

EdgeSlice: Slicing Wireless Edge Computing Network with Decentralized Deep Reinforcement Learning

Qiang Liu, Tao Han, Ephraim Moges

Electrical and Computer Engineering Department,

The University of North Carolina at Charlotte, NC, United States

{qliu12, tao.han, emoges}@uncc.edu

Abstract—5G and edge computing will serve various emerging use cases that have diverse requirements of multiple resources, e.g., radio, transportation, and computing. Network slicing is a promising technology for creating virtual networks that can be customized according to the requirements of different use cases. Provisioning network slices requires end-to-end resource orchestration which is challenging. In this paper, we design a decentralized resource orchestration system named EdgeSlice for dynamic end-to-end network slicing. EdgeSlice introduces a new decentralized deep reinforcement learning (D-DRL) method to efficiently orchestrate end-to-end resources. D-DRL is composed of a performance coordinator and multiple orchestration agents. The performance coordinator manages the resource orchestration policies in all the orchestration agents to ensure the service level agreement (SLA) of network slices. The orchestration agent learns the resource demands of network slices and orchestrates the resource allocation accordingly to optimize the performance of the slices under the constrained networking and computing resources. We design radio, transport and computing manager to enable dynamic configuration of end-to-end resources at runtime. We implement EdgeSlice on a prototype of the end-to-end wireless edge computing network with OpenAirInterface LTE network, OpenDayLight SDN switches, and CUDA GPU platform. The performance of EdgeSlice is evaluated through both experiments and trace-driven simulations. The evaluation results show that EdgeSlice achieves much improvement as compared to baseline in terms of performance, scalability, compatibility.

Index Terms—Resource Orchestration, Deep Reinforcement Learning, Network Slicing, Wireless Edge Computing

I. INTRODUCTION

The emerging use cases and heterogeneous services, e.g., Internet of things (IoT), augmented/virtual reality (AR/VR), vehicle-to-everything (V2X), and mobile artificial intelligence, drive the development and research on the 5G mobile networks [1]. Unlike the conventional services, these new services have highly diverse performance requirements such as bandwidth, delay, and reliability, which imposes a challenge for 5G to accommodate these services in terms of scalability, availability, and cost-efficiency [2].

Leveraging software defined networking (SDN) and network functions virtualization (NFV), network slicing is a promising technique to address this challenge [3]. It enables multiple logical networks, i.e., network slices, run on top of a common physical network infrastructure [4]. Network slices can be individually customized to meet various performance requirements of different network services and use cases. For example, a slice can be customized to carry IoT services that require massive connections but low data rates. At the same time,

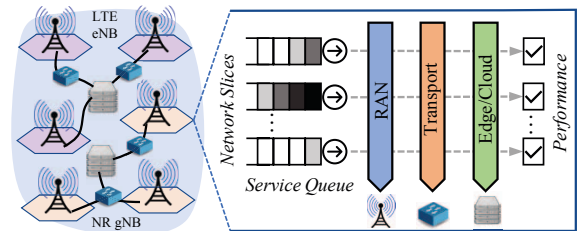


Fig. 1: An illustration of end-to-end network slicing in wireless edge computing network composed of distributed network infrastructures such as BSs and servers. Network slices require end-to-end resources from multiple technical domains, e.g., radio access network, transport network and edge/cloud computing platform, to serve their slice users.

another slice may be instantiated to support delay-sensitive services, e.g., mobile augmented reality and vehicle-to-vehicle communication. Thus, network slicing creates new network management and operation patterns and improves network performance for both the network operator and service providers in terms of network revenue, quality of service, and service autonomy.

Network operator is required to provide the performance and functional isolation to network slices [5]. The performance isolation ensures that the performance of a network slice is not affected by the operation of the other network slices. The functional isolation allows slice tenants to customize their slice's functions and resource management [6]. However, the isolation among network slices reduces the multiplexing efficiency and thus degrades the system performance [7]. It is observed that the multiplexing efficiency improves when the network resources are shared in a small time scale [7]. This observation advocates the dynamic network slicing which can dynamically change the resource allocation in network slices according to their actual needs.

Dynamic network slicing, as illustrated in Fig. 1, faces two research challenges. First, it is almost impossible to obtain the exact correlation between the resources and performance of network slices. A network slice usually requires resources from multiple technical domains such as radio access network, transport network, and edge/cloud. There are very complex tradeoffs among these resources and slice performance. For example, a short delay in the radio access network can be compensated by accelerated computation in the edge/cloud servers. Therefore, there lacks a closed-form mathematical expression that models the correlation between the resources and

performance of network slices. The existing works on multi-resource allocation usually assume that multiple resources are allocated following a certain ratio, e.g., 1 unit radio spectrum : 2 unit computing resources, which is not efficient [8], [9]. The second challenge is that the spatial diversity of mobile traffic requests the resources of network slices to be properly distributed among base stations and edge/cloud servers in different geographic locations. This further complicates the dynamic network slicing problem.

In this paper, we design *EdgeSlice*, a decentralized resource orchestration system that automates dynamic end-to-end network slicing in wireless edge computing networks. EdgeSlice introduces a novel decentralized deep reinforcement learning (D-DRL) method to efficiently orchestrate end-to-end networking and computing resources. With the D-DRL methods, the resource orchestration is carried out by a central performance coordinator and multiple decentralized orchestration agents. The orchestration agents rely on DRL to learn optimal resource orchestration policy, and the central performance coordinator coordinates the resource orchestration in the agents to ensure the service level agreements (SLAs) of network slices. To realize EdgeSlice, we also develop new radio, transport, and computing resource manager that can manage the resources at runtime according to the resource orchestration actions and instantiate network slices.

The contributions of this paper are summarized as follows:

- We design and implement EdgeSlice which is a decentralized resource orchestration system for dynamic network slicing in wireless edge computing networks. EdgeSlice automates dynamic network slicing leveraging a novel decentralized deep reinforcement learning (D-DRL) method.
- We design a new D-DRL method to automate the end-to-end resource orchestration with high efficiency. The D-DRL method is composed of a performance coordinator and multiple orchestration agents. The orchestration agent can learn the optimal resource orchestration policy under the coordination of the performance coordinator.
- We develop radio, transport and computing manager which are integrated with existing platforms: OpenAir-Interface (OAI) in radio access network, OpenDay-Light (ODL) in transport network, and CUDA GPU in edge/cloud servers. These managers enable the dynamic configuration of end-to-end resources at runtime in the EdgeSlice system.
- We build an experiential prototype and implement the EdgeSlice system. We evaluate the performance of the EdgeSlice system through both experiments using the prototype system and trace-driven network simulations.

II. EDGESLICE OVERVIEW

EdgeSlice automates dynamic network slicing in wireless edge computing networks through decentralized deep reinforcement learning. Fig. 2 outlines the design of the EdgeSlice system. To automate the network slicing process, EdgeSlice leverages machine learning, i.e., deep reinforcement learning,

entity	symbol	entity	symbol
network slice	i	resource autonomy (RA)	j
network resource	k	time interval	t
slice queue length	l	time period	\mathcal{T}
slice performance	\mathbf{U}	resource orchestration	x
min. performance	\mathbf{U}^{\min}	total resource	R^{tot}
auxiliary variable	z	dual variable	y

TABLE I: Notations throughout Sec.III

to learn end-to-end resource demands of network slices and then orchestrates the resource allocations to network slices accordingly. Owing to the temporal and spatial dynamics of the slice traffic and the complex tradeoffs between the performance of network slices and the resource orchestration, it is inefficient to use a centralized learning agent to orchestrate resource allocations to network slices. Besides, a centralized learning agent needs to obtain network performance data from all the network nodes, which introduces excessive communication overhead and delay. Toward this end, EdgeSlice introduces a new decentralized deep reinforcement learning method for network slicing in wireless edge computing networks.

We define a resource autonomy (RA) as a set of network infrastructures such as BSs and edge servers in a geographic area, and thus the network can be partitioned into multiple RAs. An orchestration agent is designed based on deep reinforcement learning to manage multi-domain resources in each RA and operates on a short timescale, e.g., seconds, to enable dynamic network slicing. The orchestration agent (detailed in Sec. IV-B) can track the network state (queue length, traffic), learn the resource orchestration policy from experience and orchestrate resources to slices autonomously.

