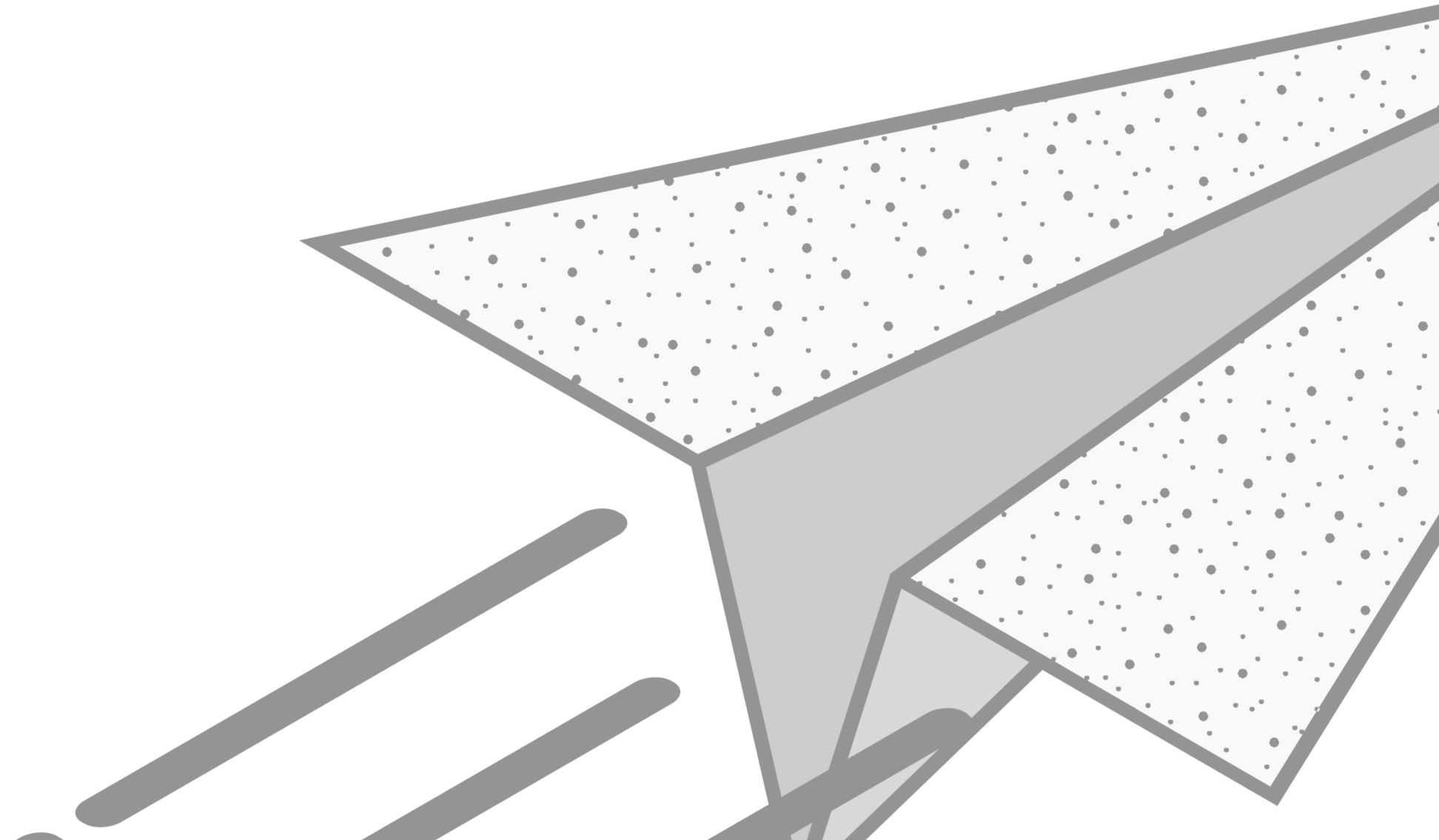


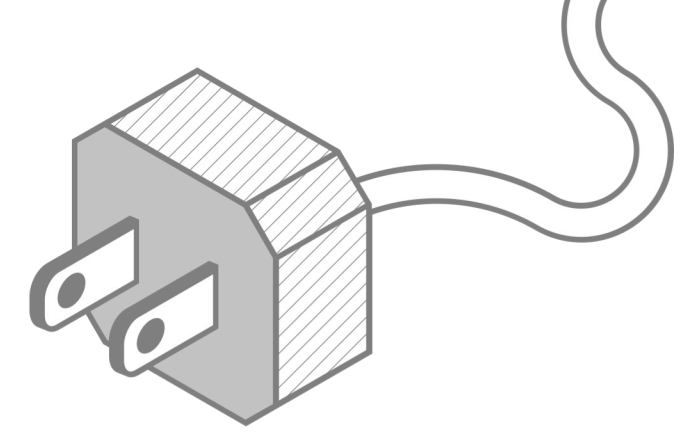
Project Proposal

結合知識蒸餾之 SAC 強化學習於 5G 動態網路資源允入控制

7113056083 楊啟弘



OUTLINE



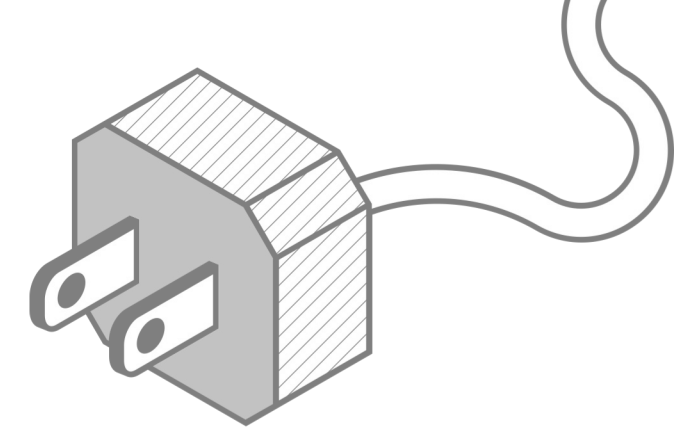
- INTRODUCTION

- RELATED WORK

