Special Topics: Machine Learning (ML) for Networking

COL867

Holi, 2025

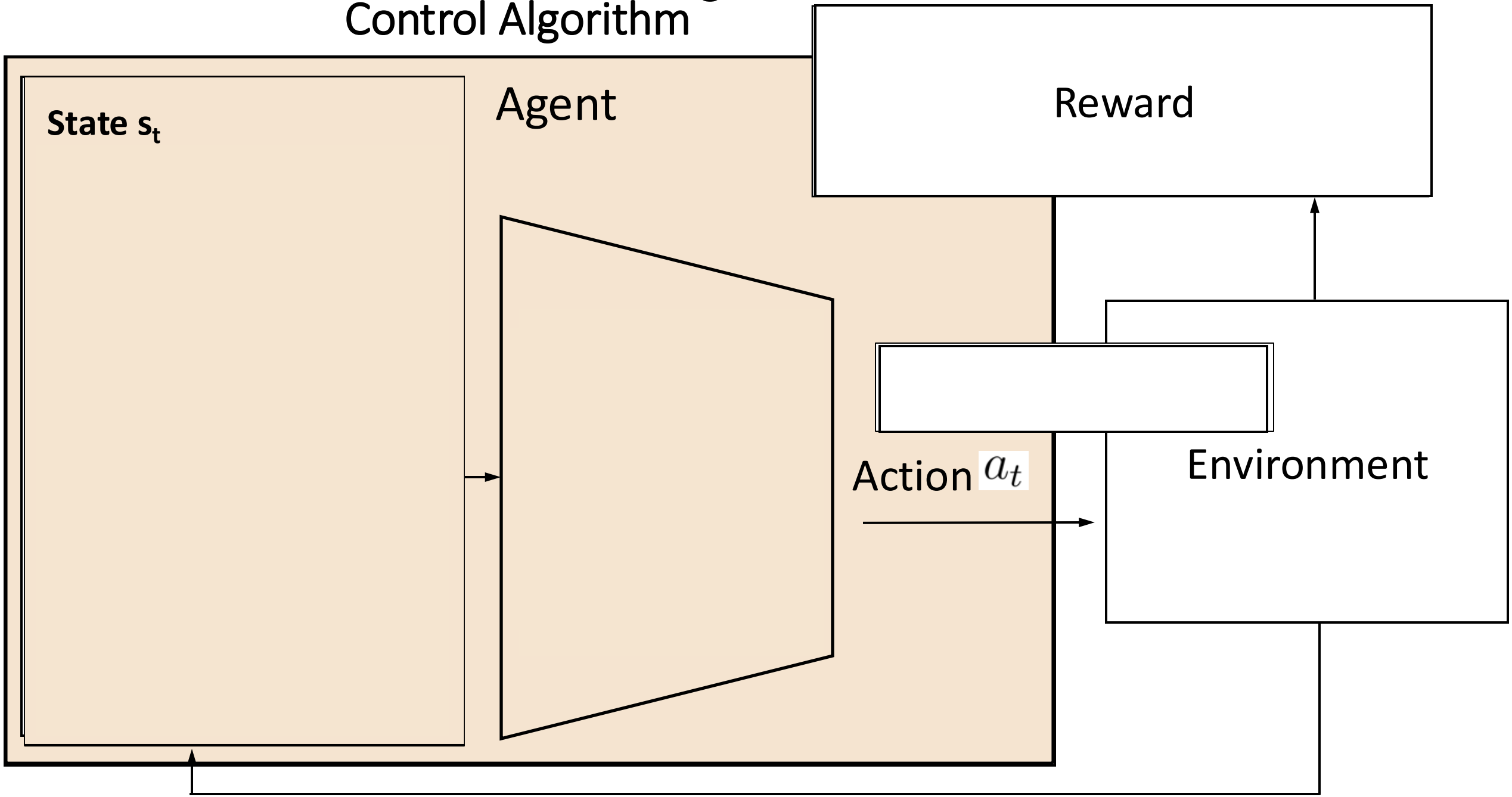
Resource Allocation

Tarun Mangla

Case Studies: ML for Specific Network Learning Tasks

- Application Classification
- Application Performance Monitoring
- Security
- **Resource Allocation**
  - Load balancing
  - Job scheduling
  - Network slicing
  - Rate control ... → Transport layer - CCA