A centralized performance coordinator is designed to coordinate the resource orchestration in all the RAs and optimizes the performance of the network on a much larger timescale. Meanwhile, the performance coordinator ensures that all the constraints related to the resource orchestration, e.g., SLAs and system capacity, are satisfied (detailed in Sec. IV-A). The performance coordinator only exchanges slight coordinating information with orchestration agents, which substantially decreases the communication overheads.

To realize EdgeSlice, resource managers, i.e., middleware, are developed to manage resources in radio access network, transport network, and edge computing servers at runtime according to the resource orchestration decision made by orchestration agents (detailed in Sec. V).

III. SYSTEM MODEL AND PROBLEM STATEMENT

To design the EdgeSlice system, we first mathematically model the wireless edge computing network and formalize the statement of end-to-end resource orchestration problem.

A. System Model

We consider an end-to-end wireless edge computing network which is composed of a radio access network (RAN) with multiple base stations (BSs), edge/cloud computing servers, and a transport network connecting the RAN and computing servers. As shown in Fig. 1, there are multiple network slices that request end-to-end resources in every RA,

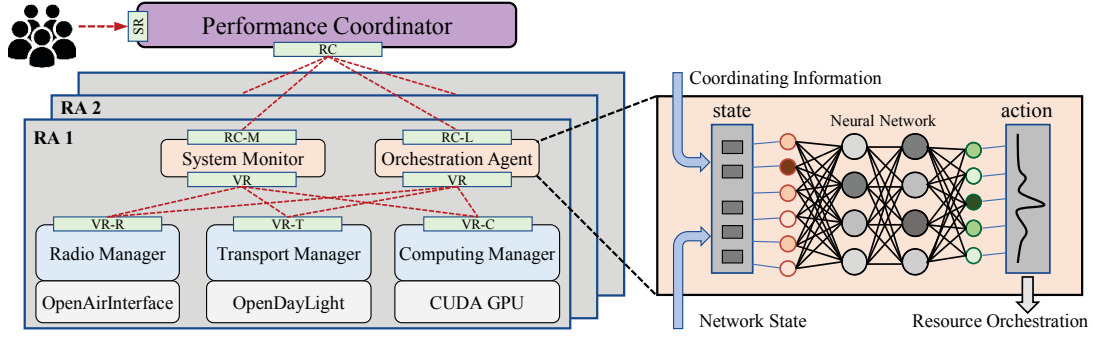


Fig. 2: The EdgeSlice System.

in order to enable seamless service coverage and support their users mobility. In each RA, network slices have service queues that buffer the arrival traffic of their slice users. We consider the network is time-slotted, and network operator can observe the performance¹ of network slices and dynamically change its resource orchestration with a minimum t time interval.

Let \mathcal{I} , \mathcal{J} and \mathcal{K} be the sets of network slices, RAs and network resources, respectively. Denote $\mathbf{x}_{i,j}^{(t)} = [x_{i,j,k}^{(t)} | \forall k \in \mathcal{K}]$ where $x_{i,j,k}^{(t)}$ is the k th resource allocated to the i th slice on the j th RA and $\mathbf{U}_{i,j}^{(t)}$ is the performance of network slice.

B. Problem Statement

The objective of network slicing is to maximize the performance of network slices in the system, and the objective of the network slicing can be expressed as $\max_{\{\mathbf{x}_{i,j}^{(t)}\}} \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \sum_{t=0}^{\tau} \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} \mathbf{U}_{i,j}^{(t)}$. As $\tau \rightarrow \infty$, the problem is an infinite time horizon stochastic programming problem. A general method to solve the problem is to transform it into a problem within a finite time period \mathcal{T} , e.g., a day [10], [11]. Hence, the resource orchestration problem is formulated as

$$\mathcal{P}_0 : \begin{aligned} & \max_{\{\mathbf{x}_{i,j} \geq 0\}} \sum_{t \in \mathcal{T}} \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} \mathbf{U}_{i,j}^{(t)} \\ & \text{s.t.} \quad (2), (3). \end{aligned} \quad (1)$$

In the context of network slicing, the resource orchestration problem subjects to two practical constraints. The first constraint is that the network-wide performance of a network slice should meet the SLA made between the slice tenant and network operator. Denote \mathbf{U}_i^{\min} as the minimum performance requirement of the i th slice according to the SLA. Thus, the performance constraint can be written as

$$\sum_{t \in \mathcal{T}} \sum_{j \in \mathcal{J}} \mathbf{U}_{i,j}^{(t)} \geq \mathbf{U}_i^{\min}, \quad \forall i \in \mathcal{I}. \quad (2)$$

The second constraint is that the resources in each RA are limited. Denote $R_j^{\text{tot}} = [r_{j,k}^{\text{tot}} | \forall k \in \mathcal{K}]$ as the total amount of each resource in the j th RA. Then, the resource allocated to network slices in the j th RA should be less than R_j^{tot} , and the constraint can be expressed as

¹Network slices could have various metrics on evaluating their performances, e.g., latency, throughput, queue status.

$$\sum_{i \in \mathcal{I}} \mathbf{x}_{i,j}^{(t)} \leq R_j^{\text{tot}}, \quad \forall j \in \mathcal{J}, t \in \mathcal{T}. \quad (3)$$

The difficulties in solving problem \mathcal{P}_0 are two-fold. First, the problem involves the end-to-end resource orchestration to network slices within each RA and the performance coordination across all RAs to maintain network-wide performance of network slices. The coupling between the intra-RA and inter-RAs resource management highly complicates the problem. Second, due to the varying network dynamics and the diversity of resource demands of network slices, the slice performance becomes a complex stochastic function. In real systems, it is almost impossible to derive an accurate mathematical model for such correlation [12]. Moreover, the resource orchestration in the network slicing system exhibits *Markovian* on serving slice users where a resource orchestration policy affects not only the current but also further network state, e.g., service queues.

IV. EDGESLICE DESIGN: COORDINATOR AND AGENTS

In this section, we present the design of performance coordinator and orchestration agents in the EdgeSlice system.

A. Performance Coordinator

Since the performance of a network slice depends on the resource orchestration in multiple RAs, the central performance coordinator is designed to coordinate the resource orchestration among RAs and thus optimizes the performance of the network slices. To design the performance coordinator, we transform problem \mathcal{P}_0 by introducing auxiliary variables $\mathcal{Z} = \{z_{i,j}, \forall i \in \mathcal{I}, j \in \mathcal{J}\}$ where

$$z_{i,j} = \sum_{t \in \mathcal{T}} \mathbf{U}_{i,j}^{(t)}, \quad \forall i \in \mathcal{I}, j \in \mathcal{J}. \quad (4)$$

Then, the constraint (2) are equivalent to

$$\sum_{j \in \mathcal{J}} z_{i,j} \geq \mathbf{U}_i^{\min}, \quad \forall i \in \mathcal{I}. \quad (5)$$

Hence, problem \mathcal{P}_0 is equivalently transformed to

$$\mathcal{P}_1 : \begin{aligned} & \max_{\{\mathbf{x}_{i,j} \geq 0, z_{i,j}\}} \sum_{t \in \mathcal{T}} \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} \mathbf{U}_{i,j}^{(t)} \\ & \text{s.t.} \quad (3), (4), (5). \end{aligned} \quad (6)$$

Problem \mathcal{P}_1 has two sets of variables, \mathcal{X} and \mathcal{Z} which are coupled by constraint (4). Next, we derive augmented Lagrangian of problem \mathcal{P}_1 as

$$\mathcal{L}_y = \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} \left(\sum_{t \in \mathcal{T}} \mathbf{U}_{i,j}^{(t)} - \frac{\rho}{2} \left\| \sum_{t \in \mathcal{T}} \mathbf{U}_{i,j}^{(t)} - z_{i,j} + y_{i,j} \right\|_2^2 \right), \quad (7)$$

where $\rho \geq 0$ is a positive constant, and $\mathcal{Y} = \{y_{i,j}, \forall i \in \mathcal{I}, j \in \mathcal{J}\}$ is the scaled dual variables. Here, the augmented Lagrangian incorporates the constraint (4) which couples variables \mathcal{Z} and \mathcal{X} .