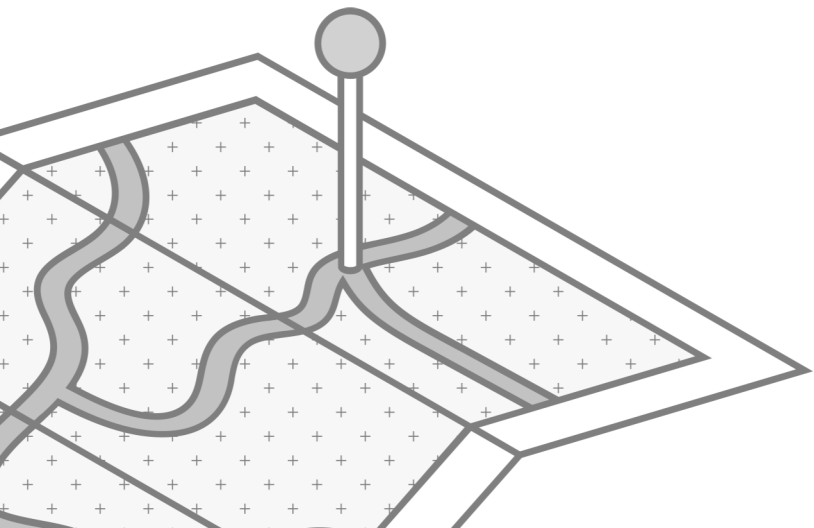
- PROPOSED DESIGN

INTRODUCTION

BACKGROUND / MOTIVATION



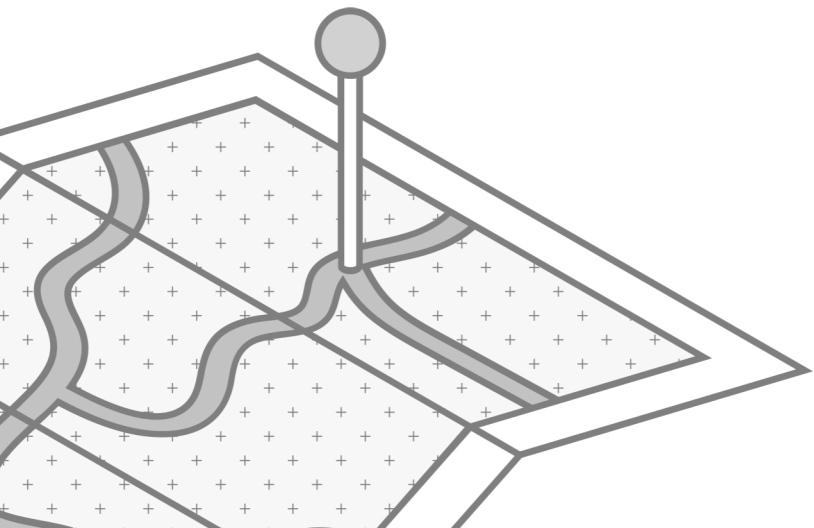
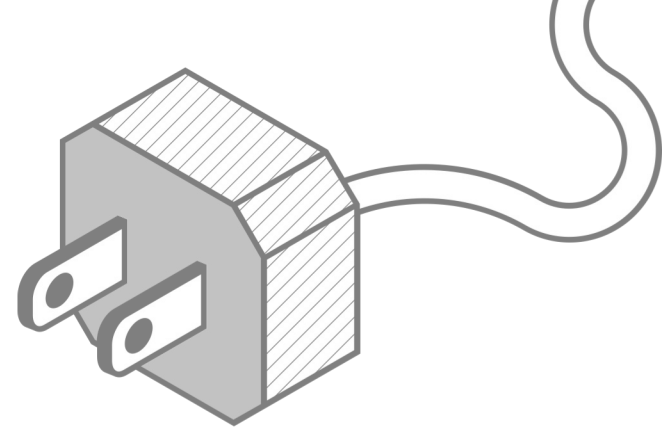
- 5G 時代連接裝置暴增，網路需求高度動態
- 必須兼顧 高效資源分配 與 服務品質 (QoS) 保障
- 強化學習 (RL) 具備自適應與學習能力，適合解決此問題