TCP Aurora: RL-based Congestion Control Algorithm

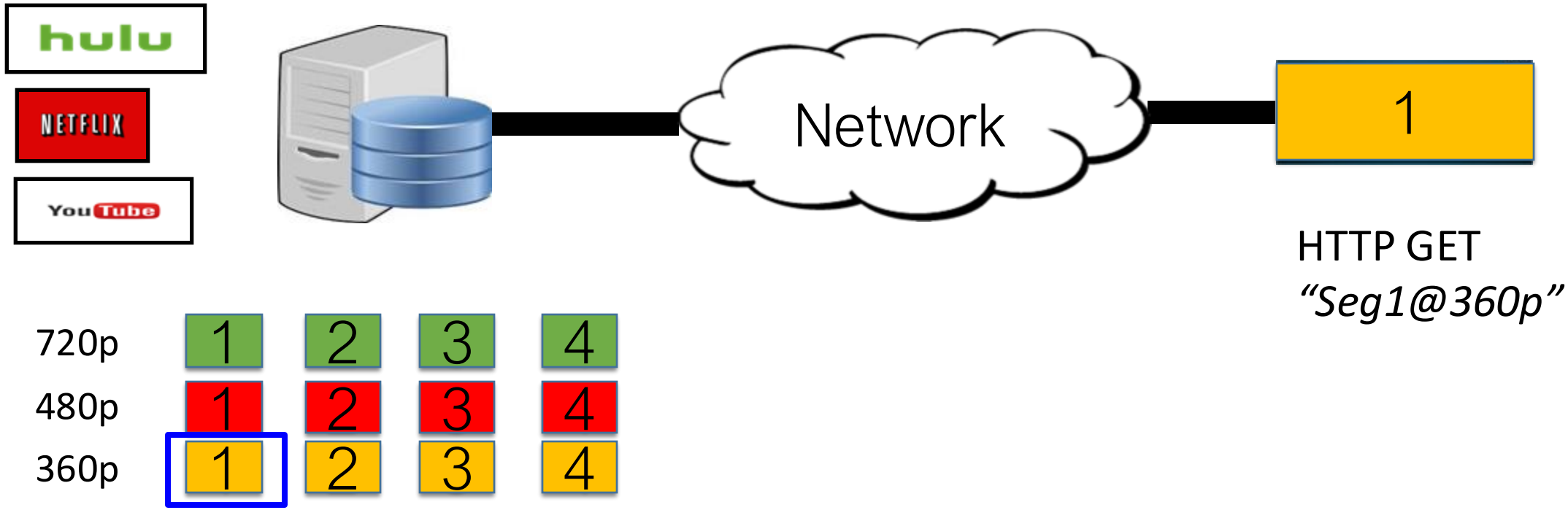


Rate Control (at End-host)

- Transport layer
- **Application layer: HTTP-based Adaptive Streaming**

Bitrate Adaptation

HTTP Adaptive Streaming (HAS) used for delivering Video over the Internet



Bitrate Adaptation

- What: Adapt player bitrate to Changing network conditions or changes in the viewing context (such as screen size, compute capability)
- Why: Support diverse clients and network conditions
- How: ??

low : 500 Kbps
Encoded
↳ 1 second

Download throughput : 1 Mbps

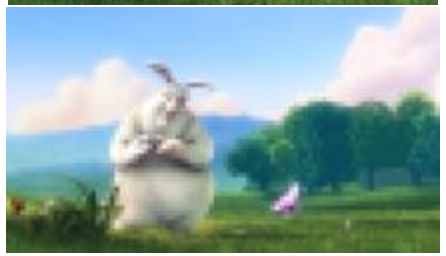
① Past chunk throughput : bps
② Remain buffer

Goal of Bitrate Adaptation

- Bitrate adaptation aims to optimize the following objective metrics:

