Inter-Slice Radio Resource Allocation: An Online Convex Optimization Approach

Tianyu Wang and Shaowei Wang

Abstract

Inter-slice radio resource allocation (IS-RRA) is a new layer of radio resource management introduced by network slicing in 5G and beyond mobile communication systems. Instead of focusing on the per-user service quality as in conventional packet schedulers, IS-RRA is required to ensure the service-level agreements on a per-slice basis, which is particularly challenging due to the diverse requirements and behaviors of network slices. In this article, we analyze the inherent limitations of the existing model-based and data-based approaches and propose a novel framework based on online convex optimization (OCO). Specifically, the proposed OCO approach is able to incorporate the offline knowledge and the online data into a general online learning framework, which overcomes the modeling difficulty and high computational complexity of the model-based approach, and avoids the blind exploration of the data-driven approach. At last, we provide the simulation results of an example scenario, which shows that the proposed OCO approach can adapt to diverse service requirements and provide comparable performance to the optimal solutions given in hindsight.

Introduction

Network slicing is envisioned as one of the disruptive technologies of 5G and beyond networks. Unlike the conventional one-size-fits-all architecture, it allows multiple self-contained logical networks, that is, network slices, to run on top of common physical infrastructures, such that the slices can provide highly optimized solutions for different vertical applications, ranging from human-centric multimedia services with high peak rates to vehicular communications with ultra-reliable and low-latency requirements, and smart city applications with massive IoT connections. From a business perspective, network slicing enables mobile operators to deploy customized services for specific market scenarios in a flexible and cost-effective way and unlocks an enterprise opportunity in the 5G era.

A generic concept of network slicing consists of three layers: the infrastructure layer, the network function layer, and the service layer [1]. The infrastructure layer refers to the physical entities and resources spanning all domains, which are virtualized and revealed to higher layers under the infrastructure as a service paradigm. The network function layer provides a variety of fine-grained virtual network functions that can be flexibly chained together to offer diverse end-to-end services, while the service layer defines the high-level description of services and provides holistic management and orchestration of slice instances. In the 4G era, network slicing has been provided in the limited form of (enhanced) dedicated core network (CN), which allows the deployment of multiple CNs over the same infrastructure. In the 5G era, 3GPP has further agreed that the radio access network (RAN) should be slice-aware, where distinct L1/L2/L3 configurations must be provided to handle slice-specific traffic.

From a resource management perspective, the fundamental challenge of network slicing lies in the dilemma that it intends to guarantee full isolation between network slices such that each can serve as an independent network, while at the same time, all slices need to share the limited infrastructure on demand so as to ensure efficient utilization. In the CN, virtual machines are widely adopted to provide an isolation-efficiency trade-off, in terms of processing, storage, and other network resources. However, in the RAN, there exists an additional dimension of allocating radio resources among network slices, which is considered to be particularly challenging due to the limited amount of radio resources and the interfering nature of the wireless environment. Therefore, inter-slice radio resource allocation (IS-RRA) becomes an open technical challenge in both industry and academia [2].

IS-RRA can adopt either the model-based approach with optimization techniques and mathematical models or the data-based approach with reinforcement learning techniques and online interactions. In this article, we analyze the advantages and drawbacks of existing model-/data-based approaches and argue that both methodologies have inherent limitations that may greatly jeopardize their success in practical deployments. To overcome the drawbacks and reap the benefits from both sides, we propose a novel framework based on online convex optimization (OCO), which is able to integrate both the offline knowledge and the online data into a general algorithmic structure [3]. Specifically, the network dynamics are continuously monitored. The corresponding data are transformed into loss functions that analytically indicate how IS-RRA would have affected the instant system performance. OCO algorithms can be employed to use the loss functions to adjust the real-time decision. An example scenario is provided to show that the proposed OCO approach can

Digital Object Identifier:
10.1109/MWC.010.2100122

The authors are with Nanjing University.

 1536-1284/20/$25.00 © 2020 IEEE

adapt to diverse service requirements and yield comparable performance to the optimal solutions given in hindsight.

The rest of the article is organized as follows. In the following section, we review three major IS-RRA schemes in the literature. Then we focus on the most popular slice scheduler scheme and analyze the advantages and drawbacks of existing model-/data-based approaches. Then, we present our proposed OCO approach and provide an example scenario. At last, we conclude the article.