According to the alternating direction method of multipliers (ADMM) method [13], problem \mathcal{P}_1 is solved by iteratively solving the following problems:

$$\mathbf{x}_{i,j} = \arg \max_{\mathbf{x}_{i,j} \in (3)} \mathcal{L}_y(\mathbf{x}_{i,j}, z_{i,j}, y_{i,j}), \quad (8)$$

$$z_{i,j} = \arg \max_{z_{i,j} \in (5)} \mathcal{L}_y(\mathbf{x}_{i,j}, z_{i,j}, y_{i,j}), \quad (9)$$

$$y_{i,j} = y_{i,j} + \left(\sum_{t \in \mathcal{T}} \mathbf{U}_{i,j}^{(t)} - z_{i,j} \right), \quad (10)$$

where problem in Eq. 8 focuses on the resource orchestration. Problem in Eq. 10 and Eq. 9 update auxiliary and dual variables, respectively, which require all the resource orchestrations in the system.

Therefore, we design the performance coordinator to solve the problem in Eq. 9 and Eq. 10 based on the resource orchestration and slice performance collected from orchestration agents in the system. Since \mathcal{X} and \mathcal{Z} are obtained, the problem in Eq. 9 is equivalent to

$$\mathcal{P}_2: \min_{\{z_{i,j}\}} \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} \left\| \sum_{t \in \mathcal{T}} \mathbf{U}_{i,j}^{(t)} - z_{i,j} + y_{i,j} \right\|_2^2 \quad (11)$$

s.t. (5).

This problem is a standard quadratic programming problem which can be solved by using convex optimization tools, e.g., CVX [14]. By solving the problem, the performance coordinator obtains auxiliary variables \mathcal{Z} and then updates dual variables \mathcal{Y} according to Eq. 10. We define the auxiliary variables \mathcal{Z} and the dual variables \mathcal{Y} as the coordinating information between the performance coordinator and orchestration agents.

B. Orchestration Agent

The orchestration agents are designed to orchestrate the end-to-end resources for network slices under the supervision of the performance coordinator, i.e., solving the problem in Eq. 8. Since the constraint of the problem only restricts the resource orchestration within a RA, it can be solved individually within each RA, i.e. decentralized. Hence, we rewrite the problem in Eq. 8 within the j th RA as

$$\mathcal{P}_3: \max_{\{\mathbf{x}_{i,j} \geq 0\}} \sum_{i \in \mathcal{I}} \sum_{t \in \mathcal{T}} \mathbf{U}_{i,j}^{(t)} - \frac{\rho}{2} \sum_{i \in \mathcal{I}} \left\| \sum_{t \in \mathcal{T}} \mathbf{U}_{i,j}^{(t)} - z_{i,j} + y_{i,j} \right\|_2^2 \quad (12)$$

s.t. (3).

The major challenge of solving the above problem is that the slice performance is very complex and without a closed-form mathematical model because of the varying network dynamic and the complicated end-to-end resource demands of network slices. Moreover, the current resource orchestration impacts both slice users in service queues and further network state. To address this challenge, we resort to deep reinforcement

learning (DRL) techniques that enable model-free machine learning [15] when designing orchestration agents.

We consider a general reinforcement learning setting where an agent interacts with an environment in discrete decision epochs. At each decision epoch t , the agent observes a state \mathbf{s}_t , takes an action \mathbf{a}_t , e.g., resource orchestration, based on its policy $\mu(\mathbf{s})$, and receives a reward $r(\mathbf{s}_t, \mathbf{a}_t)$. Then, the environment transits to the next state \mathbf{s}_{t+1} , e.g., queue status changes, based on the action taken by the agent. The objective is to find the optimal policy $\mu^*(\mathbf{s})$ mapping states to actions, that maximizes the discounted cumulative reward $\sum_{t=0}^{\infty} \gamma^t r(\mathbf{s}_t, \mathbf{a}_t)$. Here, $\gamma \in [0, 1)$ is a discounting factor.

Although DRL techniques have been extensively studied in many areas such as robotic control [16], traffic control [17], and chess games [18], the existing DRL models are not appropriate to solve problem \mathcal{P}_3 for two reasons. First, most of the DRL models are designed to solve constraint-free problems [17], [19]. However, the problem consists of multiple linear constraints. Second, the existing DRL models are unable to adjust their policies based on coordinating information from an external control [20]. However, to maintain the network-wide performance of network slices, the agent in EdgeSlice needs to orchestrate resources according to the coordinating information derived from the coordinator.

1) Design of Agents: Therefore, we design a new DRL model with customized state space, action space and reward function. In the DRL model, the constraint (3) are re-weighted and incorporated into its reward function so that the reward is affected by whether the constraints are satisfied or not. The coordinating information is augmented into state space to allow external control from the coordinator.

State Space: The state is concatenated by two parts. The first part is $[l_j^{(t)}, \forall i \in \mathcal{I}]$ which represents the current network state, i.e., queue status of network slices. The second part is $[z_{i,j} - y_{i,j}, \forall i \in \mathcal{I}]$ which is the coordinating information from the coordinator. Thus, the state in the j th RA at time interval t can be expressed as

$$\mathbf{s}_t = [l_j^{(t)}, z_{i,j} - y_{i,j}, \forall i \in \mathcal{I}]. \quad (13)$$

Action Space: The action at time interval t is defined as the resource allocations to network slices in the RA:

$$\mathbf{a}_t = [\mathbf{x}_{i,j}^{(t)}, \forall i \in \mathcal{I}]. \quad (14)$$

Reward: The reward at time interval t is defined as

$$r(\mathbf{s}_t, \mathbf{a}_t) = \sum_{i \in \mathcal{I}} \left(\mathbf{U}_{i,j}^{(t)} - \frac{\rho}{2} \left\| \mathbf{U}_{i,j}^{(t)} - \frac{1}{T} (z_{i,j} + y_{i,j}) \right\|_2^2 \right) - \beta \sum_{j \in \mathcal{J}} \left[\sum_{i \in \mathcal{I}} \mathbf{x}_{i,j}^{(t)} - R_j^{\text{tot}} \right]^+, \quad (15)$$

where $[x]^+ = \max(0, x)$, and β is a positive constant. Here, we approximate the objective function of problem \mathcal{P}_3 with identical sub-objective functions in the time domain. Moreover, we incorporate the constraints (3) into the sub-objective functions with reward shaping technique [21]. Therefore, there will be a penalty added into the reward if the constraints are violated.

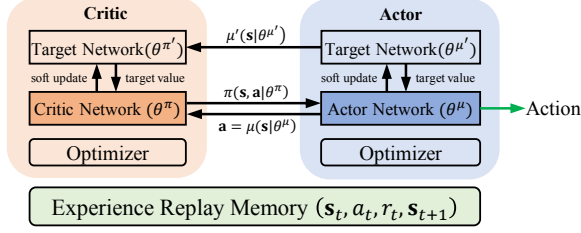


Fig. 3: The DDPG architecture.

2) *Training of Agents*: We follow deep deterministic policy gradient (DDPG), a state-of-the-art reinforcement learning technique that is capable of handling continuous and high-dimensional action spaces [15], to train our orchestration agents. As shown in Fig. 3, DDPG integrates deep Q-network (DQN) [16] and actor-critic method [22], and maintains a parameterized actor $\mu(s_t|\theta^\mu)$ and a parameterized critic $\pi(s_t, a_t|\theta^\pi)$. The critic estimates the value function of state-action pairs, and the actor specifies the current policy by mapping a state to a specific action.

The critic is implemented using a DQN. We define the value function $Q^\pi(s_t, a_t)$ as the expected discounted cumulative reward when the agent starts with the state-action pair (s_t, a_t) at decision epoch t and then acts according to the policy π . Then, the value function can be expressed as $Q^\pi(s_t, a_t) = \mathbb{E}_\pi [R_t]$, where $R_t = \sum_{k=t}^T \gamma^{(k-t)} r(s_k, a_k)$. Based on the Bellman equation [23], the optimal value function is $Q^*(s_t, a_t) = r(s_t, a_t) + \gamma \max_{a_{t+1}} Q^*(s_{t+1}, a_{t+1})$.