Minimize Re-buffering

Maximize average bitrate

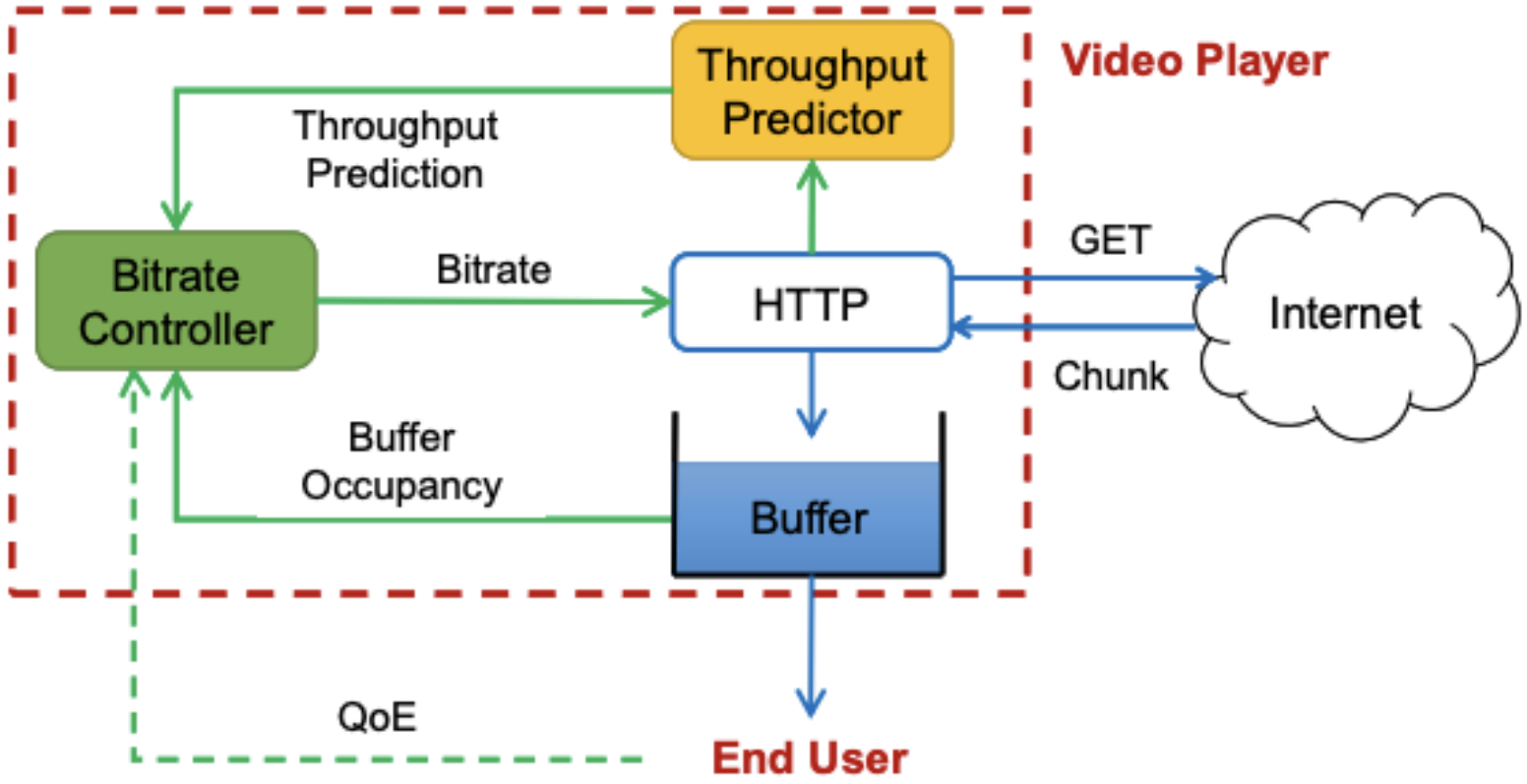


Minimize bitrate switches

Minimize startup latency

High (2Mbps
Med (1Mbps)
low (0.5Mbps)

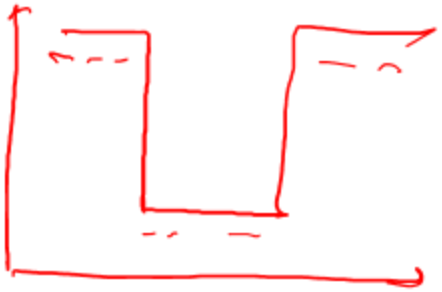
Abstract Player Model



Bitrate Adaptation

- What are the signals available to player?
  - Buffer-occupancy
  - Rate-based adaptation