Inter-Slice Radio Resource Allocation

IS-RRA addresses the problem of how to schedule radio resources in the RAN such that pre-established service-level agreements (SLAs) can be offered on a per-slice basis. The SLAs may include a variety of key performance indicators ranging from throughput, latency, reliability, and availability. Depending on the realization of network slicing in the RAN, IS-RRA can be performed at different time granularity and the slices can share the RAN with different levels of infrastructure reuse. The existing schemes can be roughly classified into three categories, that is, the dedicated allocation, the SLA-aware packet scheduler, and the SLA-based slice scheduler, which are illustrated in Fig. 1.

Dedicated Allocation

A straightforward solution is to provide each slice with a dedicated chunk of the spectrum after its creation and then keep this assignment unchanged during the entire lifecycle [4], which is referred to as the dedicated allocation. Since dynamic spectrum sharing is strictly prohibited, the network slices become physically independent subnetworks with the highest isolation level, which allows them to deploy highly customized packet schedulers on top of their dedicated spectrum without considering the behaviors of others. Therefore, the dedicated allocation scheme is suitable for life-critical services with strict reliability requirements, for example, public safety, industrial automation, and autonomous driving.

However, due to the statistical property of the wireless environment, the amount of dedicated resources for a slice needs to be large enough such that the SLAs can be fulfilled even in the worst-case scenario of the entire lifecycle. Therefore, the dedicated allocation scheme may suffer from inefficient resource utilization, especially for highly dynamic services. As shown in Fig. 1, in the dedicated allocation scheme, a large amount of time-frequency resources are committed to a slice even if they are not used most of the time.

SLA-Aware Packet Scheduler

Another solution is to implement a common packet scheduler for the users of all slices, which is referred to as the SLA-aware packet scheduler. Specifically, this scheduler is responsible for not only guaranteeing the quality of service (QoS) of each user but also enforcing the SLAs of each slice [5]. On the one hand, a common packet scheduler can fully exploit the time and frequency selectivity of wireless channels to generate the maximum multiplexing gain. On the other hand, this all-in-one-pass solution can be extremely complex due to the diverse QoS/SLA requirements in practical networks.

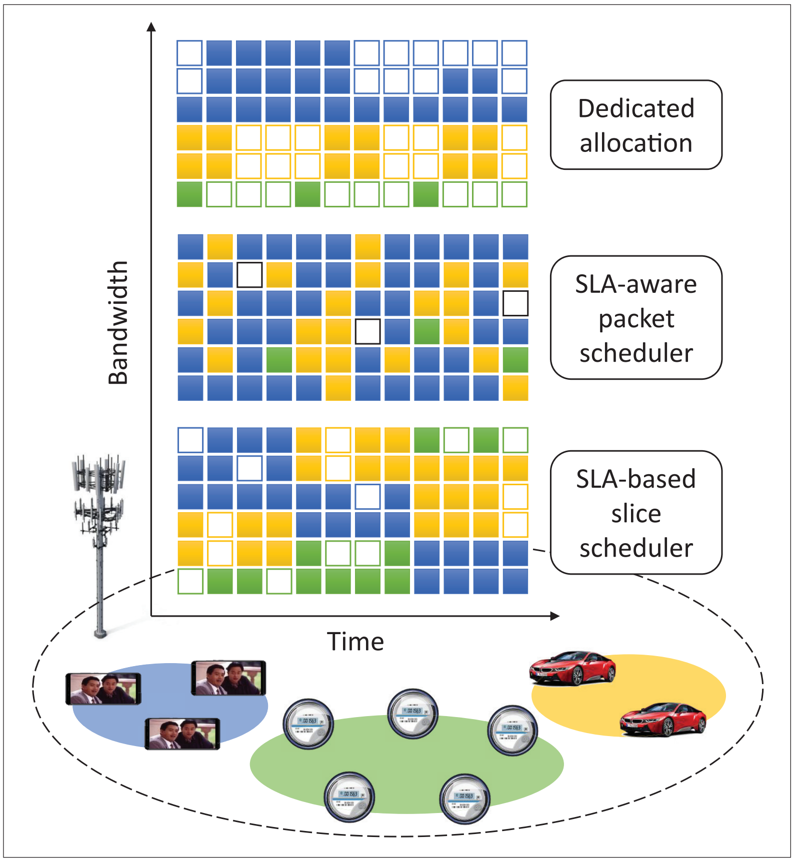

FIGURE 1. Inter-slice radio resource allocation.

