges to the network transmission. On the one hand, different services product massive data which need to be transmitted under strict throughput and delay constraint. On the other hand, the dynamic topology in mobile wireless networks improves the possibility of link error. Thus, how to improve the transmission efficiency in wireless mobile network is a research hotspot [2, 3].

The Internet Engineering Task Force (IETF) standards Multipath Transmission Control Protocols (MPTCP) in RFC8684 [4–6] which is a potential solution to mobile wireless

Received December 17, 2020; revised January 30, 2021; accepted March 7, 2021.
Communicated by Changqiao Xu.
* This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant No. 61871048, by the National Key R&D Program of China under Grant No. 2018YFE0205502.

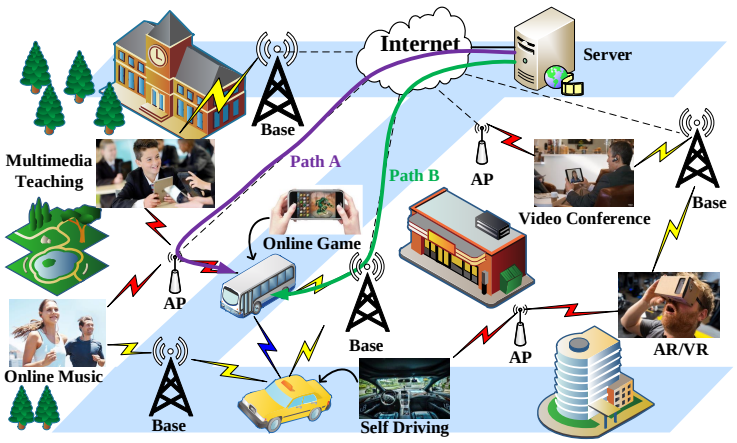


Fig. 1. Multipath transmission in wireless mobile networks.

transmission. As shown in Fig. 1, the MPTCP sender can set up sub-flows through different network interfaces. By transmitting data concurrently, MPTCP can utilize the idle bandwidth of different networks and get higher throughput. Besides, MPTCP performs more robust than single path protocol for the reason that sender can transfer the traffic from the congested sub-flow to the healthy ones. The above advantages make MPTCP increasingly used in mobile network transmission scenarios [7–9]. However, the standard still falls short in two ways:

(1) *Unawareness to the network changes*. MPTCP cannot effectively perceive the state of transmission and has no knowledge of the network changes. Thus the congestion control actions are usually inefficient and delayed.

(2) *No consideration of energy consumption.* The mobile devices are usually energy-limited. The huge energy consumption caused by multipath transmission is unfriendly to the users.

Many researcher proposed different solution to the above problems. [10] focused on the multipath videos transmission in mobile wireless network. An analytical framework was firstly designed to characterize the delay-constrained energy-quality trade-off in transmission. Then, a multipath allocation algorithm named DEAM was introduced to minimize the energy consumption while achieving the video transmission quality. However, DEAM has insufficient perception of the network condition. [11] proposed a learning-based multipath transmission control approach for heterogeneous networks which can observe the environment and adjust the congestion windows to fit different network situations. However, this solution do not consider the energy consumption and changing trend of the network. Our team previously also did a lot of works on multipath transmission optimization [12–15].

However, there is still no intelligent solution that can solve the above two problems at the same time. Thus, we propose **Q-learning based and Energy-aware Multipath Congestion Control (QE-MCC)** scheme in this paper to improve MPTCP performance in the mobile wireless networks. QE-MCC works in the transport layer and can help the sender find the Best congestion strategies for complex and dynamic networks by consid-

ering the transmission efficiency and energy consumption concurrently. The contributions of this paper can be summarized as:

- Give a comprehensive formulation to the multipath transmission system in the mobile wireless networks.
- Design a Fuzzy C-Means based clustering algorithm to simplify the high-dimensional state space.
- Propose an Energy-aware Transmission Utility Model to quantify the congestion control effect.
- Develop a Q-learning based two-layer multipath congestion control algorithm which can find the optimal actions to the dynamic and energy-limited system.
- Implement the QE-MCC in the Network Simulator 3 (NS3) and prove that QE-MCC can get better performance on throughput, delay, and energy-saving compared with standard and similar solutions.

The remainder of this paper is organized as follows. Section 2 presents the related works. Section 3 gives a brief introduction to the QE-MCC systems. The detailed QE-MCC algorithms are shown in Section 4. Section 5 displays the performance evaluation. Conclusion ans future works are discussed in Section 6.

2. RELATED WORKS

This section presents the relate works which can be classified as three subsection: (1) Multipath Transmission Control Protocols; (2) MPTCP in Mobile Wireless network; (3) Intelligent Algorithms for MPTCP.

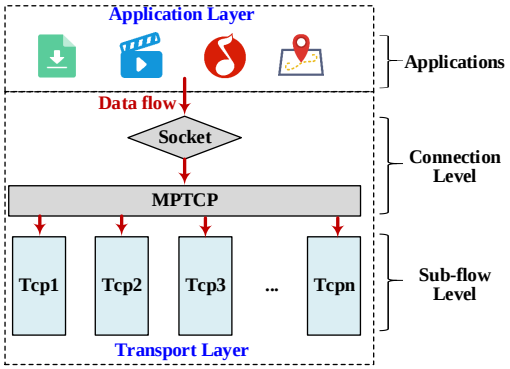


Fig. 2. The MPTCP stacks.