Rate-based adaptation



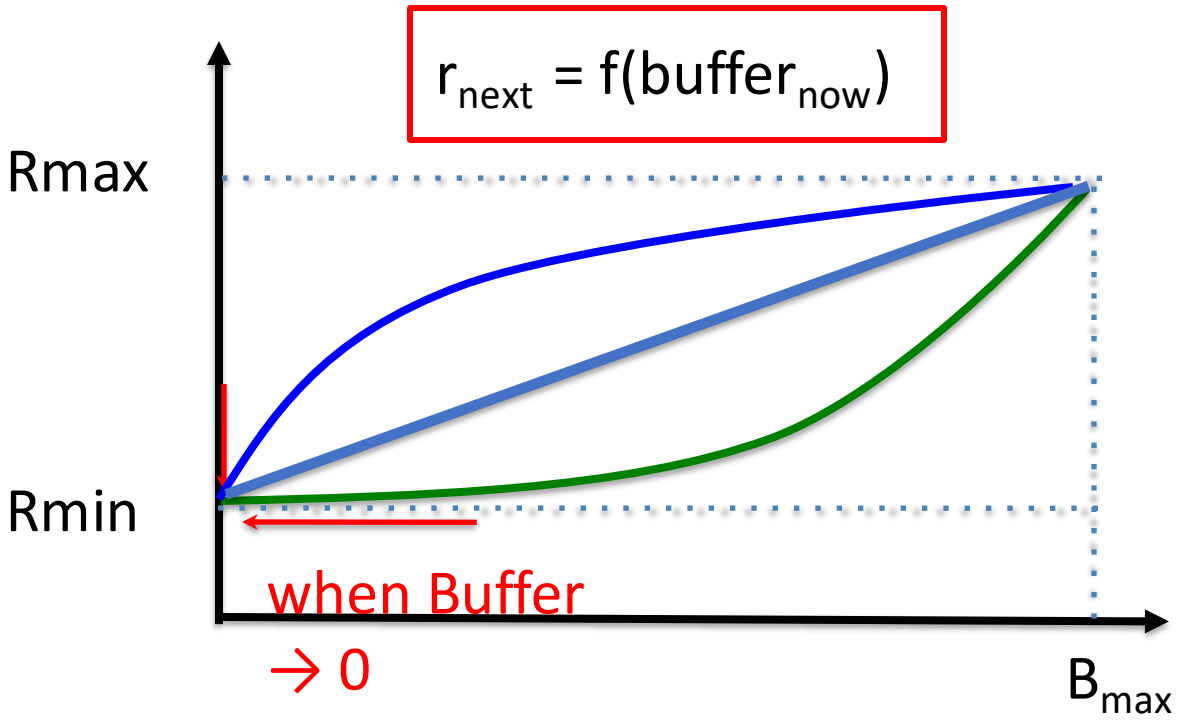
- Rate-based (use past throughput)
- Assumption/Insight: Past throughput is a good indicator of future throughput
- Example heuristic:

$$\max_{r \in R} (1+\alpha) \times r \leq T_{n-1}$$

$\alpha \geq 0$

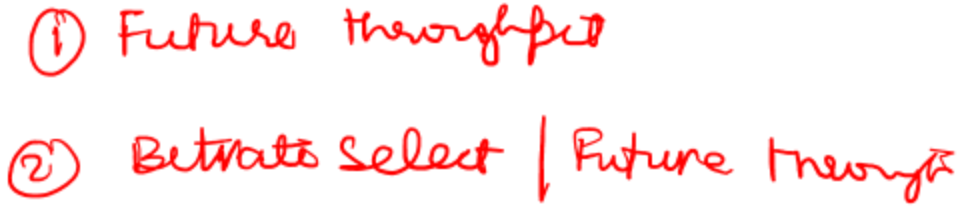
Buffer-based Adaptation (Netflix)

- Buffer-based (use current buffer occupancy)



More Systematic Framework

- **Model Predictive Control:** a control technique that uses a model to predict the future behavior of a system, and then calculates the best actions to take



- Can we use MPC for HTTP-adaptive streaming?

