An Adaptive CMT-SCTP Scheme: A Reinforcement Learning Approach

Deguang Wang, Mingze Wang, Tao Zhang, Shujie Yang,and Changqiao Xu
State Key Laboratory of Networking and Switching Technology,
Beijing University of Posts and Telecommunications, Beijing, 100876, China

With the continuous increase of end users and types of services, the scale of the network has shown explosive growth, which has brought tremendous pressure and challenges to network data transmission. How to achieve high-quality data transmission has become a core issue. Single-path transmission has been difficult to meet the above requirements. The concurrent multipath transfer extension for stream control transmission protocol (CMT-SCTP), which supports multipath and independent data streams, can solve this problem. However, the current transmission path assessment scheme has too large granularity to make full use of the resources of the transition zone. Most studies ignore the different requirements of different services, a single transmission strategy, and the lack of an intelligent dynamic adjustment mechanism. Therefore, we designed a QCMT(Q-learning based CMT-SCTP) scheduling method. This method considers the multi-dimensional characteristics of the path and the characteristic preferences of different services, periodically evaluates and trains the reinforcement learning model for service adaptation, and makes scheduling decisions dynamically. Experimental results show that dynamic scheduling based on path parameters and service preferences can reduce message delay and improve network throughput.

Index Terms—**Quality evaluation, QoS, SCTP, Concurrent Multipath Transfer, Q-learning.**

I. Introduction

With the rapid development of the Internet and the large-scale popularization of smartphones, network services are becoming more and more diverse. The increasing user demands and diversified network services have put forward higher and higher requirements for the network, such as higher bandwidth and robustness. The application of emerging technologies such as virtual reality(VR), three-dimensional (3D) multimedia and the Internet of Everything (LoE) has generated a large amount of network traffic [1]. The global mobile traffic will reach 4394EB per month in 2030 [2]. The above statistic shows the importance of improving network transmission capacity. However, the currently widely used single-path transmission technology cannot meet the above requirements. The concurrent multipath transfer (CMT) [3] mainly studies how to use multiple paths for data transmission between multi-link terminals, and realizes load balancing and bandwidth aggregation. The stream control transmission protocol (SCTP) can be regarded as the solution for CMT because of its multi-homing feature. Therefore, CMT is a promising solution to the high bandwidth and robustness challenges brought by emerging network services.

Path quality assessment is the basis of high-quality data transmission and is of great significance. Most of the existing networks use a single-path transmission technology, and the path quality assessment methods are all aimed at single-path networks and relatively simple. Existing path quality evaluation methods [4], [5] have fewer parameters and larger evaluation granularity, and are simply divided into good and bad. The above methods cannot accurately evaluate the path quality, so that the transition interval paths cannot be fully utilized. Existing researches focus on congestion control and scheduling strategies under multipath transmission, but ignore that different services have different requirements. The sender adopts a fixed transmission scheduling strategy and lacks a dynamic and intelligent scheduling mechanism, which is an important cause of congestion. Reinforcement learning (RL) is an advanced learning-by-exploration approach that can be used to solve scheduling problems. The Q-learning method is chosen in our model to generate policies and is lightweight. Moreover, RL can also handle service adaptation in complex network environments.

In this paper, based on the concurrent multipath transfer extension for stream control transmission protocol (CMT-SCTP) [6], [7], [8], [9], we proposed a QCMT (Q-learning based CMT-SCTP) scheduling method. This method establishes the path quality evaluation model of the existing multipath parallel transmission according to the delay, jitter, PLR and cwnd(congestion window). Moreover, we consider that different types of services [10], [11], [12] have different preferences for transmission paths, and use RL to generate path preference values for services. The path preference value is combined with path quality parameters to formulate transmission strategies, select suitable transmission paths for different services, use transmission paths in a reasonable and balanced manner, and improve network service quality. The main contributions of our work are as follows:

- Through the establishment of a comprehensive and accurate transmission path quality evaluation model for the dynamic multi-dimensional characteristic parameters of the transmission path, the reliability value is calculated, and more packets are allocated to the path with high reliability to ensure efficient and reliable completion of the service.
- Based on the path reliability evaluation, combined with

Manuscript received December 9, 2021; revised January 23, 2022. Corresponding author: Deguang Wang, Tao Zhang (email: wdg19@bupt.edu.cn, zhangtao17@bupt.edu.cn)

the service preference value, the transmission path is selected and integrated scheduling is performed. In this way, intelligent dynamic scheduling between services and paths is realized, transmission efficiency and quality are improved.

- The experimental results on the OMNeT++ platform show that our method has good adaptability to services, and can adjust the distribution ratio of packets on different paths as the heterogeneity between paths increases. The delay performance of the message is also superior to the existing SCTP and standard CMT methods.

This paper is organized as follows. In Section II, the related work are summarized and briefly reviewed. The basic knowledge of stream control transmission protocol (SCTP) is presented in section III. The model description and construction method and algorithm improvement are given in Section IV. The experimental simulation is carried out, and our method is compared with the standard CMT method in Section V. Finally, we summarize this paper and look forward to the future in Section VI.

II. Related work

Accurate assessment of path quality is the basis of multipath transmission scheduling strategy. Some scholars conduct research from the perspective of path quality assessment model. Jayasri *et al.* [13] proposesd an evaluation model with fewer evaluation parameters, simply dividing the paths into available and unavailable. The granularity of division is too large, which makes it impossible to fully utilize the path of the transition interval. Ansar *et al.* [14] proposed an adaptive burst protocol with two states of path quality, good or bad, with poor scalability and large granularity of evaluation results. Other scholars study the path quality prediction model. Isyaku *et al.* [15] used random forest method to establish quality prediction models, which can better evaluate the quality, but the model has high complexity, long training time and large granularity. Bote-Lorenzo *et al.* [16] proposed a vector functional link neural network (RVFL) model, the prediction effect is good when the network environment fluctuation is relatively small, but when the path parameter fluctuation is large, the prediction result is poor and the complexity of the model is high.

