# Fourth Party Logistics Routing Optimization Problem Based on Conditional Value-at-Risk Under Uncertain Environment

Guihua Bo[1]*, Qiang Liu[2], Huiyuan Shi[3], Xin Liu[4], Chen Yang[5], Liyan Wang[6]
College of Information and Control Engineering, Liaoning Petrochemical University, Fushun, Liaoning, China[1, 2, 3, 4, 5, 6]
State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang, China[3]

***Abstract*—In order to improve the level of logistics service and considering the impact of uncertainties such as bad weather and highway collapse on fourth party logistics routing optimization problem, this paper adopts Conditional Value-at-Risk (CVaR) to measure the tardiness risk, which is caused by the uncertainties, and proposes a nonlinear programming mathematical model with minimized CVaR. Furthermore, the proposed model is compared with the VaR model, and an improved Q-learning algorithm is designed to solve two models with different node sizes. The experimental results indicate that the proposed model can reflect the mean value of tardiness risk caused by time uncertainty in transportation tasks and better compensate for the shortcomings of the VaR model in measuring tardiness risk. Comparative analysis also shows that the effectiveness of the proposed improved Q-learning algorithm.**

***Keywords*—Logistics service; routing optimization; tardiness risk; conditional value-at-risk; improved Q-learning algorithm**

## I. Introduction

With the deepening of economic globalization and the intensification of competition in the logistics market, service-oriented manufacturing enterprises are pursuing more hierarchical and integrated logistics services, so the problem of third-party logistics (3PL) is becoming increasingly prominent. For example, the Toyota listed on its website that Toyota outsources its logistics services to the 3PL to focus on their core product business, on 10th May 2021 [1]. In order to obtain more orders, competition among 3PL enterprises is becoming more fierce. There is a lack of resource sharing among various 3PL service providers, which makes it difficult to accurately grasp logistics information and meet the current rapidly growing logistics demand. As a result, the traditional 3PL model cannot adapt to the pace of the times and restricts the progress of logistics globalization. Instead of relying on the 3PL providers, some large manufacturers (e.g. Haier [2] and Hisense) have recently cooperated with Cainiao Logistics (the largest fourth party logistics [4PL] provider in China) in supply chain solutions. 4PL [3] is necessary to integrate and effectively connect resources so as to achieve complementary advantages. 4PL supplier is an integrator of the supply chain that integrates and manages different resources, capabilities, and technologies of a company and complementary service providers, and performs a detailed analysis of the entire supply chain system or industry logistics system where enterprise customers are located. Thus, it provides a comprehensive solution for the design, construction, and operation of the supply chain. Many enterprises have used 4PL to complete their own logistics tasks. For example, Cainiao Logistics, the largest Chinese 4PL provider founded by the Alibaba Group, incorporates over thirty 3PL providers to serve Taobao.com and Tmall.com, two largest online markets in China [4].

With the continuous development of 4PL, many experts and scholars have begun to focus on and study various aspects of 4PL. They are committed to in-depth exploration of issues related to risk management [4], network design [5-6], combinatorial auctions [7], contract design [8-9], and routing optimization [10] in 4PL, which have proposed a series of theoretical and practical achievements.

The Fourth Party Logistics Routing Optimization Problem (4PLROP) is one of the most important problems and a hot topic about 4PL. As the creator of transportation plans, 4PL needs to optimize and design distribution routes. Based on factors such as transportation time, transportation capacity, reputation indicators, and throughput of 3PLs, it selects and allocates 3PLs that provide distribution services to achieve path optimization and select satisfactory transportation plans for enterprises.

However, in the actual delivery process, due to unpredictable reasons such as transportation, bad weather, and human error operations, all of them may cause the transit time and transportation time to be not fixed but random during the transportation process, which may lead to the risk of delivery tardiness. The impact of this randomness makes 4PL suppliers to be unable to provide timely delivery plans that satisfy customers, it may lead to additional costs and a decrease in customer satisfaction. This situation may even affect the reputation of 4PL enterprises, leading to customer churn, and having adverse effects on their long-term development. Therefore, when solving 4PLROP, the influence of the risk caused by the uncertainty cannot be ignored.

