Deep Reinforcement Learning for Multimedia Traffic Control in Software Defined Networking

Xiaohong Huang, Tingting Yuan, Guanhua Qiao, and Yizhi Ren

Abstract

Software Defined Networking (SDN) is a promising paradigm to provide centralized traffic control. Multimedia traffic control based on SDN is crucial but challenging for Quality of Experience (QoE) optimization. It is very difficult to model and control multimedia traffic because solutions mainly depend on an understanding of the network environment, which is complicated and dynamic. Inspired by the recent advances in artificial intelligence (AI) technologies, we study the adaptive multimedia traffic control mechanism leveraging Deep Reinforcement Learning (DRL). This paradigm combines deep learning with reinforcement learning, which learns solely from rewards by trial-and-error. Results demonstrate that the proposed mechanism is able to control multimedia traffic directly from experience without referring to a mathematical model.

Introduction

Software Defined Networking (SDN) [1] is a promising network architecture that implements the control plane in software. By decoupling the data plane and control plane, SDN provides a centralized view of network states and makes network management and control more flexible, consistent and holistic. It enables better traffic control and management by dynamically adapting the allocated bandwidth and path of each flow. The achievements of Google B4 [2] have shown that network performance can be improved by a network upgraded to SDN.

Nowadays, multimedia services such as audio, video and game services, are more and more popular. Multimedia traffic, which is considered to be a combination of audio, image and video, will be an integral part of future networks. Multimedia traffic poses significant challenges for service providers and network administrators because it is more sensitive to multiple metrics, such as delay, bandwidth, jitter, drop loss rate, and so on. Quality of Service (QoS) [3] is the most common indicator of overall performance, which contains multiple metrics. However, Quality of Experience (QoE) is a vital indicator to reflect the satisfaction of customers directly and intrinsically, which is more appropriate for multimedia traffic. How to improve the QoE of multimedia traffic through appropriate network traffic control is critical and still requires further study.

Network traffic control is a crucial constituent part of network management, which is the process of choosing a path and assigning bandwidth in each path. Due to stringent requirements of multimedia traffic, fine-grained and dynamic traffic control is expected. However, it is difficult in traditional distributed networks. With flexible network management capability, SDN provides great incentives for new traffic control techniques that exploit the global network view, status and flow patterns. Hence, it is of vital importance to study SDN based traffic control mechanisms to optimize the QoE of multimedia traffic. So far, few works have been done on QoE optimization based on SDN.

Traditional traffic control is usually formulated as an optimization problem [4–7]. In order to obtain appropriate solutions, it is important to formulate a good mathematical model, which is solvable and able to accord with the environment. However, this kind of traffic control has limitations to quickly adapt to network changes. Moreover, modern networks have become very complex, so it is difficult to model the network accurately. Thus, it has an undesirable performance in QoE optimization for multimedia traffic. Additionally, optimization problem solving is not flexible in the fluctuating communication environment.

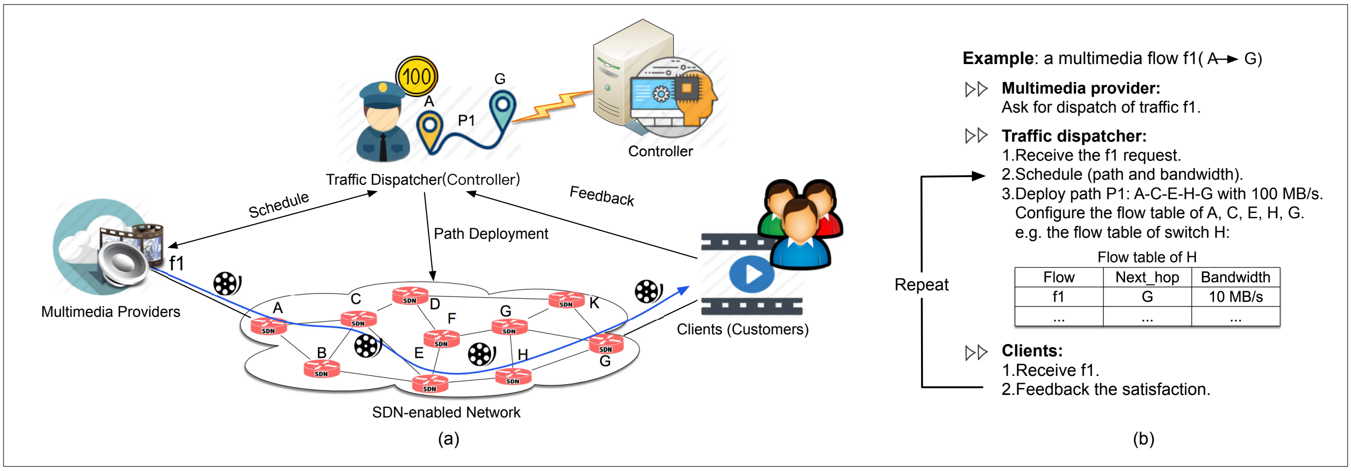
An artificial intelligence (AI) based algorithm has an outstanding advantage in solving this kind of problem. Reinforcement Learning (RL) is an AI-based algorithm that learns by trial-and-error solely from rewards of the environment. According to [8], Deep Reinforcement Learning (DRL) combines deep learning and reinforcement learning. It is able to learn for themselves from experience interacting with the environment to achieve successful strategies that lead to the greatest long-term rewards. DRL-based traffic control is just going into the scope of study, such as DRL-TE [9]. For multimedia traffic, QoE guarantee is the most important problem compared with other traffic. Multiple metrics affect QoE directly, which will address more challenges to the traffic control problem. However, there is little research that considers DRL based solutions for QoE optimization in networks.

