MARLIN: Soft Actor-Critic based Reinforcement Learning for Congestion Control in Real Networks

Raffaele Galliera
Alessandro Morelli
Roberto Fronteddu
Niranjan Suri

The University of West Florida
Institute for Human and Machine Cognition (IHMC)

Presenter: Raffaele Galliera - rg101@students.uwf.edu

The Networking Environment

- **A ″savage″ environment**
- **Highly heterogeneous**
- **Stochastic**
- **Data flows at high-speed rates**

Congestion Control

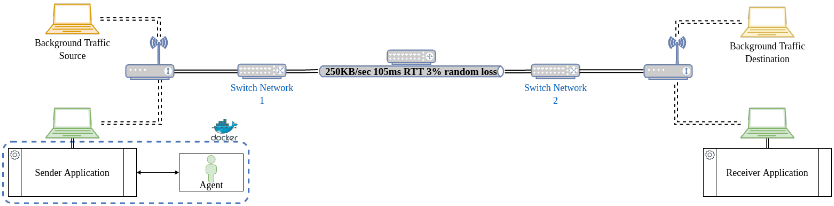
- The channel can saturate causing delays and re-transmissions
- Different heuristics are used (eg. **TCP Cubic**)
- **Congestion Window (CWND):** control over bytes allowed to be **in-flight**
- Heavy congestion can induce a link to an impracticable state

Some questions to answer...

Sim-based RL solutions presented limitations.

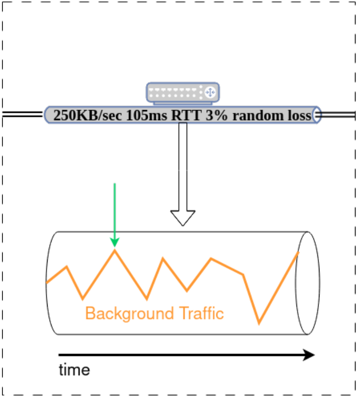
- *How can we face delays?*
- *How can we formulate a RL setting for the problem?*
- *How to deal with real, competitive traffic?*
- *Is it going to transfer to the real networks?*

The Environment


Background Traffic Source
Switch Network 1
250KB/sec 105ms RTT 3% random loss
Switch Network 2
Background Traffic Destination
docker
Sender Application
Agent
Receiver Application

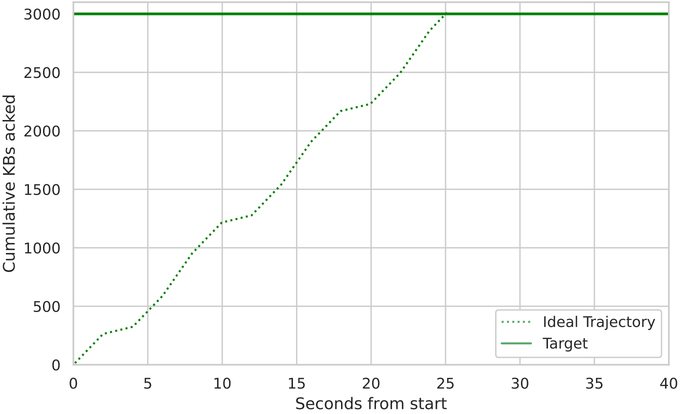
Dumbbell network.

The agent experience



An Infinite-horizon task with PEB

Evaluation

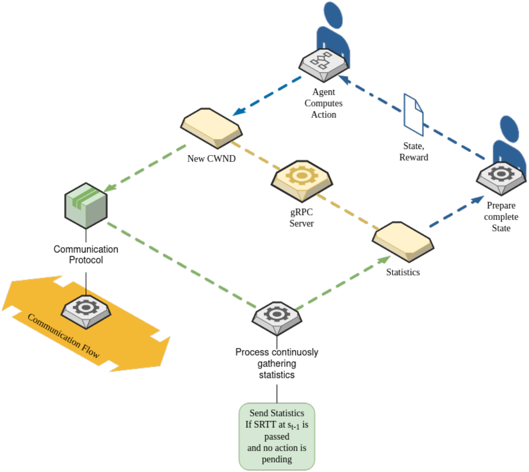


What we want the agent to do.