2.1 Multipath Transmission Control Protocols

MPTCP is an extension of single-path Transmission Control Protocols (TCP) which introduces a MPTCP layer to the transport layer. As shown in Fig. 2, the transport layer can be logically divided as connection level and sub-flow level. The data flow from the application layer will firstly be handled in the connection level where the data will segmented and numbered with Data Sequence Number (DSN). Then, these segments will be allocated to different TCP sub-flows where a Sub-flow Sequence Number (SSN) will be mapped to each segments. Finally, these segments will be transmitted by different sub-flows concurrently and independently. When these segments arriving at the receiver, the Acknowledgment (ACK) which contains the DSNs and SSNs of received segments will be fed back to the sender through the sub-flow where these segments come from. The SSNs are used to acknowledge receiving in sub-flow level while the DSNs are used to acknowledge receiving in connection level.

The congestion algorithms of MPTCP determines the transmission rate of different sub-flows and speed of recovery from congestion. The standard congestion algorithms of MPTCP are designed based on window-oriented congestion control mechanism. The most representative one is the Additive Increase Multiplicative Decrease (AIMD) which can be summarized as:

- Each ACK on sub-flow r, $cwnd_r \leftarrow cwnd_r + \frac{I_r}{cwnd_r}$.
- Each congestion on sub-flow r, $cwnd_r \leftarrow cwnd_r(1 - \frac{D_r}{cwnd_r})$.

The parameters I_r and D_r are separately the increase and decrease factor. I_r determines the cwnd increase scale when segments successfully transmitted. D_r determines the cwnd decrease scale when congestion signals (time-out, three dup Acks, $etc.$) are detected. $cwnd_r$ is the congestion window of sub-flow r.

RFC6356 [16] gives the design goals of multipath congestion control algorithms: (1) The throughput of MPTCP should be higher than the single path protocols or at least be equal; (2) MPTCP should not take more resource than the competed single flow; (3) MPTCP should move traffic off the congested sub-flow to the Balance congestion. Based on above design goals, the default MPTCP congestion control algorithm **LIA** is formulated as:

- Each ACK on sub-flow r, $cwnd_r \leftarrow cwnd_r + min(\frac{\alpha}{cwnd_{total}}, \frac{1}{cwnd_r})$.
- Each congestion signal on sub-flow$r$, $cwnd_r \leftarrow cwnd_r/2$.

The parameter α is the aggressiveness factor which can be calculated by $\alpha = cwnd_{total} \frac{\max\limits_r cwnd_r/\tau_r^2}{(\sum\limits_r cwnd_r/\tau_r)^2}$. $cwnd_{total}$ is the sum of sub-flow congestion window, which can be calculated by $cwnd_{total} = \sum_r cwnd_r$. The increase and decrease factor of LIA are separately $I_r = min(\alpha cwnd_r/cwnd_{total},\ 1)$ and $D_r = cwnd_r/2$.

2.2 MPTCP in Mobile Wireless Networks

To improve the performance of MPTCP in mobile wireless networks, many solutions have been proposed. In [17], a restricted offloading scheme was proposed to enhance the MPTCP throughput by aggregating the fifth generation (5G) new radio (NR) and LTE

bandwidth. In [18], the authors firstly developed a comprehensive approach to assess the the performance of long-lived MPTCP flows. Then, a multipath congestion control algorithm was designed based on a parallel queueing model to improve the performance of MPTCP over Cellular and WiFi networks. Lee *et al*. [19] proposed a delay-equalized scheme which can response quickly to the link state changes and achieve low end to end transmission delay by minimizing the additional reordering delay. The authors of [20] presented a receiver adaptive incremental delay algorithm named RAID to improve the MPTCP performance in high-speed mobile scenario. RAID can aggregate bandwidth for heterogeneous networks independent of accurate network quality estimation. Xue *et al*. [21] tried to solve the out-of-order problem in multipath transmission and proposed a new scheduling algorithm which can estimate the data amount sent on different sub-flow based on maximum likelihood model. However, the above solution can not solve the network state perception and energy problems at the same time.

2.3 Intelligent Algorithms for MPTCP

In recent years, the Artificial Intelligence (AI) technologies are applied to the multipath transmission control and many solutions are given. Xu *et al*. [22] proposed a deep reinforcement learning (DRL)-based control framework to make MPTCP learn the best scheduling strategies based on its own experiences. The proposed solution utilizes a flexible recurrent neural network to learn a representation for all active flows and dealing with their dynamics. The authors of [23] designed a Q-learning framework to improve the MPTCP energy efficiency in a resource-shared wireless network considering the influence of observed interface capacity and other competitors' decision. In [24], a Reinforcement Learning based Scheduler for MPTCP was proposed which can generate the control policy for packet scheduling and balance the traffic over multiple sub-flows. [25] introduced a Deep Q Network (DQN) framework to improve MPTCP performance in the asymmetric path which can get the information of each sub-flow and adaptively choose the most suitable sub-flow for transmission. In [26], Mai *et al*. focused on the MPTCP optimization in low earth orbit (LEO) satellites networks and employed the deep deterministic policy gradient to find the optimal congestion control strategies for dynamic underlying networks. However, the above solutions lacks the accurate state definition to the performance and changing trends of transmission system.

Thus, we designed an intelligent multipath congestion control algorithm for MPTCP which can effectively estimate the stability and trend parameters of transmission system and decrease the energy consumption by moving traffic to low-power sub-flows.

3. SYSTEM DESIGN

The system design of QE-MCC is shown in Fig. 3. The yellow arrows, slid black arrows and dotted arrows in Fig. 3 separately denote the data, control and inside flows. The data flow start at the sever. The MPTCP sender allocates the data to different subflows and transmit them concurrently. Transmitted through the mobile wireless network, the data flow arrive at the receiver buffer and end at the specified mobile device. The rounded rectangles denote the proposed modules. From the figure, we can know that QE-MCC contains two layer: Offline Training and Online Learning.