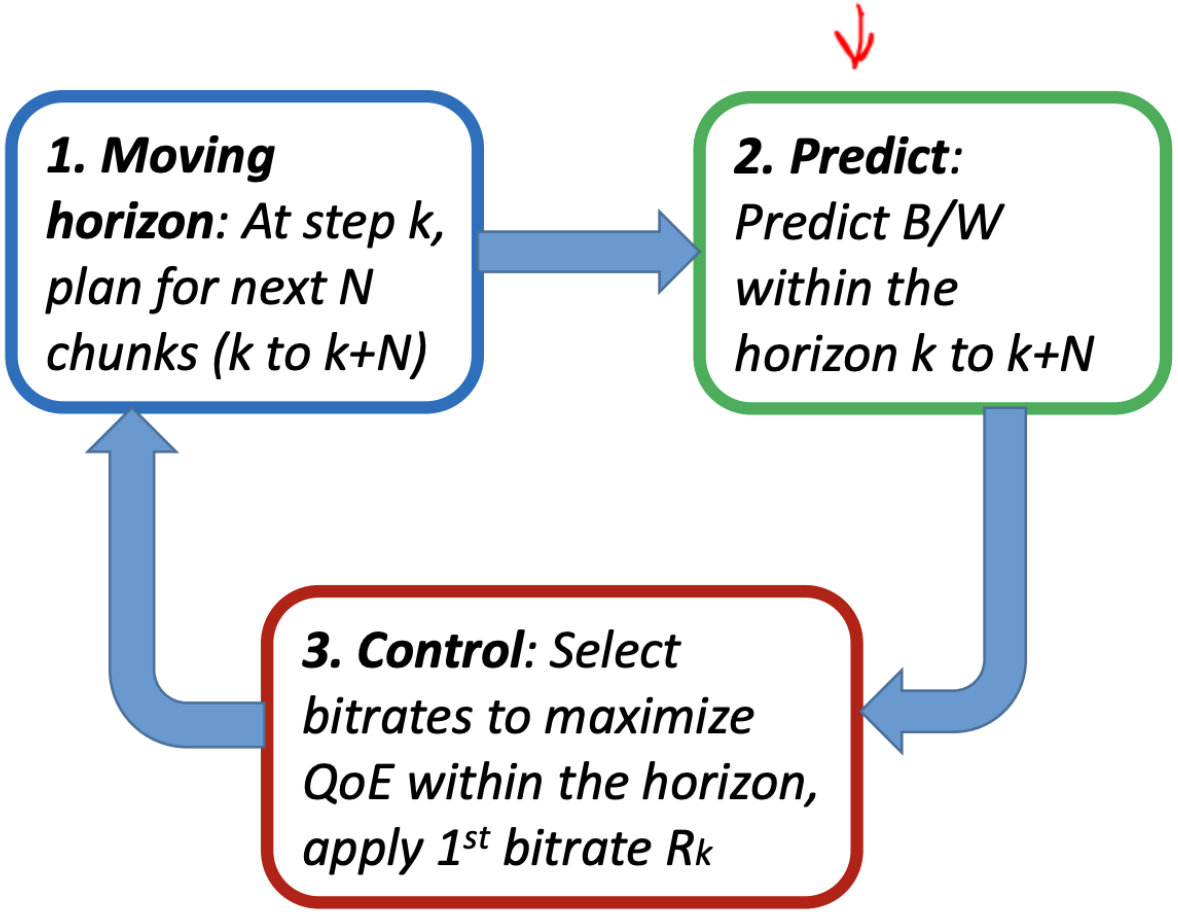
QoE Maximization Problem

$$\max_{R_1,\cdots,R_K,T_s} \quad QoE_1^K$$

last chunk : 2 Mbps

Model Predictive Control

Past Throughput



Open question: How to predict the future bandwidth?
- How to incorporate errors in the control?

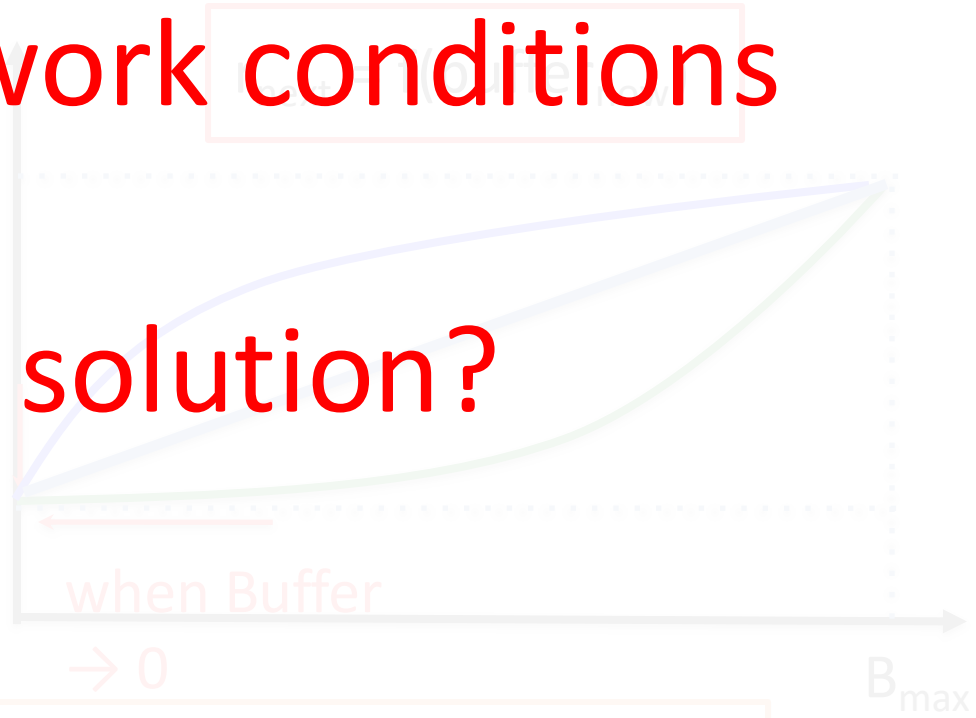
Bitrate Adaptation Heuristics

- Rate-based (use past throughput)
- Assumption/Insight: Past throughput is a good indicator of future throughput
- Example heuristic:

$$\max_{r \in R}(1+\alpha) \times r \leq T_{n-1}$$

- **Limitation: ?**

- Buffer-based (use current buffer occupancy)