To obtain the optimal policy, DQN is trained by minimizing the mean-squared Bellman error (MSBE)

$$L(\theta^\pi) = \mathbb{E} \left[(g_t - Q(s_t, a_t|\theta^\pi))^2 \right], \quad (16)$$

where θ^π are parameters of the critic network, and \mathcal{D} is a replay memory. g_t is the target value estimated by a target network, and can be expressed as

$$g_t = r(s_t, a_t) + \gamma \max_{a_{t+1}} Q(s_{t+1}, \mu(s_{t+1}|\theta^{\mu'})|\theta^{\pi'}), \quad (17)$$

where $\theta^{\pi'}$ are parameters of the target network. The target network has the same architecture as the critic network, and its parameters $\theta^{\pi'}$ are slowly updated to track that of the critic network.

The actor is implemented using another DQN which learns a deterministic policy $\mu(s_t|\theta^\mu)$ to maximize the cumulative reward of the actor, i.e., $J = \mathbb{E}_\mu [R_t]$. Since the action space is continuous, the value function is assumed to be differentiable with respect to the action. Thus, the actor network can be trained by applying the chain rule to the expected cumulative reward with respect to the actor parameters θ^μ :

$$\begin{aligned} \nabla_{\theta^\mu} J &\approx \mathbb{E} \left[\nabla_{\theta^\mu} Q(s, a|\theta^\pi) \Big|_{s=s_t, a=\mu(s_t|\theta^\mu)} \right] \\ &= \mathbb{E} \left[\nabla_a Q(s, a|\theta^\pi) \Big|_{s=s_t, a=\mu(s_t)} \cdot \nabla_{\theta^\mu} \mu(s|\theta^\mu) \Big|_{s=s_t} \right]. \end{aligned} \quad (18)$$

C. The Workflow of EdgeSlice

The workflow of the EdgeSlice system is summarized in Alg. 1. The resource orchestration starts by initializing the coordinating information, i.e., \mathcal{Z} and \mathcal{Y} . The orchestration

Algorithm 1: The EdgeSlice Resource Orchestration

Input: $\mathbf{U}_i^{\min}, \forall i \in \mathcal{I}; R_j^{\text{tot}}, \forall j \in \mathcal{J}; \rho, \beta$.

Output: $\mathcal{X}, \mathcal{Z}, \mathcal{Y}$.

```

1 Initialize  $\mathcal{Z}$  and  $\mathcal{Y}$ ;
2 while True do
3   /** optimize  $\mathcal{X}$  in each agent **/;
4   for  $j \in \mathcal{J}$  (decentralized) do
5      $\mathbf{x}_{i,j}^{(t)}, \forall i \in \mathcal{I}, t \in \mathcal{T} \leftarrow$  the  $i$ th orchestration agent;
6      $\mathbf{U}_{i,j}^{(t)}, \forall i \in \mathcal{I}, t \in \mathcal{T} \leftarrow$  the  $i$ th slice performance;
7   /** update  $\mathcal{Z}$  in the coordinator **/;
8    $z_{i,j} \leftarrow \arg \max_{z_{i,j} \in (5)} \mathcal{L}_y(\mathbf{x}_{i,j}, z_{i,j}, y_{i,j})$ ;
9   /** update  $\mathcal{Y}$  in the coordinator **/;
10   $y_{i,j} \leftarrow y_{i,j} + (\sum_{t \in \mathcal{T}} \mathbf{U}_{i,j}^{(t)} - z_{i,j})$ ;
11  /** if algorithm converges **/;
12  if convergence then
13    return  $\mathcal{X}, \mathcal{Z}, \mathcal{Y}$ ;

```

agent in each RA orchestrates resources to network slices based on its parameterized policy under the coordinating information for time intervals in \mathcal{T} . At the end of a time period \mathcal{T} , the orchestration agent collects the performance of network slices \mathbf{U} . Given \mathcal{X} and \mathbf{U} , the performance coordinator generates the coordinating information (\mathcal{Y} and \mathcal{Z}), which are fed back to orchestration agents in all RAs. It continues until the convergence of the resource orchestration.

V. EDGEslice DESIGN: RESOURCE MANAGER

In this section, we design radio, transport, and computing manager that allocates the resources orchestrated by agents to network slices at runtime, as shown in Fig. 2. These managers are integrated with OpenAirInterface (OAI), OpenDayLight (ODL), and CUDA GPU computing platform to enable dynamic configuration of resources in radio access network, transport network, and edge/cloud computing, respectively.

A. Radio Manager

The radio manager is designed to work with OpenAirInterface (OAI) to allocate radio resources to slice users in both uplink (UL) and downlink (DL) radio access network. In EdgeSlice, the total radio resources (bandwidth) can be used by a network slice is determined by the orchestration agent. Once a network slice obtains its radio resources, it allocates these resources to its users. As a result, the allocated radio resources of all slice users are known by the radio manager. Hence, the radio manager should schedule users according to their allocated resources at runtime, which is not supported by vanilla OAI.

We fulfill such functionality by developing a new user scheduling method in the MAC layer to manage physical resource blocks (PRBs) in PUSCH/PDSCH. We schedule the slice users consecutively and map their radio resources to PRBs. The users without any radio resources will not be

scheduled. To support the information exchange between the orchestration agent and the radio manager at runtime, we develop the VR-R (virtual resource - radio) and VR (virtual recourse) interfaces in the radio manager and orchestration agent, respectively. The association between a mobile user and a network slice is identified by the user's international mobile subscriber identity (IMSI). The IMSI information is extracted from the SIAP message sent from the base station to mobile management entity (MME). The information extraction does not need any modification on the mobile user's side.

B. Transport Manager

Taking advantage of the separation of data and control plane in SDN switches, we allocate the bandwidth of links between RAN and edge/cloud computing servers with an OpenDayLight [24] controller through OpenFlow (Southbound API) and RESTful (Northbound API) [25]. The OpenFlow protocol currently supports user bandwidth modification with *meters*. However, these meters and their attached flows should be deleted and reinitialize if the user bandwidth needs to be changed. As a result, when changing the user bandwidth allocation at runtime, the switch network is broken during the deletion-creation interval [26].

To enable dynamic modification of bandwidth while keeping the switches network alive, we create a new configuration that parallels with the current one when a new user bandwidth allocation is received from the orchestration agent. Only if the new configuration is available in switches, we release the current configuration to transition to the new one accordingly so that we can hide the deletion-creation interval. In addition, the information exchange between the transport manager and orchestration agent is support through the VR-T (virtual resource - transport) interface and the VR interface. The association of users and slices in the transport network are identified by using their source and destination IP addresses.

C. Computing Manager

The computing manager is designed to dynamically allocate computing resources, e.g., the number of CUDA *threads*, in the CUDA-based GPU computing platform. In the CUDA programming model, an application can launch multiple *kernels*, where every kernel can be concurrently executed by massive CUDA threads [27]. The number of threads required by a kernel is specified in its execution configuration syntax. The execution of these kernels in the kernel space follows the order of their callings in the user space. With the multiple-processes service (MPS), multiple applications or processes can share the GPU simultaneously. However, the resource scheduling strategies of user applications are nontransparent and not revealed by NVIDIA. As a result, the resource usage of user applications can not be effectively controlled.

