Advancing RAN Slicing with Offline Reinforcement Learning

Kun Yang*, Shu-ping Yeh^, Menglei Zhang^, Jerry Sydir^, Jing Yang‡, and Cong Shen*

* University of Virginia, ‡ The Pennsylvania State University, ^ Intel Corporation

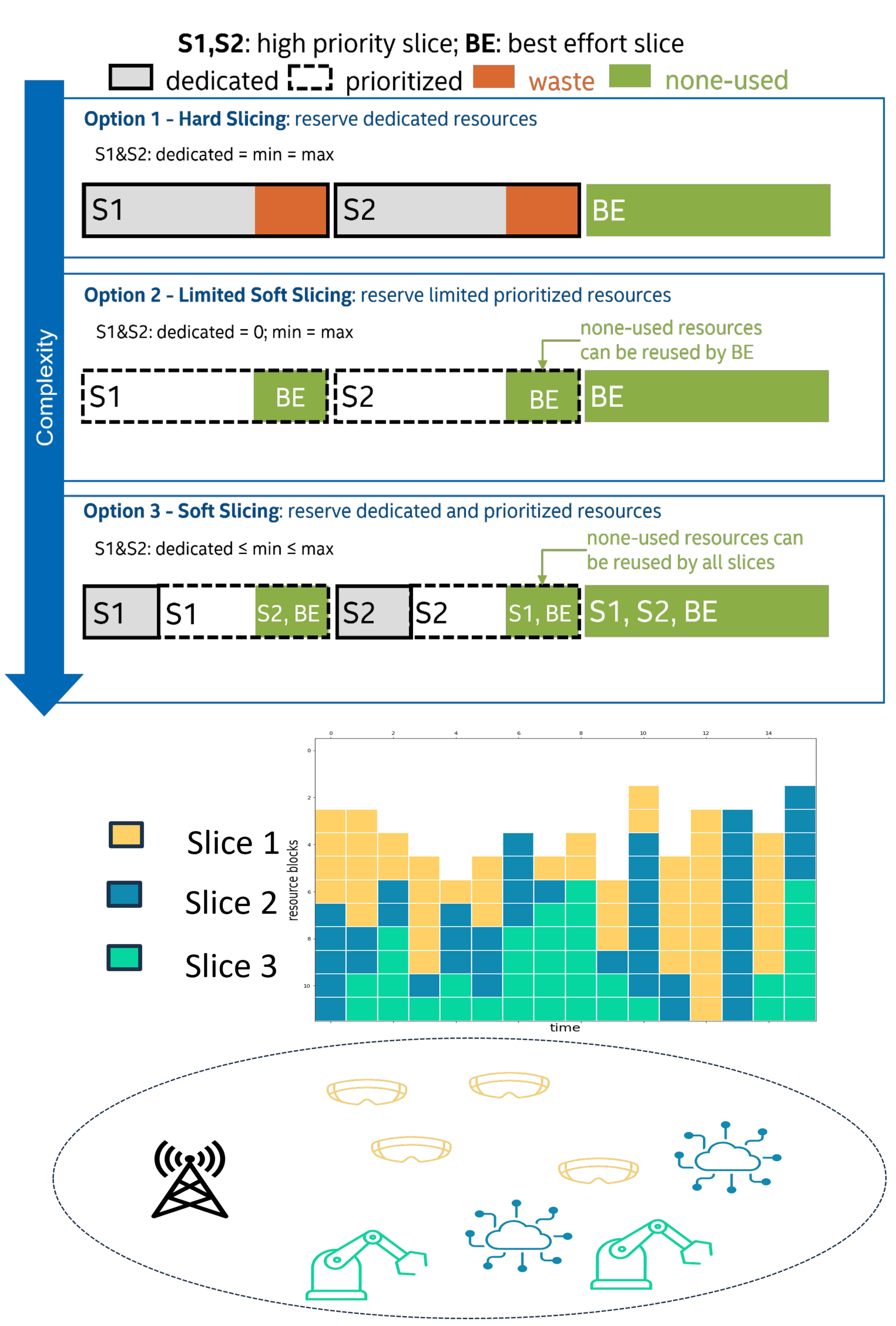
University of Virginia — SCHOOL of ENGINEERING & APPLIED SCIENCE, Charles L. Brown Department of Electrical and Computer Engineering