INTRODUCTION

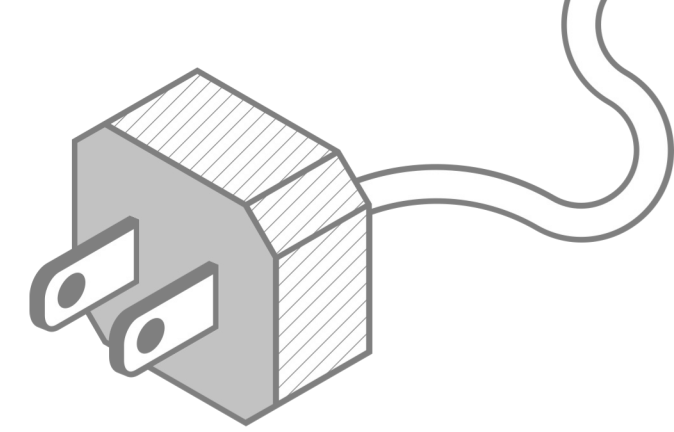
CHALLENGES




- 傳統 RAC 方法無法即時應對流量波動
- RL 訓練過程不穩、耗時高、模型大，難以部署在邊緣設備
- 在變動環境中學習效果不穩定、探索效率低

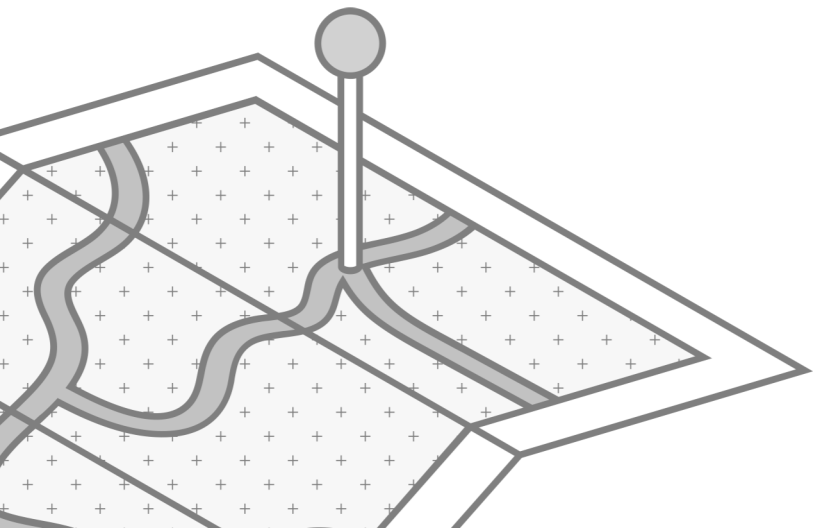


INTRODUCTION

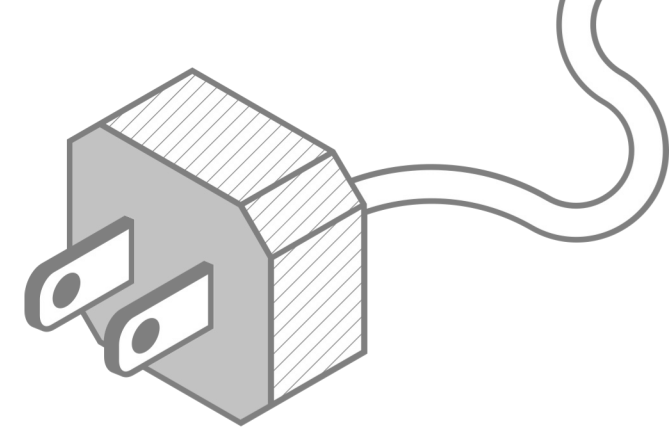
PROPOSED



- 使用 離散化 SAC 強化學習 (SACD) 建立策略
- 加入 知識蒸餾 (KD) 技術，補足 RL 的缺點：
 -  提升訓練穩定性
 -  精簡模型，利於部署
 -  加速學生模型學習、強化泛化能力



RELATED WORK



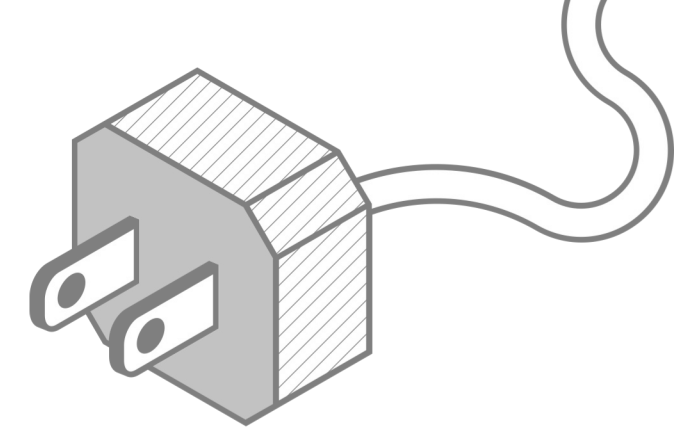
資源允入控制 (RAC) 在 5G 中的應用

- 傳統方法如 Fixed Slice、Greedy 無法適應高變化流量
- 無法兼顧不同優先級服務 (eMBB、URLLC、mMTC)
- 容易造成壅塞、延遲與服務中斷

[1] Y. Sun, M. Peng, et al., "*Application of Machine Learning in 5G Network Resource Management*", IEEE Wireless Communications, 2019.