Therefore, the SLA-aware packet scheduler scheme is suitable for slices having similar QoS requirements of their users and elastic SLA requirements that can be fulfilled in the long term. For example, a utility-based packet scheduler can be deployed for general mobile broadband slices with both guaranteed and non-guaranteed bit rate traffic. As shown in Fig. 1, in the SLA-aware packet scheduler scheme, high spectral efficiency can be achieved, while the amount of radio resources assigned to a slice is greatly influenced by the traffic fluctuations of other coexisting slices.

SLA-Based Slice Scheduler

To achieve a trade-off between slice isolation and resource efficiency, radio resources can be periodically assigned on a per-slice basis, such that the SLAs can be dynamically ensured in each allocation window, which is referred to as the SLA-based slice scheduler. Specifically, the slice scheduler can make hard decisions to assign each slice a dedicated chunk of spectrum, or soft decisions to assign each slice a certain amount of radio resources. With hard decisions, each slice can deploy an independent packet scheduler on its exclusively used spectrum in each allocation window. With soft decisions, a common low-complexity two-level packet scheduler can be introduced to maintain slice-specific schedulers and reap the inter-slice multiplexing gain at the same time [6].

Depending on the realization of network slicing, the allocation window can be configured with different time granularity from milliseconds [7–10], seconds [11, 12], to minutes [13] and hours [14].

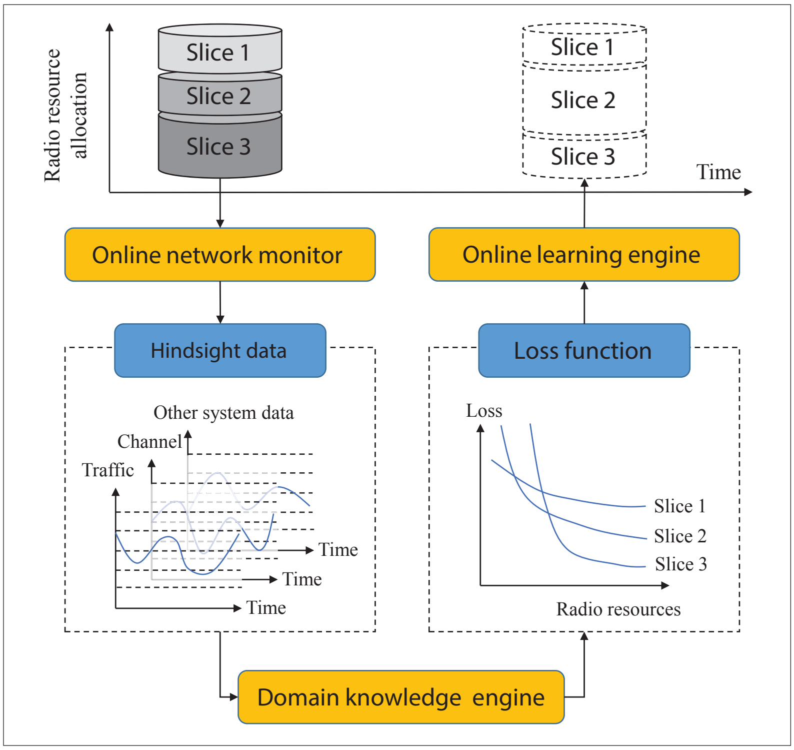

FIGURE 2. The inter-slice radio resource allocation framework based on online convex optimization.

Generally speaking, a longer allocation window implies a higher level of slice isolation and lower spectral efficiency, while a shorter one implies the opposite. The aforementioned dedicated allocation and common packet scheduler schemes can be seen as two extreme cases of the slice scheduler scheme with the allocation window size equal to infinity and zero, respectively. As shown in Fig. 1, in the SLA-based slice scheduler scheme, a trade-off between slice isolation and spectral efficiency can be achieved.

Advantages and Drawbacks of Existing SLA-based Slice Schedulers

In the literature, the SLA-based slice scheduler scheme attracts more attention due to its flexibility and simplicity as compared to the other two schemes. According to the underlying methodologies, the existing SLA-based slice schedulers can be further categorized into the model-based approach and the data-based approach. Their advantages and drawbacks are discussed below.

Model-Based Approach

In the model-based approach, the network dynamics are described by using deterministic or statistical models, for example, the arrival of data packets is typically modeled as a Poisson process. The problem formulation can be highly simplified, where the multi-dimensional SLAs can be translated into concise radio resource requirements. Depending on the time granularity of the allocation window, these requirements can be treated as instant optimization constraints of the current allocation window [8], or as long-term constraints of a continuous control process that covers a sequence of consecutive allocation windows [7, 9, 10], for which optimization methods and optimal control methods are widely adopted by the slice scheduler, respectively.