In this article, we propose an adaptive multimedia traffic control approach leveraging the

Xiaohong Huang and Tingting Yuan (corresponding author) are with Beijing University of Posts and Telecommunications, Guanhua Qiao is with University of Electronic Science and Technology of China; Yizhi Ren is with Hangzhou Dianzi University.

Digital Object Identifier:
10.1109/MNET.2018.1800097

0890-8044/18/$25.00 © 2018 IEEE


100
G
A
P1
Controller
Traffic Dispatcher(Controller)
Schedule
Feedback
Path Deployment
f1
Multimedia Providers
A
B
C
D
E
F
G
H
K
G
SDN
Clients (Customers)
SDN-enabled Network
(a)
Example: a multimedia flow f1(A → G)
Multimedia provider:
Ask for dispatch of traffic f1.
Traffic dispatcher:
1.Receive the f1 request.
2.Schedule (path and bandwidth).
3.Deploy path P1: A-C-E-H-G with 100 MB/s.
Configure the flow table of A, C, E, H, G.
e.g. the flow table of switch H:
Flow table of H
Repeat
Clients:
1.Receive f1.
2.Feedback the satisfaction.
(b)

Flow	Next_hop	Bandwidth
f1	G	10 MB/s
...	...	...

FIGURE 1. An overview of multimedia traffic control in SDN: a) the architecture of multimedia traffic control in SDN; b) an example of multimedia traffic control.

DRL technique in SDN. First, QoE is considered as the main metric for traffic control. The proposed approach is able to learn traffic control strategy directly from experience with the object of maximizing the cumulative QoE. More narrowly, it learns how to allocate an appropriate path with unequivocal bandwidth for each multimedia flow. Second, in order to quickly obtain an accurate reward from the environment, QoE mapping is built using a deep neural network for each flow. By means of QoE mapping, the relationship between QoE and metrics produced by traffic control strategies can be obtained. Therefore, QoE can be predicted according to traffic control strategy. Third, compared with Deep Q-Network (DQN), which is a DRL-based method, Deep Deterministic Policy Gradient (DDPG) [10] has an advantage in solving dynamic and continuous control problems [9]. DDPG can also improve the convergence speed compared with the traditional actor-critic method. Therefore, DDPG is chosen to solve the traffic control problem in this article.

The remainder of the article is organized as follows. The overview of the architecture and techniques for SDN-based multimedia traffic control is presented in the following section. Next, DRL-based multimedia traffic control is described. After that, numerical results compared with two baseline optimization solutions are presented. The conclusion is given in the final section.

Architecture and Techniques of Multimedia Traffic Control

SDN is able to achieve QoE-driven traffic control by dynamically allocating network resources. In this section, the overview of SDN based architecture for multimedia traffic control is presented first. Based on that, an example to illustrate the process of multimedia traffic control is described. The comparison of various traffic control strategies is presented afterward.

Architecture

The architecture of SDN-based multimedia traffic control is shown in Fig. 1a. There are four major components.

Traffic Dispatcher (Controller): The controller in an SDN can be viewed as a traffic dispatcher, that has a global view of the network states. In an SDN-enabled communication network, the path allocation, the path deployment and bandwidth allocation for traffic are accomplished by the management of the SDN controller.

SDN-Enabled Network: The network contains a set of SDN-enabled forwarding devices that are under control of the SDN controller. With SDN, multiple paths can be used to transmit traffic for one multimedia service. However, it should be noticed that multi-path scheduling leads to out-of-order packet arrival, which not only increases memory and CPU utilization but also increases latency at the receiver. Moreover, reordering fails in case of packet loss [4]. Therefore, a single-path scenario is considered in this article.

