# Multi-agent Reinforcement Learning for Improving Supply Chain Visibility in Inventory Management

Bo Zhang, Wen Jun Tan, Wentong Cai
*School of Computer Science and Engineering*
*Nanyang Technological University*
Singapore, Singapore
{bo003, wjtan, aswtcai}@ntu.edu.sg

Allan N Zhang
*Agency for Science, Technology and Research*
*Singapore Institute of Manufacturing Technology*
Singapore, Singapore
nzhang@simtech.a-star.edu.sg

***Abstract*—This paper proposes a novel approach to enhance supply chain (SC) visibility, cooperation, and performance during inventory management while effectively mitigating the risk of information leakage by leveraging machine learning techniques. The SC inventory policies are optimized using multi-agent reinforcement learning (MaRL) and SC network topological information. Furthermore, we conduct a simulation-based evaluation that demonstrates the superior performance of our method compared to alternative optimization approaches. This research effectively addresses the dual objectives of ensuring information security and achieving cost reduction in SC inventory management.**

***Index Terms*—supply chain, reinforcement learning, simulation, inventory management**

## I. Introduction

In recent years, there are increasing numbers of SC collaborations and information sharing aimed at improving overall SC operations, which have proven beneficial in enhancing operational profitability, service quality, and mitigating demand uncertainties for individual firms as well as the entire SC [1]. However, the current approaches for improving SC visibility require access to sensitive data [2]. Due to privacy concerns, SC entities are reluctant to share crucial information with other entities.

In this paper, we proposed a multi-agent reinforcement learning (MaRL) model that balances information sharing and privacy concerns. Each firm can create its own critic and policy network to maintain privacy. The policy network is used to make decisions independently based on the current SC state. The critic network assesses the value of decisions made by the policy network and guides its parameters update. Furthermore, our proposed approach incorporates topological information into the model to enhance communication efficiency and guides agents to focus on useful shared information.

The contributions of our work are as follow:

- We propose a MaRL model that incorporates SC network topology to optimize the SC inventory cost.
- Our approach facilitates distributed decision-making based on local information.
- We conduct a simulation-based evaluation that demonstrates the superior performance of our method compared to alternative optimization approaches.

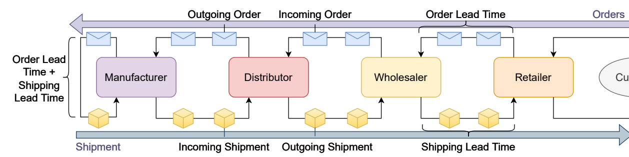


Fig. 1. A Multi-Stage Supply Chain Network.

## II. Problem Settings & Simulation Model

Fig. 1 shows a standard multi-stage SC network. We simulate the SC network using discrete event simulation and use it to evaluate the inventory policies. In the simulation, orders and shipments are scheduled as events. The customer generates the order event and periodically sends to the retailer, which sends the orders to the wholesaler. The order will then be sent to the distributor before finally reaching the manufacturer. After fulfilling the order, the manufacturer sends the shipments to the distributor, which in turn supplies to the wholesaler. The shipments will then be sent to to the retailer, before reaching the customer. In each time step of the simulation, each entity generates an order quantity based on its inventory policy and sends it to its predecessor. Its predecessor should fulfill the order to the best of its ability based on its inventory level and initiate a shipment accordingly. The objective of SC optimization is to minimize overall inventory costs by optimizing the inventory policy, thereby reducing inventory holding expenses across the entire supply chain.

## III. Method

Fig. 2 depicts the main architecture of our approach. Inspired by the Multi-Actor-Attention-Critic (MAAC) approach [3], Our model is trained in a centralized manner using individually encoded input state and incorporates SC network topological information, represented using an adjacency matrix, allowing for effective coordination among agents; whereas the decision making will be performed by each SC entity using the trained model in a decentralized fashion.

## IV. Experiment & Evaluation

In this section, we evaluate SC inventory costs of our proposed approach by using the simulation as a testbed under

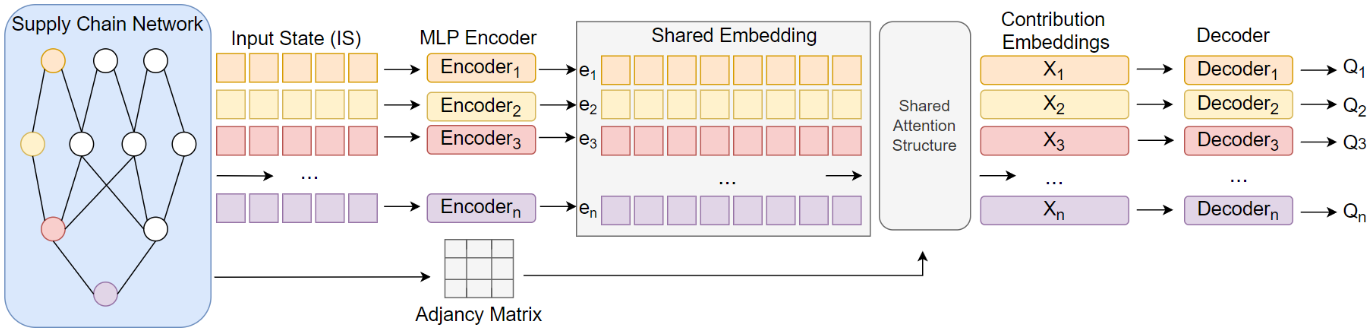


Fig. 2. Model layout.

