



Atlas: Automate Online Service Configuration in Network Slicing

Qiang Liu

University of Nebraska-Lincoln

qiang.liu@unl.edu

Nakjung Choi

Nokia Bell Labs

nakjung.choi@nokia-bell-labs.com

Tao Han

New Jersey Institute of Technology

tao.han@njit.edu

ABSTRACT

Network slicing achieves cost-efficient slice customization to support heterogeneous applications and services. Configuring cross-domain resources to end-to-end slices based on service-level agreements, however, is challenging, due to the complicated underlying correlations and the simulation-to-reality discrepancy between simulators and real networks. In this paper, we propose Atlas, an online network slicing system, which automates the service configuration of slices via safe and sample-efficient learn-to-configure approaches in three interrelated stages. First, we design a learning-based simulator to reduce the sim-to-real discrepancy, which is accomplished by a new parameter searching method based on Bayesian optimization. Second, we offline train the policy in the augmented simulator via a novel offline algorithm with a Bayesian neural network and parallel Thompson sampling. Third, we online learn the policy in real networks with a novel online algorithm with safe exploration and Gaussian process regression. We implement Atlas on an end-to-end network prototype based on OpenAirInterface RAN, OpenDayLight SDN transport, OpenAir-CN core network, and Docker-based edge server. Experimental results show that, compared to state-of-the-art solutions, Atlas achieves 63.9% and 85.7% regret reduction on resource usage and slice quality of experience during the online learning stage, respectively.

CCS CONCEPTS

- Networks → Network algorithms; Mobile networks;
- Computing methodologies → Machine learning;

KEYWORDS

Machine Learning, Network Slicing, Autonomous Management

ACM Reference Format:

Qiang Liu, Nakjung Choi, and Tao Han. 2022. Atlas: Automate Online Service Configuration in Network Slicing. In *The 18th International Conference on emerging Networking EXperiments and Technologies (CoNEXT '22), December 6–9, 2022, Roma, Italy*. ACM, New York, NY, USA, 16 pages. <https://doi.org/10.1145/3555050.3569115>

1 INTRODUCTION

Network slicing is one of the key building blocks in 5G and Beyond [14] to provide guaranteed networking performances for concurrently supporting various network services and applications, e.g.,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CoNEXT '22, December 6–9, 2022, Roma, Italy

© 2022 Association for Computing Machinery.

ACM ISBN 978-1-4503-9508-3/22/12...\$15.00

<https://doi.org/10.1145/3555050.3569115>

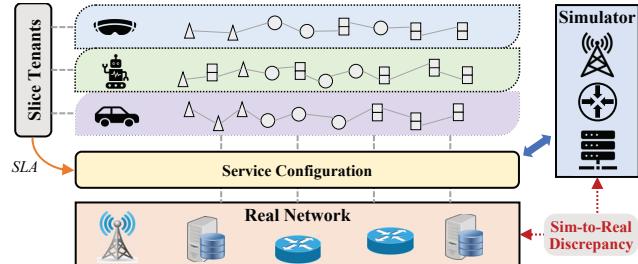


Figure 1: Illustration of online network slicing

mobile augmented reality [25], autonomous driving [51], and federated learning [47]. It enables network operators to cost-efficiently create virtual networks (*aka.* network slices) with performance and functional isolation [12] based on the common physical infrastructure. Each network slice can be highly customized according to the needs of individual slice tenants [2], e.g., throughput per slice user, delay, and quality of service (QoS). As emerging services are increasingly focusing on end-to-end performances, e.g., round-trip latency, end-to-end slicing is more than ever needed, which consists of the subnet instance [3] in radio access networks (RAN), transport networks (TN), core networks (CN), and edge networks (EN).

To initialize a network service, slice tenants make service-level agreements (SLAs) with the network operator to specify the performance requirement of its users. To maintain the performance of individual slices, as shown in Fig. 1, the network operator aims to derive a policy¹ to dynamically configure the slice's cross-domain resources and settings, e.g., radio and backhaul bandwidth, under varying network dynamics. As heterogeneous slices are with highly diverse needs with performance metrics, e.g., reliability in vehicle-to-everything and frame-per-second for video streaming [55], it is impractical to comprehensively examine individual slices before their deployments. Hence, automated network slicing is indispensable to automatically learn the actual needs of slices according to their service-level agreements (SLAs) and intelligently adjust service configuration of slices by exploiting state-of-the-art machine learning (ML) techniques [36, 49].

It is, however, challenging to obtain the automated service configuration policy in end-to-end network slicing, even if recent advances in deep neural networks (DNNs) showed promising capability in complex function approximations [27, 29, 42]. On the one hand, it is unsafe to learn the configuration policy via online interaction with real networks. The exploration during policy learning may decrease service performances occasionally [23], which results in the violation of slice SLA. On the other hand, it is insufficient to learn the configuration policy via offline interaction with network

¹We refer a policy as the mapping from network states, e.g., spatiotemporal user traffic, to service configuration actions, e.g., radio bandwidth.

simulators. The simulation-to-reality (sim-to-real) discrepancy between simulators and real networks could compromise the online performance achieved by offline policies [24], even if they seem to perform well in simulators. The sim-to-real discrepancy exists in different systems, e.g., robotics [46] and networks [29], which is found to be independent to the type of simulators [42].

In this paper, we propose *Atlas*, an integrated offline-online network slicing system, to automate the service configuration of slices, with the following novel designs.

Learning-based simulator. We design a learning-based simulator, whose simulation parameters (e.g., transmission power and operating spectrum of base stations) can be slightly adjusted to match that of real networks for reducing the sim-to-real discrepancy. Simulators are built with domain knowledge to mimic real networks in multiple aspects, e.g., parameters and protocols. We observe that the sim-to-real discrepancy is partially attributed to the inaccurate simulation parameter settings, e.g., radio channel models in simulators are usually simplified and abstracted [48]. Due to the high-dim search space and non-trivial execution time of simulators, conventional searching methods (e.g., exhaustive and grid search) fails under given time periods. We propose a new method based on Bayesian optimization to search for the optimal simulation parameters, which balances the reduction of sim-to-real discrepancy and the explainability of simulation parameters. The method is composed of Bayesian neural networks (BNN) as the approximation function, Thompson sampling for trading off exploration and exploitation, and parallel queries for accelerating the convergence with multiprocessing techniques. The obtained simulation parameters will be set in the simulator, which serves as the offline environment for offline policy training.

Offline policy training. We propose a novel offline training algorithm to automatically learn to configure while assuring the slice SLA by interacting with the augmented simulator. Due to the lack of prior models for heterogeneous slices, we exploit BNN to approximate the complex correlation between resource configurations and slice performances. To assure the slice SLA, we design an adaptive penalization method to incorporate the weighted constraint into the objective with a dynamic multiplier. As the interdependencies between consecutive configurations are weakened by large intervals [30, 42], e.g., 1 hour, we adopt Bayesian optimization with parallel Thompson sampling to search the optimal offline policy that is utilized to accelerate online learning.

Online policy learning. We propose a novel online learning algorithm to safely and sample-efficiently learn the optimal online policy and resolve the sim-to-real discrepancy within real networks. First, we adopt the sample-efficient Gaussian process as the approximation function, and use it to learn the sim-to-real discrepancy only, which is simpler than the whole correlation between resource configurations and slice performances. Second, we design a conservative acquisition function that achieves safe exploration with the guarantee of Bayesian regret bound. Third, we exploit the offline simulator to augment online transitions and update the multiplier for accelerating the online learning progress.

Contributions. To the best of our knowledge, *Atlas* is the first online network slicing system that automatically learn to configure while assuring the slice SLA. The specific contributions of *Atlas* are summarized as follows:

Performance metric	Simulator	Real Network
Average Ping Delay	34 ms	34.6 ms
UL Throughput	19.87 Mbps	17.53 Mbps
DL Throughput	32.37 Mbps	31.12 Mbps
UL Packet Error Rate	4.16E-3	9.17E-3
DL Packet Error Rate	2.05E-3	5.15E-3

Table 1: Network performance comparison (10 MHz LTE)

- We design a new parameter searching method (Sec. 4) to automatically search the simulation parameters for offline simulators to reduce the sim-to-real discrepancy.
- We design a novel offline training algorithm (Sec. 5) to automate the service configuration of slices in the augmented simulator.
- We design a novel online learning algorithm (Sec. 6) to safely and sample-efficiently learn the online policy and resolve the sim-to-real discrepancy within real networks.
- We implement *Atlas* on an end-to-end network prototype (Sec. 7) and conduct extensive experiments to evaluate *Atlas* in terms of performance assurance and sample-efficiency (Sec. 8).

2 MOTIVATION

In this section, we build a network simulator and a system prototype, evaluate the sim-to-real discrepancy in different aspects, and demonstrate the safety and sample-efficiency in online learning.

Setup. We build a system prototype to achieve end-to-end slicing with a smartphone (OnePlus 9), an eNB (OpenAirInterface [35] with Ettus USRP frontend), a SDN switch (OpenDayLight [32]), and a core network (OpenAir-CN [34]). We build a network simulator by using Network Simulator 3 (NS-3) [10], which includes mobile users, an eNB in radio access networks, a backhaul link in transport networks, and the EPC core networks. A slice application is developed in both the system prototype (Android platform) and the network simulator (emulated traffic). The simulation parameters of the simulator are matched to that of system prototype, e.g., wireless spectrum and bandwidth, user-eNB distance, application traffic and service queue. More implementation details refer to Sec. 7.

Sim-to-Real Discrepancy. We show the sim-to-real discrepancy from three perspectives, i.e., networking performances, and application performances under different user traffic and resource configurations. First, we measure networking performances in both system and simulator in Table 1. As we can observe, the system achieves slightly lower performances in most metrics, e.g., ping delay and UL packet error rate (PER). In particular, the UL and DL throughput of the system is 11.8% and 3.9% lower than that of the simulator. These discrepancies may be attributed to a variety of factors, e.g., radio channels and completion of open-source codes.

Second, we measure application performances under different user traffic. Fig. 2 shows the empirical cumulative probability function (CDF) of application latency when there is one user in the network, where the average latency in the system is 25.2% higher than that in the simulator. Besides, Fig. 3 shows the statistics of application latency under different user traffic. As we can see, the discrepancy between the simulator and the system becomes larger, e.g., mean and variance, when user traffic is increased.

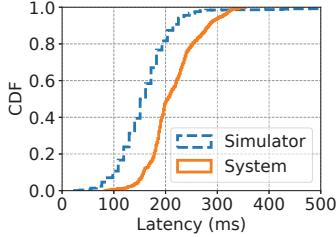


Figure 2: End-to-end latency under one slice user

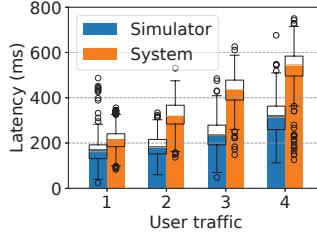


Figure 3: End-to-end latency under different user traffic

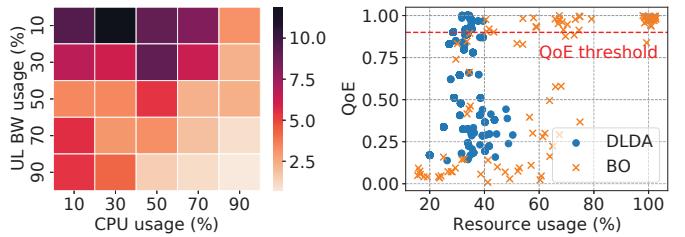


Figure 4: Heatmap of KL-divergence

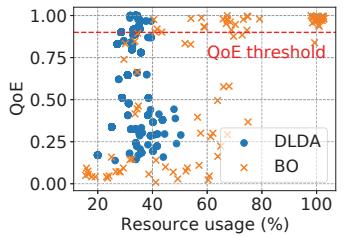


Figure 5: Footprint of online learning methods

Third, we measure application performances under different resource configurations of the slice, e.g., UL bandwidth and CPU. Here, we use the KL-divergence [20] to evaluate the difference between the distribution of application latency collected in the system and simulator. Fig. 4 shows that KL-divergence can be more than 10 under certain resource configurations, which can be interpreted as the two distributions are significantly different. We notice that KL-divergence under different traffic and resource configurations are not the same, which implies uneven sim-to-real discrepancy.

