An Actor-Critic Approach to Congestion Control

Luca Giacomoni, George Parisis

University of Sussex

Coseners 2019

Outline

TCP's lmitations

- TCP is ill-suited for a number of modern network deployments
- Congestion control policy encodes the designer's assumptions along with the objectives of the algorithm
- Assumptions and objectives evolve as new applications/services and communication technologies are introduced.

Data-Driven Congestion Control

- Machine-learning approaches have emerged as an alternative to designing fixed congestion control policies
- We propose a novel approach that integrates **deep reinforcement learning** with an **actor-critic** algorithm

TCP Deciosion Making

- Congestion control protocols (or policies) define **rules**. A **rule** is a mapping between a network state and the sender's action.

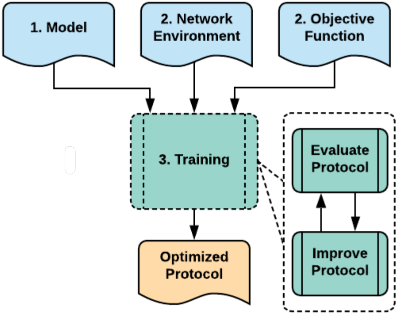
State s | Action a

Protocol Definition π

- If packet loss occurs: half transmission rate } Rule
- If packet received: congestion window = congestion window + 1/congestion window
- ... ...

$$a = \pi(s)$$

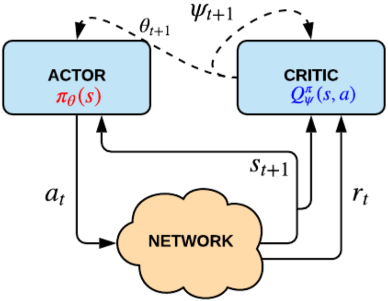
Approach



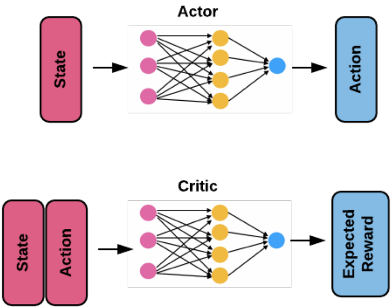
- **Outcome**: optimal **decision-making** policy based on given objective function
- Off-line training
- Model, Network Environment and Objective are arbitrary
- Optimized protocol could continue learning on-line

Actor-Critic in a glance

- Actor-critic (with Deep Reinforcement Learning) allows **continuous action** (i.e. transmission rates) and **state** spaces
- Policy π_θ is **deterministic**
- Critic Q_ψ drives actor's learning
- Training can be parallelised.



Model

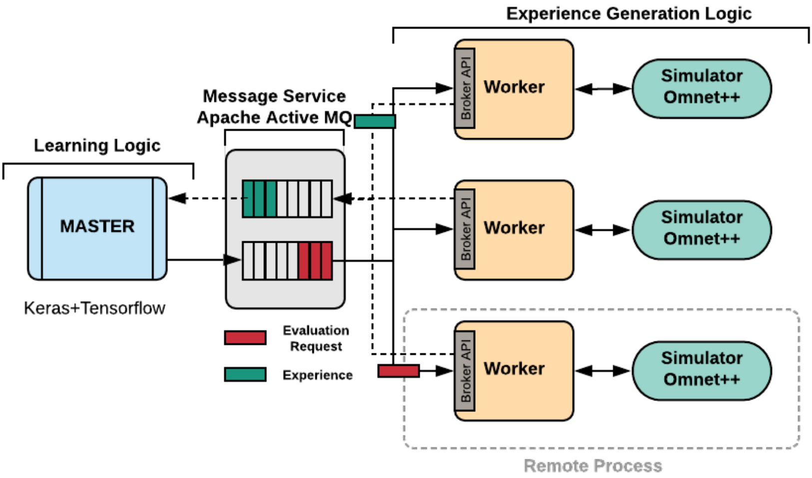


- **State**:
  - Signals observable by sender
  - E.g. RTT, loss rate, inter-arrival times of ACKs
- **Action**:
  - Transmission rate to adopt
  - Expressed as fraction of maximum congestion window
- **Objective Function**:
  - Performance metric to optimize
  - E.g. Weighted sum of throughput and delay

Challange

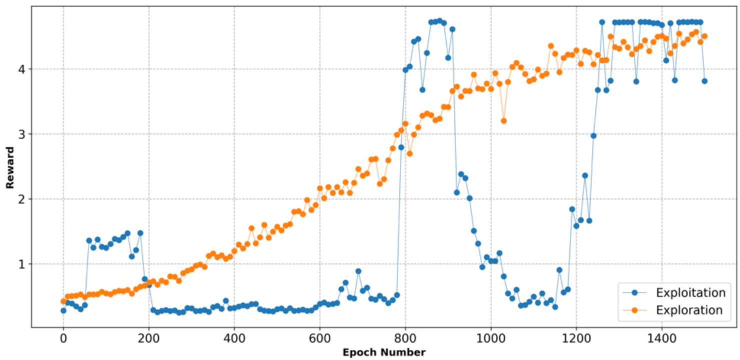
- Lack of suitable training environment
- Need to explore large state-action space
- Proposed solution: distributed and scalable training platform

Scalable and Distributed Training platform



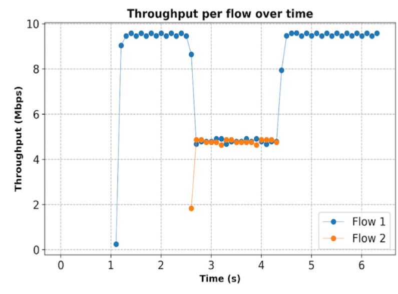
Preliminary Results

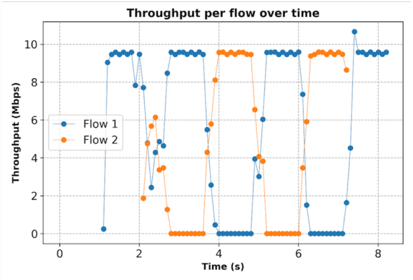
- Reward over training epoch
- Single sender
- Trained for 1500 epochs (epoch: single evaluation, improvement step)
- 12k simulations, 4h clock-time



Prelimininary Results

- Evaluation of fairness of 2 different reward functions
- Trained on fixed network conditions
- Objectives:
  1. $\sum log(throughput) - \delta \cdot log(delay)$
  2. $throughput - \delta \cdot delay$





Conclusions

- We are exploring the capabilities of actor-critic algorithm to otpimize end-host TCP-like congestion control protocol
- Actor-critic approach requires extensive exploration of the state-action space in order to converge to desired decision-making policy
- Off-line training allows to provide extra information that can be encoded into the deployable policy
- We implemented a distributed and scalable training system to evaluate (generate experience) on demand

Questions?