various existing real SC networks [4]. The SC networks are selected according to different operational complexity (CPX) [5]. The experiment results illustrate the superior performance of our proposed method compared to other approaches.

### *A. Experimental Results in Term of Operational Complexity*

We evaluate our proposed approach by comparing to base-stock (BS) policy [6] and decomposition-aggregation (DA) heuristic approaches [7] in Fig 3. In the experiments, our method is used as a normalized baseline for comparing with other approaches. The experiment results indicate that our approach is capable of reducing overall inventory costs and benefiting all entities within the various complex operational systems of the SC.

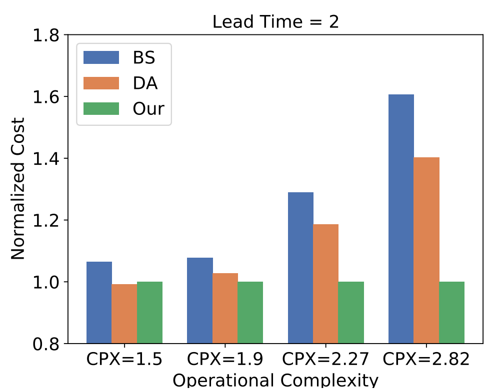


Fig. 3. Normalized cost change of different traditional optimization methods with increasing network size.

### *B. Experimental Results in Term of Convergence*

Figure 4 shows the inventory cost curves during the training process for a real SC network. Although the original MAAC approach converges faster, our approach achieves lower inventory costs. Results show that our approach can effectively balance exploration and exploitation during training, resulting in reduced inventory costs.

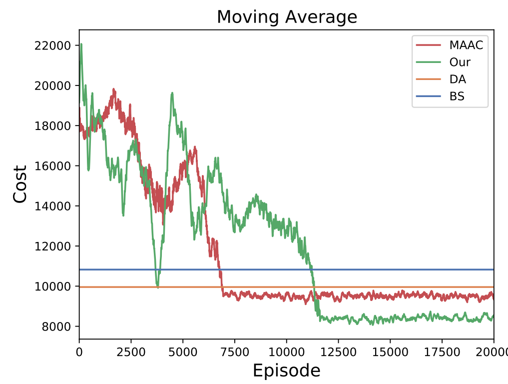


Fig. 4. Cost curves of training process

## V. Conclusions

This paper proposes a novel MaRL model that utilizes network topology to optimize inventory costs across the entire supply chain by fostering increased supply chain cooperation and decentralized decision-making. To enhance privacy protection for firms, our future research endeavors will integrate Federated Learning into our model. This approach will enable all firms to train their models locally without the necessity of sharing raw data, ensuring heightened privacy and confidentiality.

## References


[1] Panahifar, F., Byrne, P. J., Salam, M. A., and Heavey, C. (2018). Supply chain collaboration and firm's performance: the critical role of information sharing and trust. Journal of Enterprise Information Management, 31(3), 358-379.

[2] Kalaiarasan, R., Olhager, J., Agrawal, T. K., & Wiktorsson, M. (2022). The ABCDE of supply chain visibility: A systematic literature review and framework. International Journal of Production Economics, 248, 108464.

[3] Iqbal, S., & Sha, F. (2019, May). Actor-attention-critic for multi-agent reinforcement learning. In International conference on machine learning (pp. 2961-2970). PMLR.

[4] Willems, S. P. (2008). Data set—real-world multiechelon supply chains used for inventory optimization. Manufacturing & service operations management, 10(1), 19-23.

[5] Cheng, C. Y., Chen, T. L., & Chen, Y. Y. (2014). An analysis of the structural complexity of supply chain networks. Applied Mathematical Modelling, 38(9-10), 2328-2344.

[6] Graves, S. C., & Willems, S. P. (2000). Optimizing strategic safety stock placement in supply chains. Manufacturing & Service Operations Management, 2(1), 68-83.

[7] Rong, Y., Atan, Z., & Snyder, L. V. (2017). Heuristics for base-stock levels in multi-echelon distribution networks. Production and Operations Management, 26(9), 1760-1777.