From these measurements, we observe that the sim-to-real discrepancy exists in different aspects, which is non-trivial and uneven. The sim-to-real discrepancy could compromise the performance of offline policies in real networks, which needs to be resolved with different approaches, e.g., online learning.

Safety in Online Learning. We show the performance of two state-of-the-art online learning solutions, i.e., DLDA and Bayesian Optimization (BO), in terms of safety and sample efficiency. As DLDA [42] is originally designed for configuring wireless mesh networks, we modify its inputs and outputs to manage the service configuration. Fig. 5 shows the footprint of quality of experience (QoE²) (see Eq. 6) and resource usage of the slice during online learning, where the QoE requirement is 0.9. We see that both solutions can find multiple configuration actions that strike the balance between resource usage and QoE. However, most configuration actions explored by these solutions fail to meet the QoE requirement, which needs to be avoided during online learning. Therefore, it is imperative to design a safe and sample-efficient approach to resolve the sim-to-real discrepancy via online learning for network slicing.

3 SYSTEM OVERVIEW

The Atlas system includes three integrated stages, i.e., learning-based simulator, offline training, and online learning.

In the learning-based simulator stage, we use the Bayesian optimization framework to search the optimal simulation parameters for the simulator for reducing the sim-to-real discrepancy. In particular, we design a Bayesian neural network (BNN) to approximate the complex correlation between simulation parameters and the measured sim-to-real discrepancy. Besides, we design parallel Thompson sampling (PTS) to balance the exploration and exploitation, where the multiprocessing technique is leveraged to accelerate

the searching progress. The obtained optimal simulation parameters under the given real-world time (e.g., 1 hour), will be used in the augmented simulator in the following stages.

In the offline training stage, we aim to derive the optimal offline configuration policy to configure the cross-domain network resources to individual slices with the minimum resource usage. To assure the QoE requirement of slices, we design an adaptive penalization method to incorporate the constraint into the objective with a dynamic multiplier. As resource configurations are not actually implemented in real networks in this stage, intermediate violations of slice SLA are not concerned. Hence, we train the policy based on the Bayesian optimization framework with the BNN-based approximation function, where the optimistic exploration of PTS helps to achieve better performance via offline querying the augmented simulator. The obtained policy, e.g., the trained BNN, will serve as the start point and offline estimation in the online learning stage.

In the online learning stage, we aim to derive the optimal online configuration policy that resolves the sim-to-real discrepancy. As every network configuration is implemented in real networks, the safety and sample efficiency become the key concerns. Thus, we design a Gaussian process (which is sample efficient) to approximate the sim-to-real discrepancy only (which is simpler), rather than the whole correlation between resource configurations and slice QoEs. To assure the slice SLA during intermediate explorations, we design a conservative acquisition function that achieves the Bayesian regret bound. In addition, we design to update the multiplier in the augmented simulator, instead of based on limited online interactions, which accelerates the convergence of online learning.

4 LEARNING-BASED SIMULATOR STAGE

In this section, we show the design of the learning-based simulator to offline reduce the sim-to-real discrepancy.

4.1 The Problem

Network simulators. Simulators are developed to mimic the setup of real networks by exploiting domain knowledge [10, 43], e.g., protocols, scheduling, and topology. The simulation parameters of the simulator, e.g., link bandwidth, link delay, and pathloss model, are determined according to the corresponding specification or measurement in real networks. However, complex network dynamics, e.g., traffic and mobility, may lead to deviated parameter values and thus compromise the accuracy of parameters. On the other hand, abstraction mechanisms of simulators, e.g., block error rate (BLER)

²To tackle diverse performance metrics of different slices, we define a unified QoE (its value is between zero and one) to represent the slice performance for fair comparison.

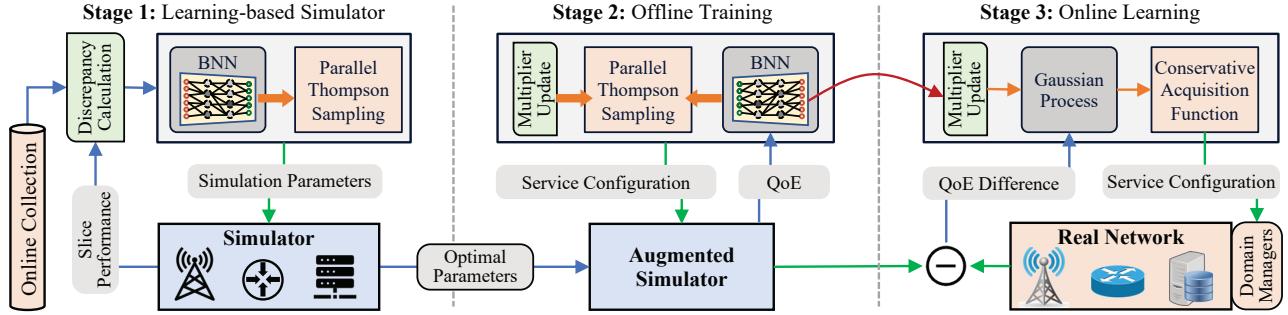


Figure 6: The system overview

mapping [17], do not have exact counterparts in real networks, whose parameters need to be set accurately.

Objective. The objective is to find the optimal simulation parameters within the given parameter space to minimize the sim-to-real discrepancy. We consider there is an online collection \mathcal{D}_r of slice performances (e.g., latency) collected from real networks³. Denote the offline collection $\mathcal{D}_s(x)$ as slice performances generated by the simulator, which is related to simulation parameters x . To evaluate the sim-to-real discrepancy, we resort to KL-divergence [20] to measure the distributional differences between the two collections. Therefore, we formulate the parameter searching problem as

$$\mathbb{P}_0 : \min_{\{x\}} \text{KL}[\mathcal{D}_r || \mathcal{D}_s(x)] \quad (1)$$

$$\text{s.t.} \quad \|x - \hat{x}\|_2 \leq H, \quad (2)$$

where KL is the KL-divergence operator, $\|\cdot\|_2$ is the l_2 -norm operator, and \hat{x} are the original simulation parameters. Here, we denote $\text{KL}[\mathcal{D}_r || \mathcal{D}_s(x)]$ as the sim-to-real discrepancy and $\|x - \hat{x}\|_2$ as the parameter distance. We introduce threshold H to prevent too large parameter distance, which assures the explainability of parameters with respect to original parameters. For instance, the sim-to-real discrepancy may be reduced to nearly zeros when all the link bandwidth are set to 10 times of the original parameter derived from technical specifications. This situation needs to be avoided, as such large parameter distance is apparently infeasible in real networks.

Challenges. The challenge of solving the above problem \mathbb{P}_0 mainly lies in the unknown (i.e., black-box) function $\text{KL}[\mathcal{D}_r || \mathcal{D}_s(x)]$, where there is no closed-form expression for modeling $\mathcal{D}_s(x)$. Without the mathematical models, traditional gradient-based methods [7], e.g., Newton's method, can hardly be applied. Due to the high-dim simulation parameters, conventional searching methods (e.g., exhaustive and grid search) fails in practice, especially under the non-trivial execution time of simulators. In other words, it is needed to design an efficient method to solve the above problem and find the optimal simulation parameters.

4.2 The Solution

We propose an offline parameter searching algorithm based on the Bayesian optimization framework to automate the parameter searching. Our method handles high-dim simulation parameters by

using a Bayesian neural network (BNN) as the approximation function, balances exploitation and exploration via Thompson Sampling, and accelerates searching progress via parallel queries.

Bayesian Optimization. Bayesian optimization [15] is a state-of-the-art global optimization framework, which generally consists of probabilistic surrogate models and acquisition functions. In each iteration, the surrogate model, e.g., Gaussian process [38], is fitted with all existing observations to regress the uncertainty of the black-box function. Then, the selected acquisition function, e.g., expected improvement (EI) and upper confidence bound (UCB) [8], determines the utility of different candidate simulation parameters x for trading off exploration and exploitation. The next simulation parameter is selected by maximizing the acquisition function, whose performance is obtained by querying the simulator. Although Gaussian process (GP) [38] shows promising performance in approximating various black-box functions, its scalability [22] is concerned due to its computation complexity $O(n^3)$, where n is the dimension of data collections. To accurately approximate the complex correlation $\text{KL}[\mathcal{D}_r || \mathcal{D}_s(x)]$, we usually need to collect thousands or more collections from the simulator before the convergence of the Bayesian optimization. The high-volume collections lead to ever-increasing execution time when fitting the GP model, which motivates us to explore more scalable approximation functions.

Bayesian Neural Network. We design the approximation function based on Bayesian neural network (BNN), which is more scalable [44] with competitive performances [18], to approximate the sim-to-real discrepancy ($\text{KL}[\mathcal{D}_r || \mathcal{D}_s(x)]$). Different from standard DNNs that generate only mean-value predictions, BNN introduces stochastic components into neural network architectures, e.g., activation or weights, to quantify the uncertainty of black-box functions. For instance, all the weights in BNNs may be represented by probability distributions, rather than having a single fixed value.

The training of BNNs aim to find the maximum a posteriori (MAP) weights expressed as $\mathbf{w}^{\text{MAP}} = \arg \max_{\mathbf{w}} \log P(\mathbf{w}|\mathcal{Y})$, where \mathbf{w} are the weights of the BNN and \mathcal{Y} is the collection of sim-to-real discrepancy. Based on the Bayes' rule, the computation of posterior needs the prior $P(\mathbf{w})$ and likelihood $P(\mathcal{Y}|\mathbf{w})$, which turns out to be impractical for large neural networks. An alternative way is the variational inference [6], which approximates the complicated posterior $P(\mathbf{w}|\mathcal{Y})$ with a simpler variational approximation, e.g., Gaussian distribution. Thus, the training of BNNs is accomplished by finding the parameter θ on a distribution on the weights $q(\mathbf{w}|\theta)$ [6] that

³We intend to pose minimal collection efforts for network operators. For example, online collections may be collected via logging the current performance achieved by existing deployed methods.

minimizes the KL-divergence between the true Bayesian posterior on the weights, i.e.,

$$\theta^* = \arg \min_{\theta} KL [q(\mathbf{w}|\theta) || P(\mathbf{w})] - \mathbb{E}_{q(\mathbf{w}|\theta)} [\log P(\mathcal{Y}|\mathbf{w})]. \quad (3)$$

Although minimizing the above function is computationally prohibitive, *Bayes-by-Backprop* [6] leverages the trick of re-parameterization and achieve the approximated loss as

$$Loss \approx \sum_{i=1}^N \log q(\mathbf{w}^i|\theta) - \log P(\mathbf{w}^i) - \log P(\mathcal{Y}|\mathbf{w}^i), \quad (4)$$

where \mathbf{w}^i denotes the Monte Carlo sample drawn from the variational posterior $q(\mathbf{w}^i|\theta)$.