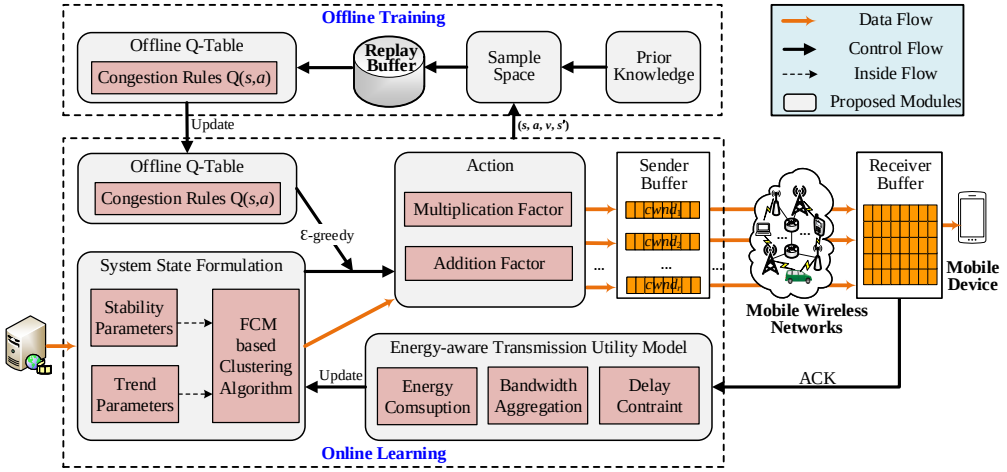


Fig. 3. The system design of QE-MCC.

The offline training layer contains three modules: Prior Knowledge, Sample Space, and Offline Q-table. The prior knowledge module provide the initial congestion control rules and the samples got under those rules. The Sample Space module is used to store the samples provide by the Prior Knowledge Prior Knowledge and Online Learning layer. The Offline Q-Table module records the congestion rules and is updated by replaying the samples.

The online learning layer contains four modules: Online Q-Table, System State Formulation, Action and Energy-aware Transmission Utility Model. The Online Q-Table module will periodically update its congestion based on the Offline Q-Table. The System State Formulation module formulate the system state based on the stability and trend parameters and simplify the higher dimensional state through the FCM based Clustering Algorithm. The Action module quantifies the cwnd adjustment with the multiplication and addition factors. The Energy-aware Transmission Utility Model evaluate the reward of congestion control actions.

The control flow can be summered as: (1) the offline training layer continuously update the congestion rules; (2) periodically transmit the congestion rules to the online learning layer; (3) find the optimal actions to current system state following the congestion rules; (4) adjust the cwnd accordingly and calculate the reward when receiving new ACKs; (5) record the congestion control logs and feed them back to the offline training layer for further training.

4. QE-MCC ALGORITHMS

In this section, the QE-MCC Algorithms are detailed which includes: System State Formulation, Energy-aware Transmission Utility Model and Q-Learning based Multipath Congestion Control algorithms.

4.1 System State Formulation

In order to formulate the states of the multipath transmission system accurately and comprehensively, the calculation process of stability and trend parameters are introduced in this subsection. The stability parameters quantify the performance of a sub-flow in current evaluation period. The trend parameters reflects the changes between different evaluation periods.

Considering a multipath connection with r sub-flows, the sub-flow set can be indicated as $R = \{1,\ 2,\ \ldots,\ r\}$. During the transmission, the sender can continuous receiving ACKs form different sub-flows in the connection level. The time between receiving these two ACKs can be defined as one evaluation period. For each sun-flow, the stability and trend parameters are maintained independently. When receiving an Acknowledged (ACK) segment in the connection level, the sender will firstly check the sub-flow number that the ACK comes from, and then update the stability and trend parameters of that sub-flow.

The stability parameters of sub-flow r can be summarized as:$cwnd_r$, RTT_r, and BW_r. $cwnd_r$ is the congestion window of sub-flow r, which is determined by the congestion control algorithm. RTT_r is Round Trip Time (RTT) of sub-flow r which can be calculated by timestamp. When sending the segments, the sender will attach the local time to the TCP head as the timestamp option. When receiving the segments, the receiver will read the timestamp option and get it back to the sender with the ACKs. Thus, the sender can get the sending time of the acknowledged segments. By varying the current system time from the sending time, one RTT sample can be gotten. BW_r is the bandwidth of sub-flow r, which can be calculated as following:

$$BW_r = \frac{Data_{ACK}}{RTT_r}. \tag{1}$$

$Data_{ACK}$ is amount of data acknowledged by one ACK. RTT_r is the sampled RTT calculated by the timestamp in that ACK.

The trend parameters of sub-flow r can be summarized as: ω_r, τ_r and, β_r . ω_r is the window parameter, which is defined as:

$$\omega_r = \frac{cwnd_{t+1,r}}{cwnd_{t,r}}. \tag{2}$$

$cwnd_{a,r}$ denotes the congestion window size of sub-flow r in ath evaluation periods. ω_r can reflect the changing trend of congestion window. τ_r is the transmission delay parameter, which is defined as:

$$\tau_r = \frac{RTT_{t+1,r}}{RTT_{t,r}}. \tag{3}$$

$RTT_{a,r}$ denotes the round trip time of sub-flow r in ath evaluation periods. tau_r can reflect the changing trend of transmission delay. β_r is the bandwidth parameter, which is defined as:

$$\beta_r = \frac{BW_{t+1,r}}{BW_{t,r}}. \tag{4}$$

$BW_{a,r}$ denotes the calculated bandwidth of sub-flow r in ath evaluation periods. β_r can reflect the changing trend of bandwidth.

Based on the above definition, the state of sub-flow can be represented as: $s_r = \{cwnd_r, RTT_r, BW_r, \omega_r, \tau_r, \beta_r\}$. The state of the MPTCP connection can be summarized as: $s = \{s_1, s_2, \ldots, s_r\}$. From the definition, we can know the dimension of s will increase with number of sub-flow. However, in the Q-learning algorithm, a high-dimensional state space can make the Q-table verbose and increase training costs. Thus, a Fuzzy C-Means (FCM) based clustering algorithm is proposed to simplify the state space.