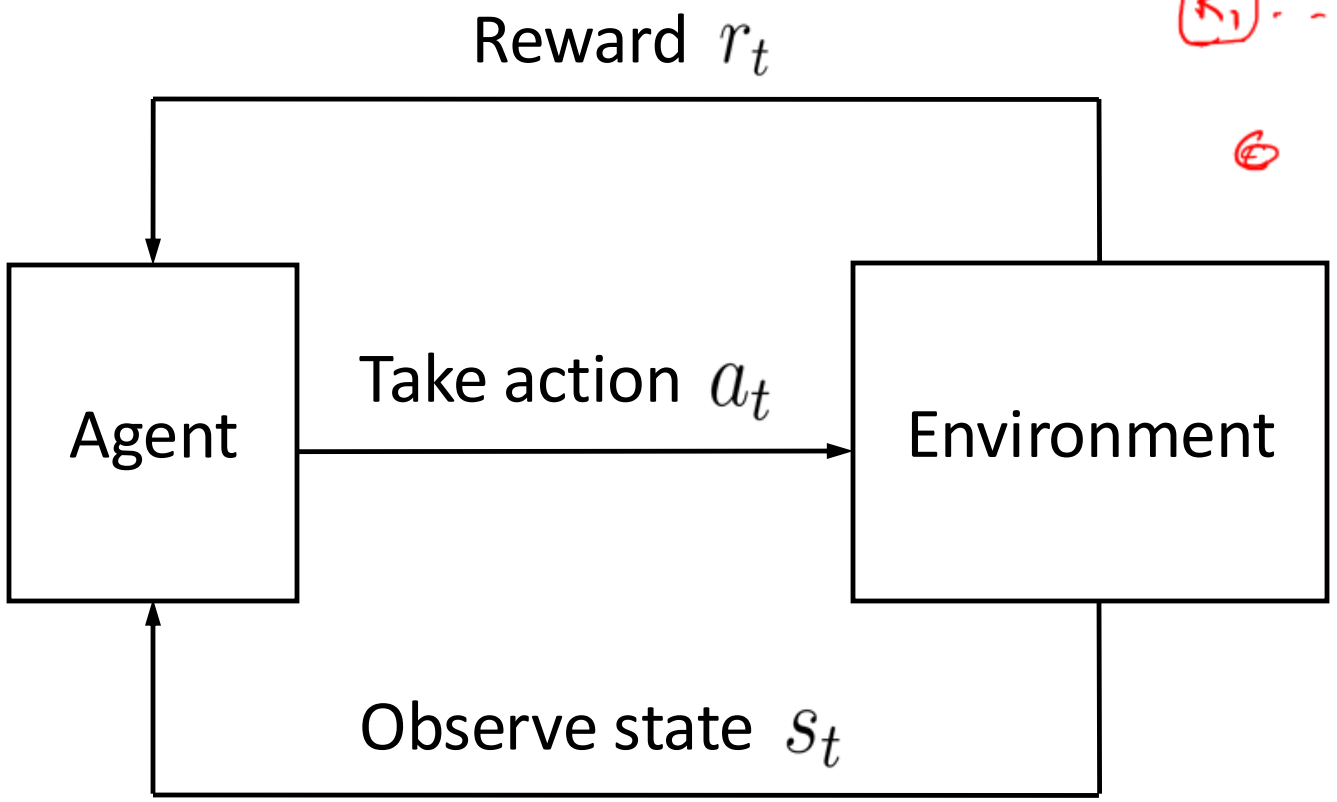


Both these heuristics have limitations

Solution: learn from video