Despite the mathematical elegance, the model-based approach suffers from the fundamental limitation of mathematical modeling, that is, the actual performance is greatly influenced by the correctness and accuracy of modeling assumptions. Considering the diversity of slices across different time and space, it can be highly difficult to determine the underlying mathematical models, for example, the data rate of virtual reality services varies greatly as the field of view changes, which is usually not Poisson and hard to predict without the application layer context. Even if the correct model is chosen, it can still be highly difficult to determine the model parameters, for example, the SLAs may not be fulfilled in the worst-case scenarios if the peak traffic rate is underestimated, while inefficient resource utilization may occur if the average traffic rate is overestimated. At last, network behaviors can be highly complex, for example, the slice-specific packet schedulers can be highly diverse with customized scheduling policies, in which case proper analytical models may not even exist. Therefore, the model-based approach may confront unexpected difficulties in practical deployments.

Another limitation is the high computational complexity that comes from the translation from multi-dimensional SLA requirements to uniform radio resource constraints, as well as the outcome optimization and optimal control problems. Such computational tasks need to be repeatedly solved in each allocation window for each active slice, which may greatly prohibit the practical deployment of the model-based approach. We note that some low-complexity slice scheduling algorithms have been proposed in the literature [7, 9]. However, these algorithms are either heuristic methods that cannot provide any theoretical performance guarantee [9], or quasi-static methods that ensure myopic optimality in stationary environments with static model parameters [7]. Therefore, it is difficult for the model-based approach to maintain low computational complexity, while at the same time, to provide guaranteed SLA performance.

Data-Based Approach

Recently, the data-based approach using machine learning techniques has attracted a lot of attention [11–13]. The basic idea is to treat IS-RRA as a reinforcement learning process, where the radio resource manager learns to improve its allocation strategy by continuously interacting with the network environment over time. In each allocation round, the resource manager decides the radio resource allocation based on the current allocation strategy, while the network environment feeds back the resulting SLA performances. Specifically, the allocation strategy can be represented by using a deep neural network, which takes the high-dimensional online data as its input and outputs the corresponding radio resource allocation. The deep neural network is updated according to the instant reward received from the wireless environment, which directly reflects the instant SLA performance of all slices. After certain rounds of exploration and feedback, the radio resource manager converges to a stable strategy that outputs high-performing radio resource allocations in the following allocation windows. Therefore, the data-based approach avoids the modeling difficulty.

The mathematical framework behind reinforcement learning is a rigorous Markov decision process with a fixed but unknown transition function. In such settings, the state-of-the-art deep reinforcement learning techniques have been shown to converge to high-performing strategies. However, in the considered IS-RRA problem, the wireless environment can be a highly dynamic system having non-stationary characteristics, for example, as the traffic pattern changes over time, the slice may require time-varying bandwidth to ensure its SLAs even if the average traffic rate is unchanged. Therefore, the underlying transition function that determines the network dynamics can be time-varying, in which case the data-based approach needs to continuously explore the allocation space online, such that the instant transition function can be captured by fine-tuning the weights and bias of the deep neural network. We note that such online explorations are random and blind in nature. Therefore, the data-based approach may suffer from severe SLA violations in practical deployments, which gives rise to the reliability issue.

IS-RRA Based on OCO

The model-based approach exploits the domain knowledge to simplify the problem formulation, while it is limited by the modeling accuracy and the computational complexity. The data-based approach avoids the difficulty of mathematical modeling, while it suffers from the reliability issue due to its random and blind interactions with the network environment. In this section, we propose a novel IS-RRA framework based on OCO, which can take advantage of both the model-based and data-based approaches by jointly exploiting the domain knowledge and online data.

OCO Framework

OCO is a continuous learning process that aims to tackle a sequence of decision-making tasks. At the beginning of each round, the learner is required to make a proactive decision for the current task, while the outcomes associated with that decision are unknown. After committing to the decision in the current round, the learner receives an instant loss function from the environment, which formulates the task performance as a function of the instant decision in the past round. In the next round, the hindsight loss functions can be utilized by the learner to adjust the real-time decision and improve the expected task performance.