The objective function of FCM algorithm can be defined as:

$$J_m(U,V) = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^m d_{ij}^2, \tag{5}$$

where matrix U is the result of the classification, and matrix V is the combination of center vector v. c is the number of clustering center, and n is the number of state point. m is the fuzzy weighted index and describes degree of fuzziness. u_{ij} represents the membership value, the membership degree of state j to class i. The value of u_{ij} need to meet the following constraint:

$$\sum_{i=1}^{c} u_{ij} = 1. \tag{6}$$

d_{ij} is the distance between point j and center point i under Euclidean distance. The meaning of the objective function J can be understood as the sum of the distances from each point to each cluster center.

The final goal of clustering is the smallest similarity within the class and the largest similarity between the classes. Thus, the clustering process is to minimize the objective function J under the constraint Eq. (6). The constrained extremum problem can be solved by the Lagrange multiplier method. The Lagrange function can be constructed as follows:

$$F = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^m d_{ij}^2 + \sum_{j=1}^{n} \lambda_j \left(\sum_{i=1}^{c} u_{ij} - 1 \right), \tag{7}$$

where λ is the restriction factor. By taking the derivative, the membership u_{ij} and center vector v can be calculated by following two equations:

$$u_{ij} = \left[\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{kj}} \right)^{\frac{2}{m-1}} \right]^{-1}, \tag{8}$$

$$v_i = \frac{\sum_{j=1}^{n} x_j u_{ij}^m}{\sum_{j=1}^{n} u_{ij}}. \tag{9}$$

Based on the above analysis, FCM based state clustering process is detailed in Algorithm 1.

The high-dimensional state s can be mapped to a one-dimensional state $\hat{s} \in \{1,2,3,...,c\}$ according to Algorithm 1. The size of Q-table size and the training cost will be reduced which is friendly to the resource-constrained mobile wireless networks.

Algorithm 1: FCM based State Clustering

Input: State samples set S, Sample size n, Number of clustering center c, Convergence criterion σ.
Output: Membership u_{ij}, Center vector v, Clustered state.
Initialization: Initialize the center matrix $V^{(0)}$, $k = 0$.
while $\left\| V^{(k+1)} - V^{(k)} \right\| > \sigma$ **do**
    $k = k + 1$;
    Compute the $u_{ij}^{(k)}$ according to Eq. (8);
    Update the $U^{(k)}$ based on the calculated $u_{ij}^{(k)}$;
    Compute the $v_i^{(k+1)}$ according to Eq. (9);
    Update the $V^{(k+1)}$ based on the calculated $v_i^{(k+1)}$;
Get the final U and V;
Let $\{1, 2, \ldots, c\}$ denotes the simplified state set;
Correspond the cluster centers to the simplified states;
for *each* $s \in S$ **do**
    Get the membership u of state s to each center vector v based on the final U and V;
    Select the center vector v with maximal membership u;
    Find the corresponding number c of the selected center vector v.
    Simplied the high-dimensional state s to one-dimensional state c.

4.2 Energy-Aware Transmission Utility Model

To evaluate control effect of QE-MCC algorithms, an energy-aware transmission utility model is proposed in this subsection which can be formulated as:

$$\mathbf{O} = U_\alpha(B_R) - \delta U_\beta(T_R) - \zeta E_R. \tag{10}$$

B_R is the total bandwidth of MPTCP connection, and can be calculated by:

$$B_R = \sum_{r \in R} BW_r. \tag{11}$$

T_R is the sum of sub-flow round trip time which can be calculated by:

$$T_R = \sum_{r \in R} RTT_r. \tag{12}$$

E_R is total energy consumption which can be calculated by:

$$E_R = \sum_{r \in R} e_r \cdot cwmd_r, \tag{13}$$

where e_r is the energy factor of sub-flow r and denotes the amount of power used to transmit one bit of data. The parameters δ and ζ in Eq. (10) are the factors of relative im-

portance which are used to balance the proportion of different parameters. The parameters α and β in Eq. (10) express the fairness-vs.-efficiency tradeoffs. The $U_x(y)$ in Eq. (10) is defined according to [27]:

$$U_p(q) = \begin{cases} \frac{q^{1-p}}{1-p}, & \text{if } p > 0 \text{ and } p \neq 0 \\ \log q, & \text{if } p = 1, \end{cases} \tag{14}$$

where p is the fairness factor and q is the evaluated variate. $p = 0$ indicates no fairness and the value of q will be maximized. $p = 1$ indicates the proportional fairness that the competitors will be treated equally in proportional. $p \to \infty$ indicates the conventional max-min fairness that trends to keep all competitors at the same level.

That tends to divide the bandwidth of a bottleneck link equally among flows, which is exactly the optimization objective of the conventional max-min fairness that will try achieve the min

The proposed transmission utility model quantify the effect of congestion control and can be used as the reward of Q-learning algorithms.

4.3 Q-Learning based Multipath Congestion Control

The Q-learning model for MPTCP congestion control is firstly introduced in this subsection. Then a two-layer learning algorithm is designed to improve congestion control efficiency.

To formulate the multipath transmission congestion control process with a Q-learning model, the key elements (agent, state, action, reward and policy) must be defined firstly.

Agent: The agent plays the role to learn and make decisions. In multipath congestion control,let the transmission controller be the agent A, which has r sub-flows. Each evaluation period defined in Subsection 4.1 can be seen as a time slot t.

State: The states are the quantitative expressions of the environment which must be discrete and finite. Based on the analysis in Subsection 4.1, the high-dimensional state $s = \{s_1, s_2, \ldots, s_r\}$ can be simplified to a one-dimensional state $\hat{s} \in \{1,2,3,...,c\}$.