There are also a large number of scholars studying transmission strategies. Verma *et al.* [17] proposed a new fast retransmission strategy based on delay, which adjusts the transmission rate of each path according to the path delay. Arianpoo *et al.* [18] introduced RL and proposed a distributed Q-learning method to improve the fairness of flows in CMT. Experiments show that it is better than the existing CMT fairness mechanism. Yu *et al.* [19] proposed a CMT scheduling strategy combined with deep RL, which improved the throughput of the network, but did not consider the different preferences of different types of services. Due to the use of neural networks, it is not convenient for large-scale deployment. Some other works [20], [21], [22] focused on the security of transmission path.

Challenges faced by existing methods are drawn from the above analysis, it can be seen that the QoS parameters are less considered in the current transmission path evaluation method. Moreover, the granularity of the evaluation result is too large, the paths in the transition interval are often ignored and cannot be used. Some models have high complexity and poor scalability, which is not conducive to large-scale deployment. Most researches focus on congestion control under multipath parallel transmission, without considering the different requirements of different services, the transmission strategy is fixed, and lacks an intelligent dynamic adjustment mechanism. The sender's scheduling strategy is an important cause of congestion. The dynamically adjusted sending strategy can reduce the occurrence of disorder and congestion.

TABLE I: Related Work Comparison

Literature	Granularity levels	QoS	Service adaptation
[13],[14]	large	✓	✗
[15],[16]	small	✓	✗
[17]	large	✗	✗
[18]	small	✓	✗
[19]	small	✓	✗
Proposed	small	✓	✓

This paper proposes a path quality evaluation model based on delay, jitter, PLR and cwnd, establishes the transmission path quality evaluation model, takes into account different types of services, and uses RL to generate service preference values in CMT. Combined with path quality parameters and service preference values, transmission strategies are formulated to select appropriate transmission paths for different services, so as to achieve the goal of using transmission paths in a balanced manner and service adaptation. The comparison between related work and the method of this paper is shown in Table I.

III. Preliminaries

This section provides a brief overview of the CMT-SCTP and RL to provide some preliminaries for our proposed model.

A. CMT-SCTP

SCTP [23], [24], [25] is a transport layer protocol formulated by the Sigtran group of IETF in October 2000. RFC 4960 defines SCTP in detail. SCTP was originally used to transmit telephone protocol (SS7) signaling messages over IP. SCTP protocol effectively combines the main advantages of the other two mainstream protocols TCP and UDP in the transport layer, and has additional support for two important protocols, namely, multi-homing and multi-streaming. It can be seen that SCTP avoids DoS attack. At the same time, the characteristics of multi-homing and multi-streaming also determine that SCTP will adopt a different transmission mode from traditional TCP and UDP and adopt multiple paths for transmission.

CMT-SCTP was proposed as an IETF draft in [26] and is being implemented in the FreeBSD SCTP and INET framework. The OMNeT++ simulator [27] includes the INET framework, the simulator is used in this article. In the current draft standard, the scheduling strategy of CMT-SCTP among multiple available paths has not been defined. CMT-SCTP in the OMNeT++ simulator schedules the available path in a round-robin fashion.

Multi-path transmission technology has many advantages, such as making full use of network path resources, improving data security and reliability, and improving network throughput. Every coin has two sides, and there are some problems with multi-path transmission that need to be solved. In multi-path parallel transmission, each path has its own independent sender cache, but the receiver cache is shared by all paths. Due to the difference between different paths, the difference will cause the blocking of the receiver's cache, which will affect the improvement of throughput in end-to-end transmission. In order to alleviate the blocking of receiving cache, it is necessary to establish an accurate quality assessment model for the path, so as to make better use of the benefits of multi-path transmission to improve the end-to-end transmission throughput.

B. Reinforcement learning

Recently, some scholars have introduced RL into the network. RL [28], [29], [30], [31] is a method to achieve a certain goal by letting the computer try constantly, get feedback from errors, and solidify the feedback returned by the incentive function into experience. It will consider the learning scenario as a Markov decision-making process [32], which is simply a circular process. The agent takes action based on the state, gets rewards, and interacts with the environment to update the strategy. It can be expressed as follows:

$$M =< S, A, P_{s,a}, R, \gamma >, \tag{1}$$

where S represents a collection of states. A represents a collection of actions. P describes the state transition matrix, $P^a_{ss'}$=$P[S_{t+1} = s'|S_t = s, A_t = a]$. R represents the reward function. γ represents the attenuation factor, $\gamma \in$ [0,1].

The optimization strategy is implemented by approximately solving the Bellman optimality equation.

$$v_*(s) = max(R^a_s + \gamma \sum_{s' \in S} P^a_{ss'} v_*(s')), \tag{2}$$

$$Q_*(s,a) = R^a_s + \gamma \sum_{s' \in S} P^a_{ss'} Q_*(s', a'), \tag{3}$$

where v describes the long-term optimization value of a state, that is, the value of this state when all possible subsequent actions are considered and the optimal actions are selected for execution. Q describes the long-term optimal value brought by being in a state and executing an action, that is, after performing a specific action in this state, consider all possible future states, and always select the optimal action to execute the long-term value brought by these states.

IV. Methodologies

In this section, we focus on the model description and construction method and algorithm improvement. We established a multi-dimensional transmission path quality evaluation model, and combined with the characteristic preferences of different services, we proposed a QCMT (Q-learning based CMT-SCTP) transmission method. The overall process of the model is shown in the Figure 1.

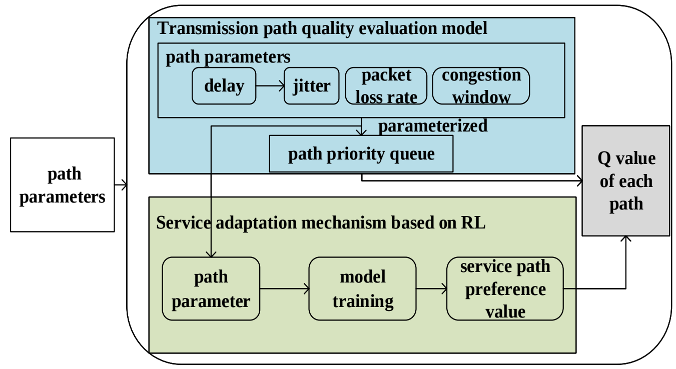


Fig. 1: The overall process of the model.