Parallel Thompson Sampling. Provided the BNN-based approximation function, we examine conventional acquisition functions, e.g., expected improvement (EI) and probability improvement (PI), and find that the expensive BNN prediction results in time-consuming maximization of acquisition function [52]. The prediction of BNN is accomplished by Monte Carlo sampling (e.g., tens of duplicate inferences), where a small number of samples fail to provide accurate uncertainty estimations.

To this end, we design parallel Thompson sampling (PTS) to address this issue and balance exploration and exploitation in the Bayesian optimization. Thompson sampling [9, 40] is a heuristic, effective and robust method, which samples the approximation function and selects the next querying point with the maximum utility. Here, we extend Thompson sampling to work with BNN-based approximation functions, and achieve parallel offline queries by leveraging the multiprocessing technique. Our basic idea is to draw an estimation of the black-box function (i.e., $KL[\mathcal{D}_r || \mathcal{D}_s(\mathbf{x})]$) by inferring the BNN only once, rather than evaluating the uncertainty via Monte Carlo sampling. To determine the next simulation parameters, we randomly sample tens of thousands of simulation parameters in the given parameter space (Eq. 2). After the one-time BNN inference on these samples, we select the next simulation parameters with the minimum weighted sim-to-real discrepancy (see following explanations), and query its actual performance in the simulator accordingly. Besides, PTS achieves parallel queries by selecting multiple simulation parameters and uses multiprocessing techniques to query the simulator simultaneously. With the parallel queries in PTS, the offline parameter searching can achieve better and more stable convergence performance (see Fig. 13).

Weighted Sim-to-Real Discrepancy. Although we limit the parameter space (i.e., Eq. 2), we still prefer to tradeoff the reduction of sim-to-real discrepancy and the modification of simulation parameters. For example, we prefer to select the simulation parameters with a smaller parameter distance among these candidates who are with the same sim-to-real discrepancy. Hence, we design to penalize the objective with a weighted parameter distance, i.e., the objective in Eq. 1 is rewritten as $KL[\mathcal{D}_r || \mathcal{D}_s(\mathbf{x})] + \alpha \|\mathbf{x} - \hat{\mathbf{x}}\|_2$ (i.e., weighted discrepancy), where α is a non-negative weighting factor. The constraint of parameter space (i.e., Eq. 2) remains.

Remark. In this stage, we design the learning-based simulator and propose a new parameter searching method (summarized in Appendix A) that efficiently finds the optimal simulation parameters to reduce the sim-to-real discrepancy. The augmented simulator (with the optimal simulation parameters) serves as the offline environment for policy training in the following stages.

5 OFFLINE TRAINING STAGE

In this section, we present the design of network configuration policy via offline training in the augmented simulator.

5.1 The Problem

Network slicing. Consider a mobile network operator aims to support a new slice requested by a slice tenant⁴, where the slice requires multiple dedicated virtual network resources in different technical domains, e.g., RAN, TN, and edge computing. The slice tenant makes a service-level agreement (SLA) with the network operator, where the SLA defines several key requirements of its service performance, e.g., latency, reliability and availability. We consider the network can be discretely configured in a time-slotted manner [30], e.g., every hour. As large configuration intervals weaken the temporal dependencies among consecutive configurations, the slice performances are mainly dependent on the current configurations [42]. The network operator can obtain network states (e.g., user traffic) and service performance of the slice at the beginning and end of each configuration interval, respectively.

Objective. We aim to derive the optimal offline policy that minimizes the network resource usage under the performance constraint of the slice by dynamically configuring the slice's resources. Denote \mathbf{a}_t as the network configuration of the slice at the time t (e.g., UL and DL bandwidth), \mathbf{s}_t as the network state (e.g., traffic), and $y(\mathbf{a}_t|\mathbf{s}_t)$ as the service performance, where $y(\cdot)$ is the unknown performance function. Therefore, we formulate the network configuration problem at the time t as follow

$$\mathbb{P}_1 : \min_{\phi} F(\phi) \quad (5)$$

$$s.t. \quad Pr(y_{\phi}(\mathbf{a}_t|\mathbf{s}_t) \geq Y) \geq E, \quad (6)$$

$$0 \leq \mathbf{a}_t \leq A, \quad (7)$$

where ϕ is the policy and A is the maximum allowable configurations, e.g., total bandwidth. The resource usage function $F(\phi) = \|\mathbf{a}_t/A\|_1$, where $|\cdot|_1$ is l_1 -norm, is developed to combine different kinds of resources, without loss of generality. The constraints [41] in Eq. 6 assure that the service performance of slices are better than the predefined threshold Y with a higher probability $E \in [0, 1]$.

Quality of Experience. We denote $Q_s(\phi) = Pr(y_{\phi}(\mathbf{a}_t|\mathbf{s}_t) \geq Y)$ as the QoE of the slice obtained in the augmented simulator. Hence, the slice SLA is assured only if its requirement E is satisfied in Eq. 6.

Challenges. The challenges of resolving the above problem \mathbb{P}_1 lie in two aspects. First, the QoE function $Q_s(\phi)$ is unknown, where offline service performance of slices can only be obtained by executing the simulator. Second, the QoE requirement needs to be maintained under different states \mathbf{s}_t , where the weighting method with fixed weights fails to achieve this constraint.

5.2 The Solution

We propose an offline network configuration algorithm to automatically learn to configure while assuring the slice SLA. First, we design to adaptively incorporate the constraint into the objective based on the Lagrangian primal-dual method [7]. Second, we leverage the Bayesian optimization framework to solve the relaxed

⁴Atlas focuses on service configuration of individual slices, and can be extended to multiple slices scenarios because of the performance and functional isolation.

problem, where BNN is used to approximate the unknown QoE function and parallel Thompson sampling is also exploited.

Adaptive Penalization Method. The idea is to convert the constrained problem \mathbb{P}_1 into an unconstrained problem, by adaptively penalizing the objective with the weighted constraint. To use the Lagrangian primal-dual method, we first build Lagrangian [7] as

$$\mathcal{L}(\mathbf{a}_t, \lambda) = F(\phi) - \lambda(Q_s(\phi) - E), \quad (8)$$

where λ is the multiplier. Then, the problem is resolved by alternatively solving the primal problem written as $\mathbf{a}_t^* = \arg \min_{\mathbf{a}_t} \mathcal{L}(\mathbf{a}_t, \lambda)$, and the dual problem $\lambda^* = \arg \min_{\lambda \geq 0} \mathcal{L}(\mathbf{a}_t, \lambda)$. The dual problem is solved by updating the multiplier with sub-gradient descent [7] as

$$\lambda = [\lambda - \varepsilon (Q_s(\phi) - E)]^+, \quad (9)$$

where $[x]^+ = \max(x, 0)$ and ε is a positive step size. In this method, the multiplier λ is increased if the slice SLA is violated, which guides the optimization in the next round.

Learn to Configure. To tackle the unknown QoE function in the primal problem, we leverage the Bayesian optimization framework to obtain the optimal offline configuration policy. In particular, a BNN is created to approximate the unknown QoE function $Q_s(\phi)$, where its inputs include the network state \mathbf{s}_t , threshold Y and network configuration \mathbf{a}_t . The exploitation and exploration are balanced with Thompson sampling, where parallel querying applies to accelerate the convergence.

Remark. In this stage, we design the offline network configuration algorithm (summarized in Appendix B) that derives the optimal offline policy to automatically learn to configure while assuring the slice SLA. The derived offline policy serves multiple purposes in the online learning stage, e.g., start point and offline acceleration.

6 ONLINE LEARNING STAGE

In this section, we present the design of network configuration policy via online learning to resolve the sim-to-real discrepancy.

6.1 The Problem

The objective is to obtain the optimal policy to minimize resource usage while meeting the slice SLA, which is the same as that in the offline training stage. The key difference is that, instead of interacting with the simulator, network configurations are queried directly to real networks in this stage. As a result, the policy safety (i.e., maintaining the slice SLA for every network configuration) and sample efficiency (i.e., the number of needed online transitions for convergence) become two critical considerations. For example, any unsatisfied QoEs in this stage are actually applied to actual slice users, and thus result in SLA violations in real networks. Besides, the algorithm is highly desired to be sample efficient in real networks, e.g., 1K transitions need more than 40 days to collect when the configuration interval is 1 hour. To evaluate the policy safety, we define the regret function with respect to the resource usage $g_n^{(u)}$ and slice QoE $g_n^{(p)}$ at the n th iteration as

$$g_n^{(u)} = \sum_{j=0}^n [F(\phi^j) - F(\phi^*)], \quad (10)$$

$$g_n^{(p)} = \sum_{j=0}^n [\max(Q(\phi^*) - Q(\phi^j), 0)], \quad (11)$$

where ϕ^* is the optimal policy, ϕ^j is the policy at the j th iteration, and $Q(\phi)$ is the slice QoE in real networks.

6.2 The Solution

We propose an online network configuration algorithm to continue learn to configure during the online learning stage, which achieves safe exploration and sample efficient approximation function. Specifically, we design the algorithm with an efficient Gaussian process model for regressing the sim-to-real performance difference, a conservative acquisition function with a Bayesian regret bound, and a offline acceleration method by using the offline simulator.

Learn Sim-to-Real Discrepancy. Gaussian process (GP) is a widely-adopted and generic model for function approximation, which constructs probabilistic models with a variety of kernel functions to regress given data collections. It is much more powerful than traditional parametric models, e.g., linear regression, and more sample efficient than deep neural networks (DNNs) [38].

Hence, we propose to create a GP model⁵ to approximate the sim-to-real performance difference only, i.e., the gap of slice QoE between the simulator and real networks. This is based on two observations. First, a GP model is insufficient to approximate the complicated QoE function of the slice as a whole (see Fig. 20 and Fig. 21). Second, the sim-to-real performance difference is easier to be learned, provided that the BNN-based approximation function in the offline training stage has been trained extensively in the augmented simulator. By using a GP model to approximate the sim-to-real performance difference, the slice QoE function in real networks is written as

$$Q(\phi) = Q_s(\phi) + G(\psi), \quad (12)$$

where $Q_s(\phi)$ is the slice QoE obtained in the simulator, and $G(\psi)$ is the QoE difference learned by the GP model (denoted by ψ).

Conservative Exploration. Existing acquisition functions (e.g., EI and PI) and Thompson sampling practically cause over-exploration and lead to intermediate SLA violations (see Fig. 22). Gaussian process upper confidence bound (GP-UCB) [45], as an acquisition function, achieves a sub-linear regret bound with strong theoretical convergence guarantees by using a hyperparameter β_t to balance exploration and exploitation. However, the hyperparameter is selected to be large to meet the regret bound requirement, which usually leads to excessive violations of slice SLAs during the online learning stage.