Clients (Customers): A customer is also an observer who gives reaction of satisfaction of multimedia services as feedback. The reaction also correspondingly reflects the performance of multimedia traffic control.

Multimedia Providers: The multimedia providers are in charge of various kinds of multimedia services, which will be transmitted over the network.

The process of multimedia traffic control in SDN with the feedback mechanism is described with an example shown in Fig. 1b. First, upon the request of flows, the controller finds paths that can satisfy requirements of each flow with a global view. Second, a decision is made according to the traffic control strategy, which includes the path chosen and bandwidth allocation for each flow. For flow f1, which is from A to G, the path A-C-E-H-G is chosen with 10 MB/s. Then, the controller gives instructions on traffic control, such as path deployment and bandwidth allocation for each multimedia flow. After that, following the instructions, corresponding forwarding tables of SDN-enabled forwarding devices are configured. In this case, the router A, C, E, H and G should be configured to support the strategy, and the flow table of H that contains item about flow f1 is also shown in Fig. 1b. Finally, the satisfaction of clients for the multimedia service is sent back to controllers as feedback. The controller updates the traffic control strategy dynamically according to the feedback.

Classification	Scheme		Metrics					Challenges		
			QoE	Load balance	Fairness	Throughput	Latency	UTM	MVR	TB
Static	SECMP [5]		×	✓	×	×	×	×	×	×
	Reservation-based [4]		×	✓	✓	✓	✓	×	✓	×
Dynamic	Optimization-model-based	NUM [6]	×	×	✓	✓	×	×	✓	✓×
		MMU [11]	×	✓	×	×	×	×	×	✓×
		ML [12]	×	×	×	×	✓	×	×	✓×
	AI-based [13]		✓	✓	✓	✓	✓	✓	✓	✓

TABLE 1. Comparison of different traffic control strategies.

Techniques

In this section, a comparison of different traffic control techniques is presented. Metrics and challenges of multimedia traffic control are shown in Table 1. Five important metrics are listed:

- *QoE:* An important metric, especially for multimedia traffic, for which the customers' satisfaction is very important.
- *Load balance:* A metric to reflect the balance of resource utilization.
- *Fairness:* Used to reflect fairness of resource allocation for each flow.
- *Throughput:* The maximum rate that the network can deliver successfully.
- *Latency:* An expression of how much time it takes for a flow to get from one designated point to another.

As shown in Table 1, the challenges contain three main aspects that are enumerated as follows:

- Unpredictable traffic matrix (UTM): Traffic matrix refers to the demands between pairs of end-points in a network, which is dynamic and difficult to predict.
- Mix of various flow with different requirements (MVR): Traffic in a network is a mix of various flows with different types, requirements and preferences.
- Traffic burstiness (TB): A large number of traffic streams with burstiness, such as variable bit rate of multimedia traffic, will bring a large fraction of the workload to the network.

We classify methods of traffic control into two main categories, static traffic control and dynamic traffic control, as shown in Table 1. The former uses fixed criteria to assign traffic to available paths. Static Equal-cost Multi-path Routing (SECMP) [5] is a popular static load balancing technique to distribute load across equal cost paths. Another example of static traffic control is reservation-based [4] traffic control, which can reserve network resources from different aspects, that is, load balance and fairness. The static approaches are simple but inflexible.

The dynamic approaches support various criteria and are able to dynamically control flows by monitoring the network status and flow state. Flows can be moved across any of the available paths with adjustable available bandwidth according to different objectives. Network Utility Maximization (NUM) [6] is a traffic control technique with the objective to maximize network utility, which can be defined to be the fairness or the throughput. MMU [11] is designed to minimize the maximum utilization of the network, which takes load balance into consideration. ML [12] is a scheduling technique to minimize deadline miss rate and lateness. The above dynamic methods are optimization-model-based, in which an accurate optimization model will be built for traffic control. AI-based traffic control [13] provides a learning-based and intelligent schedule method. Different from optimization-model-based dynamic control techniques, AI-based traffic control can learn to control the network from its own experience rather than a mathematical model.