This article adopts CVaR to describe and measure the average risk of tardiness induced by multiple factors in the real delivery process of 3PL providers in a 4PL operation. In addition, a mathematical model is proposed with CVaR minimization as the objective function and delivery cost as the constraint, and the suggested CVaR model is compared to the VaR model [11]. Then, an improved Q-learning algorithm is proposed to solve examples with different scales of the two models, and the effectiveness of the proposed CVaR model is verified. Finally, the proposed algorithm is compared with GA

embedded with Dijkstra [12] and the results verify the feasibility of improved Q-learning algorithm for solving this problem.

The principal contributions of this article are as follows:

*1)* In the context of uncertain environment, a novel 4PL route optimization problem that considers the risk of delays has been investigated;

*2)* The CVaR is employed to characterize risk, and a nonlinear programming model is established with constraint on delivery costs, aiming to minimize the risk as the objective;

*3)* An improved Q-learning algorithm is proposed to solve the model presented. Through this approach, 4PL can better adapt to the ever-changing market environment, ensuring the stability and efficiency of the supply chain.

The structure of this paper is arranged as follow: Section II gives the establishment of mathematical models and the transformations of CVaR model. Section III introduces the overall design of the proposed algorithm. Section IV performs some numerical computations. Section V finally concludes this paper.

## II. Literature Review

The current research on 4PLROP can be broadly divided into problem structure, solution approach, and distribution factors. Huang et al. [13] studied 4PLROP from single point to single point and single task, and established a mathematical model based on nonlinear integer programming and multiple graphs. Li et al. [14] studied the routing optimization problem of multi-point to multi-point 4PL systems with reliability constraints. Tao et al. [15] established a mixed integer programming model for 4PLROP from the perspective of cost discount. Hong et al. [16] studied a multi-objective transportation optimization model according to queuing theory, considering the option of 3PL providers, routes, as well as transportation methods. They used a priority based stochastic enhanced elite genetic algorithm (GA) to solve the infeasible solutions in 4PLROP. According to the prospect theory of customer psychological behavior and customer service level, Huang et al. [17] developed a nonlinear integer programming model for the design of 4PL network and offered an approximation linear approach to convert the model to an equivalent linear model so as to demonstrate the efficacy of the proposed method. Yue et al. [18] designed a particle swarm optimization with adaptive inertia weight to solve the proposed mathematical model. Lu et al. [19] designed a combination of ant colony optimization algorithm and improved grey wolf optimization algorithm to solve the 4PLROP. Huang et al. [20] studied the risk management of outsourcing logistics under the principal-agent framework from the perspective of product quality.

For the 4PLROP, most studies are based on the determination of delivery time, which assumes that the delivery cost and delivery time used in the transportation process are fixed quantities, such as references [14,15]. However, in the actual delivery process, due to unpredictable reasons such as transportation, bad weather, and human error operations, all of them may cause the transit time and transportation time to be not fixed but random during the transportation process, which may lead to the risk of delivery tardiness. The impact of this randomness makes 4PL suppliers to be unable to provide timely delivery plans that satisfy customers, it may lead to additional costs and a decrease in customer satisfaction. This situation may even affect the reputation of 4PL enterprises, leading to customer churn, and having adverse effects on their long-term development. Therefore, when solving 4PLROP, the influence of the risk caused by the uncertainty cannot be ignored.

The important question is how to define, measure, and control the risk to improve their logistics service quality, which is in the best interest of 4PL and creates a win-win for both parties. This is the focus of our paper. Value-at-Risk (VaR) was used to measure the time risk in Reference[13,21], the VaR model only considered the possible tardiness time that will not exceed VaR with the confidence level, but it did not take into account the extreme events (when the amount of tardiness time exceeds the VaR value), in which the tardiness risk mean value should be considered.