To this end, we propose a clipped randomized GP-UCB (cRGPU-UCB) [5] as the acquisition function to assure conservative exploration while guaranteeing the Bayesian regret bound. Specifically, cRGPU-UCB evaluates the utility of configurations via $\mu_t(\mathbf{a}_t) + \sqrt{\beta_t} \cdot \sigma_t(\mathbf{a}_t)$, where $\mu_t(\cdot)$ and $\sigma_t(\cdot)$ is the mean and standard deviation function, respectively. Note that $\mu_t(\cdot)$ and $\sigma_t(\cdot)$ are estimated by using both offline BNN and online GP model (Eq. 12). The hyperparameter β_t is sampled from a distribution, instead of calculating to be fixed in GP-UCB, at each iteration. The hyperparameter is obtained by $\beta_t = \Gamma(\kappa_t, \rho)$, where Γ is the Gamma distribution, and

$$\kappa_t = \log \left((n^2 + 1) / \sqrt{2\pi} \right) / \log (1 + \rho/2), \quad (13)$$

⁵The total number of online transitions is limited to hundreds, which alleviates the concern of GP scalability.

and ρ is a scaling parameter. The distributional hyperparameter β_t allows greater freedom to select smaller β_t , as compared to GP-UCB, which contributes to maintaining the Bayesian regret bounds⁶. As the sampled hyperparameter still can reach up to hundreds, we clip the hyperparameter for conservative exploration in practice. In other words, the actual β_t is expressed as $\beta_t = \text{clip}(\beta_t, 0, B)$, where B is the upper bound of hyperparameter. Here, both ρ and B can be adjusted by individual slice tenants to tradeoff potential performance improvements and possible risks of SLA violations.

Offline Acceleration. To maintain the slice SLA, we use the adaptive penalization method (see Sec.5.2) to dynamically penalize the objective in Eq. 5 for both the offline training and online learning stage. In online learning, the Lagrangian in Eq. 8 is rewritten as

$$\mathcal{L}(\mathbf{a}_t, \lambda) = F(\phi) - \lambda(Q_s(\phi) + G(\psi) - E). \quad (14)$$

In the offline training stage, its low convergence rate [7] is hidden by using parallel queries to collect multiple slice QoEs under different network configurations in simulators. In the online learning stage, however, the single query in real networks causes insufficient updates on the multiplier λ . As a result, inappropriate multipliers usually lead to changing resource usage and frequent violations of slice SLA, e.g., too small multipliers may fail to penalize the objective for assuring the slice SLA.

To this end, we propose to exploit the augmented simulator to update the multiplier for accelerating the convergence of online algorithm in the online learning stage. Our basic idea is to update the multiplier multiple times by estimating the slice QoE with Eq. 12, where the $Q_s(\phi)$ is obtained in the augmented simulator and $G(\psi)$ is predicted by the current online GP model ψ . Hence, the update of multiplier in Eq. 9 is rewritten as

$$\lambda = [\lambda - \varepsilon(Q_s(\phi) + G(\psi) - E)]^+. \quad (15)$$

In other words, the slice QoE obtained from real networks serves more to regress the GP model ψ . As more online transitions are collected, a more accurate GP model can be achieved, which helps to estimate actual sim-to-real performance differences.

Remark. In this stage, we design the online algorithm (summarized in Appendix C) that derives the optimal online policy to resolve the sim-to-real discrepancy via online learning with real networks. The policy is composed of two models, i.e., the offline BNN ϕ learns offline estimation of slice QoE $Q_s(\phi)$ and online GP ψ learns only sim-to-real discrepancy $G(\psi)$.

7 SYSTEM IMPLEMENTATION

In this section, we present implementation of end-to-end slicing prototype (Fig. 7), network simulator and algorithms.

7.1 The System

RAN. We implement the radio access network based on OpenAir-Interface (OAI) [35] with 4G LTE. The eNB operates at band 7 with 10MHz radio bandwidth, i.e., 50 physical resource blocks (PRBs). The RAN is hosted in an Intel i7 desktop with a low-latency kernel of Ubuntu 18.04, which connects an Ettus USRP B210 as the RF front-end. The distance between the eNB antenna and stationary

⁶The detailed description of Bayesian regret bounds can be found in [5, 45].

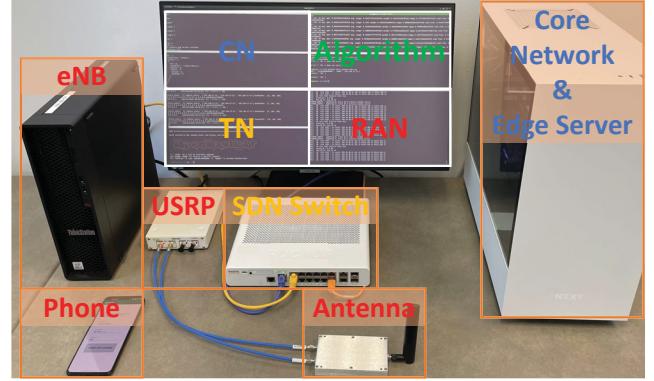


Figure 7: Overview of system prototype

smartphone is 1 meter. To enable the network slicing capability, we develop the radio domain manager with FlexRAN support [13].

UEs. We use a OnePlus 9 5G smartphone (Qualcomm Snapdragon 888 and Android 11) as the user to connect with the eNB. We develop an Android application⁷ for the smartphone, which continuously sends frames (540p) to the edge server in CN. The server processes the frame with a feature extraction algorithm (ORB [39]) and then feeds the results back. We limit the number of on-the-fly frames (i.e., the frames with no results back yet) for the purpose of congestion control. The performance metric of this application is end-to-end latency. To emulate varying user traffic, we control the number of on-the-fly frames, e.g., we may set it as four to emulate the traffic from four users.

TN. We implement the transport network based on OpenDay-Light (ODL) [32] with OpenFlow 1.3. We use a Ruckus ICX 7150-C12P as the SDN switch to connect the eNB and CN, where each port has 1Gbps capacity. To enable the network slicing capability, we develop the transport domain manager by dynamically modifying the link bandwidth with *meters* API in OpenFlow protocol [31].

CN. We implement the core network based on OpenAir-CN [34] with the control and data plane separation architecture (CUPS). The network functions, e.g., HSS, MME, SPGW-C, and SPGW-U, are deployed with Docker containers. To enable the network slicing capability, we develop the core domain manager by mapping users' destination SPGW-U to that of slices. Thus, each slice has an isolated SPGW-U container, while the other network functions are shared.

EN. We implement the edge computing network based on Docker container [33], which virtualizes the computing resources and provides isolation for edge servers. For the sake of simplicity, we co-locate the edge server of slices with their SPGW-U containers. The edge server serves the slice users, where the performances can be retrieved with a minimum 60-second interval via the developed REST API. Besides, we develop the edge domain manager to dynamically manage the CPU ratio of edge servers, via *docker update*.

Configuration Space. We develop 6-dim network configuration actions in both RAN, TN and EN in the prototype, see Table 2.

⁷Atlas makes no prior assumptions on the performance metric of slices and is compatible with other applications.

Configuration	Meaning	Range
<i>bandwidth_ul</i>	maximum uplink PRBs	[0, 50]
<i>bandwidth_dl</i>	maximum downlink PRBs	[0, 50]
<i>mcs_offset_ul</i>	uplink MCS offset [24]	[0, 10]
<i>mcs_offset_dl</i>	downlink MCS offset [24]	[0, 10]
<i>backhaul_bw</i>	transport bandwidth (Mbps)	[0, 100]
<i>cpu_ratio</i>	CPU ratio of docker	[0, 1.0]

Table 2: Network configuration space

Parameters	Meaning
<i>baseline_loss</i>	base loss in pathloss model (dBm)
<i>enb_noise_figure</i>	noise by non-ideal transceivers (dBm)
<i>ue_noise_figure</i>	noise by non-ideal transceivers (dBm)
<i>backhaul_bw</i>	additional transport bandwidth (Mbps)
<i>backhaul_delay</i>	additional transport delay (ms)
<i>compute_time</i>	additional server compute time (ms)
<i>loading_time</i>	additional loading time in UE (ms)

Table 3: Simulation parameter space

7.2 The Simulator

The network simulator is developed based on Network Simulator 3 (NS-3) [10], which is an extensively adopted platform and capable of conducting various network simulations. We develop 7-dim simulation parameters in the NS-3 simulator, see Table 3. The cellular network is developed based on the LENA project with 4G LTE, where we adopt the *LogDistancePropagationLossModel* pathloss model and no fading model. The transport network is simulated by a *p2p* link, whose bandwidth and delay are matched to experimental measurements in the prototype. We develop the edge computing module to allow queue-based computing simulation (matched to the prototype), where the computing delay is sampled from the experimental collections (81ms mean and 35ms std). The Android application is also replicated in NS-3, which matches the traffic pattern and uplink transmission size (28.8kb mean and 9.9kb std). All other settings are thoughtfully examined and configured to match that in the prototype, e.g., MAC scheduler algorithm, antenna type and gain, frequency band, and distance between eNB and smartphones. The simulation results are obtained by reading the tracer including not only end-to-end latency of every frame, but also transmission and computing details, e.g., queuing time, computing time, and uplink and downlink transmission time.

7.3 The Algorithm

We develop BNNs with PyTorch 1.5, where neural networks use 4-layer fully connected layers, i.e., 128x256x256x128, with *ReLU* activation functions [16]. We adopt the *Adadelta* optimizer with the initial learning rate of 1.0, where the learning rate is decayed by using the *StepLR* scheduler with gamma 0.999 and batch size 128.

We develop the GP model using *sklearn* toolkit [37] with the *GaussianProcessRegressor* module. We adopt the *Matern* kernel with $v = 2.5$, which is a generalization of the radial-basis function (RBF) kernel. Besides, target values are normalized by removing the mean and scaling to unit-variance for better regression performance.

To evaluate the computation complexity of *Atlas*, we use the *cProfile* tool in a desktop with AMD Ryzen 5 3600 and 32G DDR4 RAM. In learning-based simulator stage, we obtain the computation time of 22.27s with 214201 calls per iteration, and 0.4GB memory usage. In the offline training stage, we obtain the computation time of 27.23s with 211665 calls per iteration, and 0.5GB memory usage. In the online learning stage, we obtain the computation time of 16.99s with 357292 calls per iteration, and 1.1GB memory usage.

8 PERFORMANCE EVALUATION

In this section, we conduct extensive network simulations and experiments to evaluate the performance of *Atlas*. The step size of dual problem updates in Eq. 15 is $\varepsilon = 0.1$. The scaling factor ρ is 0.1 [5] and the clipping value B is 10 to prevent too large exploration. We determine the fixed weight α is 7 to balance the reduction of sim-to-real discrepancy and the parameter distance. The application related parameters, i.e., $E = 0.9$, and $Y = 300ms$, are set according to capability of the prototype, e.g., RAN throughput. The number of offline and online iterations is 1000 and 100, respectively. The simulation and experimental time of each network configuration are 60 seconds, for collecting statistical performance. We compare *Atlas* with the following solutions.

- *Baseline*. The baseline uses the Bayesian optimization with GP model and expected improvement (EI) acquisition function to online learn in real networks directly.
- *DLDA*. The DLDA [42] is a state-of-the-art online learning solution, which transfers the offline knowledge from a teacher model to the student model via interacting with real networks. As DLDA is designed to improve the prediction accuracy, we modify it to choose the configuration with minimum resource usage while meeting the QoE requirement. This is completed by randomly sampling 10K configurations⁸ from the configuration space.
- *VirtualEdge*. The VirtualEdge [26] uses a GP model to online learn the unknown slice QoE function, and relies on a predictive gradient descent method to update the current configuration under accumulative online interactions.

8.1 Learning-based Simulator

In the learning-based simulator stage, we focus on the metric of sim-to-real discrepancy under given number of iterations.