To address this issue, we develop a kernel-split mechanism to control the GPU computing resources by managing the maximum concurrent number of threads occupied by every user application. The kernel-split mechanism splits a kernel that requests a large number of threads into multiple small

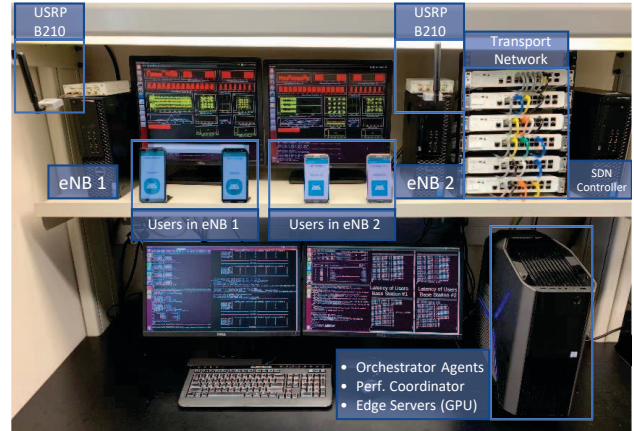


Fig. 4: The overview of prototype.

and consecutive kernels with a specific number of threads. We heavily modify the kernels of user applications to dynamic split the kernels according to the user's virtual resources at runtime. Since the execution of kernels are in-order and consecutive, the number of threads occupied by a user application always less than its virtual resources. We develop the VR-C (virtual resource - computing) interface in the computing manager for exchanging information with the orchestration agent. The association between a mobile user and the network slice is identified by the IP address.

D. System Monitor

The system monitor is designed to collect information of network state, e.g., traffic load and slice performance, by using a dataset. The database also records the user-slice association based on the users' IMSIs and IP addresses. The system monitor uses the VR interface to communicate with radio, transport and computing manager.

The RC (resource coordination) interface is developed to allow the central performance coordinator to communicate with orchestration agents and system monitors through the RC-L (resource coordination - learning) and RC-M (radio coordination - monitoring), respectively. The SR (slice request) interface is developed to enable the slice tenants to request and configure their network slices. For example, slice tenants can make and modify their service-level agreements (SLAs) with network operator. The SLAs will be enforced during the resource orchestrations.

VI. SYSTEM IMPLEMENTATION

A. Hardware Details

We develop a prototype of the EdgeSlice system as depicted in Fig. 4. It is composed of a RAN with 2 eNodeBs, a transport network with 6 OpenFlow switches, a core network, and 2 edge servers with CUDA GPUs. The details of hardware are summarized in Table II. To eliminate the co-channel interference, eNodeBs are operating at different frequency bands, i.e., LTE Band 7 and Band 38. We configure the band

Component	Hardware	Software
UEs	4x Samsung smartphones with band selection capability	Android 7.0
eNodeBs	2x Intel i5 Computer with low-latency kernel 3.19	OpenAirInterface (OAI) [28]
RF Front-End	2x Ettus USRP B210	N/A
Transport	6x OpenFlow 1.3 Ruckus switches	OpenDayLight-Boron [24]
Core Network	Intel i7 desktop computer	openair-cn [29]
Edge Servers	2x NVIDIA GEFORCE GTX 1080Ti	CUDA 9.0 [27]

TABLE II: Details of the Prototype

selection option on smartphones so that the users in eNodeB 1 and 2 can only search for band 7 and band 38, respectively.

In the prototype, there are 2 RAs, 2 slices and 4 mobile users (1 user per slice per RA), where a RA is the set of an eNodeB, an edge server and a transport link. The orchestration agents and performance coordinator are implemented in the core network (Alienware R7 desktop) with Python 3.5. The optimization toolbox used in the performance coordinator is CVXPY 1.0 [30]. The radio manager is deployed in every eNodeB. The transport manager is deployed on an individual desktop computer. The computing manager is implemented on the edge server for every RA. Both eNodeBs are with 5MHz (25 PRBs) wireless bandwidth. The total bandwidth between an eNodeB and its corresponding edge server is 80Mbps. The total amount of the computing resource for each RA is 51200 CUDA threads.

We implement orchestration agents with Tensorflow 1.10 [31]. We use a 2-layer fully-connected neural network in both actor and critic networks. Both layers adopt Leaky Rectifier [32] activation functions with 128 neurons. In the output layer, we use *sigmoid* [32] as the activation function. On training orchestration agents, we conduct extensive and empirical tunings on the hyper-parameters. We randomly generate $z_{i,j} - y_{i,j}$ between 0 and R_j^{tot} to train the agents under different coordinating information. The parameter $\beta = 20$ to have sufficient weight on enforcing the total orchestrated resources constraint (3). The learning rates of both actor and critic networks are 0.001. The batch size is 512. The total training step is 1E6. The discounted factor for cumulative reward is $\gamma = 0.99$. We add the decaying Gaussian noise on actions during the training phase for balancing the exploitation and exploration. The noise starts from $\mathcal{N}(0, 1)$ and decays with factor 0.9999 per update step.

B. Simulated Network Environment

The orchestration agents are trained offline by using a simulated network environment as shown in Fig. 5. In the environment, we implement a first-in first-out (FIFO) queue for services in individual network slices, and the performance function of each slice can be customized. In each time interval, the traffic, i.e., service tasks, in the network slices is generated according to real network traffic traces [33]. The service time of each task is determined by the end-to-end resource orchestration.

With the simulated network environment, we generate the training dataset by traversing all possible orchestration actions using the grid search method for radio, transport and computing resources, respectively. Due to the large number

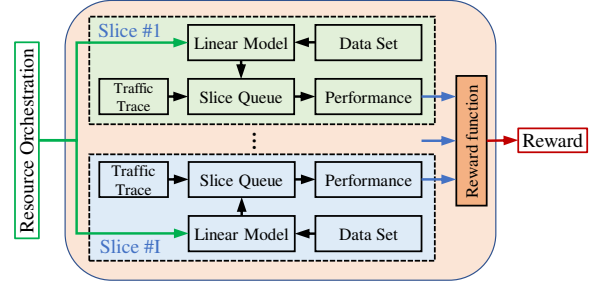


Fig. 5: The simulated network environment.

of orchestration actions, we conduct the experiments with resource granularity 10% for all the resources, which means the dataset only contains discrete orchestration actions. During the training of agents, it may produce orchestration actions that are not contained in the training dataset. To solve this problem, we build a linear regression model with *scikit-learn* [34] tool to approximate the correlations between orchestration actions and the slice performance. Given a resource orchestration action such as $[12, 38, 22]\%$, we use adjacent orchestration actions in the dataset, e.g., $[10, 30, 20]\%$ and $[10, 40, 20]\%$, to fit the linear model. Once the linear model is fitted, it makes the prediction for the service time under the orchestration action. The service time determines the traffic departure in service queues. At the end of each time interval, the reward is derived based on the performances of all network slices and the design of reward function in Eq. 15.

VII. PERFORMANCE EVALUATION

In this section, we evaluate the performance of EdgeSlice with both prototype experiments and network simulations. At each time interval, the i th slice on the j th RA reports its performance to orchestration agent according to $\mathbf{U}_{i,j}^{(t)} = -(l_{i,j}^{(t)})^\alpha, \forall i \in \mathcal{I}, j \in \mathcal{J}, t \in \mathcal{T}$, where $\alpha = 2$ and $l_{i,j}^{(t)}$ is the queue length. Note that the performance function is defined to evaluate whether EdgeSlice can learn the optimal resource orchestration policy. In other words, neither the performance coordinator or orchestration agent know the closed-form expression of the performance function. Besides, various performance functions are evaluated in simulations. The performance requirements of slices are defined as $\mathbf{U}_i^{\min} = -50, \forall i \in \mathcal{I}$ and $\rho = 1.0$ [35].

A. Mobile Application

To evaluate the system performance, we develop a mobile application which offloads computation tasks to the edge/cloud

servers. Here, the computation tasks are the video analysis based on the YOLO object detection framework [36]. The basic procedures of these applications are: 1) a user sends a video frame with a specific resolution to server and waits to receive the processed results; 2) the server receives the frame from the user and executes the YOLO algorithm with a specific computation model to analyze the frame; 3) the server sends the analysis results back to the user. The mobile application can use different frame resolutions, e.g., 100x100, 300x300 to 500x500, and select computation models, e.g., YOLO 320x320, YOLO 416x416 to YOLO 608x608. Here, the application with a higher frame resolution has heavier transmission traffic, and the application with a larger computation model requires a more intensive computation workload.