The process of the whole method is shown in Figure 1. First, the path quality evaluation parameters are calculated by the path transmission feedback round-trip time (RTT), jitter, PLR and cwnd parameters to evaluate the transmission quality L of the path. Then combine RL to perform adaptation training for different types of services to obtain the service preference value P of each path, and multiply L by P to obtain Q, and distribute the traffic according to the Q value.

A. Transmission path quality evaluation model

In the network resource scheduling, the reliability value of the transmission path [6] is calculated by the dynamic characteristic parameters of the transmission path, and then the path with high reliability is allocated more service data to ensure the efficient and reliable completion of the service.

In the selection of path quality evaluation parameters, round trip time (RTT) is usually one of the most important parameters in path quality evaluation. Round trip delay describes the speed at which each transmitted data is successfully received, including the transmission time of transmitted data, the data processing time at the receiving end and the return time of ACK confirmation message at the receiving end; Jitter represents the stability of the transmission path; Packet loss ratio (P_l) describes the success rate of data transmission; cwnd describes whether congestion occurs in the current network environment state and the transmission quality characteristics of the current path. Therefore, round-trip delay (RTT), jitter(J), PLR (P_l) and congestion window ($cwnd$) are selected As the parameter variable for calculating the path quality evaluation parameters, the equation established by using the above relationship is as follows:

$$L = \frac{\overline{cwnd} \times J_{max}}{\overline{RTT} \times J}(e^(1 - \sqrt{P_l}) - 1), \tag{4}$$

where $\overline{cwnd}$,$\overline{RTT}$,J and P_l respectively represent the current path cwnd, the mean value of round-trip delay, path transmission jitter and PLR respectively, J_{max} is the maximum value of jitter in all paths in this evaluation. The reason why the average value is used to calculate the quality evaluation parameters of the path is that if the quality evaluation is performed before each transmission, it will seriously affect the efficiency. Due to the uncertainty of network environment change, the smaller the jitter value, the more stable the path is and the better the quality is. When the PLR is 1, the path quality evaluation parameter is 0, indicating that the path is completely unavailable and should be discarded. Selecting an appropriate cycle for quality evaluation can achieve a relative balance between network transmission performance and transmission efficiency. The evaluation cycle will be described later.

The average round trip delay ($\overline{RTT}$) is obtained as follows:

At time t_1, the sender sends probe packet to the receiver, and the sender records the sending time as st_1. at time t_2, the sender receives the confirmation information from the receiver, and records the current time as st_2 to calculate the avarage round-trip delay:

$$RTT = st_1 - st_2. \tag{5}$$

According to equation 2, calculate the average size in an evaluation cycle.

$$\overline{RTT} = \frac{\sum_{i=1}^{n} RTT_i}{n}. \tag{6}$$

The calculation method of jitter is as follows:

$$J = \sqrt{\frac{\sum_{i=1}^{n}(RTT_i - \overline{RTT})^2}{n-1}} / \overline{RTT}. \tag{7}$$

According to equation 4, the jitter is calculated in one evaluation cycle.

The method to obtain PLR is as follows:

$$P_l = \frac{P_s - P_r}{P_s}, \tag{8}$$

where P_s means the number of data packets sent by the sender and P_r is the number of data packets received by the receiver. The number of data packets not successfully sent in an evaluation cycle is used as the PLR of the current path.

The average $\overline{cwnd}$ is obtained as follows:

$$\overline{cwnd} = \frac{\sum_{i=1}^{n} cwnd_i}{n}. \tag{9}$$

Cwnd is an important control variable in congestion control. As a regulating factor for SCTP to avoid network congestion, it can dynamically reflect the quality of the current path. Congestion control window controls the transmission volume of path data. When congestion packet loss occurs to the path, the value of cwnd is halved. The purpose of this is not only to alleviate the current network congestion, At the same time, it can indicate the transmission quality of the current path; At the same time, when the window is too small and the path has random packet loss, the quality evaluation value of the path can be affected by the PLR. Therefore, the cwnd size is an important parameter in the path quality evaluation parameters.

The calculation method of cwnd size of cwnd is that after the sender receives the confirmation information from the receiver or sends congestion packet loss, the sender will update the current cwnd value, and also calculate the average value of path cwnd in an evaluation cycle.

Path quality evaluation takes path return time, jitter, PLR and cwnd as path evaluation parameters under the condition of certain bandwidth, and then reasonably selects the evaluation cycle.

For the selection of evaluation cycle, if the interval of evaluation cycle is too small (for example, less than RTT), the path quality cannot be accurately reflected; If the selection is too large, it can not reflect the dynamic change of path quality. In order to better select the evaluation cycle, the evaluation cycle interval is determined by using the method of confidence interval [33] estimation. The confidence interval means the estimation interval of the overall parameters constructed by sample statistics. In statistics, the confidence interval of a probability sample is the interval estimation of a population parameter of this sample. The confidence interval shows the extent that the real value of a parameter has a certain probability of falling around the measurement results. It gives the degree of confidence of the measured value of the measured parameter, that is, the required "certain probability", which becomes the confidence level. Confidence limits are the two ends of the confidence interval. For the estimation of a specific case, the higher the confidence level, the larger the corresponding confidence interval.

The sample acquisition method of confidence interval is as follows:

(a) When the sender sends the first packet, the current record time is the start time.

(b) In case of packet loss, record the current time as the end time, immediately retransmit the data packet and enter the next step.

(c) At this time, the sample of the evaluation cycle is the difference between the end time and the start time, and it is recorded as a sample value. Records are cleared and sample data is retrieved.

After obtaining the sample according to the above method, the calculation method is as follows:

Obtained multiple evaluation cycle samples $\{x_1, x_2, ..., x_n\}$. Calculate the sample mean value of the evaluation period through equation 10:

$$\overline{X_n} = \frac{\sum_{i=1}^{n} x_i}{n}, \tag{10}$$

where x_i is the sample value of the successful data transmission time, n is the number of samples taken in the cycle, and $\overline{X_n}$ is the average value of the obtained samples.