Searching Progress. Fig. 8 shows the searching progress of different methods, where the first 100 iterations are purely exploration. The average weighted discrepancy is calculated by averaging the weighted sim-to-real discrepancy in each iteration. Our method finds better simulation parameters, which reduce 24.5% average weighted discrepancy than that of the GP-based approach. In the experiments, conducting a 60-second simulation in NS-3 consumes average 27.8 seconds real-world time⁹.

Table 4 shows the details of best simulation parameters obtained by different methods, where the order of simulation parameters refers to Table 3. The original simulator uses the default simulation

⁸As the dimensional of the configuration space is six, we believe 10K samples are sufficient to seek the optimal action under controlled accuracy.

⁹Leveraging the multiprocessing technique, parallel querying of NS-3 with 16 processes achieves nearly the same real-world time with single query.

Methods	Sim-to-Real Discrepancy	Parameter distance	Best simulation parameters
Original Simulator	1.38	0	[38.57, 5.0, 9.0, 0.0, 0.0, 0.0, 0.0]
Aug. Simulator, GP	0.31	0.16	[38.57, 1.44, 7.48, 5.07, 9.23, 6.02, 6.47]
Aug. Simulator, Ours	0.26	0.12	[38.76, 0.68, 8.93, 5.03, 8.93, 2.16, 3.10]

Table 4: Details of offline learning-based simulator

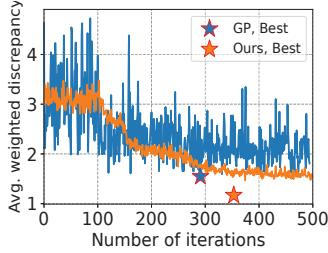


Figure 8: Searching progress under different methods

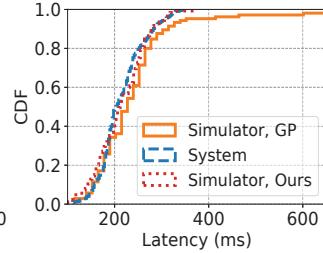


Figure 9: CDF of latency under different methods

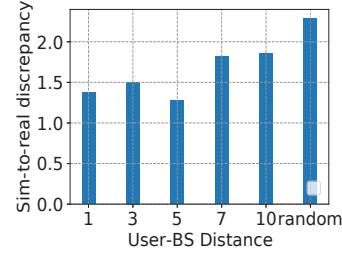


Figure 10: Sim-to-real discrepancy under user mobility

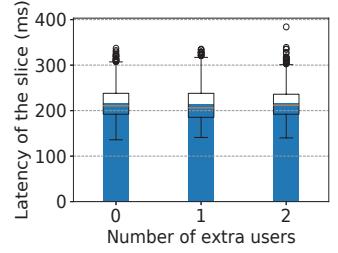


Figure 11: Slice latency under extra mobile users

parameters, where *ReferenceLoss* is 38.57 in the *LogDistancePropagationLossModel* pathloss model in NS-3, and the UE and eNB noise figure are 9.0 and 5.0, respectively. As a result, the original simulator has zero parameter distance, while gets 1.38 sim-to-real discrepancy. The GP-based method obtains 0.31 sim-to-real discrepancy under 0.16 parameter distance, with heavily increased *loading_time*, *backhaul_bw* and *compute_time*. In contrast, our method achieves 0.26 sim-to-real discrepancy (i.e., 81.2% reduction than original simulator) with 0.12 parameter distance. As we calculate the parameter distance with L-2 norm, the actual parameter difference is even smaller, which helps to maintain the explainability of simulation parameters. Fig. 9 shows the cumulative probability of latency under the best simulation parameters obtained by different methods. We can see that the CDF achieved by the GP-based approach has a long tail, which is worse than that of our method. As compared to the CDF of original simulator (see Fig. 2), the sim-to-real discrepancy is substantially reduced with the learning-based simulator.

Network Dynamics. Fig. 10 shows the sim-to-real discrepancy obtained under different user mobility. In general, we see the sim-to-real discrepancy increases under a larger line-of-sight distance between the user and the base station. This may be attributed to the disparity of the radio channel model, where the pathloss model in the simulator fails to represent the real channel dynamics in experiments, especially under random walk scenarios. Besides, to evaluate the end-to-end performance isolation among network slices, we dynamically attach new users, generate traffic (i.e., YouTube), and detach the users. Fig. 11 shows that the latency performance of the slice is very stable no matter how many extra users are in the network. This is attributed to the performance isolation achieved in both RAN, TN, CN, and EN in the developed system prototype.

Pareto Boundary. Fig. 12 shows the Pareto boundary achieved by our method via varying the weight α . Provided that the sim-to-real discrepancy is 0.26 when $\alpha = 7$, it can be further reduced to 0.21 at the cost of more than 0.2 parameter distance. On the other hand, given the maximum 0.1 parameter distance, the lowest sim-to-real

discrepancy can reach 0.4, which still achieves a 71.0% reduction as compared to that of the original simulator. Besides, our method allows the customized weight to balance sim-to-real discrepancy and the parameter distance.

Parallel Queries. Fig. 13 shows the searching progress of our method under different number of parallel queries. When the number of parallel queries is only one, i.e., single query per iteration, the average discrepancy curve is similar to that of the GP-based method (see Fig. 8). As the number of parallel queries increases, our method can achieve lower discrepancy, which suggests the necessity of parallel queries for Thompson sampling to train BNNs.

Discrepancy Reduction. Fig. 14 shows the sim-to-real discrepancy achieved by our method under different user traffic. Here, we emulate a maximum of four mobile users in the slice, due to the limitation of hardware capability and software stability, e.g., constrained uplink throughput, in the system prototype. Note that, the optimal simulation parameters are derived only based on the user traffic 1, and apply to the learning-based simulator for all user traffic simulations. With the simulation parameters in Table 4, the sim-to-real discrepancy is reduced to 81.2%, 56.7%, 43.6%, and 61.6% under the user traffic 1, 2, 3, and 4, respectively. This result suggests that the sim-to-real discrepancy may share some common patterns under different scenarios. Fig. 15 shows that the sim-to-real discrepancy is reduced substantially (79.3% in average) for almost all the resources. Another interpretation of these figures is that, the sim-to-real discrepancy is not identical and not even under different scenarios. As it is impractical to have online collections under all possible scenarios, the sim-to-real discrepancy will still exist between the augmented simulator and real networks.

8.2 Offline Training

In the offline training stage, we evaluate the achieved resource usage and slice QoE of the policy under given number of iterations.

Training Progress. Fig. 16 shows the training progress of our method, including the average resource usage and QoE, where the

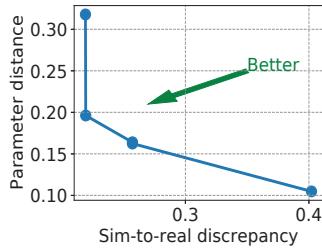


Figure 12: Pareto boundary of augmented simulator

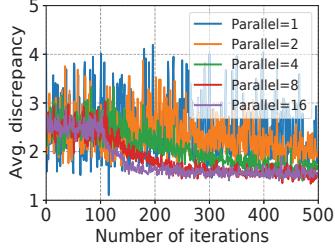


Figure 13: Searching progress with parallel queries



Figure 14: Discrepancy reduction under user traffic

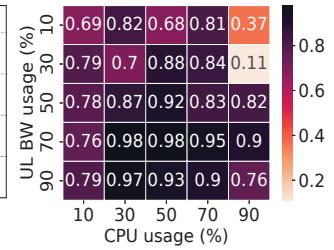


Figure 15: Discrepancy reduction (1.0 means 100%) under resources

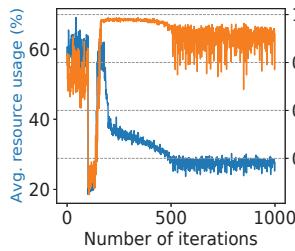


Figure 16: Training progress of our method

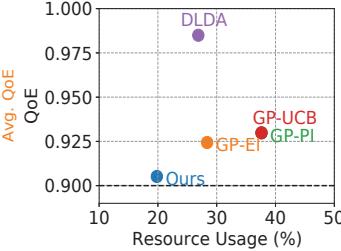


Figure 17: Performance of different methods

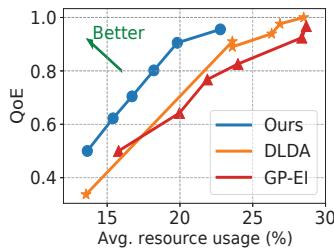


Figure 18: Pareto boundary under different availability

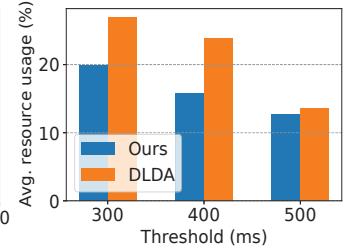


Figure 19: Average usage under different thresholds

first 100 iterations are purely exploration. We observe that resource usage is gradually decreased when the average QoE is above the requirement ($E = 0.9$), e.g., between the iteration 300 to 500. Then, both the average QoE and resource usage converge, where the found optimal policy achieves stable average performance and the oscillations are mainly from Thompson sampling. The total real-world time consumption for 1000 iterations is 7.72 hours, given the average of 27.8 seconds for NS-3 simulations.

Performance Comparison. Fig. 17 depicts the QoE and resource usage under the best network configuration policy obtained by different methods. We see that our method achieves the best performance, where the QoE is 0.905 and the resource usage is 19.81%. When user traffic is 1, the best configuration actions are 9 and 3 uplink and downlink PRBs, 6.2 Mbps backhaul bandwidth and 0.8 CPU ratio at the edge server, where the uplink and downlink MCS offsets are zeros. In contrast, DLDA achieves 0.98 QoE at the cost of 26.87% resource usage, which fails to balance the QoE and the resource usage. The other GP-based methods, e.g., GP-EI and GP-PI, maintain more than 0.92 QoE while using up to 37.62% resources.

Pareto Boundary. Fig. 18 shows the Pareto boundary of different methods, which are obtained by varying the QoE requirement E . It can be seen that our method outperforms the other methods, in both QoE requirement and resource usage, where the slice SLAs on these curves are met. For training the DLDA, we collect the offline dataset with 4096 configuration actions¹⁰ by grid searching the configuration space (Eq. 7), where each dimension is with 4 different values (i.e., $[0.0, 0.3, 0.6, 0.9]$). We notice that DLDA has a huge leap of QoE from 0.33 to 0.89, which may be attributed to the

¹⁰We collect the slice performances in 60 seconds for each configuration action, which consumes approximately 68.5 hours in total.

coarse-grained dataset and implies that the grid searching method fails in handling high-dimensional action spaces.

Slice Requirement. Fig. 19 shows the achieved performance by our method and DLDA under different latency thresholds Y . Our method obtains lower resource usage under all the scenarios with satisfied thresholds. Besides, we see that the difference in resource usage between our method and DLDA shrinks when the threshold increases. This is because we set a minimum of 6 uplink and 3 downlink PRBs for maintaining radio connectivities of users, where these resources may be sufficient to satisfy the loosen thresholds.

8.3 Online Learning

In the online learning stage, we evaluate the regret of resource usage and slice QoE under given number of iterations.

Training Progress. Fig. 20 and Fig. 21 show average resource usage and slice QoE achieved during 100 online interactions, respectively. The very first online configuration action is the optimal one obtained in the offline learning stage, if applicable, for all methods. Recall that our method achieved 0.905 QoE in the augmented simulator (see Fig. 17), it turns out to be 0.65 QoE in real networks, which implies the noticeable sim-to-real discrepancy.