Conditional Value-at-Risk (CVaR) is a risk measurement tool proposed by Rockafellar et al. [22] on the basis of VaR, which considers the tardiness risk mean value. It is mainly used in combinatorial optimization, setting risk limits, resource allocation, and financial supervision and credit risk measurement of various financial regulators on relevant enterprises and institutions. In recent years, it has been widely used in inventory management optimization [23], supply chain [24], selection of fourth party logistics suppliers, network design and other fields for risk measurement and optimization [25].

In summary, this paper adopts a new risk measure tool, CVaR, to measure the delay risk, sets up a stochastic programming mathematical model, and designs an improved Q-learning algorithm. The aim is to help 4PL to provide an optimal supply chain distribution solution and improve the level of logistics service.

## III. Mathematical Model

### A. Problem Description

In this section, the 4PLROP with consideration of delay risk is described, and the notations used throughout the paper are introduced.

A manufacturing company (such as Haier) wants to invest in designing a distribution route to deliver its products and services from plant to customer through DCs and 3PL providers to reduce costs and improve customer satisfaction. As a result, it employs a 4PL provider to offer a comprehensive supply chain solution. Specifically, manufacturing companies are investors. A 4PL provider needs to help investors integrate 3PL providers, select the number and location of DCs, and complete the distribution of product from plants to customer.

The 4PLROP requires not only the selection of the route from the plant to the customer, but also the determination of 3PL providers which provide the delivery service. That increases the difficulty of solving 4PLROP.

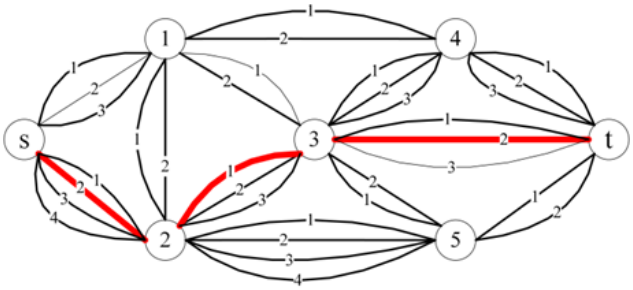


Fig. 1. Multiple graph for the 7-node problem.

A multiple graph, shown in Fig. 1, is used to describe the potential distribution network and demonstrate 4PLROP, where **V** represents the node cities, and **E** represents the edges. Among all nodes, indicates the supply city, indicates the target city, and others indicate the transit cities. All nodes have attributes such as time, cost, carrying capacity and reputation. In addition, there may be several edges between any two nodes because multiple 3PL suppliers may offer delivery services for any two cities. And each edge represents a 3PL, and the numeral on the edge is the serial number of 3PL.

Transportation time management is especially critical for 4PL providers. Traffic congestion, adverse weather, a surge in holiday order volumes, and node transfers can all cause uncertainty in time, leading to the risk of tardiness. Delayed delivery can directly affect customer interests, reduce the reputation of 4PL companies, and lead to customer loss, indirectly affecting the long-term development of 4PL supplier enterprises. Therefore, in order to improve customer satisfaction, 4PL suppliers need to consider the tardiness risk caused by time uncertainty. If the cost is within customer's budget, the less risk of tardiness, the better. Therefore, 4PL suppliers should monitor the tardiness risk mean value that may occur during transportation in real-time. Based on the above considerations, CVaR is introduced to quantify the average level of tardiness risk in delivery path, and a model with minimized CVaR and delivery cost as the constraint is developed.