Action: The actions in the Q-learning model can be seen as the agent reactions to the observed state. To be compatible with the default LIA algorithm, the actions for QE-MCC is also designed based on cwnd. In each time slot, the cwnd adjustments which can be formulated as:

$$cwnd_{t+1} = \phi \cdot cwnd_t + \varphi, \tag{15}$$

where ϕ is the multiplication factor and φ is the addition factor. According to the definition, the action for sub-flow r can be denoted as: $a_r = (\phi_r, \varphi_r)$. The action for the MPTCP connection can be denoted as:

$$a = (a_1, a_2, a_3, ..., a_r). \tag{16}$$

Reward: The reward can be seen as the effect of taking an action in a given state which is related to the final goals. The reward in Q-learning model must be non-aftereffect which means it only depends on the current state and action, and is independent of previous states and actions. In QE-MCC, the reward r is defined based on the energy-aware

transmission utility model proposed in Subsection 4.2. By setting the $p = 1$, the Eq. (14) can be denoted as: $\log q$. According to Eq. (10), the reward of QE-MCC can be denoted as:

$$r = \log(B_R) - \delta \log(T_R) - \zeta \log(E_R). \tag{17}$$

Policy: The policy π is the rules set of how to select the action in each state which directly influences the learning results. As the objective of QE-MCC is to maximize the discounted rewards, π is defined as ε-greedy. Under this policy, agent will randomly select actions with ε probability and select actions following the greedy rules with $1-\varepsilon$ probability.

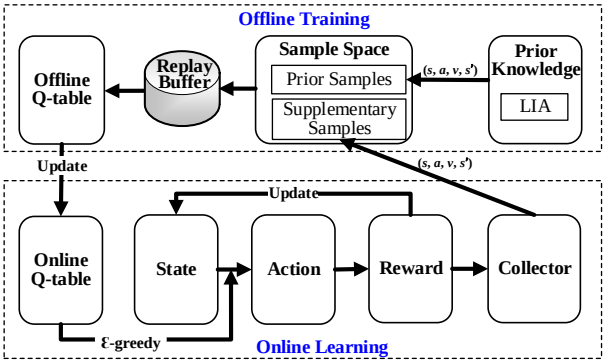


Fig. 4. The two-layer Q-learning framework.

Based on the above definitions, a Two-Layer Q-learning (TLQ) framework is displayed in Fig. 4. The multipath congestion control contains two processes: the offline training and online learning.

In the offline training layer, LIA algorithm is selected as the prior knowledge to avoid making bad decisions at the beginning. Based on the prior knowledge, prior samples (s,a,r,s') can be produced, where s' is the latest state after taking action a in state s. In the replay buffer, the samples will be analyzed. Q-learning algorithms usage the Q-table to record the $Q(s,a)$ for each state and action. When new sample is replayed, the offline Q-table will be updated as following:

$$Q(s,a) \leftarrow Q(s,a) + \theta[r + \omega Q(s',a') - Q(s,a)], \tag{18}$$

where $Q(s,a)$ is estimated reward of taking action a in state s. a' is the action that can get maximal reward in the next state s'. $\theta \in (0,1)$ is the learning rate and $\omega \in (0,1)$ is the discount rate. The closer θ gets to 1, the learning algorithm puts more emphasis on new knowledge instead of previous training. The closer ω gets to 1, the learning algorithm puts more emphasis on future rewards. Periodically, the values in the offline Q-table will be transmitted to the online Q-table. The update period can be defined as M.

In the online learning layer, the MPTCP sender will collect the parameters formulate in Subsection 4.1 and determine the state of current system. Then, the optimal action can be selected based on the online Q-table following the ε-greedy policy. After adjusting the cwnd according to the selected action, reward will be got when new ACK are received.

Algorithm 2: TLQ algorithm

Input: Prior Samples, Update period M.
Output: Congestion control strategies.
Initialization: Initialize the offline Q-table and online Q-table with the prior knowledge.
for *each M* **do**
    Select a sample (s, a, r, s') from the sample space;
    Replay the samplel;
    Update the offline Q-table according to the Eq. (18);
Update the online Q-table based on the offline Q-table;
while *transmission not end* **do**
    Calculate the stability and trend parameters;
    Obtain the system state s based on the FCM clustering algorithm;
    Chose the optimal action a based on the online Q-table;
    Adjust the cwnd according to the Eq. (15);
    Calculate the reward r according to Eq. (17);
    Get the new state s';
    Transmit the log (s, a, r, s') to the sample space;

Then the collected parameters will be recalculated and the system state wil also be updated. Finally, the congestion control log will be collected as the supplementary samples and be transmitted to the sample space.

Overall, the offline training layer can periodically supply an available Q-tale to the online learning layer and reduce the computation delay caused by online training. The online learning layer executes the trained congestion strategies and produces new supplementary samples to the offline training layer. By replaying these samples, the values in the offline Q-table be more accurate. The TLQ algorithm is formulated in Algorithm 2.

5. PERFORMANCE EVALUATION

5.1 Simulation Setup

To evaluate the performance of proposed solution, we implement QE-MCC and compared solutions: LIA [16], SmartCC [11], and DEAM [10] in the NS-3.29 platform based on the MPTCP implementation published in [28]. The simulation topology is shown in Fig. 5. A most common scenario in our life is simulate where a mobile device access the sever through both Wi-Fi and Cellular interfaces. The parameter settings are shown in Table 1.

To simulate Internet traffic, a background flow is injected into the network. Having the results shown in [13], the background flow has the following details: 49% are 44 bytes, 1.2% are 576 bytes, 2.1% are 628 bytes, 1.7% are 1300 bytes and 46% are 1500 bytes. These packets are also carried by different transmission protocols: 90% by TCP and 10% by UDP.