Motivation

- RL can solve sequential decision-making problems like RRM
- Network slicing is designed to handle different services -> create heterogenies data.
- Traditional methods struggle **to adapt between distinct services.**
- Online RL needs **extra exploration and training** for a new environment/service requirement.
- Data with ***good coverage (hetero data sources)*** can help offline RL training.

Slice Type	Data Rate	Capacity	Latency
eMBB	**Very High**	High	Low
URLLC	Moderate	Moderate	**Ultra-low**
mMTC	Low	High	Moderate

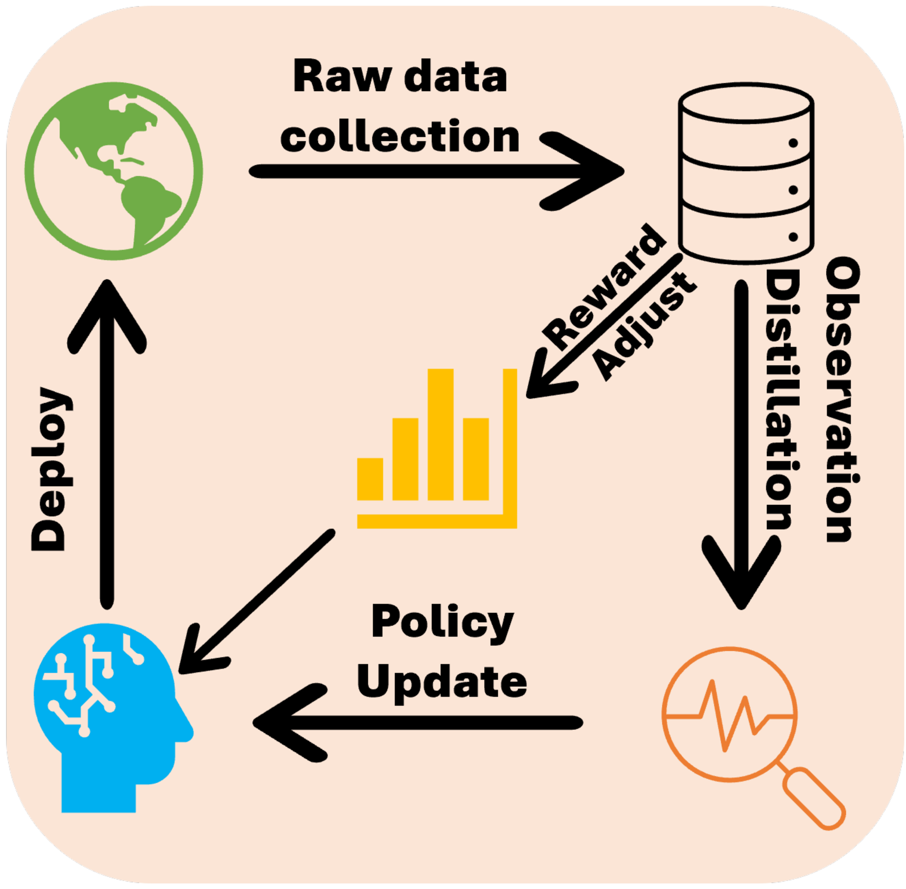
Environment Setting



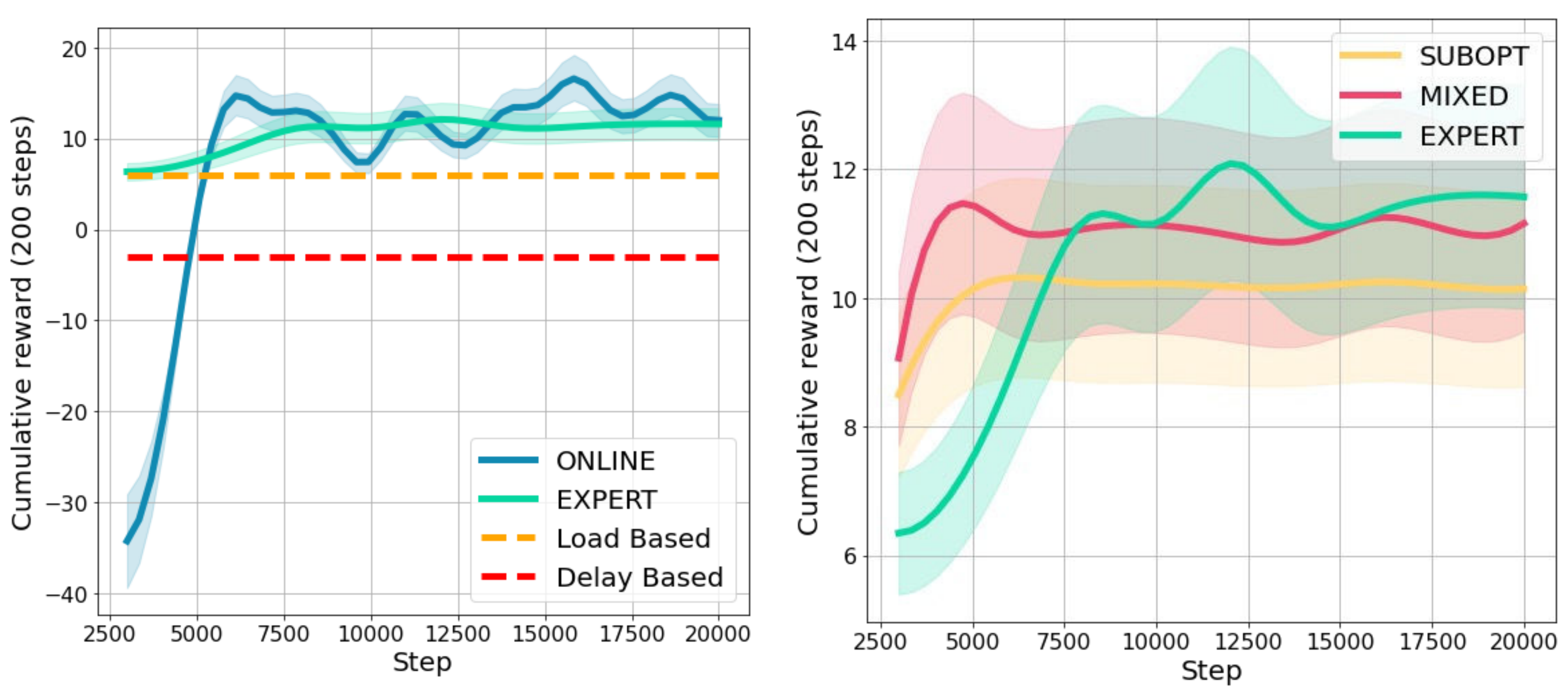
- **Two prioritized slices**
- One best effort slice (Background)
- 1 Cell with 30 users:
  - **Service Level Agreement (SLA):**
    - **Reduce delay violation rate**
    - Maintain received (rx) traffic
- **Objective:**
  - Allocate resource blocks for prioritized slices
  - **Meet SLA**