TABLE I. THE DEFINITION OF PARAMETERS AND VARIABLES

| Variables | Definition |
|---|---|
| $r_{ij}$ | The quantity of 3PLs that offer transportation services between node cities $i$ and $j$ (i.e. the quantity of edges connecting two nodes) |
| $C_{ijk}$ | The transportation cost required for the $k$-th 3PL supplier between node cities $i$ and $j$ |
| $T_{ijk}$ | Random transportation time required for the $k$-th 3PL supplier between node cities $i$ and $j$ |
| $C_j^{'}$ | Transfer cost required when passing through node city $j$ |
| $T_j^{'}$ | Random transit time required when passing through node city $j$ |
| $R$ | A path containing a set of nodes and edges, i.e. $R=\{\upsilon_s,2,\upsilon_2,1,\upsilon_3,2,\upsilon_t\}$ can be used to represent the red path in Fig. 1. |

In order to set up the mathematical model, the definition of the parameters and variables are listed in Table I, the decision variables are defined as follows:

$$x_{ijk}(R)=\begin{cases}1 & \text{The } k\text{-th edge between nodes } i \\ & \text{and } j \text{ belongs to path } R \\ 0 & \text{others}\end{cases} \tag{1}$$

$$y_j(R)=\begin{cases}1 & \text{Node } j \text{ belongs to path } R \\ 0 & \text{others}\end{cases} \tag{2}$$

where $x_{ijk}(R)$ is used to determine whether the 3PL supplier provides a delivery task between cities $i$ and $j$, and $y_j(R)$ indicates whether node city $j$ provides a transfer task.

### B. Mathematical Model Based on CVaR Criterion

The VaR model [21] only considered the possible tardiness time that will not exceed VaR with the confidence level $\beta$, but it did not take into account the extreme events (remaining $1-\beta$ when the amount of tardiness time exceeds the VaR value), the tardiness risk mean value is generated. Therefore, this paper adopts CVaR to improve the objective function of the model at the confidence level $\beta$, as shown in (3), the tardiness risk mean value generated when the tardiness time exceeds VaR, i.e. minimizing CVaR, is calculated to determine the delivery path with the lowest average tardiness risk. Therefore, the following mathematical model is established:

$$\min \quad CVaR_{\beta}(\Delta T) \tag{3}$$

$$\text{s.t.} \quad \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{r_{ij}} C_{ijk}x_{ijk}(R)+\sum_{j=1}^{n} C_j^{'}y_j(R)\le C_0 \tag{4}$$

$$\Delta T=\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{r_{ij}}\tilde{T}_{ijk}x_{ijk}(R)+\sum_{j=1}^{n}\tilde{T}_j x_{ijk}(R)\right)-T_0 \tag{5}$$

$$R=\left\{\upsilon_s,\cdots,\upsilon_i,k,\upsilon_j,\cdots,\upsilon_k\right\}\in G \tag{6}$$

$$x_{ijk}(R),y_j(R)\in\{0,1\} \tag{7}$$

Eq. (3) represents the objective function, which minimizes CVaR, where indicates the level of confidence, indicating the customer's level of risk aversion. The delivery cost constraint shown in (4) is the highest cost which is acceptable to the investor. Eq. (5) states the delayed time that the total delivery time needed on the route exceeds the due date, which is a random variable. Eq. (6) represents the constraint on the path, ensuring that it is a legitimate connecting path from the origin city to the target city. And then the constraints on decision variables are represented by (7).

### C. Transformation of Mathematical Models

Theorem 1. The linear combination of a finite number of independent normal distribution random variables still obeys normal distribution. For a set of random variables $X_1,X_2,...,X_k,...,X_n$, if $X_1\sim N(\mu_1,\ \delta_1^2)$, $X_2\sim N(\mu_2,\ \delta_2^2),...,\ X_k\sim N(\mu_k,\ \delta_k^2),...,\ X_n\sim N(\mu_n,\ \delta_n^2)$, there is $\sum_{k=1}^{n}X_k\sim N(\sum_{k=1}^{n}\mu_k,\sum_{k=1}^{n}\delta_k^2)$.

Theorem 2. When the random variable follows the normal distribution, then $CVaR(\Delta T) = E(\Delta T) + c_1(\beta) \times STD(\Delta T)$ , where $E(\Delta T)$ refers to the expectation of random variable, $STD(\Delta T)$ refers to the standard deviation of random variable, $c_1(\beta) = \varphi(\phi^{-1}(\beta)) \times (1-\beta)^{-1}$ , $\phi^{-1}(\square)$ refers to the inverse function of standard normal distribution function, and $\varphi(\square)$ refers to the probability density of standard normal distribution [26].