RL Keypoints

- **Action space:**$[-1, 1]$ tweaking the CWND
- **Non-blocking** communication
- **Learning algorithm:** Soft Actor-Critic (SAC)
- **PEB** during training, file transfer for testing
- **Episodic training**

Agent-Protocol communication



Statistics → State → Action → Reward

State

	Feature	Description		Statistic
1	Current cwnd	Current cwnd	1	Last
2	KBs Sent	Amount of KB sent *	2	Mean
3	New KBs sent	Amount of KB acked *	3	STD
4	Acked KBs	Amount of KB acked *	4	Min
5	Packets sent	Packets sent *	5	Max
6	Retransmissions	Number of packets retransmitted *	6	EMA
7	Instantaneous Throughput	Throughput *	7	Difference from Previous
8	Instantaneous Goodput	Goodput *		
9	Unacked KBs	Amount of KBs in flight		
10	Last RTT	Last rtt detected *		
12	Min RTT	Min rtt *		
12	Max RTT	Max rtt *		
13	SRTT	Smoothed rtt *		
14	VAR RTT	rtt variance *		
		* During the last rtt timeframe		

Every feature has 7 nested statistics with a 10 observations history.

Reward (1/2)

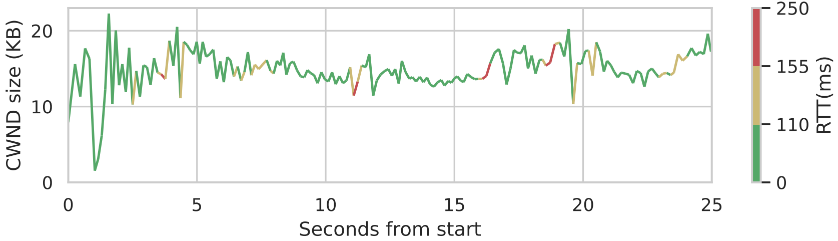
$$r_t = -\frac{target_t}{target_t + acked\_kilobytes_t^{cumulative}} \tag{1}$$

Reward (2/2)

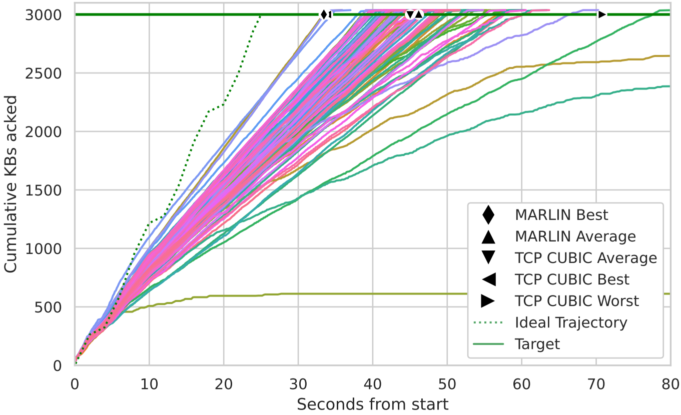
$$r_t = -\frac{target_t}{target_t + acked\_kilobytes_t^{cumulative} * (1 - penalties)} \tag{2}$$

$$penalties = \begin{cases} \alpha \frac{rtt_{diff}}{rtt_{min}^{ema}}, & \text{if } \frac{rtt_{diff}}{rtt_{min}^{ema}} < 1 \\ 0.99, & \text{otherwise} \end{cases} \tag{3}$$

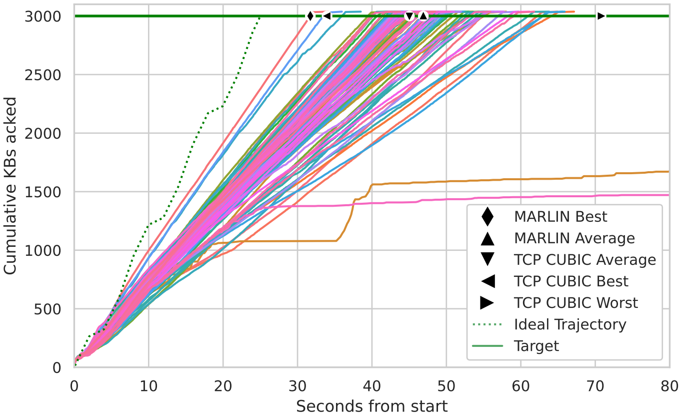
Agent in action



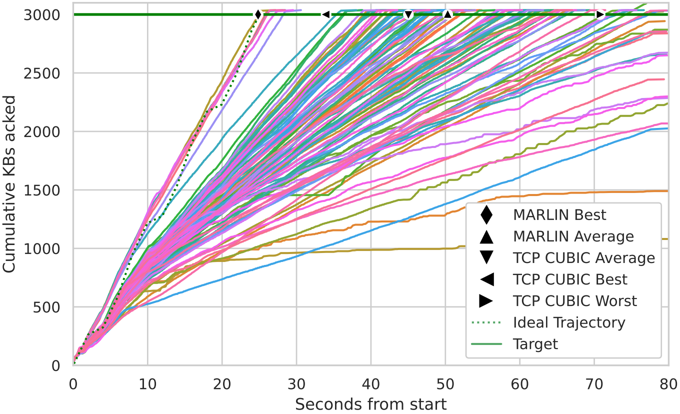
Results (1/3) - Single Traffic Pattern



Results (2/3) - Traffic Pattern Permutations



Results (3/3) - Alternative Reward



…more questions for future research.