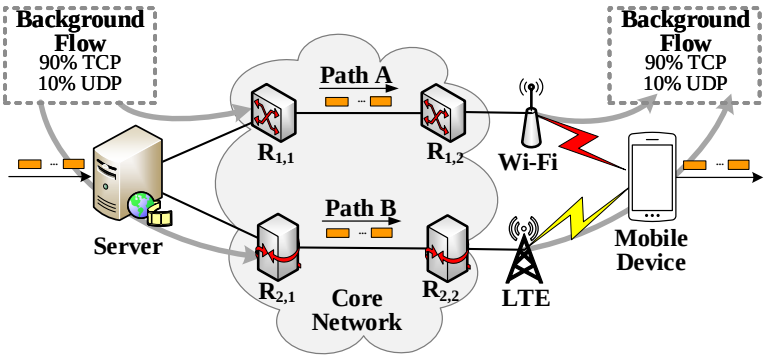


Fig. 5. Simulation topology.

Table 1. Path parameter settings.

Parameters	Path A	Path B
Wireless technology	Wi-Fi	Cellular
Access bandwidth	2-10Mbps	10Mbps
Access link delay	10-20ms	10-20ms
Core network delay	50-70ms	80-120ms
Energy factor	1μJ/bit	2μJ/bit
Packet-loss rate	0.05-0.1	0.04-0.08

In the tests, the multiplication factor ϕ and addition factor φ are chosen from a set of discrete values: $\phi \in \{0.2, 0.5, 1, 2\}$ and $\varphi \in \{0, \pm 1, \pm 2\}$. There are 20 available actions in total. The clustering center c is set as 4. The learning rate $\theta = 0.5$, discount rate $\omega = 0.2$ in Eq. (18), greedy factor $\varepsilon = 0.1$ in the ε-greedy policy.

5.2 Performance Analysis

Before the transmission, the offline Q-Table was firstly trained based on the prior knowledge. To estimate the training effect, we introduce the normalized reward, which can be calculated by: $(r_a - r_p)/r_a$. The parameters r_a and r_p are separately the actual reward and predicted reward.

Fig. 6 shows the normalized reward vs. training time. From Fig. 6, we can know that the normalized reward grows quickly in the first 15 minutes. The normalized reward reaches about 0.78 when being trained 15 minutes. Then the normalized reward is growing more slowly. When being trained 20 minutes, the normalized reward is about 0.9. IWhen being trained 35 minutes, the normalized reward is about 0.95.

Based on the prior offline Q-Table, the transmission simulation begins. During the simulation, the server sent data to the mobile device through two sub-flow concurrently. The multipath transmission lasted for 600 seconds. The throughput, delay, and energy consumption data of different solutions were recorded and analyzed.

Fig. 7 illustrates the CDF of average throughput. From the figure, we can know that the average throughput is mainly higher than the other solutions. About 46 percent of the average throughput samples for QE-MCC are higher than 10Mbps while the compared

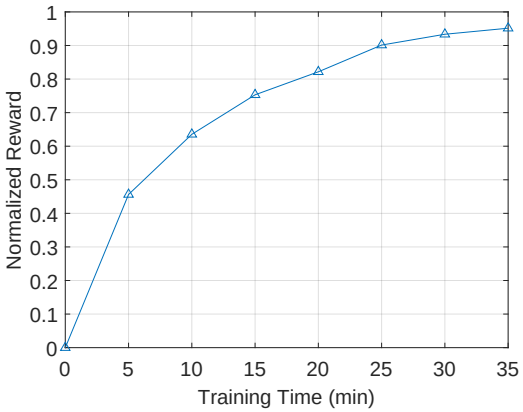


Fig. 6. The normalized reward vs. training time.

SmartCC, DEAM, and LIA solutions are separately about 31, 11, and 3 percent. This is because that QE-MCC can comprehensively formulate the stability and trend parameters of transmission system. The accurate system estimation can help the learning algorithms find the optimal congestion control actions more easily.

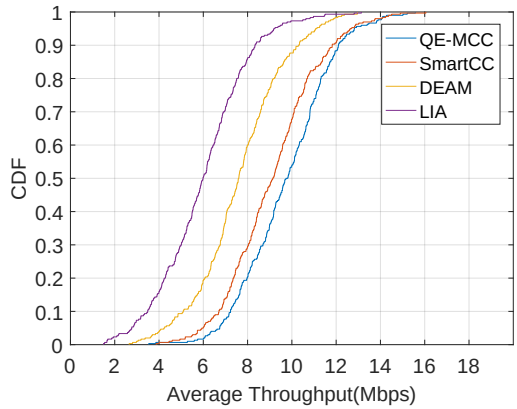


Fig. 7. The CDF of average throughput.

Fig. 8 illustrates the quantitative analysis of different sub-flows. From the figure, we can know that the total average throughput of QE-MCC is about 9.8Mbps, while compared SmartCC, DEAM, and LIA solutions are separately about 9.1Mbps, 7.7Mbps, 6.0Mbps. The total average throughput of QE-MCC is about 7.69%, 27.27%, and 63.33% higher than the compared SmartCC, DEAM, and LIA solutions. From the average throughput of subflows, we can know that QE-MC trends to move more traffic to the Wi-Fi flow. This is because Wi-Fi flow is more energy-saving than the cellular flow. The energy factor in the reward calculation make the agent prefer to send more data on Wi-Fi flow.

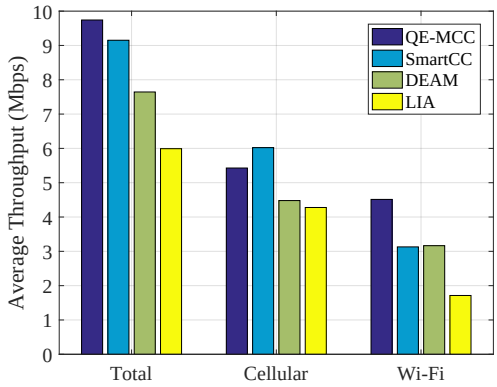


Fig. 8. The quantitative analysis of average throughput.