[2] Almerexhi, Hind, et al. "*Are these comments triggering? predicting triggers of toxicity in online discussions.*" Proceedings of the web conference 2020.

RELATED WORK



強化學習應用於 5G 資源管理

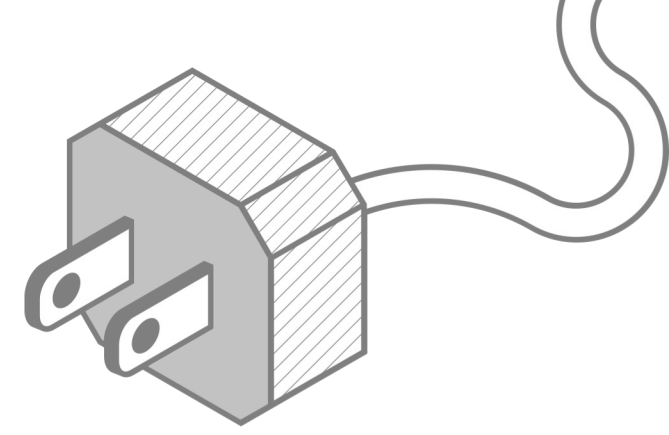
- DRL 被廣泛用於網路切片、排程與資源分配
- 常見方法：DQN、PPO、DDPG、SAC
- 具備自適應能力，但存在收斂慢、訓練不穩等問題

[3] H. Ye, G. Y. Li, "*Deep Reinforcement Learning for Resource Allocation in V2V Communication*", IEEE TWC, 2018.

[4] X. Zhang et al., "*A Deep Q-learning Network for 5G Network Slicing Resource Allocation*", IEEE Access, 2020.