Table 1 shows the comparison of different scheduling techniques in traffic control. The first comparison is made from the metrics' point of view. From the table, it is easy to find that static approaches and optimization-model-based dynamic approaches take only one or two metrics to build the traffic control model. QoE can be regarded as a combination of various metrics, which include those obtained from the experience evaluations made by human users, and those obtained from explicit functions of measurable and controllable parameters related to the network. Static approaches and optimization-model-based dynamic approaches fail to support QoE optimization, because it is very hard to control the traffic based on the real-time collection of users' satisfaction. However, AI-based traffic control approaches are able to learn from experience interacting with the environment and adapt the traffic control based on the feedback of QoE. The second comparison is made from the challenges' point of view. Static approaches are not designed to handle dynamic traffic patterns, therefore they are limited in terms of UTM, MVR and TB. Optimization-model-based dynamic approaches rely on accurate and mathematically solvable models, which also limit their application in complex networks with UTM, MVR and TB characteristics. As for AI-based dynamic approaches, they enable model-free traffic control. Therefore, they are able to deal with complex networks with unpredictable behaviors and mixed traffic patterns.

Deep Reinforcement Learning for Multimedia Traffic Control

In this section, we present our design for multimedia traffic control in SDNs with the objective to optimize QoE. First, we propose a DRL-based

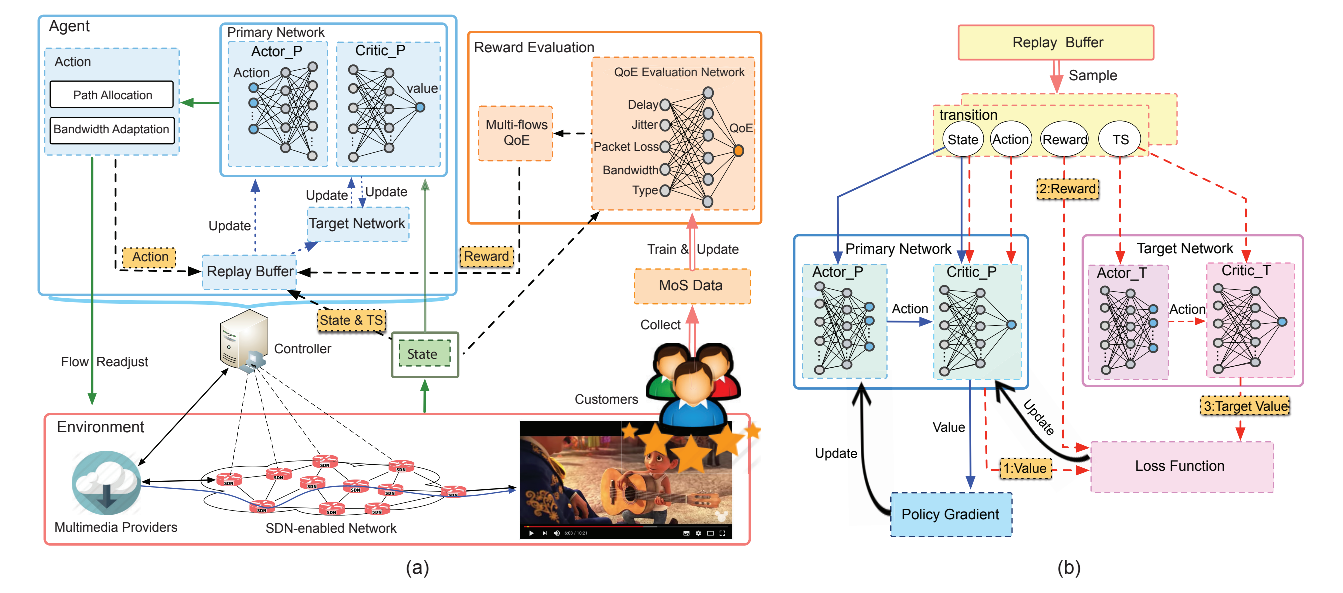


FIGURE 2. DRL-based multimedia traffic control: a) architecture; b) updating of the primary network using DDPG.

architecture for multimedia traffic control and describe the components in the architecture. After that, multimedia traffic control solutions with DRL are outlined.

DRL Components in SDN

The DRL-based multimedia traffic control architecture is shown in Fig. 2a. A standard reinforcement learning setup consists of two essential parts: an intelligent agent and the observed environment. Moreover, we model this problem as a Markov decision process (MDP) with a state space, an action space, and instantaneous reward function. Therefore, the architecture includes five components that are described as follows.

Environment: According to the architecture shown in Fig. 1a, the environment of SDN for multimedia traffic control comprises three parts: an SDN-enabled network, multimedia providers and customers. The SDN-enabled network consists of forwarding devices under the control of an SDN controller, in which the traffic control policy is deployed. The multimedia providers offer customers multimedia services, such as video, audio, and so on. The customers give feedback about multimedia services to reflect their satisfaction.

Agent: An agent is used to interact with the environment. Upon observing the consequences, it learns to alter its behavior and action in response to the received reward. When the DRL is applied to the system, the SDN controller (can be seen as an agent) serves as the centralized control to collect states, make decisions and take actions. The controller has a global view of the network to obtain the environment state. Based on observation, it can carry out a series of actions to react to the current state and offer a flexible way of policy deployment.