Fig. 9 illustrates the comparisons of CDF of average round trip time. From the figure, we can know that the average round trip time of different solutions similar. QE-MCC performs a little better than the other solutions. About 78 percent of the RTT samples are shorter than 90ms while the compared SmartCC, DEAM, and LIA solutions are separately about 65, 55, and 50 percent. This is because that QE-MCC trends to move more traffic from the cellular to the Wi-Fi sub-flow. The Wi-Fi sub-flow has lower core network delay which can decrease the average RTT. But Wi-Fi sub-flow has packet-loss rate which will increase the average RTT. Overall, the final average RTT of different solutions are not that different.

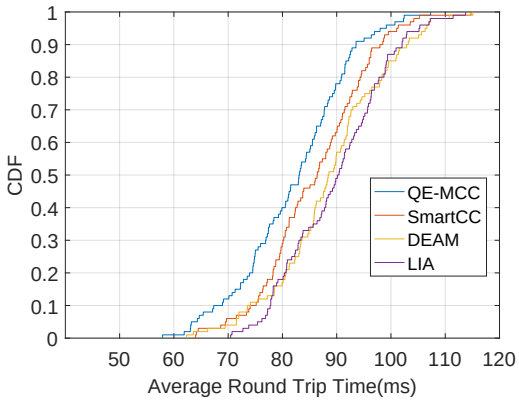


Fig. 9. The CDF of average RTT.

Fig. 10 illustrates the comparison of energy consumption. We transmitted a 500MB file over the proposed network topology. From the figure, we can know that the energy consumption of QE-MCC is lower than the other solutions. After sending all the data, QE-MCC consumed about 770J while the compared SmartCC, DEAM, and LIA solutions are separately about 830J, 790J, and 860J. QE-MCC can save about 7.21%, 2.53%,

and 10.46% energy than SmartCC, DEAM, and LIA. This is because that QE-MCC take the energy consumption in consideration when calculating the reward. The learning algorithms can find some more energy-saving congestion control strategies.

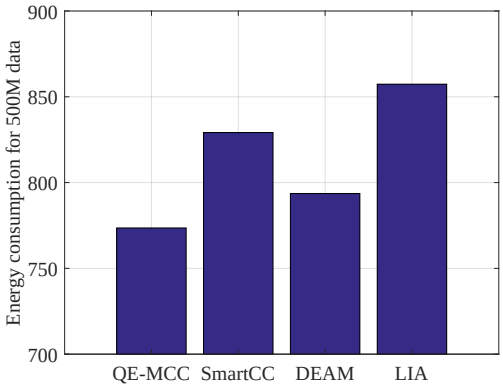


Fig. 10. The comparison of energy consumption.

6. CONCLUSIONS AND FUTURE WORKS

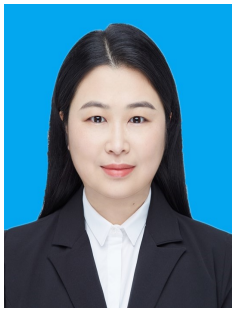
This paper proposes a Q-Learning based and energy-aware multipath congestion control scheme for MPTCP in mobile wireless networks. To address the problem insufficient state awareness, the stability and trend parameters are introduced to give a comprehensive formulation to the multipath transmission system. Then a Fuzzy C-Means based clustering algorithm is proposed to simplify the high-dimensional state samples. After that, an Energy-aware Transmission Utility Model is designed to quantify congestion control effect. Finally, Q-learning based two-layer multipath congestion control algorithm is presented. The simulation results show that QE-MCC outperforms than other standard and similar solutions in throughput, delay, and energy-saving. In the future, we will try introduce more intelligent algorithms to multipath congestion control.

REFERENCES

1. X. Ge, L. Pan, Q. Li, G. Mao, and S. Tu, "Multipath cooperative communications networks for augmented and virtual reality transmission," *IEEE Transactions on Multimedia*, Vol. 19, 2017, pp. 2345-2358.
2. Y. Thomas, M. Karaliopoulos, G. Xylomenos, and G. C. Polyzos, "Low latency friendliness for multipath TCP," *IEEE/ACM Transactions on Networking*, Vol. 28, 2020, pp. 248-261.
3. P. Hurtig, K. Grinnemo, A. Brunstrom, S. Ferlin, O. Alay, and N. Kuhn, "Low-latency scheduling in MPTCP," *IEEE/ACM Transactions on Networking*, Vol. 27, 2019, pp. 302-315.