streaming sessions in actual network conditions

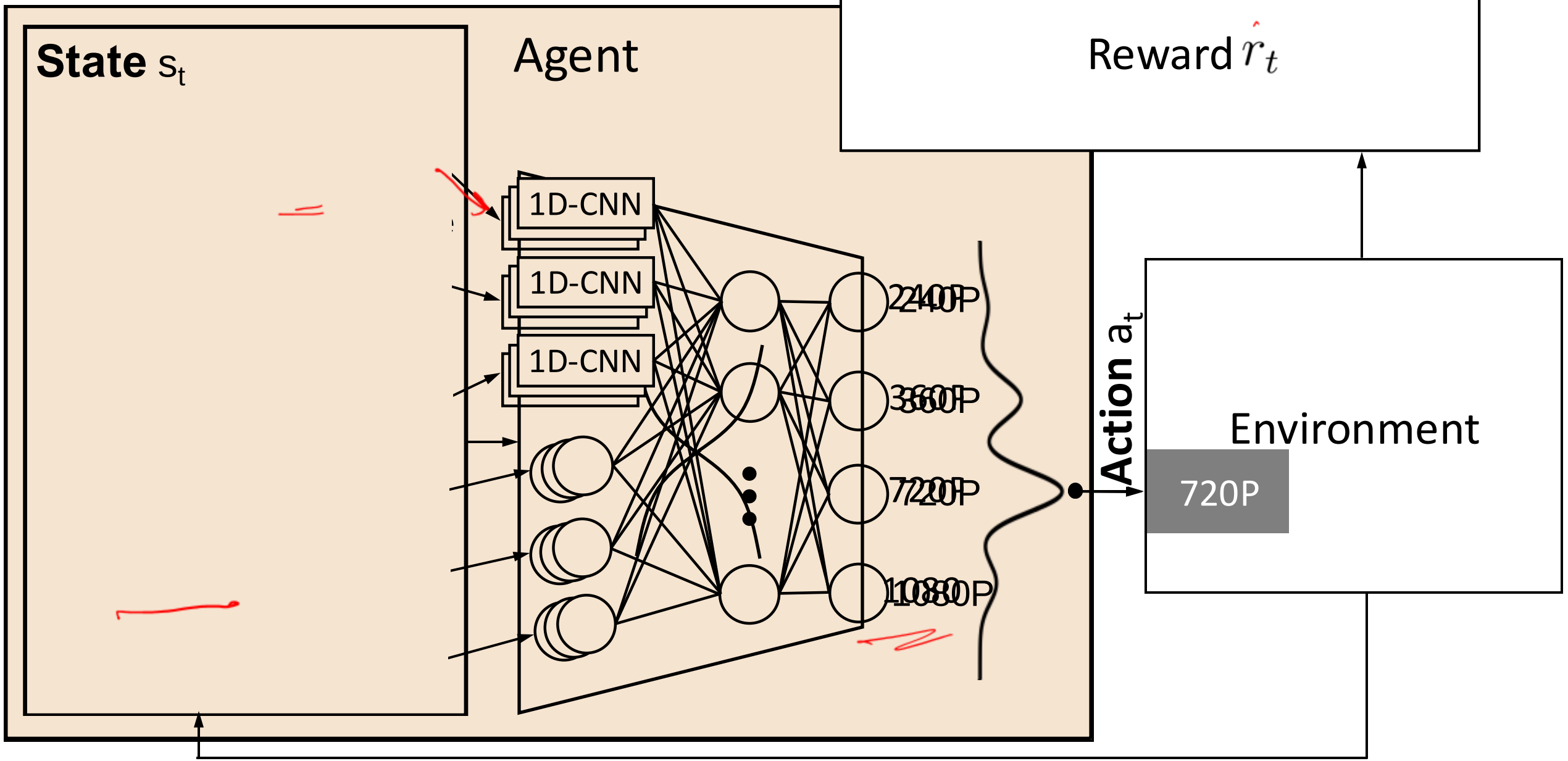
What kind of ML solution?