B. Comparison Algorithms

In the performance evaluation, we compare the EdgeSlice resource orchestration with the following algorithms:

Traffic-Aware Resource Orchestration (TARO): TARO is the baseline algorithm in which all the resources are proportionally shared by slices according to the current queue length. In other words, $\mathbf{x}_{i,j}^{(t)} = R_j^{tot} \cdot l_{i,j}^{(t)} / \sum_{i \in \mathcal{I}} l_{i,j}^{(t)}, \forall j \in \mathcal{J}$. This sharing scheme applies to all the RAs in the system.

EdgeSlice-Non-Traffic (EdgeSlice-NT): EdgeSlice-NT is a simplified version of EdgeSlice in which the orchestration agent manages resources only based on the coordination information from the performance coordinator. Therefore, the state space of the orchestration agent of EdgeSlice-NT is $\mathbf{s}_t = [z_{i,j} - y_{i,j}, \forall i \in \mathcal{I}]$. In other words, EdgeSlice-NT does not use queue length of network slices as the state in the DRL model. By comparing EdgeSlice and EdgeSlice-NT, we can evaluate the impact of the state space design, i.e. whether including traffic load or not, on the performance of network slices.

C. Experimental Results

Here, we present the experimental results and evaluate the performance of the EdgeSlice system from different angles. In the experiment, there are 2 slices, 2 RAs and 3 types of resources. The mobile application in the first slice uses 500x500 frame resolution and selects YOLO 320x320 as the computation model. This application represents the type of applications that have heavy transmission traffic load and moderate computation workload. The mobile application in the second slice uses 100x100 frame resolution and selects YOLO 608x608 as the computation model. This application represents the type of applications that have light transmission traffic load and intensive computation workload.

In the experiments, the time interval t is 1 second and the time period \mathcal{T} is composed of 10 time intervals. During the time intervals, the task arrival of network slices follow the Poisson process with average arrival rate² 10.

²The slice traffic is normalized based on the hardware capability of the prototype such as bandwidth and GPU on accommodating the mobile applications.

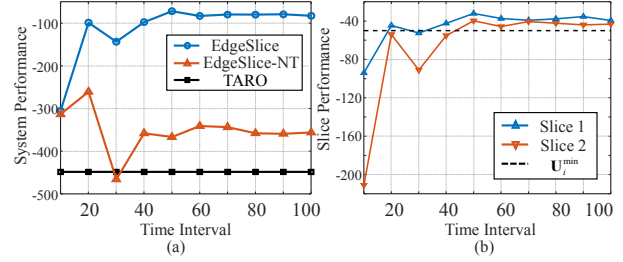


Fig. 6: The convergence of algorithms: (a) system performance vs. time interval; (b) slice performance vs. time interval.

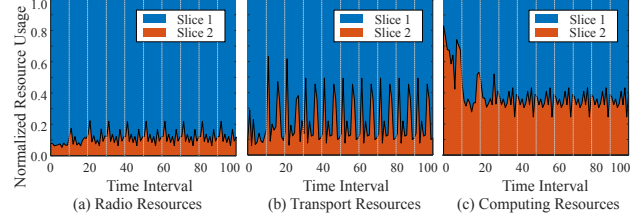


Fig. 7: The multiple resource orchestrations of EdgeSlice: (a) radio resource; (b) transport resource; (c) computing resource.

1) Convergence: In the EdgeSlice system, the performance coordinator coordinates multiple orchestration agents via the coordinating information $[z_{i,j} - y_{i,j}, \forall i \in \mathcal{I}]$. We first evaluate how fast the interaction between the coordinator and orchestration agents can converge. As depicted in Fig. 6 (a), both EdgeSlice and EdgeSlice-NT are able to converge after several time periods. This result also reveals that orchestration agents can effectively orchestrate resources to slices under different coordinating information. EdgeSlice obtains 3.69x and 2.74x improvement on the system performance as compared to TARO and EdgeSlice-NT, respectively. The performance gain over TARO proves that EdgeSlice can effectively learn the optimal resource orchestration policy based on the current network state and the coordinating information. The performance gain over EdgeSlice-NT indicates that observing the traffic load of slices by orchestration agents can significantly improve the system performance. In addition, as shown in Fig. 6 (b), the EdgeSlice system ensures that both network slices meet their minimum performance requirements.

Fig. 7 shows the normalized usage of multiple resources, i.e., radio, transport and computing resources, with the EdgeSlice system. In the experiments, slice 1 has a higher demand of radio and transport resources and a lower demand of computing resources than slice 2 does. Hence, we observe that EdgeSlice allocates more radio and transport resources to slice 1 (blue area). Since slice 2 serves compute-intensive applications, it requires more computing resources. Therefore, in the beginning, slice 2 is allocated more computing resources. Later, EdgeSlice observes that the performance requirement of slice 1 cannot be met although it is allocated almost all the radio and transport resources. Thus, EdgeSlice starts to allocate more computing resources to slice 1 and then the resource orchestration converges. Moreover, we observe the resources orchestrations becomes stable after 6 interactions, which corresponds to the observations in Fig. 6 (a).

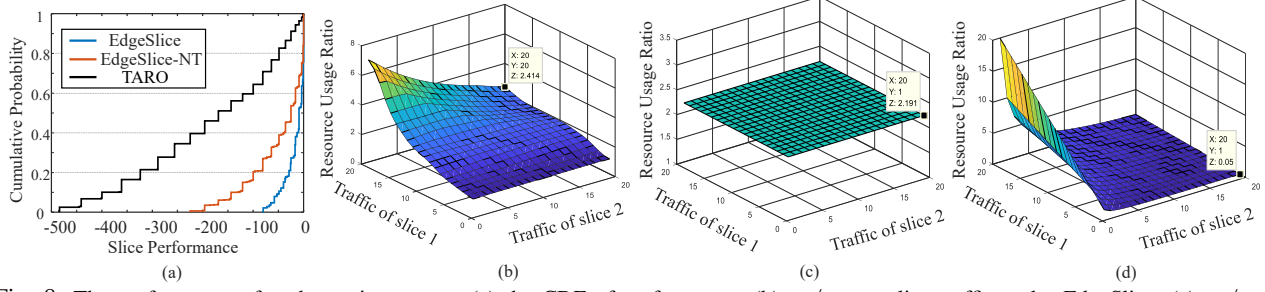


Fig. 8: The performance of orchestration agents: (a) the CDF of performance; (b) η_1/η_2 vs. slice traffic under EdgeSlice; (c) η_1/η_2 vs. slice traffic under EdgeSlice-NT; (d) η_1/η_2 vs. slice traffic under TARO.

2) *Resource Orchestration*: We evaluate the orchestration agent without any central coordination to understand its resource orchestration policy. Fig. 8 (a) depicts the cumulative distribution function (CDF) of the slice performance under randomly generated slice traffic loads. We can see that EdgeSlice substantially outperforms both TARO and EdgeSlice-NT in terms of the slice performance. For example, 80% of the slice performance is larger than -30 using EdgeSlice while it is only 11% and 55% using TARO and EdgeSlice-NT, respectively. The performance difference between EdgeSlice and EdgeSlice-NT is smaller than that shows in Fig. 6 (a). The reason is that the performance deficiency of the orchestration agent in EdgeSlice-NT accumulates during the iterative interactions between the agents and the coordinator.

Fig. 8 (b)-(d) show the average resource usage ratio between slice 1 and slice 2 obtained by using EdgeSlice under different traffic loads. The average resource usage of a slice is calculated as $\eta_i = \sum_{k \in \mathcal{K}} x_{i,j,k} / r_{j,k}^{tot}$. It can be observed that EdgeSlice allocates resources to slices based on both traffic load and the application's resource needs in different domains. For example, when traffic loads of slice 1 and slice 2 are 20 and 5, respectively, the average resource usage ratio is about 5. This example shows the traffic-awareness of EdgeSlice. Since the orchestration agent in EdgeSlice-NT does not learn the slice traffic load in the resource orchestration, the resource usage ratio is a constant as shown in Fig. 8 (c). TARO allocates resources purely based on the slice traffic and is not aware of the actual resource needs from each domain. The resource usage ratio with TARO is shown in Fig. 8 (d). The comparison between EdgeSlice and TARO shows that EdgeSlice is aware of the multi-domain resource needs of an application. These results validate that orchestration agents of EdgeSlice are able to autonomously orchestrate end-to-end resources under varying slice traffic.