In order to save storage space and calculation efficiency, the average value can be calculated by the iterative method of equation 11, which can avoid storing all sample values:

$$\overline{X_{n+1}} = \frac{\overline{X_n} \times n + x_{n+1}}{n+1}, \tag{11}$$

where $\overline{X_{n+1}}$ is the sample mean calculated by iterative method based on the historical record, and $\overline{X_n}$ is the sample mean before the current transmission data. x_{n+1} is the sample value of the current transmission data, n is the number of samples in the previous time, and the sample standard deviation is calculated according to equation 12:

$$S_n = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{X_n})^2}{n-1}}, \tag{12}$$

where S_n is the standard deviation of the sample, $\overline{X_n}$ is the sample mean, x_i is the sample value, and n is the number of samples.

Similarly, in order to improve efficiency and reduce the amount of calculation, use the iterative method of equation 13 to obtain the standard deviation:

$$S_{n+1} = \sqrt{\frac{S_n^2 \times (n-1)}{n} + \frac{(x_{n+1} - \overline{X_n})^2}{n+1}}, \tag{13}$$

where S_{n+1} is the sample standard deviation calculated by iterative method based on historical records, S_n is the sample standard deviation obtained in the previous time, x_{n+1} is the current sample value, $\overline{X_n}$ is the sample mean value of the previous time, and n is the number of samples in the previous time. According to the central limit theorem [34], the quality evaluation period c is calculated by using the confidence interval method from the calculated sample mean and sample standard deviation:

$$P\{\overline{X} - Z_{1-\alpha/2} \times \frac{S}{\sqrt{n}} < c < \overline{X} + Z_{1-\alpha/2} \times \frac{S}{\sqrt{n}}\} = 1-\alpha, \tag{14}$$

where S is the standard deviation of samples, $\overline{X}$ is the average of the samples, n is the number of samples, α is the significant level,and $Z_{1-\alpha/2}$ can be obtained by looking up the table of confidence level. Some confidence level parameters are given in the following table II:

TABLE II: Confidence level and corresponding α and Z values

confidence level	$\alpha/2$	$Z_{1-\alpha/2}$
80%	0.1	1.282
90%	0.05	1.645
95%	0.025	1.96
98%	0.01	2.326
99%	0.005	2.576

The path quality evaluation parameters are calculated through the information round-trip time (RTT), jitter, PLR and cwnd of path transmission feedback to evaluate the transmission quality of the path, and the evaluation cycle is calculated by using the method of confidence interval.

In the transmission path quality evaluation model, first collect the delay characteristics, jitter, PLR and cwnd of the transmission path, parameterize the path into the model, and get a path priority queue according to the quality evaluation results of each path.

B. Service adaptation mechanism based on RL

After the transmission path quality evaluation, the path priority queue is obtained, but different services have different requirements for the transmission path in terms of delay, jitter and PLR. RL model is used to adapt flexibly to different services, generate path preference values, and select paths that are more suitable for the type of service in combination with path quality. The variables in Equation 1 are defined in this method as follows:

Service type is defined as state, there are n types of services, then $S = \{s_1, s_2, ..., s_n\}$.

Selecting path is defined as action, there are m paths, then $A = \{a_1, a_2, ..., a_m\}$.

P describes the state transition matrix, $P_{ss'}^{a}$ $=P[S_{t+1} = s^{'}|S_t = s, A_t = a]$.

R represents the reward function, $R(s, a)$ describes the reward for doing action a in state s, $R(s, a)$ $=E[R_{t+1}|S_t = s, A_t = a]$. The delay, PLR, and cwnd of the i-th path are represented by d_i, p_i, and c_i respectively.

$$R = X \times R_d + Y \times R_p + Z \times R_c \tag{15}$$

X, Y, Z are state indicator functions,

$$X = \begin{cases} 1, & \text{state } s \text{ is a delay-sensitive service,} & (16a) \\ 0, & \text{not,} & (16b) \end{cases}$$

$$Y = \begin{cases} 1, & \text{state } s \text{ is a plr-sensitive service,} & (17a) \\ 0, & \text{not,} & (17b) \end{cases}$$

$$Z = \begin{cases} 1, & \text{state } s \text{ is a cwnd-sensitive service,} & (18a) \\ 0, & \text{not,} & (18b) \end{cases}$$

The reward function of the service sensitive to delay R_d is set as follows:

$$R_d = \begin{cases} \frac{\sum_{i=1}^{m} d_i}{d_a}, & d_a \leq min\{d_1, ..., d_m\}, & (19a) \\ -1, & d_a > min\{d_1, ..., d_m\}, & (19b) \end{cases}$$

The reward function of the service sensitive to PLR R_p is set as follows:

$$R_p = \begin{cases} \frac{\sum_{i=1}^{m} p_i}{p_a}, & p_a \leq min\{p_1, ..., p_m\}, & (20a) \\ -1, & p_a > min\{p_1, ..., p_m\}, & (20b) \end{cases}$$

The reward function of the service sensitive to cwnd R_c is set as follows:

$$R_c = \begin{cases} \frac{c_a}{\sum_{i=1}^{m} c_i}, & c_a \geq max\{c_1, ..., c_m\}, & (21a) \\ -1, & c_a < max\{c_1, ..., c_m\}, & (21b) \end{cases}$$

Considering the efficiency and complexity, we choose the Q-learning [35] algorithm in RL. In the model, the service type is the state s in RL, the action selects the path according to the current state (service), and the reward function is set to judge whether the action (path selection) is the best and return the corresponding reward according to the service characteristics. The pseudo code of the model is as follows:

The Q-learning method is convergent, and QCMT is based on Q-learning, so it can be proved to be convergent mathematically. The detailed convergence proof of Q-learning can be found in literature [36]. Through the RL model, for each different service, a path preference value $\{p_1, p_2, ..., p_n\}$ will be generated for the existing n paths, and then the Q value of