[5] T. Haarnoja et al., "*Soft Actor-Critic: Off-Policy Maximum Entropy Deep RL with a Stochastic Actor*", ICML, 2018.

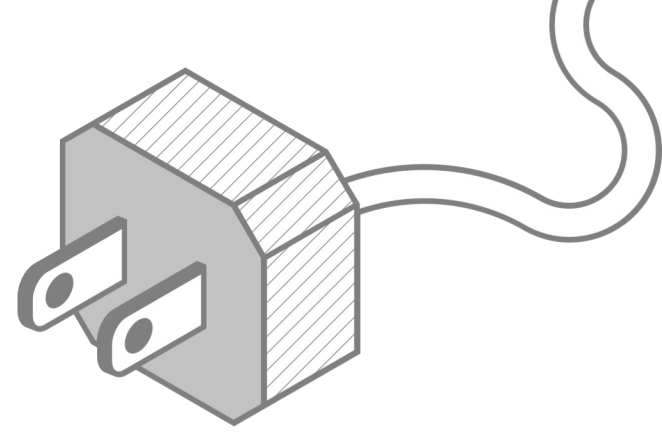
RELATED WORK



知識蒸餾在強化學習中的應用

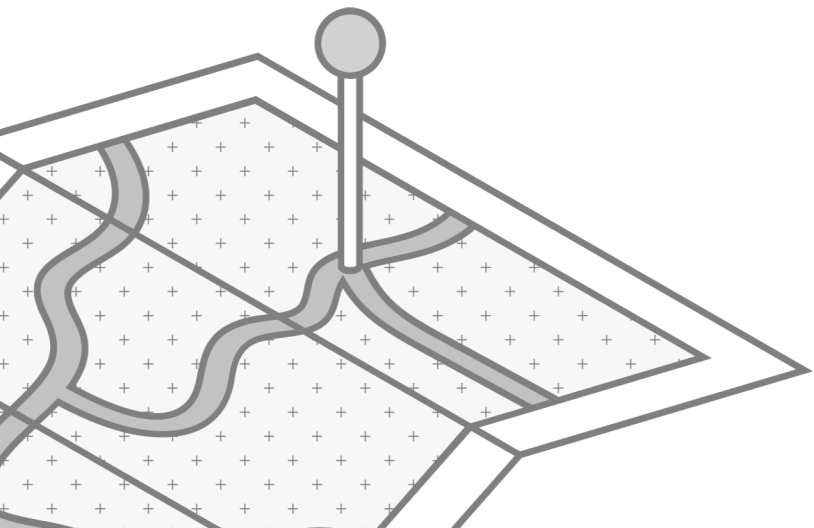
- 將教師策略壓縮轉移至輕量學生模型
- 補足 RL 模型大、樣本效率低、部署難等缺點
- 加速學習、提升泛化能力，適合即時應用與邊緣裝置

PROPOSED DESIGN

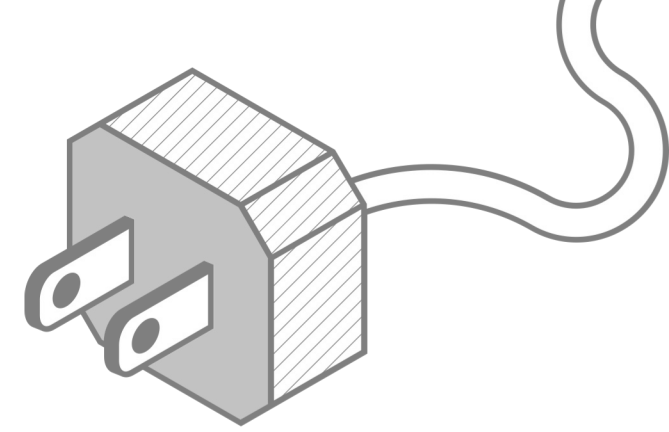


系統架構總覽

- 輸入：目前可用網路切片數量、不同行為服務請求（eMBB/URLLC/mMTC）狀態
- 核心決策器：Actor-Critic 強化學習架構
- 優化模組：知識蒸餾（KD）機制 → 輕量學生模型