This article assumes the random variable $\tilde{T}_{ijk} \sim N(\mu_{ijk}, \delta_{ijk}^2)$ and $\tilde{T}_j \sim N(\mu_j, \delta_j^2)$ in (5), so the objective function (3) can be converted to (8). Consequently, the proposed CVaR model includes (8), (4), (5), (6) and (7).

$$\min\Big(\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k}^{r_{ij}}\mu_{ijk}x_{ijk}(R) + \sum_{j=1}^{n}\mu_j y_j - T_0\Big) + c_1(\beta) \times \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k}^{r_{ij}}\delta_{ijk}^2 x_{ijk}^2(R) + \sum_{j=1}^{n}\delta_j^2 y_j^2(R)} \tag{8}$$

## IV. Algorithm Design

At present, the solution of the 4PLROP mainly includes the branch and bound, cut plane method, and other accurate solution algorithms that can only solve the optimization problem meeting specific conditions, as well as GA [27], particle swarm optimization algorithm [18] and ant colony algorithm [19], harmony search algorithm [13] and other simulation natural processes to form intelligent optimization algorithms. In solving 4PLROP, intelligent optimization algorithms can be roughly separated into two types, one approach is using repair strategies to repair illegal roads and obtain legitimate paths, which may result in the loss of good solution information during the repair process; another method is to utilize intelligent algorithms to construct simple graphs, followed by exact algorithms to discover the shortest path across the simple graph, which will waste a lot of storage space and computational time. Therefore, the Q-learning algorithm is used to solve the 4PL path problem in this article, directly resolving the proposed mathematical model on multiple graphs by setting state action pairs and reward values.

The Q-learning algorithm is proposed by Watkins [28] which involves the interaction between intelligent agents and the environment, constantly trying and learning, and ultimately obtaining one or more excellent behavior strategies. In a typical Q-learning algorithm, an intelligent agent decides to execute an action on the basis of the current state as well as past empirical knowledge. After executing the action, the agent transits to the following state according to a certain state transition strategy and obtains a return value. The Q-learning algorithm establishes a Q table based on the agent's state space and action space, which stores the corresponding Q values of all state action pairs. At the same time, the intelligent agent can be punished or rewarded by designing a reward function. When the action selected by the intelligent agent has an advantage in the environment, the state action pair can receive positive rewards from the environment, and the corresponding Q value of the state action pair will continue to increase. When the action selected by the intelligent agent is at a disadvantage in the environment, its state action pair will receive negative rewards from the environment, and its corresponding Q value will continue to decrease. The Q function is used to obtain the anticipated return value of a specific state action, which is updated in each training round and gradually approaches the optimal value. As shown in (9), the Q function is viewed as past-experienced knowledge that the agent has acquired, which is constantly updated and improved.

$$Q^*(s,a) \leftarrow Q(s,a) + \alpha\left[R(s,a,s') + \gamma \max Q(s',a') - Q(s,a)\right] \tag{9}$$

where $\gamma \in (0,1)$ is the discount factor, representing the degree of impact of the next state's Q value on the corresponding Q value of the current state, that is, the importance of immediate and future benefits. The larger the $\gamma$ , the greater the weight given to prior experiences. The smaller the $\gamma$ , the more emphasis is placed on immediate benefits R. When $\gamma = 0$ ,it means only focusing on current interests and not considering future interests. When $\gamma = 1$ , it indicates a focus on past experience and future interests. Parameter $\alpha$ is the learning rate, which represents the learning speed of the agent throughout the entire learning process. Its range is $\alpha \in (0, 1)$ , and it determines the degree that the new information obtained by the agent in the environment covers the old experience. The larger the $\alpha$ , the less effective it is to retain the previous training. The smaller the $\alpha$ , the more the effect of previous training will be retained.