As shown in Fig. 2, the proposed OCO framework consists of three major functional modules, that is, the online network monitor, the domain knowledge engine, and the online learning engine. Specifically, the online network monitor stores the online data generated in the last allocation window, including the traffic information, the channel information, the scheduling details, and other performance-related data, which can be used to reproduce the network behaviors in hindsight. The domain knowledge engine transforms such hindsight data into slice-specific loss functions, each of which formulates the instant relationship between the slice performance and the amount of exclusively assigned radio resources. The online learning engine utilizes the instant slice-specific loss functions to adjust the real-time radio resource allocation.

In OCO, the decisions are constrained by a convex set and the loss functions are required to be convex functions over the decision set. In the IS-RRA problem, the radio resource allocation set is convex due to the total bandwidth constraint, while the slice-specific loss functions can be arbitrary depending on the slice characteristics and the tenant's business model. Nevertheless, convex loss functions can still be appropriate based on the following observations. First, there usually exists an upper bound for the slice performance, for example, the maximum cumulative rate due to admission control, the minimum packet delay due to physical layer protocols, and the minimum SLA violation ratios due to the statistical wireless environment. Thus, as the amount of dedicated radio resources increases, the slice performance strictly improves and asymptotically converges to the upper bound, which roughly implies a convex trend for the performance loss. Second, the marginal revenue brought by certain performance improvement usually decreases as the slice performance becomes higher, for example, logarithmic utility functions are widely used in the literature of network slicing [7]. Thus, if the tenant's revenue is taken into account, convex loss functions are appropriate assumptions. In the proposed OCO framework, we define the loss in an allocation window as the sum loss of all slices, which is convex when all slice-specific loss functions are convex. Therefore, the proposed OCO framework is justified.

The OCO framework intends to minimize the cumulative loss in the long term, which however is extremely difficult since the exact loss functions can only be achieved in hindsight. Instead, an appropriate metric for OCO is the difference between the cumulative loss incurred by the online learning engine and that of a benchmark solution, which is referred to as the regret in the literature [3]. Depending on which benchmark algorithm is considered, a variety of regrets can be defined. For example, the classic regret is defined as the gap to the optimal static decision, while the dynamic regret is defined as the gap to the optimal decision sequence with certain regularity.

Note that the benchmark algorithms are offline algorithms that require the loss functions of all rounds in advance, which makes them impossible to be implemented as online learners. In the proposed OCO framework, the radio resource allocation can be adjusted online by exploiting the hidden structure of previous loss functions. Specifically, OCO algorithms can guarantee a sublinear regret, that is, the excessive loss compared to the benchmark algorithm increases sublinearly with time. It implies that on average OCO algorithms can perform as well as the benchmark algorithm as the time horizon goes to infinity. Classical OCO algorithms include the online gradient descent and the online Newton step, which utilize the first and second-order derivatives of loss functions, respectively, to ensure bounded regrets in an arbitrary environment [3]. Advanced OCO algorithms, such as the scale-free online gradient descent (SOGD), run multiple classic OCO learners in parallel and dynamically combine their decisions to achieve lower regret bounds [15].

The model-based approach exploits the domain knowledge to simplify the problem formulation, while it is limited by the modeling accuracy and the computational complexity. The data-based approach avoids the difficulty of mathematical modeling, while it suffers from the reliability issue due to its random and blind interactions with the network environment.

The OCO approach can achieve comparable performance to the optimal static and optimal dynamic solutions given in hindsight. Therefore, our proposed OCO approach can effectively track the network dynamics by using the online network monitor and efficiently allocate the radio resources in each allocation window by using the online learning engine.

At last, we note that the system dynamics are predictable to some extent, for example, the channel parameters have high correlation coefficients within the coherence time. The predictability of wireless networks provides additional network data in the future, which can be further leveraged to design OCO algorithms with lower regret bounds. Therefore, network prediction techniques can be efficiently incorporated to promote the proposed OCO approach.

Advantages of OCO

Compared with the model-based and data-based approaches, the proposed OCO approach integrates both the offline knowledge and the online data into a general algorithmic framework, which can reap the benefits from both sides. The advantages are summarized as follows.

Modeling Simplicity: Instead of formulating a general IS-RRA problem as in the model-based approach, the domain knowledge engine focuses on the instant system behaviors, which can be faithfully reproduced by the hindsight data. Therefore, at least in principle, the modeling process of the OCO approach can be simplified and the modeling accuracy can be improved. For example, the packet arrival rate of the Poisson traffic model can be directly calculated by using the hindsight data instead of being predicted or given beforehand. In addition, for highly complex systems without proper mathematical models, the domain knowledge engine can deploy a well-trained deep neural network to infer the proper loss functions for the specific hindsight data, which completely avoids the difficulty of mathematical modeling.