D. Simulation Results

We set up network simulations to evaluate EdgeSlice in terms of scalability and working with different training techniques and performance functions. In the simulation, there are 5 slices, 10 RAs, and 3 types of resources. The applications served by the network slices randomly select the frame resolutions, e.g., 100x100, 300x300, or 500x500, and computation models, e.g., 320x320, 416x416, 608x608. We use the network trace from an Italy mobile network over the Province of

Trento [33] to generate the traffic in network slices. The network trace contains 154.8M entries with a minimum 10 minutes time interval collected in December 2013. Each entry includes the counts of phone calls, SMS, Internet traffic, and the geographic square area id. We obtain the average calling traffic in 24 hours under different geographic areas, and use them for the traffic of network slices. In the simulation, the time interval t is 1 hour and the time period \mathcal{T} is composed of 24 time intervals.

1) *Scalability of EdgeSlice*: We evaluate the scalability of EdgeSlice by varying the number of slices and RAs. As shown in Fig. 9 (a), both EdgeSlice and EdgeSlice-NT maintain similar performance per RA as the number of RAs increases, while the performance per RA of TARO decreases substantially. This result indicates the EdgeSlice agents learn much superior resource orchestration policy than TARO in each RA. Besides, EdgeSlice is capable of scaling to large network sizes without noticeably sacrificing system performance. Fig. 9 (b) shows the performance per slice versus different number of network slices. As the number of slices increases, the system performance decrease because the resource demand is higher and the average allocated resources of slices are reduced. Nevertheless, EdgeSlice is still able to obtain a better performance than the others. These results validate the scalability of the EdgeSlice system.

2) *Training Techniques of Agents*: We study the impact of various techniques on training the orchestration agents in the EdgeSlice system. As depicted in Fig. 10 (a), the system performance drops remarkably when the training steps of agent is insufficient such as 1E5. In general, a learning-based agent with a large number of training steps has better performance than that with a small number of training steps. We can see that the performance of EdgeSlice and EdgeSlice-NT can be worse than that of TARO if the number of training steps is 1E5 or less. This means that if the agent is not well trained, it could lead to very poor performance. Moreover, various techniques, e.g., SAC [37], PPO [38], TRPO [39], and VPG [40], have been proposed to improve the performance of agents. We evaluate the system performance of EdgeSlice under different training techniques as shown in Fig. 10 (b). The training setting and hyper-parameters are the same as mentioned in Sec. VI. The orchestration agent trained using DDPG exhibits better performance than that trained by the other techniques. These results show the importance of the training techniques

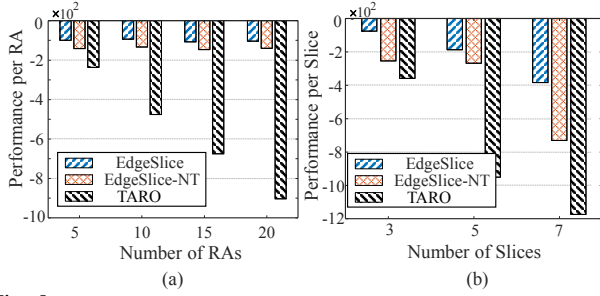


Fig. 9: The scalability of EdgeSlice: (a) performance per RA vs. the number of RAs; (b) performance per slice vs. the number of slices.

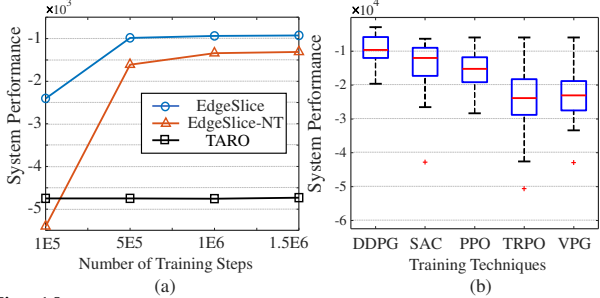


Fig. 10: The impact of training techniques: (a) system performance vs. the number of training steps of orchestration agents; (b) system performance vs. various training techniques.

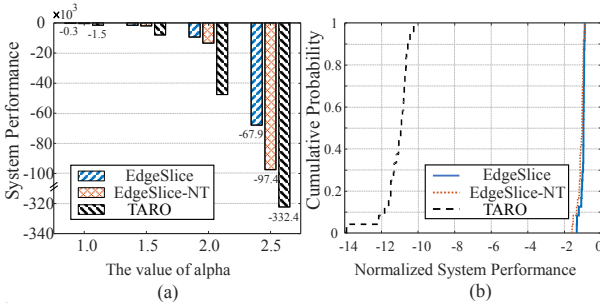


Fig. 11: The compatibility of EdgeSlice: (a) system performance vs. the value of α , (b) CDF of normalized system performance.

in developing the EdgeSlice system.

3) *Handling different performance functions:* We evaluate the performance of EdgeSlice under different performance functions of network slices. As shown in Fig. 11 (a), we vary the value of α in the performance function. The large α indicates slice reports worse performance under the same queue length. The EdgeSlice outperforms the others for all conditions, which implies EdgeSlice can automatically learn superior resource orchestration policy under varying performance functions. Furthermore, we define another performance function as the negative service time of slice users without considering traffic in slice queue. As shown in Fig. 11 (b), EdgeSlice and EdgeSlice-NT achieve almost the same system performance. Because we intentionally eliminate the impact of slice traffic on the slice performance function. As a result, the network state, i.e., queue length, observed by EdgeSlice is not helpful on learning the correlations. In contrast, the performance of TARO is much worse. These results indicate that when the performance function is less dependent on the network state, learning-based EdgeSlice and EdgeSlice-

NT still have much performance gain over TARO. These results verify the capability of EdgeSlice on handling various performance functions of slices.

VIII. RELATED WORK

This work relates to resource management in network slicing and deep reinforcement learning for networking problems.

Resource Management in Network Slicing: The resource management problem in network slicing has been extensively studied with the goal to maximize the system performance. Caballero *et al.* [8] constructed a network slicing game in which tenants are selfish to maximize its own performance. The authors proved that this game with such strategic behavior converges to a Nash equilibrium for elastic traffic. Halabian *et al.* [9] showed that non-collaborative slices in the system compromise the fairness performance when maximizing the overall system performance and proposed a distributed solution by extending the Dominant Resource Fairness (DRF) framework. To exploit the statistical multiplexing gain of slices, Sciancalepore *et al.* [41] designed STORNS that optimizes the admission control of slices with considering per slice SLA requirement by leveraging stochastic geometry theory. Salvat *et al.* [11] developed an end-to-end resource orchestration system, formulated an orchestration problem to maximize the revenue in network slicing, and proposed an optimal Benders decomposition method and a heuristic method. Foukas *et al.* [5] developed an efficient RAN slicing system that enables the dynamic and real-time virtualization of base stations and slices customization to meet slices' service demands. However, the fundamental assumption of these works is that the resource demands of slices and their performance functions are known as closed-form mathematical expressions to network operator. In contrast, the EdgeSlice system proposed in the paper enables a model-free resource orchestration solution.

Deep Reinforcement Learning (DRL) in Networking: Machine learning techniques such as deep learning and reinforcement learning gain significant popularity in solving resource management problems in mobile network for coping the complicated network dynamics. Mao *et al.* [20] designed DeepRM with the DQN technique to optimize the admission control and resource orchestration of users. They obtained a considerable reduction on the average slowdown of user tasks as compared to heuristic solutions. Xu *et al.* [17] utilized the state-of-the-art DDPG technique to solve the traffic engineering (TE) networking problem, i.e., allocating the bandwidth of network links, and obtained significant end-to-end latency reduction and performance improvement under the unknown performance function. Bega *et al.* [42] proposed DeepCog with deep learning techniques to forecast the network capacity within an individual slice and achieve the balance between resource over-provisioning and service request violations. Yang *et al.* [43] proposed an adaptive reinforcement learning based approach for microservice workflow system that enables model-free resource allocation and improves the response time of microservices. However, these existing works advocate the centralization of resource management by using a central

agent. Although their solutions may be applied to perform the resource orchestration, the centralized resource orchestration is highly complex for wireless edge computing networks. Different from these methods, the EdgeSlice system enables a decentralized resource orchestration.