Reinforcement Learning

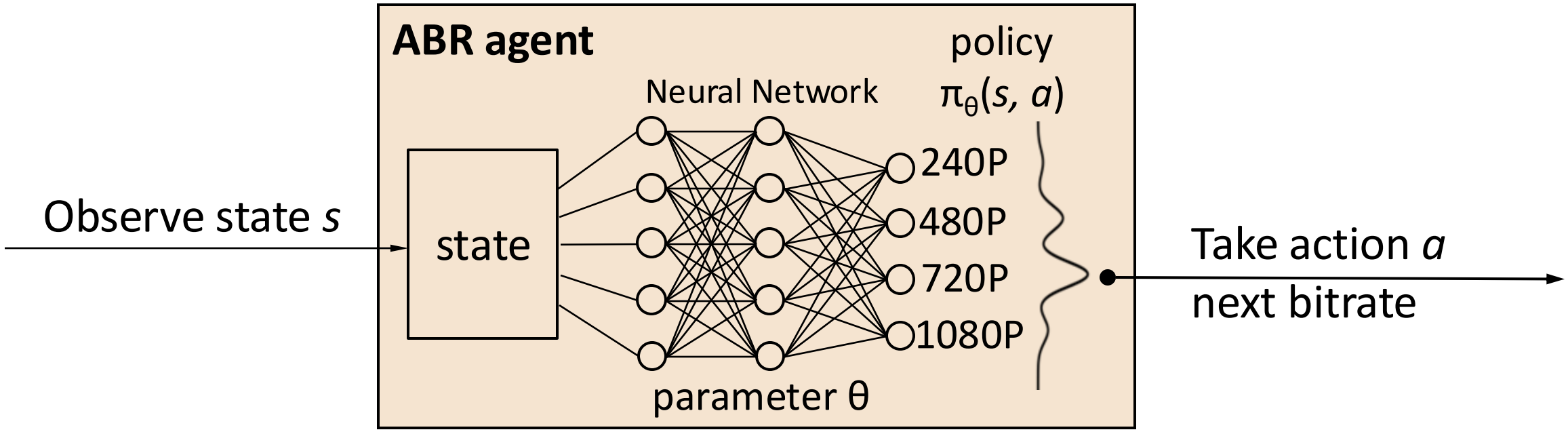


Goal: maximize the cumulative reward $\sum_t r_t$

Pensieve Design



How to Train the ABR Agent



Collect experience data: trajectory of [state, action, reward]

Training: $\theta \leftarrow \theta + \alpha \nabla_\theta \mathbb{E}_{\pi_\theta} \left[\sum_t r_t \right]$

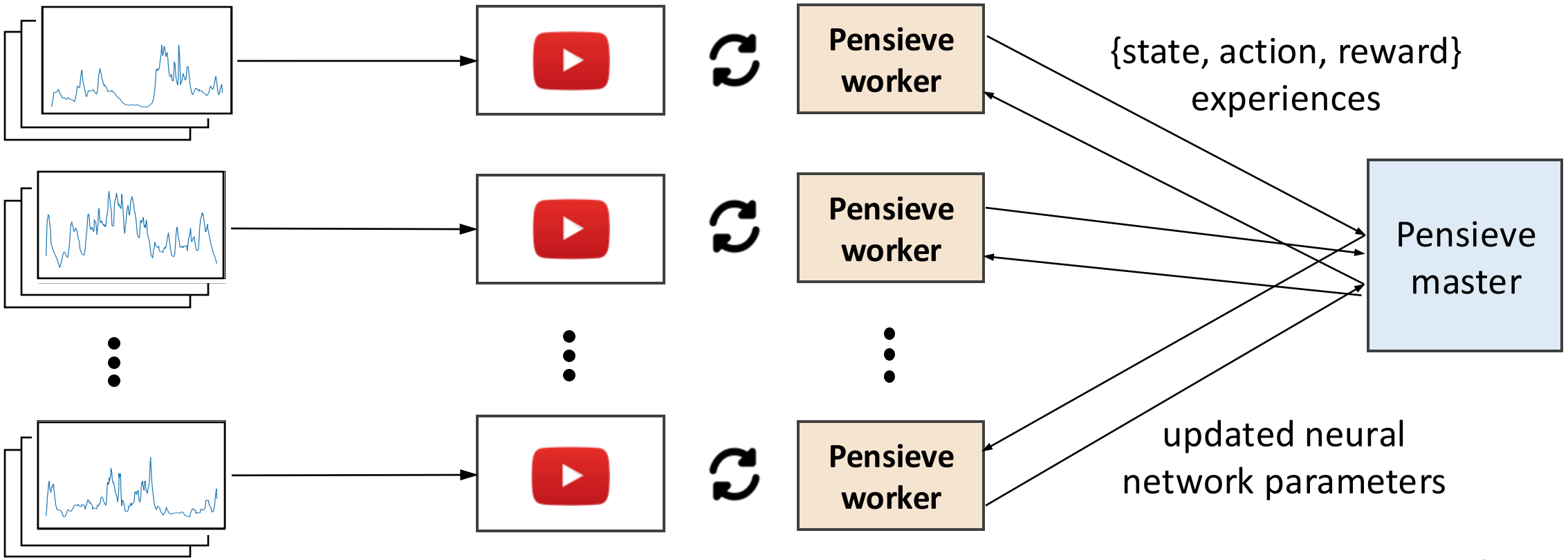
estimate from empirical data

Pensieve Training System

Large corpus of network traces
cellular, broadband, synthetic

Video playback
Fast chunk-level simulator

Model update
TensorFlow



① Simulator
② Emulation
③ Real-world

QoE Breakdown

Pensieve reduces rebuffering by 10-32% over second best algorithm

Summary

- Pensieve uses Reinforcement Learning to generate ABR algorithms
- Optimizes bitrate selection by directly learning from different network conditions
- Uses trace-driven simulation to experience many network conditions
- Open question: Can it generalize to diverse network conditions?

Summary

- **Module 1**: Case studies of network learning problems
  ① Security ② Application perf monitor
  ③ Resource Allocation ④ Traffic classification
- **Module 2:** Deep learning-based pipelines for network learning
  - **Generalized packet representation**
  - Foundation model
- **Next class: Generalized packet representation**
  - New Directions in Automated Traffic Analysis
- **No class on Friday. We will meet on Tuesday**