Low Computational Complexity: Instead of optimizing each allocation window as in the model-based approach, OCO algorithms focus on the cumulative loss in the long term. In each round, the real-time radio resource allocation takes only one step forward in the direction that decreases the hindsight loss. This one-step update can be efficiently calculated due to the convexity of loss functions. Note that the step forward may result in an infeasible allocation, in which case the OCO algorithms project the allocation back to the feasible set, that is, finds the closest point in the convex hull formed by the total bandwidth constraint. This projection operation is equivalent to a convex quadratic programming problem, which can be efficiently solved by interior-point methods with logarithmic complexity. Therefore, the overall computational complexity is low.

Robustness in Non-Stationary Environments: The sublinear regret bound guarantees that the time average performance of OCO algorithms is at least as good as the high-performing benchmark algorithm in any arbitrary environment in the long term. In other words, OCO algorithms can gradually learn the environmental changes and adjust the radio resource allocation accordingly based on the derivatives of hindsight loss functions. Therefore, the OCO approach avoids the random and blind interactions in the data-based approach and provides robustness in non-stationary environments.

Challenges of OCO

Although the proposed OCO approach has the above advantages as compared to the existing model-based and data-based approaches, there still exists several practical concerns that should be carefully addressed.

Allocation Frequency: In principle, the allocation window should be as short as possible such that fine-grained network dynamics can be learned to exploit inter-slice multiplexing gain. However, a short window size implies a high allocation frequency, which leads to increased online computational complexity. Therefore, there exists a trade-off between system performance and computational complexity. Also, as the window size decreases, the instant SLA performance will be greatly influenced by the outcomes of the previous window. Therefore, the resulting Markov property may greatly impact the corresponding performance.

Loss Functions: For slices with multi-dimensional SLA requirements, the slice-specific loss functions should be carefully designed to balance all considered performance dimensions. Also, loss functions of different slices should be comparable such that slice fairness can be achieved when the sum loss is minimized. One possible principle is to evaluate the SLA violation ratio, which is always between 0 and 1 for all slices. The loss function parameters can be determined online by using lookup tables or a deep neural network, which outputs proper parameters for the past allocation window. In addition, the system loss can be defined as the weighted sum loss of all slices, such that different slices can be assigned with different weights according to their priorities.

Additional Constraints: Due to the complexity of network slicing, there may exist additional allocation restrictions except for the total bandwidth constraint. For example, there may exist a minimum amount of bandwidth for an industrial IoT slice in the corresponding SLAs, such that strict service reliability can be provided. Thus, the feasible allocation set can be non-convex and even time-varying. In such cases, the projection operation should be revisited to maintain low computational complexity, and the corresponding OCO algorithms need to be modified to ensure sublinear regrets.

Example

In this section, we provide an example of the proposed OCO approach. Specifically, we consider an LTE cell where two slices share a total of 20 MHz downlink bandwidth. The allocation window is set as 1 s. We assume each slice has 10 active users that are randomly distributed within the range of 100 m. Also, the slices deploy a proportional fair scheduler to address the Poisson traffic, where the packet delay budgets are given by 5 ms and 10 ms for slices 1 and 2, respectively. The SLAs are defined as the maximum packet loss ratio (PLR) within an allocation window, which are given by 0.01 and 0.02 for slices 1 and 2, respectively. The total traffic rate is assumed to be constant during the entire simulation, while the instant traffic rate of a slice can fluctuate. Here, we adopt the SOGD algorithm for the online learning engine.

In Fig. 3, we show the cumulative distribution function of the PLRs of both slices, where the total traffic rate is 80 Mb/s and the traffic rate of each slice fluctuates within a range 0.3 times of the total traffic rate. As we can see, the curve of slice 1 is above slice 2, indicating that the proposed OCO approach guarantees a lower PLR for slice 1 as compared to slice 2. In addition, we see that both slices are equally treated, as their target PLRs 0.01 and 0.02 are fulfilled with an approximately equal probability 0.88. Therefore, we show that the proposed OCO approach can provide differentiated services, while at the same time, ensure slice fairness in terms of their SLAs.