We see our method searches for the optimal online policy under 40% resource usage, while the resulted slice QoEs are around the requirement ($E = 0.9$). As compared to other methods in Table 5, our method achieves the lowest regret for both average resource usage (63.9% reduction than DLDA) and slice QoE (85.7% reduction than DLDA) in 100 iterations. In other words, our method uses only 3.16% more resource usage and obtains 0.077 less slice QoE on average, as compared to the best policy in 100 iterations. This result validates the efficacy of our method in terms of policy safety and sample

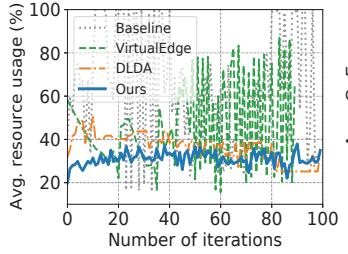


Figure 20: Avg. usage under different methods

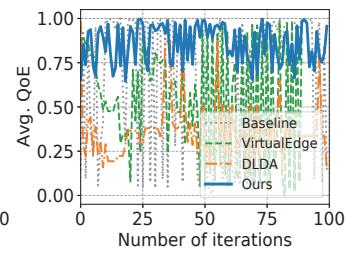


Figure 21: Avg. QoE under different methods

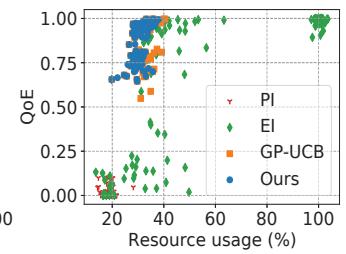


Figure 22: Our method with different acquisition functions

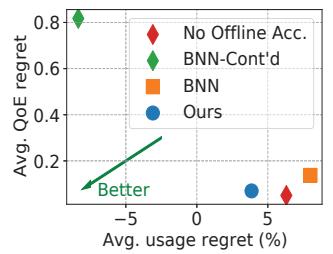


Figure 23: Our method with different online models

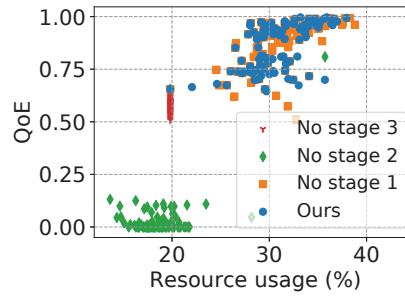


Figure 24: Impact of individual components in our method

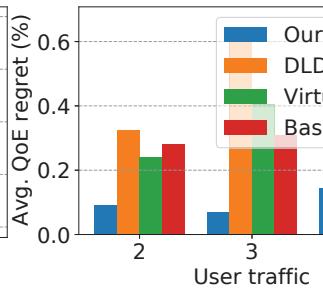


Figure 25: Avg. QoE regret under different user traffic

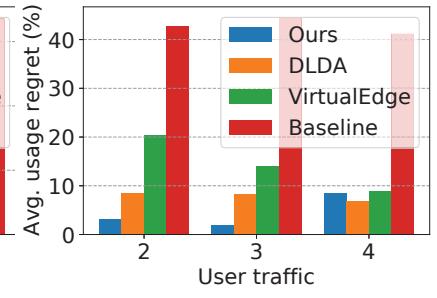


Figure 26: Avg. usage regret under different user traffic

Methods	Avg. usage regret (%)	Avg. QoE regret	Offline queries
Baseline	35.83	0.31	0
VirtualEdge	16.06	0.34	0
DLDA	8.79	0.54	0
Ours	3.17	0.077	20×100

Table 5: Details of online learning under different methods

efficiency. Although our method needs $N = 20$ offline queries with the augmented simulator after each online action, it does not affect and delay the behavior of online network configuration, especially when configuration intervals are usually tens of minutes or even hours in real networks [30, 42].

Acquisition Function. Fig. 22 shows the footprint of our method under different acquisition functions, which is obtained by scattering the achieved resource usage and slice QoE in each iteration. Our proposed conservative acquisition function (Eq. 13) outperforms classic acquisition functions, which explores configuration actions with lower resource usages while approximating the slice QoE requirement. Also, we see that GP-UCB has nearly comparative performance, which usually uses more network resources to satisfy the slice QoE requirement. As the only difference is acquisition function, this result justifies the safe exploration of Atlas.

Approximation Function. Fig. 23 shows the performance of our method under different approximation functions, e.g., BNN and GP. When using BNN to approximate the sim-to-real performance difference ($G(\psi)$ in Eq. 12), we see the regret of resource usage and

slice QoE is increased by 107.6% and 96.5%, respectively. This can be attributed to the poor sample efficiency of BNN, as 100 online collections are insufficient to train the BNN. Alternatively, if we use the offline trained BNN to continue learning in the online learning stage (i.e., BNN-Cont'd), we see the average QoE regret soared. Besides, if we do not exploit the offline acceleration, we observe average usage regret is increased by 63.5% consequently.

Individual Components. Fig. 24 shows the footprint of our method when individual stages are absent. When there is no online learning stage, we see the resource usage keep constant while the QoE is around 0.65, where the sim-to-real discrepancy remains. When there is no offline training stage, all the policies have to be learned via directly interacting with real networks. As a result, it is no surprise that the early-stage performance is poor, which justifies the importance of offline training. Without the learning-based simulator, the QoE performance is worsen due to the large sim-to-real discrepancy between original simulator and real networks.

Dynamic Traffic. Fig. 25 and Fig. 26 show the average slice QoE and resource usage under different user traffic, where the latency threshold $Y = 500ms$. We see our method achieves the lowest regret of both resource usage and slice QoE for almost all user traffic. Although DLDA has a slightly lower regret of average resource usage when user traffic is 4, its average QoE regret is much higher than that of ours.

9 RELATED WORK

Network Slicing. Network slicing systems have been increasingly studied to support heterogeneous applications in terms of flexibility and cost-efficiency [4, 11, 13, 26, 41]. Foukas *et. al.* [12] designed

Orion as the first RAN slicing system based on FlexRAN [13], which enables on-the-fly RAN virtualization with both performance and functional isolation. Marquez *et. al.* [30] empirically demonstrated that dynamic resource orchestration improves the efficiency of resource multiplexing in RAN. Liu *et. al.* [26] proposed VirtualEdge that automatically learns to orchestrate cross-domain resources for maximizing the performance of slices. Salvat *et. al.* [41] proposed to multiplex network resources via overbooking for concurrently supporting multiple slices with two new resource provisioning algorithms. D’Oro *et. al.* [11] proposed SI-EDGE that enables joint network-edge slicing, where near-optimal algorithms are designed to instantiate slices under constrained resources in edge nodes. These works, however, mainly focused on network slicing in individual domains with known slice resource demands, which do not apply to network configuration in end-to-end slicing systems.

Machine learning for Networking: Recent advances in machine learning techniques provoke the interest in applying ML in handling complex networking problems [1, 21, 23, 50, 53]. Microscope [54] introduced an ML-based decomposition method to efficiently estimate the service demand of slices, which deals with complex spatiotemporal features in aggregated traffic. EdgeSlice [27] handled the multi-user and multi-domain resource orchestration problems by using a decentralized deep reinforcement learning (DRL) algorithm for network slicing. The constrained resource allocation problems are studied in network slicing [19, 27], where different techniques, e.g., reward shaping and interior-point method [28], are utilized to maintain the performance requirement of slices. These works, however, primarily focused on the offline training and online deployment strategy, which suffer the sim-to-real discrepancy when applying offline policies to real networks.

Sim-to-Real Discrepancy. The sim-to-real discrepancy has been increasingly unveiled for online system design. Mao *et. al.* [29] identified the non-negligible performance difference between the high-fidelity simulator and real data center networks under the identical policy. OnRL [55] is proposed to address the sim-to-real discrepancy in real-time mobile video telephony, via online DRL within real networks, where an individualized hybrid learning method is designed to counter the potential performance degradation. OnSlicing [24] resolved the cross-domain resource orchestration problem in online network slicing by designing a proactive baseline switching mechanism for near-zero SLA violations. These works focus on time-correlated problems with small time scales (e.g., seconds or minutes) and unfortunately fail in network configuration problems with large time scales. Shi *et. al.* [42] designed a transfer learning DNN-based algorithm to bridge the sim-to-real gap by using both offline and online datasets (obtained by grid searching), which fails in terms of performance assurance and sample efficiency.

10 DISCUSSION

In this section, we discuss the scalability, adaptability and generalizability of the proposed Atlas system.

Scalability. As Atlas tackles the service configuration of individual network slices, it can seamlessly accommodate dynamic slice admission and removal. When a new slice is admitted and launched, an individualized Atlas will be initialized to build its learning-based simulator, offline train the policy, and online learn to configure

continually. As there are infrastructure changes, e.g., installing more antennas and higher bandwidth switches, the corresponding parts in the learn-based simulator will be updated. The simulation parameters will be continually searched based on its last optima if needed. The offline policies will be fine-tuned in the updated simulator, and the online learning stage continues uninterruptedly as the GP model learns only the sim-to-real discrepancy. All these procedures will be offline executed and accelerated if applicable, without the need for additional online transitions.

Adaptability. For parameter and configuration space changes, Atlas can reuse previous experience (e.g., buffered transitions) to accelerate the training and finetuning of new approximation functions (i.e., BNNs and GPs). If network dynamics are observable, e.g., spatiotemporal traffic changes, the topology and settings in the simulator can be updated accordingly. Then, the learning components, e.g., simulation parameters and approximation functions, at the learning-based simulator stage and offline training stage will be refined with the new updated simulator. During the online learning stage, the learned GP model regarding the sim-to-real discrepancy may be further continuously updated with new interactions. If network dynamics are not observable, e.g., instant link failure, Atlas will recognize it as the fluctuations of simulation-to-reality discrepancy and rely on the sample efficient Gaussian process model to learn and adapt continually.

Generalizability. The proposed methods in Atlas have the potential to be generalized to tackle sim-to-real discrepancy and online learning problems in other network systems, e.g., datacenter networks and wireless mesh networks [42]. This is attributed to the minimal assumptions of Atlas, summarized as follows: 1) there is a queryable simulated environment to imitate the real-world system. 2) there is an online dataset collected directly from the real-world system. 3) the real-world system can be queried, where network states and generated results are observable.

11 CONCLUSION

In this paper, we proposed an online network slicing system (i.e., Atlas) to automate the service configuration of slices. We design three interrelated stages in Atlas to tackle the sim-to-real discrepancy for obtaining the optimal network configuration policy in real networks. First, we designed a learning-based simulator to reduce the sim-to-real discrepancy with a new parameter searching method. Second, we designed a novel offline training algorithm to train the policy in the augmented simulator. Third, we designed a novel online learning algorithm to safely learn the policy and resolve the sim-to-real discrepancy with real networks. We develop an end-to-end network prototype and evaluated Atlas in terms of performance assurance, sample efficiency, and resource usage.

ACKNOWLEDGMENTS

We would like to thank the anonymous shepherd and reviewers for their insightful comments and valuable guidance in refining this paper.

This work is partially supported by the US National Science Foundation under Grant No. 2212050, No. 2147624, No. 2147821, and No. 2147623.