### A. Action State Settings

This article applies Q-learning algorithm to the 4PLROP, treating the selection of actions as a 3PL supplier selection problem, and treating node cities as different states. When the current node is $s$ , the 3PL suppliers related to the node cities can be represented by action spaces $A = \{a_1, a_2, ..., a_k, ..., a_K\}, k = 1, 2, ..., K$ , with every state action pair matching a Q value.

Taking 7 nodes as an example, as shown in Fig. 1, the initial node can be regarded as the initial state $s$ . The nodes connecting the initial node are Transport Node 1 and Transport Node 2, respectively. There are three 3PL suppliers that can be selected from the initial node to Transport Node 1, with serial numbers 1, 2, and 3. Actions can be set to $a_1$, $a_2$, $a_3$ , respectively. There are four 3PL suppliers that can be selected from the initial node to Transport Node 2, with serial numbers 1, 2, 3, and 4, respectively, The action sequence numbers can be set to $a_4$, $a_5$, $a_6$, $a_7$ , and there are 7 corresponding actions that can be selected in the initial state. They are set to jump to the next state, namely transfer node 1, when the initial node selection action is $a_1$, $a_2$, $a_3$ , and also set to jump to the next state, namely transfer node 2, when the initial node selection action is $a_4$, $a_5$, $a_6$, $a_7$ . Other node action sets can be set by analogy. The selected the action and its corresponding next state for 7-node problem is shown in Table II.

TABLE II. 7-NODE PROBLEM ACTIONS AND CORRESPONDING NEXT STATES SETTINGS

| Selected Action | The Corresponding Next State |
|---|---|
| $A=\{a_1,a_2,a_3\}$ | Transit Node City 1 |
| $A=\{a_4,a_5,\cdots,a_9\}$ | Transit Node City 2 |
| $A=\{a_{10},a_{11},a_{14},a_{15},a_{16}\}$ | Transit Node City 3 |
| $A=\{a_{12},a_{13},a_{21},a_{22},a_{23}\}$ | Transit Node City 4 |
| $A=\{a_{17},a_{18},a_{19},a_{20},a_{24},a_{25}\}$ | Transit Node City 5 |
| $A=\{a_{26},a_{27},\cdots,a_{33}\}$ | Transit Node City t |

### B. Exploration Strategy

- When the intelligent agent interacts with the environment for learning, while selecting the known action with the maximum reward value, it is also necessary to ensure that more experience can be learned in the unknown environment, laying the foundation for obtaining more cumulative rewards. Therefore, it is necessary to set appropriate exploration strategies to achieve the optimal training effect. The exploration strategy adopted in this article is ε-greedy strategy.
- The mathematical description of the ε-greedy strategy is as follows:

$$\pi(a,s)=\begin{cases}\arg\max Q(s,a) & 1-\varepsilon \\ a_{random} & \varepsilon\end{cases} \tag{10}$$

- For (10), it can be understood as a certain probability ε. Randomly select the actions that can be selected in the current state, with 1-ε probability selects the action corresponding to the maximum Q value in the current action.

### C. Construction of Reward Function

Due to the fact that the Q-learning algorithm is on the basis of the Markov Decision Process model, a more computationally efficient discrete reward and penalty function is adopted. Eq. (11) indicates that when there is a connection between two cities and it is not the endpoint, the reward is 1. When there is no connection between two cities, the reward is -1. When there is a connection between two cities and it is the endpoint, the reward is 100.

$$r(s,a)=\begin{cases}1 & \text{i. j has a connection point j is not a termination node} \\ -1 & \text{i. j is not connected} \\ 100 & \text{i. j is connected and j is the termination node}\end{cases} \tag{11}$$