Experiment Process & Result

* **Observation distill** & **reward adjustment** enable ***flexibility***

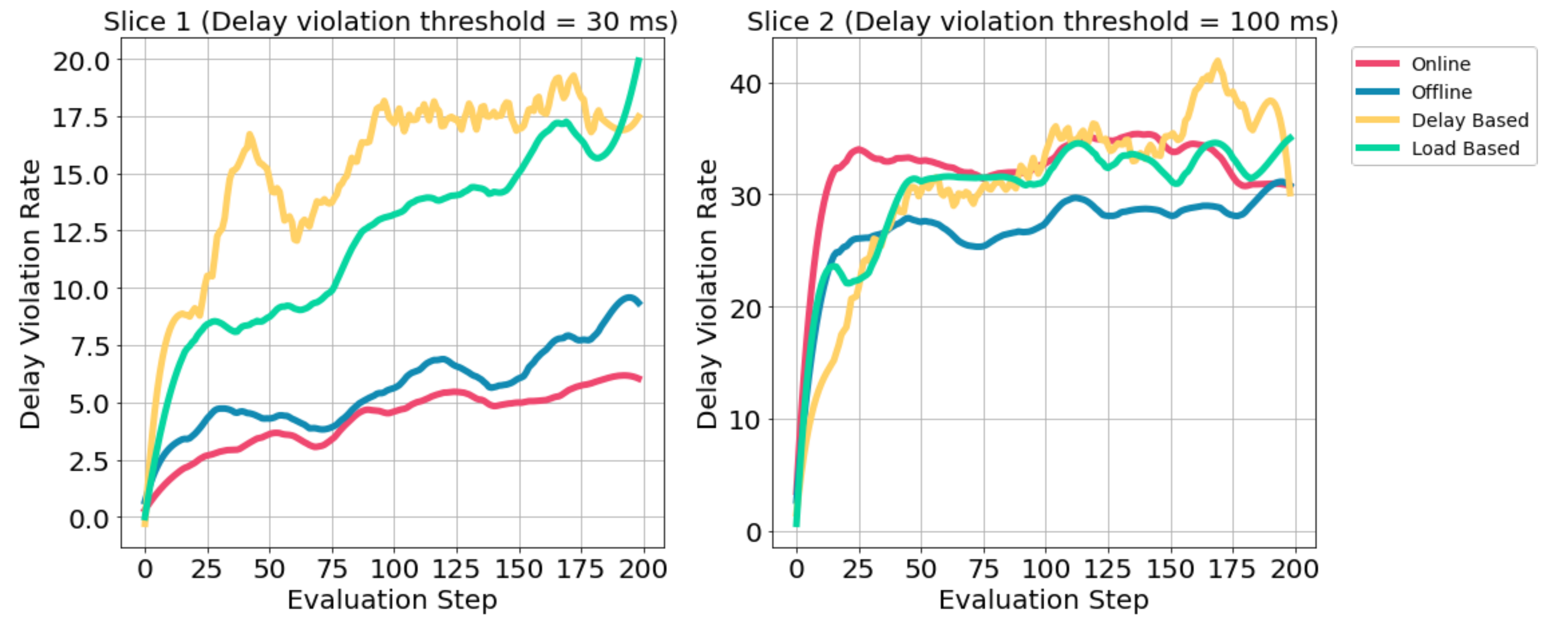


- Training from ***pure expert data*** can help offline RL outperform the online behavior policy.
- With **mixed suboptimal** datasets, the offline RL recovers online RL performance.



Adjust SLA and Objective

- Offline RL can handle different SLAs **without retraining**.



- With *reward adjustment*, offline RL can retrain policies to handle different SLA requirements.

SLA requirement	Delay violation rate	Total Throughput	Resource Usage
Delay	**6.5 ± 3.5**	52.48 ± 13.65	49.15
Throughput	9.1 ± 4.4	**58.68 ± 11.23**	49.35
Resource	7.3 ± 4.1	51.44 ± 12.68	**48.89**

Conclusion & Future work

- Offline RL is able to recover online-level policies with **mixed suboptimal dataset**.
- With **reward adjustment and observation distillation,** offline RL can adjust to different SLAs **without additional data collection.**
- Future question: Can offline RL algorithms handle different SLA **without retraining**?

Acknowledgement

The work is partially supported by the US National Science Foundation under awards **SWIFT-2029978** and **SII-2132700**