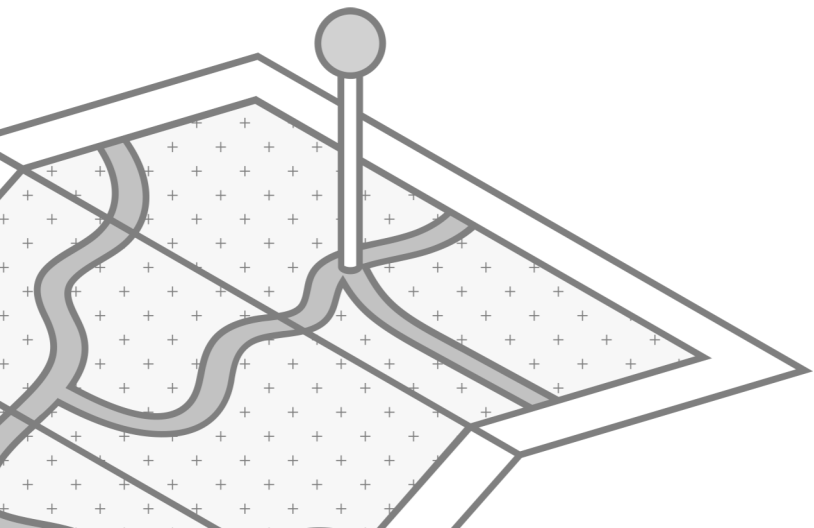


PROPOSED DESIGN

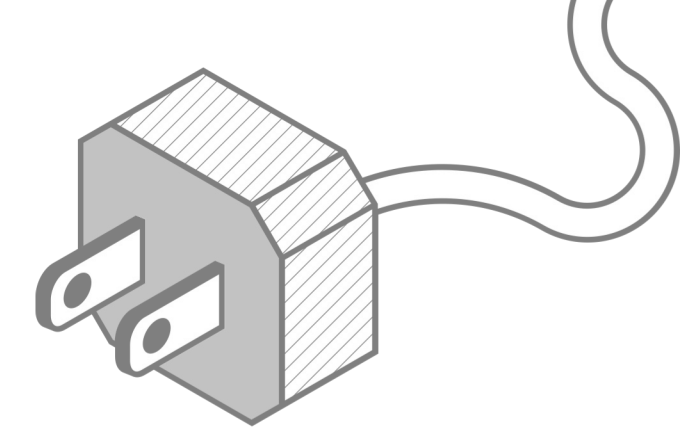


Markov Decision Process

- 狀態 S ：資源剩餘、隊列狀況、請求屬性
- 動作 A ：允入/拒絕/排隊，及切片分配選擇
- 獎勵 R ：成功接入高優先級流量為正獎勵，壅塞為懲罰






PROPOSED DESIGN

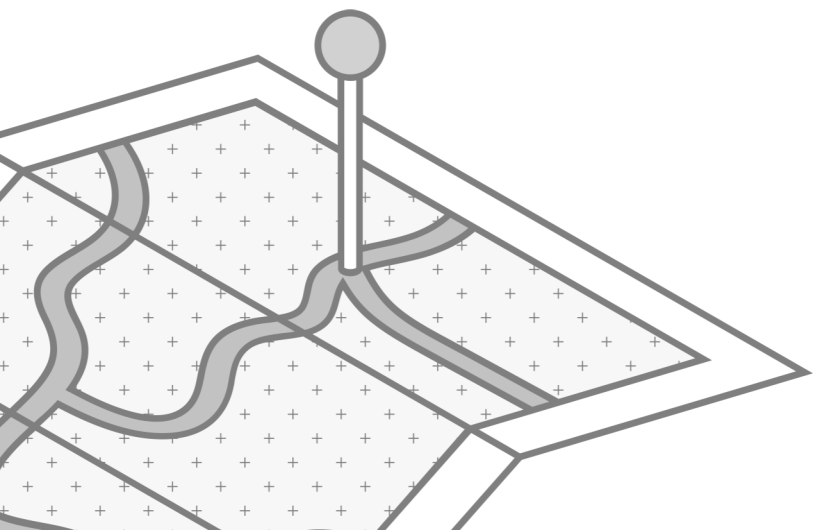


知識蒸餾 (KD) 整合設計

- 教師模型：完整 SAC 訓練出來的高效策略
- 學生模型：模仿教師行為分布與 Q 值估計
- 學習目標：降低策略差異與 KL 散度，達到「輕量 + 精準」效果

- 效果：

-  模型參數量大幅下降
-  加快推論與部署速度
-  保留策略效能



THANKS

R

O

U

LISTENLING