Algorithm 1 QCMT

/** **Path quality evaluation** **/
Periodically obtain the delay, calculate jitter from delay,PLR and cwnd of each path;
Calculate the average of RTT, cwnd;
Save the above data for the next iteration calculation;
Get the path quality L according to Equation 4;
/** **Service adaptation mechanism based on RL** **/
Q-learning model parameters: step size $\alpha \in$ (0,1],small ϵ >0
Initialize policy $Q(s, a)$, for all states and actions $s \in S$,$a \in A(s)$,arbitraily except the $Q(terminal, \cdot) = 0$
Loop **for** each episode:
    Initialize S
    Loop **for** step of episode:
        Choose action(chosing path) a based on state s using policy derived from Q(e.g.,ϵ-greedy)
        Take action a, obtain R, $s^{'}$
        $Q(s, a) \leftarrow Q(s, a)$+$\alpha$[R+ $\gamma\ max_a Q(s^{'}, A)$-$Q(s, a)$]
        $s \leftarrow s^{'}$
    until s is terminal;
    end for
end for
Get and use the latest Q-table;

service adaptation will be obtained by combining the L value of each path.

$$Q_i = P_i \times L_i, (i = 1, 2, ..., n). \tag{22}$$

Based on the path quality assessment parameter L, multiplied by the service preference value P to get the Q value, each service gets the Q value of each path that is adapted to the service, and each service distributes packets according to the Q value of each path.

V. Performance evaluation

In order to evaluate the service adaptation and message latency on multiple paths, we carried out a series of simulation experiments to test the model performance. We used the OMNeT++ simulator [27], [37] deployed with the INET framework. The simulator also contains a fully verified SCTP protocol model that supports CMT.

This paper divides services into three categories, sensitive to delay, PLR, and cwnd, respectively. At the same time, since the latency of a single message or message flow is an important performance indicator of traffic, the delay between the sender and the receiver should be as low as possible. Therefore, we conducted four sets of comparative experiments to test our proposed method, which are the adaptation rate of the above three types of services and the message delay performance of data transmission.

Figure 2 shows the simulated network connected via a dual-homing network, which contains a sender and a receiver.

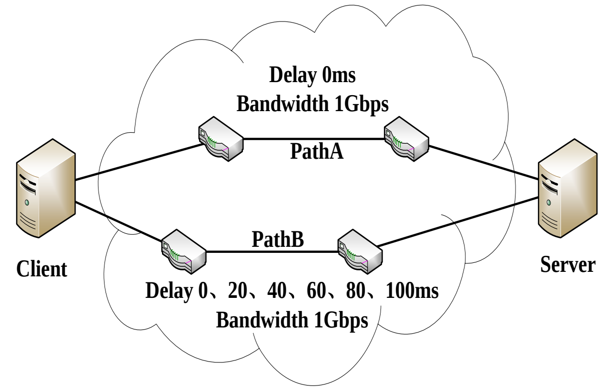


Fig. 2: Network scenario.

These two paths pass through two routers, and each router has a bandwidth bottleneck. This is a simply equipped network topology, but the sender node regards the complex network just as a single network path with specific characteristics, it can be used to evaluate the sender's scheduling. We send data in one direction on a network without competing traffic, and use homogeneous and heterogeneous network paths to evaluate scheduling on lossless paths. This scenario is designed to simulate how SCTP is currently used for signaling traffic in a private network. In a private network, the traffic is controlled and data loss is minimal. The max capacity of the bottleneck link is 1Gbps. Although the bandwidth of the transmission path is very large, the available capacity may be much less due to the size of the current cwnd. This is also in line with the fact [38] that in actual network scenarios, the cwnd can better measure the maximum throughput of the path than the path bandwidth. Services that are sensitive to the cwnd can be considered bandwidth sensitive services. In all experiments, the path inherent delay of path A in Figure 2 is set to 0 milliseconds, while in different experiments, the inherent delay of path B ranges from 0 milliseconds to 100 milliseconds. The PLR of Path A and Path B are both 0 by default. In addition, in order to simulate the real network environment more realistically, SACK delay and Nagles algorithm were started in all experiments. At the same time, each path will have additional delays such as processing, queuing and sack in the transmission process.

First, we designed three sets of experiments to verify the service adaptation performance of QCMT through the proportion of data packets sent on path B of three different types of services that are sensitive to delay, PLR, and cwnd.

Experiment 1, delay sensitive service adaptation capability experiment, the parameters are the same as Figure 2, and the inherent delay range of path B is 0 milliseconds to 100 milliseconds. Figure 3 shows the experimental results.

Figure 3 shows the adaptive performance of different methods of delay-sensitive services. In the experiment, the inherent delay of path A is always 0, and the delay of path B continues to increase. SCTP selects path B to send data packets by default, and does not have service adaptation at all. The standard CMT under OMNeT++ adopts the round-robin scheduling strategy, so the allocation ratio of packets on path

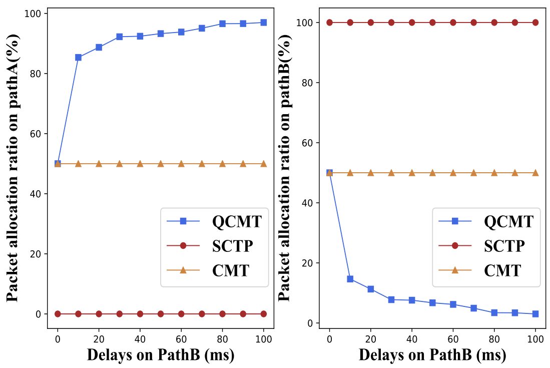


Fig. 3: Packets distribution ratio diagram of delay sensitive service on path A and B.

A and path B is always 50%. In the QCMT method, for delay-sensitive services, paths with large delays are allocated fewer data packets than paths with small delays, and an appropriate distribution ratio of data packets is obtained through training. As can be seen from Figure 3, as the delay of path B increases, the proportion of data packets allocated to path B decreases.

Experiment 2, PLR sensitive service adaptation capability experiment, the parameters are the same as Figure 2, and the PLR of path B varies from 0‰ to 10‰.