4. A. Ford, C. Raiciu, M. J. Handley, O. Bonaventure, and C. Paasch, "TCP extensions for multipath operation with multiple addresses," RFC 8684, https://rfc-editor.org/rfc/rfc8684.txt
5. A. Elgabli and V. Aggarwal, "Smartstreamer: Preference-aware multipath video streaming over MPTCP," *IEEE Transactions on Vehicular Technology*, Vol. 68, 2019, pp. 6975-6984.
6. S. Ferlin, S. Kucera, H. Claussen, and O. Alay, "Mptcp meets fec: Supporting latency-sensitive applications over heterogeneous networks," *IEEE/ACM Transactions on Networking*, Vol. 26, 2018, pp. 2005-2018.
7. J. Xu, B. Ai, L. Chen, L. Pei, Y. Li, and Y. Y. Nazaruddin, "When high-speed railway networks meet multipath tcp: Supporting dependable communications," *IEEE Wireless Communications Letters*, Vol. 9, 2020, pp. 202-205.
8. S. R. Pokhrel, M. Panda, and H. L. Vu, "Fair coexistence of regular and multipath tcp over wireless last-miles," *IEEE Transactions on Mobile Computing*, Vol. 18, 2019, pp. 574-587.
9. J. Zhao, J. Liu, H. Wang, C. Xu, W. Gong, and C. Xu, "Measurement, analysis, and enhancement of multipath tcp energy efficiency for datacenters," *IEEE/ACM Transactions on Networking*, Vol. 28, 2020, pp. 57-70.
10. J. Wu, R. Tan, and M. Wang, "Energy-efficient multipath tcp for quality-guaranteed video over heterogeneous wireless networks," *IEEE Transactions on Multimedia*, Vol. 21, 2019, pp. 1593-1608.
11. W. Li, H. Zhang, S. Gao, C. Xue, X. Wang, and S. Lu, "Smartcc: A reinforcement learning approach for multipath tcp congestion control in heterogeneous networks," *IEEE Journal on Selected Areas in Communications*, Vol. 37, 2019, pp. 2621-2633.
12. C. Xu, T. Liu, J. Guan, H. Zhang, and G. Muntean, "Cmt-qa: Quality-aware adaptive concurrent multipath data transfer in heterogeneous wireless networks," *IEEE Transactions on Mobile Computing*, Vol. 12, 2013, pp. 2193-2205.
13. C. Xu, P. Wang, C. Xiong, X. Wei, and G. Muntean, "Pipeline network coding-based multipath data transfer in heterogeneous wireless networks," *IEEE Transactions on Broadcasting*, Vol. 63, 2017, pp. 376-390.
14. C. Xu, Z. Li, L. Zhong, H. Zhang, and G. Muntean, "Cmt-nc: Improving the concurrent multipath transfer performance using network coding in wireless networks," *IEEE Transactions on Vehicular Technology*, Vol. 65, 2016, pp. 1735- 1751.
15. C. Xu, Z. Li, J. Li, H. Zhang, and G. Muntean, "Cross-layer fairness-driven concurrent multipath video delivery over heterogeneous wireless networks," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 25, 2015, pp. 1175-1189.
16. C. Raiciu, M. J. Handley, and D. Wischik, "Coupled congestion control for multipath transport protocols," RFC 6356, https://rfc-editor.org/rfc/rfc6356.txt.
17. C. Lee, S. Song, H. Cho, G. Lim, and J. Chung, "Optimal multipath tcp offloading over 5g nr and lte networks," *IEEE Wireless Communications Letters*, Vol. 8, 2019, pp. 293-296.
18. S. R. Pokhrel and M. Mandjes, "Improving multipath tcp performance over wifi and cellular networks: An analytical approach," *IEEE Transactions on Mobile Computing*, Vol. 18, 2019, pp. 2562-2576.

19. C. Lee, J. Jung, and J. Chung, "Deft: Multipath tcp for high speed low latency communications in 5G networks," *IEEE Transactions on Mobile Computing*, 2020, p. 1.
20. Y. Zhang, P. Dong, S. Yu, H. Luo, T. Zheng, and H. Zhang, "An adaptive multipath algorithm to overcome the unpredictability of heterogeneous wireless networks for high-speed railway," *IEEE Transactions on Vehicular Technology*, Vol. 67, 2018, pp. 11 332-11 344.
21. 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, 2018, pp. 1521-1534.
22. Z. Xu, J. Tang, C. Yin, Y. Wang, and G. Xue, "Experience-driven congestion control: When multi-path tcp meets deep reinforcement learning," *IEEE Journal on Selected Areas in Communications*, Vol. 37, 2019, pp. 1325-1336.
23. M. J. Shamani, S. Rezaei, G. Jourjon, and A. Seneviratne, "Mptcp energy enhancement paradox: A q-learning approach," in *Proceedings of the 27th Interna- tional Telecommunication Networks and Applications Conference*, 2017, pp. 1-4.
24. H. Zhang, W. Li, S. Gao, X. Wang, and B. Ye, "Reles: A neural adaptive multipath scheduler based on deep reinforcement learning," in *Proceedings of IEEE Conference on Computer Communications*, 2019, pp. 1648-1656.
25. J. Luo, X. Su, and B. Liu, "A reinforcement learning approach for multipath tcp data scheduling," in *Proceedings of IEEE 9th Annual Computing and Communication Workshop and Conference*, 2019, pp. 0276-0280.
26. T. Mai, H. Yao, Y. Jing, X. Xu, X. Wang, and Z. Ji, "Self-learning congestion control of mptcp in satellites communications," in *Proceedings of the 15th International Wireless Communications Mobile Computing Conference*, 2019, pp. 775-780.
27. K. Winstein and H. Balakrishnan, "Tcp ex machina: Computer-generated congestion control," in *Proceedings of ACM SIGCOMM 2013*, 2013, pp. 123-134.
28. N. Kashif, "Mptcp implemention in ns3," https://github.com/Kashif-Nadeem/ns-3-dev-git.



Jiuren Qin received the B.S. degree from Beijing University of Posts and Telecommunications, China, in 2014. She is currently pursuing the Ph.D. degree with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications. Her research interests include multimedia communication, multipath transmission protocol.

Kai Gao received the B.S. degree from Hebei University of Engineering, China, in 2014. He is currently pursuing the Ph.D. degree with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications. His research interests include multipath transmission protocol, software defined network and game theory.

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. Her research interests include communication networks, computer system and architecture, and mobile Internet technology.

Shujie Yang is the corresponding author of this paper. He received the Ph.D. degree from the Institute of Network Technology, Beijing University of Posts and Telecommunications, Beijing, China, in 2017, where he is currently a Lecturer with the State Key Laboratory of Networking and Switching Technology. His major research interests are in the areas of wireless communications and wireless networking.