In Fig. 4, the average SLA violation ratios are given for networks where the total traffic rate and the fluctuation range are set as (80, 0.3), (75, 0.3) and (80, 0.1), respectively. The optimal static solution maintains a static assignment with the minimal cumulative loss in the entire simulation horizon, and the optimal dynamic solution dynamically chooses the optimal decision in each allocation window. As we can see, there exists a trade-off between the maximal traffic rate and the maximal fluctuation range for a certain SLA violation ratio, which indicates the fundamental trade-off between spectral efficiency and slice isolation. In addition, we see that the OCO approach can achieve comparable performance to the optimal static and optimal dynamic solutions given in hindsight. Therefore, our proposed OCO approach can effectively track the network dynamics by using the online network monitor and efficiently allocate the radio resources in each allocation window by using the online learning engine.

Conclusions

In this article, we have shown that the existing model-/data-based methodologies for IS-RRA have inherent limitations that may greatly jeopardize their success in practical deployments. Then, we propose a novel OCO framework that is capable of integrating both offline knowledge and online data into a general algorithmic framework. The proposed OCO approach simplifies the mathematical modeling, maintains a low computational complexity, and provides theoretical robustness in non-stationary environments. An example scenario shows that the OCO approach can provide differentiated services and slice fairness at the same time, and achieve comparable performance to the optimal solutions given in hindsight. Although there exist practical challenges for the application of the proposed OCO approach, we believe it can be a promising mathematical framework for IS-RRA in future mobile networks.

Acknowledgment

This work was partially supported by the National Natural Science Foundation of China under Grants 61801208, 61931023, and U1936202.

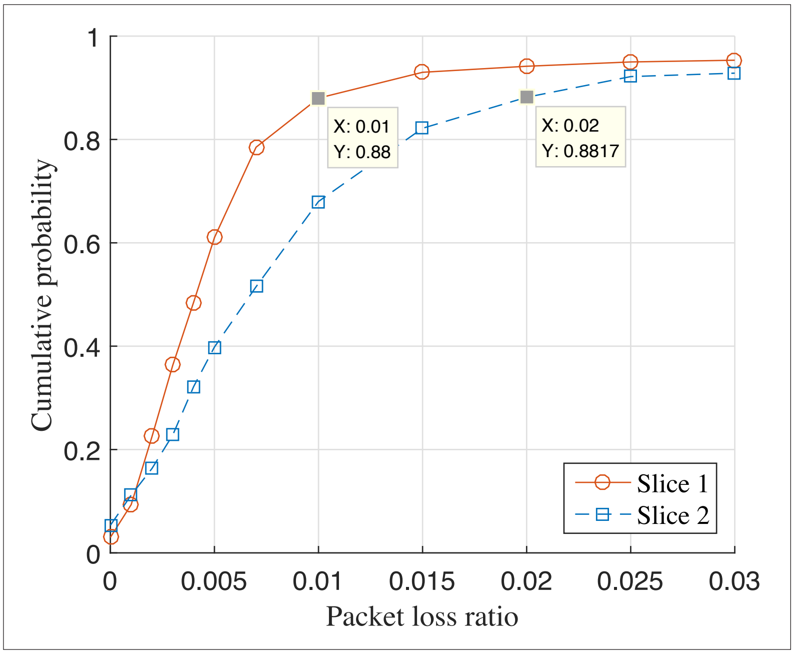

FIGURE 3. The cumulative distribution function of the packet loss ratio for the proposed OCO approach.

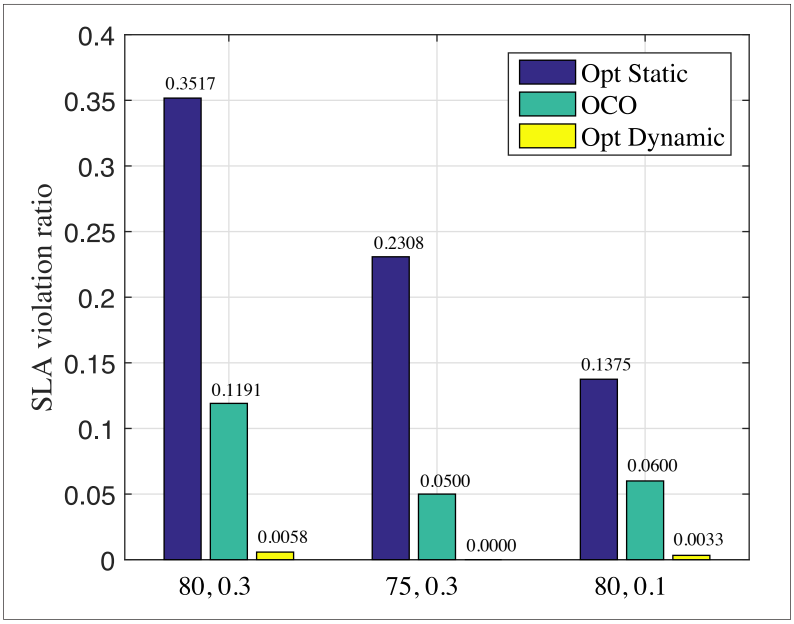

FIGURE 4. Average SLA violation ratios in different settings.