REFERENCES

[1] Shivang Aggarwal, Urjit Satish Sardesai, Viral Sinha, Deen Dayal Mohan, Moinak Ghoshal, and Dimitrios Koutsonikolas. 2020. LiBRA: learning-based link adaptation leveraging PHY layer information in 60 GHz WLANs. In *Proceedings of the 16th International Conference on emerging Networking EXperiments and Technologies*. 245–260.

[2] GSM Association. 01 June 2021. Generic Network Slice Template, Version 5.0. <https://www.gsmarena.com/newsroom/wp-content/uploads//NG.116-v5.0-7.pdf>

[3] Idris Badmus, Marja Matimikko-Blue, and Jaspreet Singh Walia. 2019. Network slicing management technique for local 5G micro-operator deployments. In *2019 16th International Symposium on Wireless Communication Systems (ISWCS)*. IEEE, 697–702.

[4] Dario Bega, Marco Gramaglia, et al. 2019. A Machine Learning approach to 5G Infrastructure Market optimization. *IEEE Transactions on Mobile Computing* (2019).

[5] Julian Berk, Sunil Gupta, Santu Rana, and Svetha Venkatesh. 2020. Randomised gaussian process upper confidence bound for bayesian optimisation. *arXiv preprint arXiv:2006.04296* (2020).

[6] Charles Blundell, Julien Cornebise, Koray Kavukcuoglu, and Daan Wierstra. 2015. Weight uncertainty in neural network. In *International conference on machine learning*. PMLR, 1613–1622.

[7] Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. 2004. *Convex optimization*. Cambridge university press.

[8] Eric Brochu, Vlad M Cora, and Nando De Freitas. 2010. A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *arXiv preprint arXiv:1012.2599* (2010).

[9] Olivier Chapelle and Lihong Li. 2011. An empirical evaluation of thompson sampling. *Advances in neural information processing systems* 24 (2011).

[10] A discrete event network simulator for internet systems. 2022 [Online]. NS-3.

[11] Salvatore D’Oro, Leonardo Bonati, et al. 2020. SI-EDGE: Network Slicing at the Edge. *arXiv preprint arXiv:2005.00886* (2020).

[12] Xenofon Foukas, Mahesh K Marina, and Kimon Kontovasilis. 2017. Orion: RAN slicing for a flexible and cost-effective multi-service mobile network architecture. In *Proceedings of the 23rd annual international conference on mobile computing and networking*. 127–140.

[13] Xenofon Foukas, Navid Nikaein, Mohamed M Kassem, Mahesh K Marina, and Kimon Kontovasilis. 2016. FlexRAN: A flexible and programmable platform for software-defined radio access networks. In *Proceedings of the 12th International on Conference on emerging Networking EXperiments and Technologies*. 427–441.

[14] Xenofon Foukas, Georgios Patounas, et al. 2017. Network slicing in 5G: Survey and challenges. *IEEE Communications Magazine* 55, 5 (2017), 94–100.

[15] Peter I Frazier. 2018. A tutorial on Bayesian optimization. *arXiv preprint arXiv:1807.02811* (2018).

[16] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep learning*. MIT press.

[17] Josep Colom Ikuno, Martin Wirllich, and Markus Rupp. 2010. System level simulation of LTE networks. In *2010 IEEE 71st Vehicular Technology Conference*. IEEE, 1–5.

[18] Laurent Valentin Jospin, Hamid Laga, Farid Boussaid, Wray Buntine, and Mohammed Bennamoun. 2022. Hands-on Bayesian neural networks—A tutorial for deep learning users. *IEEE Computational Intelligence Magazine* 17, 2 (2022), 29–48.

[19] Sami Khairy, Prasanna Balaprakash, Lin X Cai, and Yu Cheng. 2020. Constrained deep reinforcement learning for energy sustainable multi-UAV based random access IoT networks with NOMA. *IEEE Journal on Selected Areas in Communications* 39, 4 (2020), 1101–1115.

[20] Solomon Kullback. 1997. *Information theory and statistics*. Courier Corporation.

[21] Jinsung Lee, Sungyong Lee, Jongyun Lee, Sandesh Dhawaskar Sathyanarayana, Hyoyoung Lim, Jihoon Lee, Xiaoqing Zhu, Sangeeta Ramakrishnan, Dirk Grunwald, Kyunghan Lee, et al. 2020. PERCEIVE: deep learning-based cellular uplink prediction using real-time scheduling patterns. In *Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services*. 377–390.

[22] Haitao Liu, Yew-Soon Ong, Xiaobo Shen, and Jianfei Cai. 2020. When Gaussian process meets big data: A review of scalable GPs. *IEEE transactions on neural networks and learning systems* 31, 11 (2020), 4405–4423.

[23] Qiang Liu, Nakjung Choi, and Tao Han. 2021. Constraint-Aware Deep Reinforcement Learning for End-to-End Resource Orchestration in Mobile Networks. In *2021 IEEE 29th International Conference on Network Protocols (ICNP)*. IEEE, 1–11.

[24] Qiang Liu, Nakjung Choi, and Tao Han. 2021. OnSlicing: online end-to-end network slicing with reinforcement learning. In *Proceedings of the 17th International Conference on emerging Networking EXperiments and Technologies*. 141–153.

[25] Qiang Liu and Tao Han. 2018. Dare: Dynamic adaptive mobile augmented reality with edge computing. In *2018 IEEE 26th International Conference on Network Protocols (ICNP)*. IEEE, 1–11.

[26] Qiang Liu and Tao Han. 2019. VirtualEdge: Multi-domain resource orchestration and virtualization in cellular edge computing. In *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 1051–1060.

[27] Qiang Liu, Tao Han, and Ephraim Moges. [n. d.]. EdgeSlice: Slicing wireless edge computing network with decentralized deep reinforcement learning. In *2020 IEEE 40th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 234–244.

[28] Yongshuai Liu, Jiaxin Ding, and Xin Liu. 2020. A Constrained Reinforcement Learning Based Approach for Network Slicing. In *2020 IEEE 28th International Conference on Network Protocols (ICNP)*. IEEE, 1–6.

[29] Hongzi Mao, Malte Schwarzkopf, et al. 2019. Learning scheduling algorithms for data processing clusters. In *Proceedings of the ACM Special Interest Group on Data Communication*. ACM, 270–288.

[30] Cristina Marquez, Marco Gramaglia, Marco Fiore, Albert Banchs, and Xavier Costa-Perez. 2018. How should I slice my network? A multi-service empirical evaluation of resource sharing efficiency. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*. 191–206.

[31] Nick McKeown, Tom Anderson, Hari Balakrishnan, Guru Parulkar, Larry Peterson, Jennifer Rexford, Scott Shenker, and Jonathan Turner. 2008. OpenFlow: enabling innovation in campus networks. *ACM SIGCOMM computer communication review* 38, 2 (2008), 69–74.

[32] Jan Medved, Robert Varga, et al. 2014. Opendaylight: Towards a model-driven SDN controller architecture. In *IEEE WoWMoM 2014*. IEEE, 1–6.

[33] Dirk Merkel. 2014. Docker: lightweight linux containers for consistent development and deployment. *Linux journal* 2014, 239 (2014), 2.

[34] OpenAirInterface Software Alliance. Openair-cn repository. <https://gitlab.eurecom.fr/oi/openair-cn>. 2017.

[35] OpenAirInterface Software Alliance. OpenAirInterface repository. <https://gitlab.eurecom.fr/oi/openairinterface5g>. 2017.

[36] Lei Pang, Chungang Yang, Danyang Chen, Yanbo Song, and Mohsen Guizani. 2020. A survey on intent-driven networks. *IEEE Access* 8 (2020), 22862–22873.

[37] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.

[38] Carl Edward Rasmussen. 2003. Gaussian processes in machine learning. In *Summer school on machine learning*. Springer, 63–71.

[39] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. 2011. ORB: An efficient alternative to SIFT or SURF. In *2011 International conference on computer vision*. Ieee, 2564–2571.

[40] Daniel J Russo, Benjamin Van Roy, Abbas Kazerouni, Ian Osband, Zheng Wen, et al. 2018. A tutorial on thompson sampling. *Foundations and Trends® in Machine Learning* 11, 1 (2018), 1–96.

[41] Josep Xavier Salvat, Lanfranco Zanzi, et al. 2018. Overbooking network slices through yield-driven end-to-end orchestration. In *ACM CoNEXT*. ACM, 353–365.

[42] Junyang Shi, Mo Sha, and Xi Peng. 2021. Adapting Wireless Mesh Network Configuration from Simulation to Reality via Deep Learning based Domain Adaptation. In *18th USENIX Symposium on Networked Systems Design and Implementation (NSDI 21)*. 887–901.

[43] Discrete Event Simulator. 2022 [Online]. OMNeT++.

[44] Jasper Snoek, Oren Rippel, Kevin Swersky, Ryan Kiros, Nadathur Satish, Narayanan Sundaram, Mostofa Patwary, Mr Prabhat, and Ryan Adams. 2015. Scalable bayesian optimization using deep neural networks. In *International conference on machine learning*. PMLR, 2171–2180.

[45] Niranjan Srinivas, Andreas Krause, Sham M Kakade, and Matthias Seeger. 2009. Gaussian process optimization in the bandit setting: No regret and experimental design. *arXiv preprint arXiv:0912.3995* (2009).

[46] Jie Tan, Tingnan Zhang, Erwin Coumans, Atil Iscen, Yunfei Bai, Danijar Hafner, Steven Bohez, and Vincent Vanhoucke. 2018. Sim-to-real: Learning agile locomotion for quadruped robots. *arXiv preprint arXiv:1804.10332* (2018).

[47] Nguyen H Tran, Wei Bao, Albert Zomaya, Minh NH Nguyen, and Choong Seon Hong. 2019. Federated learning over wireless networks: Optimization model design and analysis. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. IEEE, 1387–1395.

[48] Esa Tuomaala and Haiming Wang. 2005. Effective SINR approach of link to system mapping in OFDM/multi-carrier mobile network. In *2005 2nd Asia Pacific Conference on Mobile Technology, Applications and Systems*. IEEE, 5–pp.

[49] Luis Velasco, Marco Signorelli, Oscar González De Dios, Chrysa Papagianni, Roberto Bifulco, Juan Jose Vegas Olmos, Simon Pryor, Gino Carrozzo, Julius Schulz-Zander, Mehdi Bennis, et al. 2021. End-to-End Intent-Based Networking. *IEEE Communications Magazine* 59, 10 (2021), 106–112.

[50] Haoyu Wang, Zetian Liu, and Haiying Shen. 2020. Job scheduling for large-scale machine learning clusters. In *Proceedings of the 16th International Conference on emerging Networking EXperiments and Technologies*. 108–120.

[51] Jiadai Wang, Jiajia Liu, and Nei Kato. 2018. Networking and communications in autonomous driving: A survey. *IEEE Communications Surveys & Tutorials* 21, 2 (2018), 1243–1274.

[52] James Wilson, Frank Hutter, and Marc Deisenroth. 2018. Maximizing acquisition functions for Bayesian optimization. *Advances in neural information processing systems* 31 (2018).

[53] Francis Y Yan, Hudson Ayers, Chenzhi Zhu, Sadjad Fouladi, James Hong, Keyi Zhang, Philip Levis, and Keith Winstein. 2020. Learning in situ: a randomized experiment in video streaming. In *17th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 20)*. 495–511.

[54] Chaoyun Zhang, Marco Fiore, Cezary Ziemlicki, and Paul Patras. 2020. Microscope: mobile service traffic decomposition for network slicing as a service. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. 1–14.