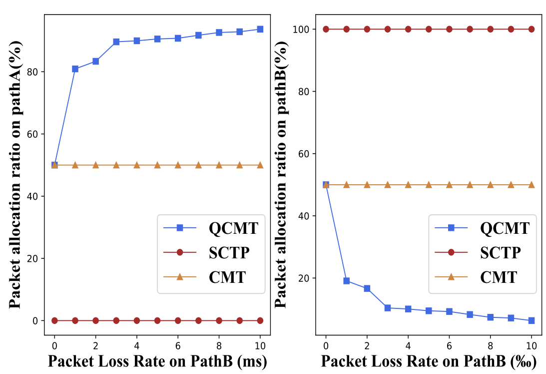


Fig. 4: Packets distribution ratio diagram of PLR sensitive service on path A and B.

Figure 4 shows the adaptive performance of different methods for PLR-sensitive services. In the experiment, the PLR of path A is always 0, and the PLR of path B gradually increases. SCTP uses a single path to send data packets, and selects path B to send data packets by default, regardless of the increasing PLR of path B. The CMT under OMNeT++ adopts a round-robin scheduling strategy and does not perceive changes in path quality. Therefore, the allocation ratio of packets on path A and path B is always 50%. In the QCMT method, for PLR-sensitive services, paths with small PLRs are allocated more data packets than paths with large PLRs, and a reasonable distribution ratio of data packets is obtained through training, providing more reliable data transmission. As can be seen from Figure 4, as the PLR of path B increases, the proportion of data packets allocated to path B is decreasing, and the proportion of data packets on path A is gradually increasing.

Experiment 3, bandwidth sensitive service adaptation capability experiment, the parameters are the same as Figure 2, the cwnd of path A remains unchanged, while the size of path B is getting smaller and smaller. The difference in the size of the cwnd between path B and path A is from 10% to 99%.

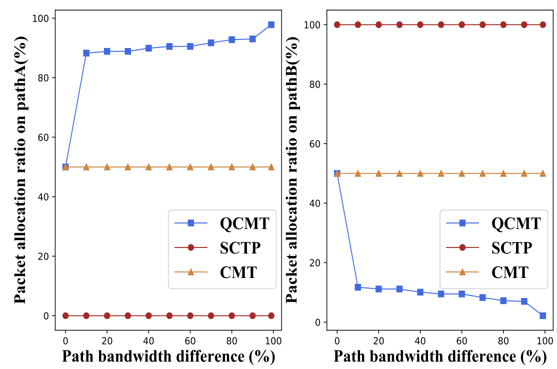


Fig. 5: Packets distribution ratio diagram of bandwidth sensitive service on path A and B.

It can be seen from Figure 5 that the adaptability performance of different methods to cwnd-sensitive services. In the experiment, the bandwidth difference between path A and path B gradually increases. SCTP only uses a single path and uses path B to send packets by default. The CMT adopts the round-robin scheduling strategy and does not consider changes in path parameters. Therefore, the allocation ratio of packets on path A and path B is always equal to half. In the QCMT method, for cwnd-sensitive services, the path with larger bandwidth after training has a larger proportion of data packets and a reasonable allocation ratio of data packets is obtained. As can be seen from Figure 5, as the bandwidth ratio of path B decreases compared to path A, the proportion of data packets allocated to path B decreases, and the proportion of data packets on path A gradually increases, there is always a dynamic and appropriate distribution ratio.

The latency of the message is an important aspect to evaluate the network transmission performance. Therefore, in order to compare the message latency performance of different transmission methods in different degrees of homogeneous and heterogeneous networks. We performed simulation experiments using the topology diagram in Figure 2. We extend the experiment to different network configurations for delay. In this experiment, the delay of path A has been kept at 0ms, while the delay of path B is variable, and the inherent delay ranges from 20ms to 100ms. The sender periodically sends different types of service packets and calculates the delay, and takes the average value to compare the performance of SCTP, standard CMT and QCMT. The experimental results are shown in Figure 6.

The performance of different methods on message latency can be seen from Figure 6. In the experiment, the delay of path B is gradually increasing, and the heterogeneity of path A and path B is gradually increasing. Since SCTP only uses a single path and selects path B by default, the message latency has always been large. CMT adopts the round-robin scheduling

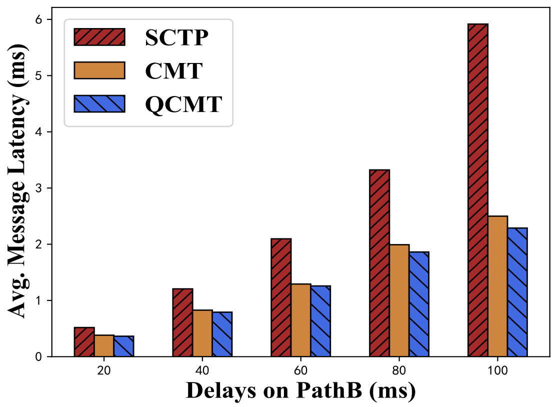


Fig. 6: Avarage message latencis for different pathB delays.

strategy, but does not take into account the change of path parameters, so the distribution ratio of data packets on path A and path B is always equal to half, but because two paths are used at the same time, the message latency is compared with that of sctp much lower. In the QCMT method, after training, different types of services get a reasonable packet distribution ratio according to the path parameters. In the experiments, different types of services all get a good packet allocation ratio. Overall, with the increase of path B delay, the message latency of QCMT method is always better than that of sctp and standard cmt methods, and with the increase of path A and path B heterogeneity, the advantage is more obvious.

The experimental results in this section confirm that when traffic is transmitted to varying degrees on heterogeneous network paths, the QCMT proposed in this paper can adapt to services compared to SCTP and standard CMT. During the training process of QCMT, it will choose a suitable path according to the preferences of different types of services. When training completed, poorer quality paths are assigned less packets. And as the quality difference between different paths increases, the allocation ratio is also adjusted accordingly. At the same time, the most important message delay indicator in the transmission algorithm, QCMT through intelligent adjustment, effectively reduces the message delay, has always been better than sctp and standard CMT, and with the increasing heterogeneity of the path, the advantages become more obvious.