State: The state in DRL reflects the situation of the environment. For the problem of multimedia traffic control, the state of the environment refers to the state of flows. It covers several metrics including the allocated bandwidth, the delay, the jitter and the packet loss rate of flows. These metrics have different physical units (for example, bandwidth in B/s and delay in ms), which affects the learning efficiency. Thus, one method to deal with this problem is normalization, which normalizes each dimension of one state to a united scope.

Action: The objective of the agent is to map the space of states to the space of actions, and determine the optimal traffic control policy. In the area of multimedia traffic control, an action includes two portions: the path chosen and bandwidth adaption for all multimedia flows. As shown by the analysis in the previous chapter, a flow can be allocated with a certain routing path and an explicit bandwidth. Additionally, the available actions have to meet network constraints; for example, the allocated bandwidth of a flow should be within its demand range, which includes the lower-bound and the upper-bound. Moreover, the actions should be adjusted according to the feedback.

Reward: Based on the current state and action, the agent obtains a reward from the feedback of the environment. In this article, the purpose of traffic control is to improve the satisfaction degree of multimedia service; therefore, QoE is used as feedback. The mean opinion score (MOS) [14] can be used to evaluate QoE, which is the most common measurement for the quality of multimedia services. However, it is difficult to obtain MOS in real time, because it is unrealistic to interact with customers frequently. Compared with other methods, such as multiple linear regression and machine learning, deep learning can capture deep features and has an advantage in prediction accuracy [15]. Thus, in this article, a multi-layer deep neural network is used to map the network and application metrics to MOS. With the evaluation model, it is able to get the MOS quickly based on the state of flows.

Neural Networks in DDPG

Multimedia traffic control is a continuous problem, because its action space includes the bandwidth allocation of each flow, which is a continuous variable. DDPG, an action-critic method, is a preponderant method to solve it. Thus, a replay buffer, which is used to store the learning experience, can be used in DDPG for learning in each batch. This method is called experience replay, which benefits DDPG to learn across a set of uncorrelated transitions.

In DDPG, there are two kinds of networks in an agent, that is, the primary network and the target network, as shown in Fig. 2a

Primary Network: Used to determine an action based on the current state with the corresponding critic values. Thus, its input is the current state, and its output is an action. It consists of two deep neural networks, namely the actor network and the critic network.

Actor Network (Actor_P): The actor neural network is used to explore the policy, which specifies the current policy by deterministically mapping state to a specific action. Thus, the input of the neural network is the state, and the output is the action.

Critic Network (Critic_P): The critic neural network estimates the performance of Actor_P, and then it provides the critic value, which helps the actor to learn the gradient of the policy. The input of the critic network is the state from the environment and the action determined by Actor_P. Its output is the corresponding critic value.

For example, the state is the normalized metrics of each flow including the bandwidth, the delay, the jitter and the packet loss rate. The Actor_P determines a path with the reserved bandwidth for each multimedia flow (output) according to its state (input). Then, the Critic_P gives the critic value (output) according to the state and the action from the Actor_P (input). The critic value, for example from –3 (bad action) to 3 (excellent action), is used as a guide for updating the Actor_P.

Target Network: Used to generate the target value to train the Critic_P. Its input is the transformed state (TS), which follows the state transition rule, and output is the target value. It is an earlier snapshot of the primary network. Thus, it has same components with the primary network including a target actor network (Actor_T) and a target critic network (Critic_T). The input and output of the Actor_T and the Critic_T are similar with those of the Actor_P and the Critic_P. The main difference between the target network and the primary network is that the input of the target network is not the current state, but the TS.

QoE Evaluation Network: A flow-based multi-layer neural network that is used for MOS estimation. The delay, jitter, packet loss rate, allocated bandwidth and type of flows are chosen as the input of this neural network, because they are important metrics, which directly reflect QoE. However, the input is not restricted to these five factors. Based on the features of neural networks, it can be easily extended to support more factors by increasing its input neurons. The output of this neural network is the estimated MOS.

Multimedia Traffic Control with DDPG

Using Fig. 2a, the process of multimedia traffic control with DDPG in SDN (the solid line) is introduced as follows.

A replay buffer, which is used to store the learning experience, can be used in DDPG for learning in each batch. This method is called experience replay, which benefits DDPG to learn across a set of uncorrelated transitions.

Step 1: The information of the environment state is collected by the SDN controller, which is sent to the Actor_P.

Step 2: The agent uses the neural network of Actor_P to learn how to react to the current state, and determines an action based on its knowledge.