References

[1] X. Foukas *et al.*, "Network Slicing in 5G: Survey and Challenges," *IEEE Commun. Mag.*, vol. 55, no. 5, May 2017, pp. 94–100.

[2] H. Zhang *et al.*, "Network Slicing Based 5G and Future Mobile Networks: Mobility, Resource Management, and Challenges," *IEEE Commun. Mag.*, vol. 55, no. 8, Aug. 2017, pp. 138–45.

[3] E. Hazan, "Introduction to Online Convex Optimization," *Found. Trends Opt.*, vol. 2, no. 3-4, Aug. 2016, pp. 157–325.

[4] N. Nikaein *et al.*, "Network Store: Exploring Slicing in Future 5G Networks," *Proc. ACM MobiArch'15*, Paris, France, Sept. 2015.

[5] S. Mandelli *et al.*, "Satisfying Network Slicing Constraints via 5G MAC Scheduling," *Proc. IEEE INFOCOM'19*, Paris, France, Apr. 2019.

[6] A. Ksentini and N. Nikaein, "Toward Enforcing Network Slicing on RAN: Flexibility and Resources Abstraction," *IEEE Commun. Mag.*, vol. 55, no. 6, June 2017, pp. 102–08.

[7] R. Kokku *et al.*, "NVS: A Substrate for Virtualizing Wireless Resources in Cellular Networks," *IEEE/ACM Trans. Netw.*, vol. 20, no. 5, Oct. 2012, pp. 1333–46.

[8] C. Chang and N. Nikaein, "RAN Runtime Slicing System for Flexible and Dynamic Service Execution Environment," *IEEE Access*, vol. 6, Jun. 2018, pp. 34 018–42.

[9] T. Guo and A. Suarez, "Enabling 5G RAN Slicing with EDF Slice Scheduling," *IEEE Trans. Veh. Tech.*, vol. 68, no. 3, Jan. 2019, pp. 2865–77.

[10] A. Papa *et al.*, "Optimizing Dynamic RAN Slicing in Programmable 5G Networks," *Proc. ICC'19*, Shanghai, China, May 2019.

[11] C. Qi *et al.*, "Deep Reinforcement Learning with Discrete Normalized Advantage Functions for Resource Management in Network Slicing," *IEEE Commun. Lett.*, vol. 23, no. 8, Aug. 2019, pp. 1337–41.

[12] Y. Hua *et al.*, "GAN-Powered Deep Distributional Reinforcement Learning for Resource Management in Network Slicing," *IEEE JSAC*, vol. 38, no. 2, Feb. 2020, pp. 334–49.

[13] V. Sciancalepore, X. Costa-Perez, and A. Banchs, "RL-NSB: Reinforcement Learning-Based 5G Network Slice Broker," *IEEE/ACM Trans. Netw.*, vol. 27, no. 4, Aug. 2019, pp.

1543–57.
[14] C. Marquez *et al.*, "Resource Sharing Efficiency in Network Slicing," *IEEE Trans. Netw. Service Manag.*, vol. 16, no. 3, Sept. 2019, pp. 909–23.
[15] L. Zhang *et al.*, "Adaptive Regret of Convex and Smooth Functions," *Proc. ICML'19*, Los Angeles, United States, June 2019.

Biographies

Tianyu Wang [S'11, M'16] received the Ph.D. degree from Peking University, Beijing, China, in 2016. He is currently an associate researcher with the School of Electronic Science and Engineering, Nanjing University, China. He has published more than 40 IEEE journal and conference papers and received the Best Paper Award from IEEE ICC'15, IEEE GLOBECOM'14, and ICST ChinaCom'12. His current research interests include network slicing and machine learning in wireless networks.

Shaowei Wang [S'06, M'07, SM'13] received the Ph.D. degree from Wuhan University, Wuhan, China, in 2006, and joined the School of Electronic Science and Engineering at Nanjing University, Nanjing, China, as a faculty member in the same year, where he is currently a full professor. From 2012 to 2013, he was a visiting scholar/professor with Stanford University, Stanford, CA, USA, and The University of British Columbia, Vancouver, BC, Canada. His research interests include communications and networking, operations research and machine learning.