VI. Conclusion and Future Work

In this paper, we proposed a service-oriented adaptive multipath parallel transmission method, QCMT (Q-learning based CMT-SCTP). This method considers the multi-dimensional characteristics of the path and characteristic preferences of different services, periodically carries out the path quality assessment and the service-oriented RL model, and makes the scheduling decision dynamically. The experimental results show that the algorithm can perform path selection well according to services preference, and intelligent dynamically adjust according to the path. At the same time, because QCMT has good service adaptation performance for services, it also has a good performance on message delay, which is better than SCTP and standard CMT. As the path heterogeneity increases, the effect is more significant. In the future, in addition to considering path quality and service preference, scheduling policies can also be generated by combining the priorities of different services.

References

[1] Shahid Mumtaz, Josep Miquel Jornet, Jocelyn Aulin, Wolfgang H Gerstacker, Xiaodai Dong, and Bo Ai. Terahertz communication for vehicular networks. *IEEE Transactions on Vehicular Technology*, 66(7), 2017.

[2] Anastasia Yastrebova, Ruslan Kirichek, Yevgeni Koucheryavy, Aleksey Borodin, and Andrey Koucheryavy. Future networks 2030: Architecture & requirements. In *2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, pages 1–8. IEEE, 2018.

[3] Ping Dong, Bin Song, Hongke Zhang, and Xiaojiang Du. Improving onboard internet services for high-speed vehicles by multipath transmission in heterogeneous wireless networks. *IEEE Transactions on Vehicular Technology*, 65(12):9493–9507, 2016.

[4] Shumayla Yaqoob, Ata Ullah, Muhammad Akbar, Muhammad Imran, and Mohsen Guizani. Fog-assisted congestion avoidance scheme for internet of vehicles. In *2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC)*, pages 618–622. IEEE, 2018.

[5] Mohammad Mahdi Tajiki, Behzad Akbari, Mohammad Shojafar, Seyed Hesomodding Ghasemi, Mahdi Latifi Barazandeh, Nader Mokari, Luca Chiaraviglio, and Michael Zink. Cect: computationally efficient congestion-avoidance and traffic engineering in software-defined cloud data centers. *Cluster Computing*, 21(4):1881–1897, 2018.

[6] Changqiao Xu, Tianjiao Liu, Jianfeng Guan, Hongke Zhang, and Gabriel-Miro Muntean. Cmt-qa: Quality-aware adaptive concurrent multipath data transfer in heterogeneous wireless networks. *IEEE transactions on mobile computing*, 12(11):2193–2205, 2012.

[7] Jiyan Wu, Bo Cheng, Ming Wang, and Junliang Chen. Energy-aware concurrent multipath transfer for real-time video streaming over heterogeneous wireless networks. *IEEE Transactions on circuits and systems for video technology*, 28(8):2007–2023, 2017.

[8] Zhanpeng Li, Mangui Liang, and Hongyu Liu. A selection method of path for concurrent multipath transfer based on network layer. In *2020 International Conference on Computer Engineering and Application (ICCEA)*, pages 559–564. IEEE, 2020.

[9] Lal Pratap Verma, Neelaksh Sheel, and Chandra Shekhar Yadev. Concurrent multipath transfer using delay aware scheduling. In *Innovations in Computational Intelligence and Computer Vision*, pages 247–255. Springer, 2021.

[10] Honghao Gao, Yueshen Xu, Yuyu Yin, Weipeng Zhang, Rui Li, and Xinheng Wang. Context-aware qos prediction with neural collaborative filtering for internet-of-things services. *IEEE Internet of Things Journal*, 7(5):4532–4542, 2019.

[11] Lianyong Qi, Yi Chen, Yuan Yuan, Shucun Fu, Xuyun Zhang, and Xiaolong Xu. A qos-aware virtual machine scheduling method for energy conservation in cloud-based cyber-physical systems. *World Wide Web*, 23(2):1275–1297, 2020.

[12] Wenbo Zhang, Yue Liu, Guangjie Han, Yongxin Feng, and Yuntao Zhao. An energy efficient and qos aware routing algorithm based on data classification for industrial wireless sensor networks. *IEEE Access*, 6:46495–46504, 2018.

[13] T Jayasri and M Hemalatha. Link quality estimation for adaptive data streaming in wsn. *Wireless Personal Communications*, 94(3):1543–1562, 2017.

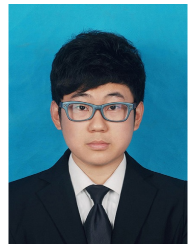
[14] Zeeshan Ansar and Waltenegus Dargie. Adaptive burst transmission scheme for wsns. In *2017 26th International Conference on Computer Communication and Networks (ICCCN)*, pages 1–7. IEEE, 2017.

[15] Babangida Isyaku, Kamalrulnizam Abu Bakar, Mohd Soperi Mohd Zahid, Eman H Alkhammash, Faisal Saeed, and Fuad A Ghaleb. Route path selection optimization scheme based link quality estimation and critical switch awareness for software defined networks. *Applied Sciences*, 11(19):9100, 2021.

[16] Miguel L Bote-Lorenzo, Eduardo Gómez-Sánchez, Carlos Mediavilla-Pastor, and Juan I Asensio-Pérez. Online machine learning algorithms to predict link quality in community wireless mesh networks. *Computer Networks*, 132:68–80, 2018.