Can we benefit from:

- *going from sim to real?*
- *using episode truncation to speed up training?*
- *Formulate a better state?*
- *integrating other ML models into the pipeline?*
- *letting the agent decide* ***when*** *to take an action?*

GitHub Repository

Backup - Hyperparameters

Hyperparameter	Value
Training steps	1×10^6
History length	10
Training episode length	200
Learning rate	3×10^{-4}
Buffer size	5×10^5
Warm-up (learning starts)	1×10^4 steps
Batch size	512
Tau	5×10^{-3}
Gamma	0.99
Training Frequency	$1/episode$
Gradient Steps	-1 (same as episode length)
Entropy regularization coefficient	″auto″ (Learned)
MLP policy hidden layers	$[400, 300]$

Table: Hyperparameters used in our experiment.

Backup - Reward

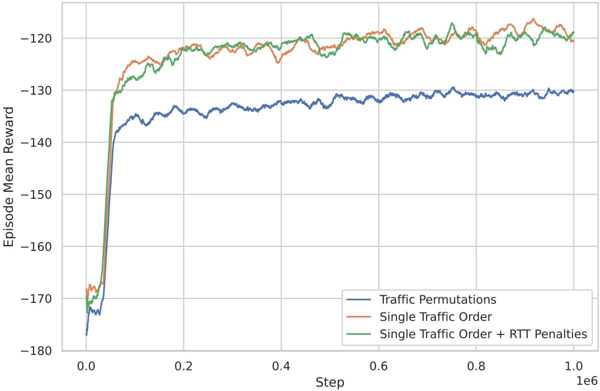


Figure: Rolling mean during training

Backup - Alpha

$$\alpha = \begin{cases} 1, & \text{if } |\frac{rtt_{diff}}{rtt_{min}^{ema}}| > 0.6 \\ 0.5, & \text{if } 0.1 < |\frac{rtt_{diff}}{rtt_{min}^{ema}}| \leq 0.6 \\ 0.3, & \text{if } 0.05 < |\frac{rtt_{diff}}{rtt_{min}^{ema}}| \leq 0.1 \\ 0.1, & \text{otherwise} \end{cases} \tag{4}$$

Backup - Traffic Patterns

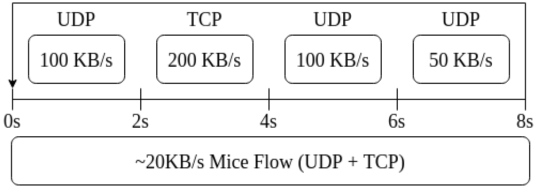


Figure: Generated Traffic Patterns

Backup - Traffic Patterns 2

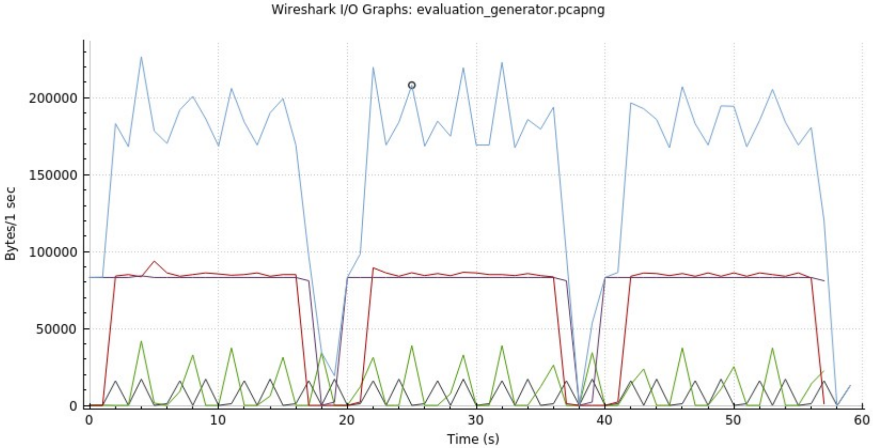


Figure: Generated Traffic Patterns