Step 3: The environment receives instructions about the determined action from the agent; devices in the environment are redeployed following the instructions, that is, SDN switches are deployed with new flow table items.

Step 4: After that, the state will transit from one to another called TS. The environment transition function is a law of the dynamics of the network state.

Step 5: Based on the new state, a reward is given using the QoE evaluation model. The mean MOS value of all flows is used as the reward of an action.

Step 6: The transition, which contains the state, the action, the reward and TS, is stored in the replay buffer. The transition storage procedure is shown using the dotted line with a label.

Learning Traffic Control with DDPG

The primary network, the target network and QoE evaluation networks are three important parts of DRL for multimedia traffic control. Deep neural networks are introduced in these networks, whose parameters are updated according to learning. To make efficient use of hardware optimizations, DDPG explores policy with off-policy algorithms in mini-batches [10], rather than on-line. For the problem of multimedia traffic control, it can be considered as a discrete-time event environment. At each time step, Actor_P and Critic_P are updated by sampling a minibatch from the replay buffer. Their updating using DDPG is introduced as follows, which is shown in Fig. 2b.

First, to update networks in the primary network (Actor_P and Critic_P), some samplings are selected from the replay buffer randomly or with priority [9]. A sampling is made up of a series of transitions, which contains the state, the action, the reward and TS.

Actor_P Updating: The process is shown with solid lines in Fig. 2b. The input of Actor_P is the state of a transition. Its output is an action determined by current parameters of this neural network. A critic value is provided to this action based on the present state using current Critic_P. Then, Actor_P updates the policy in a direction that would increase the probability to take a good action with a high critic value, and vice versa. The policy gradient method, which relies upon optimizing parametrized policies with respect to the expected return by gradient descent, is used for Actor_P updating.

Critic_P Updating: For the critic network, DQN, which combines Q-learning and deep learning, is used to guide the actor to choose a good policy. The Critic_P can be trained and updated by minimizing the loss function [10]. The

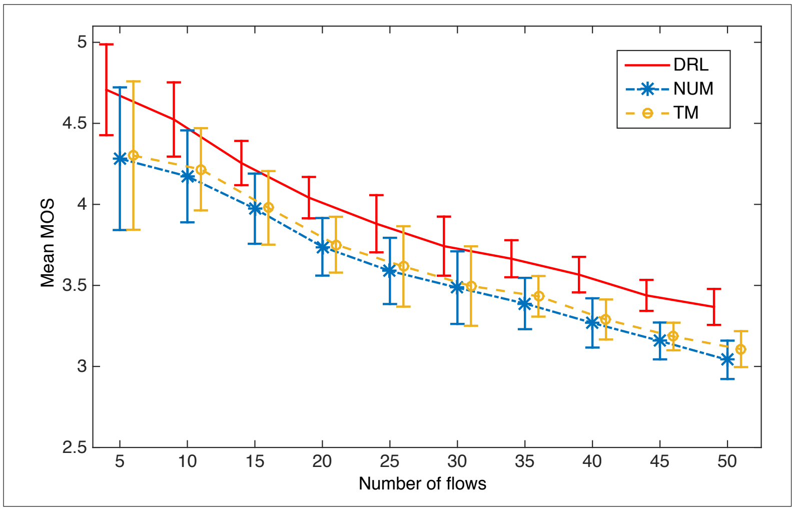


FIGURE 3. Performance of mean MOS value with different methods over the INDIA35 topology.

loss function is made up of three key points: the critic value from Critic_P, reward of a sampling, and the target value from the target network, shown in Fig. 2b with dotted lines. The loss function is defined as the residual sum of squares with these three parts.

The weights of Actor_P and Critic_P are updated simultaneously after the policy gradient and loss function of samplings are obtained.

Second, the way to update the target network and QoE evaluation networks is shown as follows.

Target Network Updating: The target network can be defined as an old version of the primary network, whose weights of neural networks are almost fixed. They are only periodically or slowly updated to the primary network, such as in a large constant number of steps. Additionally, it can also be modified using "soft" target updates, rather than directly copying the weights, which is proposed in [10]. This method is more stable than the first update method, but may be with slower learning.

QoE Evaluation Networks Updating: The reward of the environment comes from the QoE evaluation model, which is formulated as a neural network with raw MOS data collected from customers off-line. However, there is usually a time gap for the reward network to update after collecting enough new MOS data. Thus, the reward network should be updated periodically, slowly and off-line.

With the updating methods mentioned above, the controller can learn gradually how to react to obtain the most reward. It means the controller can learn how to improve the QoE of the multimedia flows by learning.