[17] Lal Pratap Verma, Varun Kumar Sharma, and Mahesh Kumar. New delay-based fast retransmission policy for cmt-sctp. *International Journal of Intelligent Systems and Applications*, 10(3), 2018.
[18] Nasim Arianpoo and Victor CM Leung. A smart fairness mechanism for concurrent multipath transfer in sctp over wireless multi-hop networks. *Ad Hoc Networks*, 55:40–49, 2017.
[19] Chengxiao Yu, Wei Quan, Deyun Gao, Yuming Zhang, Kang Liu, Wen Wu, Hongke Zhang, and Xuemin Shen. Reliable cybertwin-driven concurrent multipath transfer with deep reinforcement learning. *IEEE Internet of Things Journal*, 8(22):16207–16218, 2021.
[20] Changqiao Xu, Tao Zhang, Xiaohui Kuang, Zan Zhou, and Shui Yu. Context-aware adaptive route mutation scheme: A reinforcement learning approach. *IEEE Internet of Things Journal*, 8(17):13528–13541, 2021.
[21] Tao Zhang, Xiaohui Kuang, Zan Zhou, Hongquan Gao, and Changqiao Xu. An intelligent route mutation mechanism against mixed attack based on security awareness. In *2019 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6. IEEE, 2019.
[22] Tao Zhang, Changqiao Xu, Bingchi Zhang, Xiaohui Kuang, Yue Wang, Shujie Yang, and Gabriel-Miro Muntean. Dq-rm: Deep reinforcement learning-based route mutation scheme for multimedia services. In *2020 International Wireless Communications and Mobile Computing (IWCMC)*, pages 291–296. IEEE, 2020.
[23] Lyndon Ong, John Yoakum, et al. An introduction to the stream control transmission protocol (sctp). Technical report, RFC 3286 (Informational), May, 2002.
[24] Shruti Saini and Ansgar Fehnker. Evaluating the stream control transmission protocol using uppaal. *arXiv preprint arXiv:1703.06568*, 2017.
[25] Imtiaz Ali Halepoto, Muhammad Sulleman Memon, Nazar Hussain Phulpoto, Ubaidullah Rajput, and Muhammad Yaqoob Junejo. On the use of multipath transmission using sctp. *IJCSNS*, 18(4):58, 2018.
[26] M Tuexen, R Stewart, P Natarajan, J Iyengar, N Ekiz, T Dreibholz, M Becke, and P Amer. Load sharing for the stream control transmission protocol (sctp). *IETF Draft, Individual Submission, Draft-Tuexen-Tsvwg-Sctp-Multipath-16*, 2018.
[27] Johan Eklund, Karl-Johan Grinnemo, and Anna Brunstrom. Using multiple paths in sctp to reduce latency for signaling traffic. *Computer Communications*, 129:184–196, 2018.
[28] Joohyun Shin, Thomas A Badgwell, Kuang-Hung Liu, and Jay H Lee. Reinforcement learning–overview of recent progress and implications for process control. *Computers & Chemical Engineering*, 127:282–294, 2019.
[29] Martin Gregurić, Miroslav Vujić, Charalampos Alexopoulos, and Mladen Miletić. Application of deep reinforcement learning in traffic signal control: An overview and impact of open traffic data. *Applied Sciences*, 10(11):4011, 2020.
[30] Kaiqing Zhang, Zhuoran Yang, and Tamer Başar. Multi-agent reinforcement learning: A selective overview of theories and algorithms. *Handbook of Reinforcement Learning and Control*, pages 321–384, 2021.
[31] Yaodong Yang and Jun Wang. An overview of multi-agent reinforcement learning from game theoretical perspective. *arXiv preprint arXiv:2011.00583*, 2020.
[32] Zichang He and Wen Jiang. An evidential markov decision making model. *Information Sciences*, 467:357–372, 2018.
[33] David W Hosmer and Stanley Lemeshow. Confidence interval estimation of interaction. *Epidemiology*, pages 452–456, 1992.
[34] Sang Gyu Kwak and Jong Hae Kim. Central limit theorem: the cornerstone of modern statistics. *Korean journal of anesthesiology*, 70(2):144, 2017.
[35] Christopher JCH Watkins and Peter Dayan. Q-learning. *Machine learning*, 8(3-4):279–292, 1992.
[36] Francisco S Melo. Convergence of q-learning: A simple proof. *Institute Of Systems and Robotics, Tech. Rep*, pages 1–4, 2001.
[37] Thomas Dreibholz. Multi-path transport with omnet++ and the inet framework. 2017.
[38] Hilal H Nuha and Sidik Prabowo. Tcp congestion window analysis of twitter with exponential model. In *2018 6th International Conference on Information and Communication Technology (ICoICT)*, pages 61–65. IEEE, 2018.

Deguang Wang received his B.S. degree in Computer Science from Nanjing University of Posts and Telecommunications, Nanjing, China, in 2019. He is currently working toward a Master in School of Computer Science, Beijing University of Posts and Telecommunications, Beijing, China. His research interests include network transmission, and reinforcement learning.



Mingze Wang received his B.S. degree in material forming and control engineering from Shenyang Ligong University in 2016. He is currently working toward a master's degree in School of Computer Science, Beijing University of Posts and Telecommunications, Beijing, China. His research interests include cloud computing, moving target defense, and reinforcement learning.

Tao Zhang received his B.S. degree in Internet of Things Engineering from Beijing University of Posts and Telecommunications, Beijing, China, and Queen Mary University of London, London, UK in 2018. He is currently working toward a Ph.D in School of Computer Science, Beijing University of Posts and Telecommunications, Beijing, China. His research interests include network security, moving target defense, and reinforcement learning. He was a recipient of the Best Paper Award of NaNA 2018 and IWCMC 2021.

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

Chaoqiao Xu received the Ph.D. degree from the Institute of Software, Chinese Academy of Sciences (ISCAS) in Jan. 2009. He was an Assistant Research Fellow and R&D Project Manager in ISCAS from 2002 to 2007. He was a researcher at Athlone Institute of Technology and Joint Training PhD at Dublin City University, Ireland during 2007- 2009. He joined Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in Dec. 2009. Currently, he is a full Professor with the State Key Laboratory of Networking and Switching Technology, and Director of the Network Architecture Research Center at BUPT. His research interests include Network Security, Mobile Communications, and Future Internet Technology. He has published over 160 technical papers in prestigious international journals and conferences, including IEEE Comm. Surveys & Tutorials, IEEE Wireless Comm, IEEE Comm. Magazine, IEEE/ACM ToN, IEEE TMC etc. He has served a number of international conferences and workshops as a Co-Chair and Technical Program Committee member. He is currently serving as the Editor-in-Chief of Transactions on Emerging Telecommunications Technologies (Wiley).