[55] Huanhua Zhang, Anfu Zhou, Jiamin Lu, Ruoxuan Ma, Yuhua Hu, Cong Li, Xinyu Zhang, Huadong Ma, and Xiaojiang Chen. 2020. OnRL: improving mobile video telephony via online reinforcement learning. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. 1–14.

A PARAMETER SEARCHING METHOD

The parameter searching algorithm for learning-based simulator is summarized in Alg. 1. First, tens of thousands of simulation parameters are sampled from the parameter space (i.e., Eq. 2). Second, the next query is selected by minimizing the weighted discrepancy, where the sim-to-real discrepancy $KL[\mathcal{D}_r || \mathcal{D}_s(\mathbf{x})]$ is estimated by the BNN for all the sampled parameters. Third, the actual values of sim-to-real discrepancy is obtained via querying the offline simulation. Fourth, the transition $\langle \mathbf{x}, KL[\mathcal{D}_r || \mathcal{D}_s(\mathbf{x})] \rangle$ are stored for BNN training. Note that these above steps are executed in parallel with multiprocessing technique.

Algorithm 1: The Parameter Searching Method

Input: $H, \alpha, \hat{\mathbf{x}}$

```

1 while True do
2   for  $n = 0, 1, \dots$  (Parallel) do
3     Sample  $\mathbf{x}$ ;
4     Estimated sim-to-real discrepancy  $KL[\mathcal{D}_r || \mathcal{D}_s(\mathbf{x})]$ 
        from BNN for all samples;
5     Get the next query
6      $\mathbf{x} \leftarrow \arg \min (KL[\mathcal{D}_r || \mathcal{D}_s(\mathbf{x})] + \alpha |\mathbf{x} - \hat{\mathbf{x}}|_2)$ ;
7     Obtain sim-to-real discrepancy  $KL[\mathcal{D}_r || \mathcal{D}_s(\mathbf{x})]$  by
        querying simulator;
8     Store transition  $\langle \mathbf{x}, KL[\mathcal{D}_r || \mathcal{D}_s(\mathbf{x})] \rangle$ ;
9     Train the BNN with new added transitions;
10    if Convergence then
11      break;

```

B THE OFFLINE ALGORITHM

The offline network configuration algorithm is summarized in Alg. 2. First, tens of thousands of network configuration actions are sampled from the configuration space (i.e., Eq. 7). Second, the next query is selected by minimizing the Lagrangian, where the resource usage is calculated in $F(\phi)$ and the slice QoE is estimated by the BNN for all the sampled actions. Third, the actual values of slice QoE is obtained via querying the offline simulation. Fourth, the transition $\langle \mathbf{a}_t, F(\phi), Q_s(\phi) \rangle$ are stored for BNN training. Note that these above steps are executed in parallel with multiprocessing technique. Then, the multiplier is updated by averaging the results from parallel queries.

C THE ONLINE ALGORITHM

The online network configuration algorithm is summarized in Alg. 3. First, the multiplier in the online learning stage is initialized with the final multiplier in the offline training stage. Then, we update the multiplier by interacting with the offline simulator for N times (Line 4–10). Specifically, we calculate the Lagrangian for all sampled configuration actions with both the offline BNN and the online GP model. The next network configuration is selected by minimizing the Lagrangian. We update the multiplier with $\lambda \leftarrow$ Eq. 15 by querying offline simulator and estimating GP model. Next, we online query real networks, when it needs to, e.g., configuration

Algorithm 2: The Offline Algorithm

Input: ϵ, E, Y, A ,

- 1 Initialize parameters $\lambda \leftarrow 0$;
- 2 **while** *True* **do**
- 3 **for** $n = 0, 1, \dots$ (*Parallel*) **do**
- 4 Sample \mathbf{a}_t , calculate $F(\phi)$ for all samples;
- 5 Estimated $Q_s(\phi)$ from BNN for all samples;
- 6 Calculate Lagrangian in Eq. 8;
- 7 Get the next query $\mathbf{a}_t \leftarrow \arg \min L$;
- 8 Obtain slice QoE by querying the simulator;
- 9 Store transition $< \mathbf{a}_t, F(\phi), Q_s(\phi) >$;
- 10 Train the BNN with new added transitions;
- 11 Update multiplier $\lambda \leftarrow$ Eq. 9;
- 12 **if** *Convergence* **then**
- 13 **break**;
- 14 **return** ϕ ;

interval is 1 hour. The new obtained online transitions are used to update the GP model for approximating the sim-to-real performance differences.

Algorithm 3: The Online Algorithm

Input: ϵ, Y, E, A, B, N

- 1 Initialize multiplier λ from offline stage;
- 2 **while** *True* **do**
- 3 **for** $n = 0, 1, \dots N$ **do**
- 4 Sample \mathbf{a}_t , calculate $F(\phi)$ for all samples;
- 5 Estimated $Q(\phi)$ from BNN $Q_s(\phi)$ and GP model $G(\psi)$ for all samples with Eq. 12;
- 6 Calculate Lagrangian in Eq. 14;
- 7 Get the next query $\mathbf{a}_t \leftarrow \arg \min L$;
- 8 Obtain $Q_s(\phi)$ by querying simulator;
- 9 Estimate $G(\psi)$ with ψ ;
- 10 Update multiplier $\lambda \leftarrow$ Eq. 15;
- 11 **if** *Time to online query* **then**
- 12 Apply \mathbf{a}_t to real networks;
- 13 Calculate $G(\psi) = Q(\phi) - Q_s(\phi)$;
- 14 Train ψ with new online transitions;
- 15 **if** *Convergence* **then**
- 16 **break**;
- 17 **return** ψ ;

D ARTIFACT APPENDIX**D.1 Abstract**

This section describes the instructions for performing artifact evaluation for this paper. In the paper, we achieve *Atlas* with three integrated stages, i.e., learning-based simulator, offline training, and online learning.

D.2 Artifact check-list

- **Algorithm:** Algorithm 1, 2, 3 described in the paper
- **Program:** Python scripts, NS-3 simulator, OpenDayLight, executable files in OpenAirInterface RAN and CORE, Android applications
- **Compilation:** g++ 8.0, gcc 8.0 or higher
- **Data set:** own dataset provided in open-source codes.
- **Run-time environment:** At least two desktop computers, one for RAN and one for CN&Edge. OS can be either Ubuntu 18.04 or 20.04. Docker containers is needed in CN desktop.
- **Hardware:** For network simulator: No restriction on hardware; For network prototype: Intel i7 or above CPU, NI USRP B210, Ruckus ICX 7150-C12P, OnePlus 9 5G smartphone, Antenna compatible with LTE B7 frequency band.
- **Execution:** sole user is preferred
- **Metrics:** runtime log in NS-3, performance reported by smartphones.
- **Output:** saved experiment results in pickle format, which will be used for following figure plotting.
- **Experiments:** see README in the open-source codes.
- **How much disk space required (approximately)?:** 1GB for data and trained model, the disk space for software installation are not counted.
- **How much time is needed to prepare workflow (approximately)?:** For network simulator: less than 30 minutes; For network prototype: several hours if not more.
- **How much time is needed to complete experiments (approximately)?:** For stage 1: 3 5 hours, under 16 threads; For stage 2: 3 5 hours, under 16 threads; For stage 3: 1 3 hours.
- **Publicly available?:** Yes.

D.3 Descriptions

- *main_simulator.py* is the main file for conducting experiments for the stage 1: learning-based simulator.
- *main_offline.py* is the main file for conducting experiments for the stage 2: offline training.
- *main_online.py* is the main file for conducting experiments for the stage 3: online learning.
- *plot_*.py* are mainly for plotting the results based on completed experiments.
- *system.py* is the main file for connecting with network prototype.
- *simulator.py* is the main file for connecting with network simulator.

D.3.1 How to access. The codes are open-sourced available in <https://doi.org/10.5281/zenodo.7262492>, where more detailed instructions can be found in <https://github.com/int-unl/Atlas.git>.

D.3.2 Hardware dependencies. At least two desktop computers with Intel i7 or above CPUs, one for RAN and one for CN&Edge. OS can be either Ubuntu 18.04 or 20.04. Docker containers is needed in CN desktop.

D.3.3 Software dependencies. For executing the codes: Ubuntu 20.04, Python 3.6.9, PyTorch 1.10.2, scipy 1.5.4, sklearn 0.24.2, numpy 1.19.5, pickle 4.0, CUDA is not required.

For RAN host in the network prototype: Ubuntu 18.04, low-latency kernel; For CN and Edge in the network prototype: Ubuntu 20.04.

D.3.4 Data sets. We collect our own dataset from the real-world network prototype. The file is in *app_eval/* folder, which is mainly used for the stage 1: learning-based simulator.

D.4 Network Simulator and Prototype Build

Network Simulator:

- Install NS-3 3.36, either with official instruction or provided bash file.
- Add additional files (i.e., `edge/`) in NS-3 `contrib/` folder, and rebuild.
- Add additional files (i.e., `main.cc`) in NS-3 `scratch/` folder, and rebuild.
- Validate the network simulator can be connected by `simulator.py` script.

Network Prototype:

- Install OpenAirInterface RAN with official instruction.
- Install OpenAirInterface CORE with official instruction, where the dockerized network functions are required, e.g., SPGW-U and HSS.
- Modify SPGW-C and rebuild to enable core network slicing, validate it redirects specific mobile users to corresponded SPGW-Us.
- Connect smartphones with RAN with programmed USIM card to match the PLMN and other parameters, validate they can access to the RAN and the SPGW-U docker.
- Install the provided Android application to smartphones (Android 11).
- Install the provided edge server application in individual SPGW-U dockers, validate smartphones can connect with their servers with periodically performance updates.
- Install FlexRAN controller 2.0 with official instruction, validate its slicing capability when changing PRB allocation on RAN to different mobile users.
- Initialize the SDN switch, install OpenDayLight to connect the switch, validate the provided `tn_server.py` can connect OpenDayLight.
- Connect SDN switch between the RAN and CORE desktop, validate it can enforce bandwidth allocation to between mobile users and their SPGW-U dockers.
- As everything is well validated individually, close all scripts and programs.
- The order of running the network prototype is: OpenDayLight – Transport controller – FlexRAN controller – CORE – edge server applications – RAN – Configure slicing parameters – Smartphone (disable airplane mode and open the app) – start `system.py`.
- The demo procedures of bringing up the network prototype is online: <https://youtu.be/-AFw17ANBN8>

- Run `plot_online.py` to reproduce the figures in the paper, after all corresponded experiments are done.

D.5 Experiment workflow

- Download open-source codes
- Install package dependencies, e.g., `sklearn`, `scipy`, `torch`, and more.
- **Stage 1: learning-based simulator.**
 - Run `main_simulator.py`, where the arguments vary according to different experiments.
 - Run `plot_simulator.py` to reproduce the figures in the paper, after all corresponded experiments are done.
 - Update the searched optimal simulation parameter to `parameter.py`.
- **Stage 2: offline training.**
 - Run `main_offline.py`, where the arguments vary according to different experiments.
 - Run `plot_offline.py` to reproduce the figures in the paper, after all corresponded experiments are done.
 - Update the searched optimal resource configuration to `parameter.py`.
- **Stage 3: online learning.**
 - Bring up the network prototype (see above), validate it is live and can be connected by the `system.py` script.
 - Run `main_online.py`, where the arguments vary according to different experiments.