Numerical Results

In this section, extensive simulations are conducted to evaluate the performance of the proposed DRL-based architecture. Based on the proposed architecture, three parts are implemented, including the environment, the agent, and the reward evaluation. Additionally, the state is implemented in the environment part and the action is implemented in the agent part. For the environment, it is implemented using OMNET++, which is a modular, component-based C++ network simulator. It generates the state of delay, jitter and packet loss rate of flows in the simulated network to the agent. The agent is implemented using Python with Keras, which is a high-level neural network API capable of running on top of TensorFlow. For the reward evaluation, the collected data includes the bandwidth, the latency, the packet loss rate and the jitter of flows in a real network for a long time. The MOS data for off-line QoE evaluation network training, which ranges from 1 (Bad) to 5 (Excellent), is measured by a subjective method. The values of MOS are ranked by the viewers after they finish watching videos. The neural networks are initialized with Xavier initialization, which uses Gaussian distribution to generate the initial value of the weights.

We compared our proposed method with two baseline solutions in terms of QoE as follows:

- Throughput maximization (TM): Its objective is to obtain maximum throughput with traffic control.
- NUM: Its objective is to maximize network utility with traffic control, which is defined as a logarithmic function of allocated bandwidth. The solutions can be obtained by solving the convex programming problem.

The value of mean MOS can reflect the average satisfaction of resource allocation. Thus, MOS is selected as a crucial performance metric for comparison.

Figure 3 shows the mean MOS value versus the number of flows with different traffic control methods. The topology INDIA35 is used for simulation, which is a well known network topology from SNDlib. It has 35 nodes and 80 links with randomly generated capacity between 15 MB/s and 25 MB/s. The simulation is done multiple times with different sets of flows, which are randomly generated. This figure shows 95 percent confidence interval. The corresponding simulation results show that DRL has better performance in terms of mean MOS compared with the others, because the other two methods only consider bandwidth allocation in traffic control. As shown in this figure, the mean value of MOS is decreased with the number of flows, because when more flows exist in the network, there is more competition for network resources, such as bandwidth. Thus, the bandwidth of each flow is decreased with more flows deployed. Additionally, the quality of paths is decreased because of congestion.

In addition, we show the performance of our proposed mechanism in different topologies including INDIA35, GERMANY50 and TA2. GERMANY50 has 50 nodes and 88 links, and TA2 has 65 nodes and 108 links. This simulation is done multiple times with different sets of flows, which are randomly generated. Compared with TM, Fig. 4 shows the MOS improvement of DRL with 95 percent confidence interval. As expected, we can see that DRL-based multimedia traffic control significantly improves the MOS on all the three topologies. The improvement is about 8 percent on average compared with TM.

In conclusion, the simulation results show that the proposed method of multimedia traffic control has good performance in QoE improvement compared with other strategies.

Conclusion

SDN is proposed as a promising paradigm, which can provide centralized network management and traffic control. In order to improve user satisfaction for multimedia traffic, a DRL-based network traffic control architecture is proposed for traffic control in SDN. With DRL, the proposed architecture enables model-free traffic control, which is able to learn directly from experience and make decisions quickly. The results show that the proposed method can significantly improve the QoE of multimedia traffic compared with other strategies.

Acknowledgments

This article is supported by the National Key R&D Program of China (2016YFE0133000): EU-China study on IoT and 5G (EXCITING).

References

[1] N. Feamste, J. Rexford, and E. Zegura, "The Road to SDN: An Intellectual History of Programmable Networks," *SIGCOMM Comput. Commun. Rev.*, vol. 44, no. 2, Apr. 2014, pp. 87–98.

[2] S. Jain *et al.*, "B4: Experience with a Globally-Deployed Software Defined WAN," *SIGCOMM Comput. Commun. Rev.*, vol. 43, no. 4, Aug. 2013, pp. 3–14.

[3] Y. Zhang *et al.*, "Home M2M Networks: Architectures, Standards, and QoS Improvement," *IEEE Commun. Mag.*, vol. 49, no. 4, Apr. 2011, pp. 44–52.

[4] M. Noormohammadpour and C. S. Raghavendra, "Datacenter Traffic Control: Understanding Techniques and Trade-Offs," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 2, May 2018, pp. 1492–1525.

[5] A. Greenberg *et al.*, "VL2: A Scalable and Flexible Data Center Network," *Commun. ACM*, vol. 54, no. 3, Mar. 2011, pp. 95–104.

[6] A. Zhou *et al.*, "Joint Traffic Splitting, Rate Control, Routing, and Scheduling Algorithm for Maximizing Network Utility in Wireless Mesh Networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 4, Apr. 2016, pp. 2688–2702.