IX. CONCLUSION

In this paper, we have designed EdgeSlice, a new decentralized resource orchestration system, to automate dynamic network slicing in wireless edge computing networks. To realize EdgeSlice, we have developed a novel decentralized deep reinforcement learning method which consists of a central performance coordinator and multiple orchestration agents. The orchestration agent learns the optimal resource orchestration policy for network slicing under the coordination of the central performance coordinator. We have also designed new radio, transport and computing resource manager that enable dynamic configuration of end-to-end resources at runtime. We have developed a prototype of EdgeSlice with OpenAirInterface (OAI) in radio access network, OpenDayLight (ODL) in transport network, and CUDA GPU computing in edge/cloud servers. The performance of EdgeSlice has been validated through both prototype implementation and network simulations.

ACKNOWLEDGEMENT

This work is partially supported by the US National Science Foundation under Grant No. 1731675, No. 1810174, and No. 1910844.

REFERENCES

- [1] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 1617–1655, 2016.
- [2] X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, "Network slicing in 5G: Survey and challenges," *IEEE Communications Magazine*, vol. 55, no. 5, pp. 94–100, 2017.
- [3] J. Ordonez-Lucena, P. Ameigeiras *et al.*, "Network slicing for 5G with SDN/NFV: Concepts, architectures, and challenges," *IEEE Communications Magazine*, vol. 55, no. 5, pp. 80–87, 2017.
- [4] I. Afolabi, T. Taleb *et al.*, "Network slicing and softwarization: A survey on principles, enabling technologies, and solutions," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 2429–2453, 2018.
- [5] X. Foukas, M. K. Marina, and K. Kontovasilis, "Orion: RAN slicing for a flexible and cost-effective multi-service mobile network architecture," in *ACM MobiCom*, 2017, pp. 127–140.
- [6] P. Rost, C. Mannweiler *et al.*, "Network slicing to enable scalability and flexibility in 5G mobile networks," *IEEE Communications magazine*, vol. 55, no. 5, pp. 72–79, 2017.
- [7] C. Marquez, M. Gramaglia, M. Fiore, A. Banchs, and X. Costa-Perez, "How should I slice my network?: A multi-service empirical evaluation of resource sharing efficiency," in *MobiCom*. ACM, 2018, pp. 191–206.
- [8] P. Caballero, A. Banchs, G. De Veciana, and X. Costa-Perez, "Network slicing games: Enabling customization in multi-tenant mobile networks," *IEEE/ACM Transactions on Networking*, 2019.
- [9] H. Halabian, "Distributed resource allocation optimization in 5G virtualized networks," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 3, pp. 627–642, 2019.
- [10] P. Kall *et al.*, *Stochastic programming*. Springer, 1994.
- [11] J. X. Salvat *et al.*, "Overbooking network slices through yield-driven end-to-end orchestration," in *ACM CoNEXT*. ACM, 2018, pp. 353–365.
- [12] H. Mao, M. Schwarzkopf *et al.*, "Learning scheduling algorithms for data processing clusters," in *Proceedings of the ACM Special Interest Group on Data Communication*. ACM, 2019, pp. 270–288.
- [13] S. Boyd, N. Parikh *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [14] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge university press, 2004.
- [15] T. P. Lillicrap, Hunt *et al.*, "Continuous control with deep reinforcement learning," *arXiv preprint arXiv:1509.02971*, 2015.
- [16] V. Mnih, K. Kavukcuoglu, Silver *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, p. 529, 2015.
- [17] Z. Xu, J. Tang *et al.*, "Experience-driven networking: A deep reinforcement learning based approach," in *IEEE INFOCOM*. IEEE, 2018, pp. 1871–1879.
- [18] D. Silver, T. Hubert *et al.*, "A general reinforcement learning algorithm that masters chess, shogi, and go through self-play," *Science*, vol. 362, no. 6419, pp. 1140–1144, 2018.
- [19] L. Chen, J. Lingys *et al.*, "AuTO: scaling deep reinforcement learning for datacenter-scale automatic traffic optimization," in *SIGCOMM 2018*. ACM, 2018, pp. 191–205.
- [20] H. Mao, M. Alizadeh *et al.*, "Resource management with deep reinforcement learning," in *ACM HotNets*. ACM, 2016, pp. 50–56.
- [21] S. Griffith *et al.*, "Policy shaping: Integrating human feedback with reinforcement learning," in *Advances in neural information processing systems*, 2013, pp. 2625–2633.
- [22] V. R. Konda and J. N. Tsitsiklis, "Actor-critic algorithms," in *Advances in neural information processing systems*, 2000, pp. 1008–1014.
- [23] R. Bellman, "Dynamic programming," *Science*, vol. 153, no. 3731, pp. 34–37, 1966.
- [24] J. Medved, R. Varga, A. Tkacik, and K. Gray, "Opendaylight: Towards a model-driven SDN controller architecture," in *IEEE WoWMoM 2014*. IEEE, 2014, pp. 1–6.
- [25] N. McKeown, T. Anderson *et al.*, "Openflow: enabling innovation in campus networks," *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 2, pp. 69–74, 2008.
- [26] O. S. Specification, "Openflow switch specification version 1.5.1," <https://www.opennetworking.org/wp-content/uploads/2014/10/openflow-switch-v1.5.1.pdf>, 2013.
- [27] C. Nvidia, "Nvidia CUDA C programming guide," *Nvidia Corporation*, vol. 120, no. 18, p. 8, 2011.
- [28] OpenAirInterface Software Alliance. OpenAirInterface repository. <https://gitlab.eurecom.fr/oai/openairinterface5g>, 2017.
- [29] OpenAirInterface Software Alliance. Openair-cn repository. <https://gitlab.eurecom.fr/oai/openair-cn>, 2017.
- [30] S. Diamond and S. Boyd, "CVXPY: A Python-embedded modeling language for convex optimization," *Journal of Machine Learning Research*, vol. 17, no. 83, pp. 1–5, 2016.
- [31] M. Abadi, P. Barham *et al.*, "Tensorflow: A system for large-scale machine learning," in *12th USENIX OSDI*, 2016, pp. 265–283.
- [32] I. Goodfellow *et al.*, *Deep learning*. MIT press, 2016.
- [33] T. Italia, "Telecommunication activity dataset," <https://dandelion.eu/datagems/SpazioDati/telecom-sms-call-internet-tn/description/>, 2013.
- [34] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [35] M. Hong and Z.-Q. Luo, "On the linear convergence of the alternating direction method of multipliers," *Mathematical Programming*, vol. 162, no. 1–2, pp. 165–199, 2017.
- [36] J. Redmon *et al.*, "You only look once: Unified, real-time object detection," in *IEEE CVPR*, 2016, pp. 779–788.
- [37] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," *arXiv preprint arXiv:1801.01290*, 2018.
- [38] J. Schulman, F. Wolski *et al.*, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.
- [39] J. Schulman, S. Levine, P. Abbeel, M. Jordan, and P. Moritz, "Trust region policy optimization," in *International Conference on Machine Learning*, 2015, pp. 1889–1897.
- [40] R. S. Sutton, D. A. McAllester *et al.*, "Policy gradient methods for reinforcement learning with function approximation," in *Advances in neural information processing systems*, 2000, pp. 1057–1063.
- [41] V. Sciancalepore, M. Di Renzo, and X. Costa-Perez, "STORNS: Stochastic radio access network slicing," *arXiv:1901.05336*, 2019.
- [42] D. Bega, M. Gramaglia *et al.*, "Deepcog: Cognitive network management in sliced 5G networks with deep learning," 2019.
- [43] Z. Yang *et al.*, "MIRAS: Model-based reinforcement learning for microservice resource allocation over scientific workflows," in *IEEE ICDCS*. IEEE, 2019, pp. 122–132.