MULTI-AGENT META REINFORCEMENT LEARNING FOR PACKET ROUTING IN DYNAMIC NETWORK ENVIRONMENTS

Shan Sun, Mariam Kiran (advisor)

DAPHNE @ESnet

I. INTRODUCTION

Motivation

- Communication networks are **highly dynamic** (non-stationary).
- A **central controller** may be costly to control all devices and prone to malicious attacks.

Goal

- Efficiently minimize the **average packet completion time** while reducing **packet loss** across complex network topologies.
- Build a **decentralized solution**.

Overview

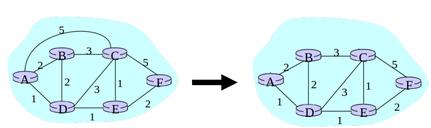
- We address this with a **multi-agent meta reinforcement learning algorithm** (MAMRL) by using policy optimization algorithm and learning a well-generalized policy initialization that can be quickly adapted to a different scenario with a few gradient steps.
- We test the proposed framework with **static** and **dynamically changing network topologies** respectively. We compare results to baseline controllers.

Keywords

Meta learning, multi-agent reinforcement learning, dynamic environment, packet routing, decentralized solution.

II. PROBLEM DEFINITION

- **Network:** A possibly **time-varying communication network**, which includes several routers and several links, is considered in this work. The solution works at the level of **core routers** of the Internet.



- **Routing:** Each packet in the network is originated from a router and destined for another router through available links. The queue of router follows the first-in-first-out criterion.
- **Objective:** The problem aims at finding the optimal transmission path between source and destination routers to minimize the **average packet completion time** while reducing **packet loss**.

III. INTELLIGENT CONTROLLER

Intelligent network controller (MAMRL):

- Policy optimization
- Decentralized solution
- Meta learning

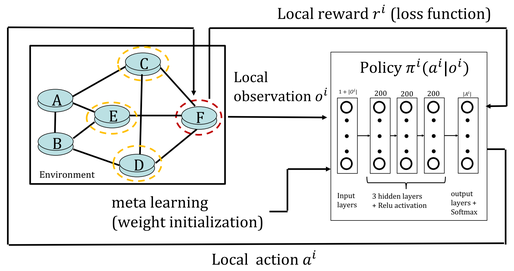


Figure 1: The framework of the packet routing problem.

Policy optimization:

- **Observation** o^i**:** 1) destination router of first packet in the local queue; 2) the latest five step actions taken by router i; 3) the router which has the longest queue among all the neighbor router of router i.
- **Action** a^i**:** next hop of current packet in the local queue.
- **Reward estimate** r^i**:** an estimate of the average of **all** x^i. x^i is negative average completion time of all the packets delivered to router i plus negative number of packet loss occurred in router i.

III. INTELLIGENT NETWORK CONTROLLER (CONT.)

Decentralized solution:

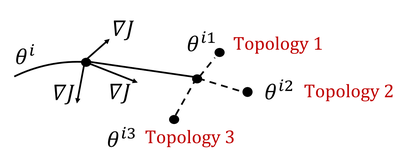
The goal of the network routing problem is to minimize the average packet completion time of the whole network while reducing packet loss. For each router, in order to get the **global** reward estimate r^i using only **local** information, we adapt the following **dynamic consensus algorithm** to estimate the global reward value,

$$r_t^i = x_t^i - y_t^i, \qquad y_{t+1}^i = \sum_{j \in \mathcal{N}_i} (r_t^i - r_t^j) + y_t^i, \tag{1}$$

- $\mathcal{N}_i$ is the **neighborhood** set of router i.
- It can be proved that $r_t^i \to \frac{1}{n} \sum_{\text{all } k} x_t^k$ within a few time steps as long as the network is connected [1].

Meta learning:

In order to address the packet routing in a **dynamic** environment, other than the policy optimization loop, we leverage the meta learning algorithm in [2] into the multi-agent packet routing case. As shown in the following figure, a good policy model parameter θ^i should be **close** to all the **optimal** parameters of different environments which makes θ^i the **best parameters initialization** that can quickly adapt to different environments.



[1]F. Chen, Y. Cao, and W. Ren. "Distributed average tracking of multiple time-varying reference signals with bounded derivatives." IEEE Transactions on Automatic Control, 2012.

[2] Finn, Chelsea, Pieter Abbeel, and Sergey Levine. "Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks." ICML. 2017.

IV. RESULTS

We test the MAMRL algorithm in packet routing problem with the **static topology** and **dynamically changing topologies** respectively.

Results for static topology

we compare the results of using MAMRL and the shorted path algorithm (SPA) at various network load levels.

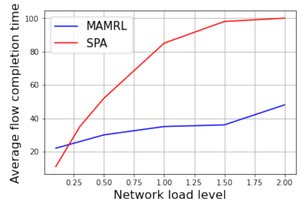


Figure 2: Average packet completion time results for static topology.

Results for dynamic topology

We compare results of using the following three controllers: 1) training the policy from the initialization parameters obtained by MAMRL; 2) training the policy from randomly initialized weights; 3) shortest path algorithm (SPA).

At the beginning of the experiment, the routers are in a communication network presented in Figure 1, then the link between router A and router B disappear at a certain moment.

Since the SPA algorithm relies on the topology of the network and cannot be adaptive, if the network topology changes, the SPA controller would keep sending packets to the failed link which causes huge packet loss. Reinforcement learning algorithms are model-free controllers and the policy of the model-free controller can be improved by interacting with the environment.

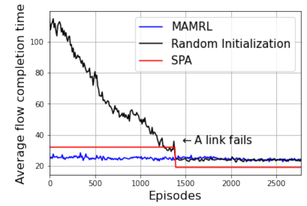


Figure 3: Average packet completion time results for dynamic topologies.

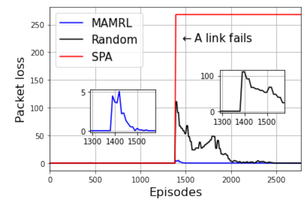


Figure 4: Packet loss results for dynamic topologies.

V. CONCLUSION

When there is static topology, the MAMRL performs better than SPA at high network load level.
When there is network topology change, **1)** model-free reinforcement learning algorithm can reduce the packet loss significantly while offering comparable average packet completion time compared to the shortest path algorithm, and **2)** the MAMRL can adapt to the new topology within much fewer episodes compared to non-meta reinforcement learning.

ACKNOWLEDGEMENT

We would like to express our gratitude to Dr. Bashir Mohammed and Prof. Wei Ren for their useful suggestions and critiques of this research work. This work was supported by the U.S. Department of Energy, Office of Science Early Career Research Program for 'Large-scale Deep Learning for Intelligent Networks' Contract no FP00006145.