[7] Y. Zhang *et al.*, "Cognitive Machine-to-Machine Communications: Visions and Potentials for the Smart Grid," *IEEE Netw.*, vol. 26, no. 3, June 2012, pp. 6–13.

[8] K. Arulkumaran *et al.*, "Deep Reinforcement Learning: A Brief Survey," *IEEE Signal Process. Mag.*, vol. 34, no. 6, Nov. 2017, pp. 26–38.

[9] Z. Xu *et al.*, "Experience-Driven Networking: A Deep Reinforcement Learning Based Approach," *Proc. IEEE INFOCOM*, Honolulu, HI, USA, Apr. 2018.

[10] T. Lillicrap *et al.*, "Continuous Control with Deep Reinforcement Learning," *Proc. ICLR*, San Juan, Puerto Rico, May 2016.

[11] S. Agarwal, M. Kodialam, and T. V. Lakshman, "Traffic Engineering in Software Defined Networks," *Proc. IEEE INFOCOM*, Turin, Italy, Apr. 2013, pp. 2211–19.

[12] L. Chen *et al.*, "Scheduling Mix-flows in Commodity Datacenters with Karuna," *Proc. ACM SIGCOMM*, Florianopolis, Brazil, Aug. 2016, pp. 174–87.

[13] H. Luo and M.-L. Shyu, "Quality of Service Provision in Mobile Multimedia-A Survey," *Human-Centric Comput. Inf. Sci.*, vol. 1, no. 5, Nov. 2011, pp. 1–15.

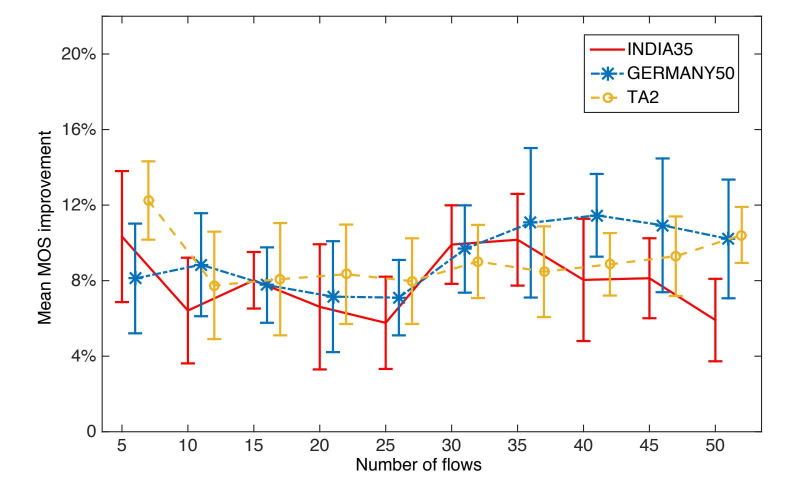


FIGURE 4. Performance of MOS improvement with different topologies.

[14] A. Khan, L. Sun, and E. Ifeachor, "QoE Prediction Model and Its Application in Video Quality Adaptation over UMTS Networks," *IEEE Trans. Multimedia*, vol. 14, no. 2, Apr. 2012, pp. 431–42.

[15] Y. Lecun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature*, vol. 521, no. 7553, May 2015, pp. 436–44.

Biographies

Xiaohong Huang (huangxh@bupt.edu.cn) received her B.E. degree from Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2000, and the Ph.D. degree from Nanyang Technological University, Singapore in 2005. She joined BUPT in 2005 and is now an associate professor of the Network and Information Center in the Institute of Network Technology of BUPT. She has published more than 50 academic papers. Her current interests are Internet architecture and software defined networking.

Tingting Yuan (yuantingting@bupt.edu.cn) is a Ph.D. candidate at the Institute of Network Technology, Beijing University of Posts and Telecommunications (BUPT), Beijing, China. Her current research interests are in computer networks and next-generation networks, including software defined networking, artificial intelligence, 5G and so on.

Guanhua Qiao (qghuestc@126.com) is currently working toward the Ph.D. degree at the School of Information and Communication Engineering, University of Electronic Science and Technology of China (UESTC). His current research interests include scheduling of mobile edge computation, deep reinforcement learning-based resource management and network optimization.

Yizhi Ren (renyz@hdu.edu.cn) is currently an associate professor with the School of Cyberspace, Hangzhou Dianzi University, China. He received a B.Sc. degree in computer science from Anhui Normal University, China, in 2004, and the M.Eng. and Ph.D. degrees in computer software and theory from Dalian University of Technology, China in 2006 and 2010, respectively. From 2008 to 2010, he was a research fellow at Kyushu University, Japan. His current research interests include network security, complex data/networks and artificial intelligence.