Considering that the size of the reward is related to the mean and variance in the objective function, while also meeting certain cost constraints, the reward value function (12) can be constructed to ensure that the agent does not violate the constraints and obtains the optimal delivery path for the objective value. As shown in (13), $\omega_1$ is related to the delivery cost related to the selected 3PL provider and transportation node, and the smaller the delivery cost, the larger the reward value obtained; $\omega_2$, shown in (14), is related to the mean of the random time related to the selected 3PL provider and transit node. The smaller the mean of the random time, the greater the reward value obtained; as shown in (15), $\omega_3$ is related to the variance of the random time related to the selected 3PL provider and transit node. While the variance of the random time is smaller, the reward value obtained is larger, where $k_1$ and $k_2$ are the weighting coefficients of the reward function.

$$r=\omega_1 r(s,a)+\omega_2 r(s,a)+\omega_3 r(s,a) \tag{12}$$

$$\omega_1=\frac{k_1}{C_{ijk}+C_j} \tag{13}$$

$$\omega_2=\frac{k_2}{\mu_{ijk}+\mu_j} \tag{14}$$

$$\omega_3=\frac{1-k_1-k_2}{\delta_{ijk}^2+\delta_j^2} \tag{15}$$

### D. Model Training

The agent is trained by designing a Q table, in which each row represents all the states that the agent can choose, each column represents the actions that the agent can perform in the corresponding state, each state represents different city nodes in multiple graphs, and each action represents different 3PL suppliers in multiple graphs. Initially, set all states in the Q table to 0, and then calculate the reward values obtained by executing different actions (selecting different suppliers) based on the reward matrix established by the reward function. Use (9) to update the values of each element in the Q table. Treat each iteration as a training session for the agent. For each training session, the agent attempts to reach the destination node from the initial node, and after each action, updates the elements in the Q table.

### E. Improved Q-learning Algorithm Process and Steps

When the improved Q-learning algorithm is used to solve 4PLROP, first initialize the elements in the matrix using a reward function based on existing data. Due to the existence of several various 3PL providers across two node cities, there is one state corresponding to multiple. Consequently, it is necessary to set the corresponding actions for each state, and then train and update the matrix Q through the setting of matrix R and related parameters. Finally, the optimal path planning can be obtained based on the Q table. The specific steps for solving the proposed model using the improved Q-learning algorithm are as follows:

Step 1: Preprocess the known factors of the problem.

Step 2: Import known information into Matlab.

Step 3: Initialize the parameters $\gamma$, $\alpha$ and Q table, set the initial and final states, and generate a reward matrix using (12) based on existing data.

Step 4: Initialize the state to the initial node.

Step 5: Utilize ε-greedy strategy selection action (the optional 3PL supplier corresponding to the state).

Step 6: Execute the action $a$ (select a 3PL supplier from the current node), transfer to a new state $s'$ (next node city), and update the Q table based on the reward matrix R and related parameter settings.

Step 7: Determine whether $s'$ is in a terminated state. If not, proceed to step 5. Else, proceed to step 8.

Step 8: Determine whether the training frequency has been reached. If not, continue with step 4, otherwise continue with step 9.

Step 9: After training, output the Q table.

Step 10: Output the optimal delivery plan based on the Q table.

## V. Experimental Results and Analysis

The improved Q-learning algorithm is used to solve different scale examples in this section to analyse the algorithm performance and the model effectiveness. The proposed CVaR model effectiveness is verified by comparing and analyzing the solution results of the two models. The algorithm is implemented using software MATLAB and runs in the Intel (R) Core (TM) i7-2600 @ 3.40GHz environment.

### A. Parameter Testing Analysis

By conducting extensive experimental simulations to test parameters, the method is to observe the impact of certain parameters change on the solution results while keeping other parameters constant. The experimental results demonstrate that the optimal parameters are $\gamma = 0.8$, $\alpha = 0.9$, $episode = 100$, $\varepsilon = 0.95$.

On the basis of the data in Table III, through repeated experiments on the weighting coefficients $k_1$ and $k_2$ in reward functions of different scales, the improved Q-learning algorithm performs best.

### B. Model Performance Analysis

Verify the effectiveness of the proposed CVaR model and improved Q-learning algorithm by solving several different scale examples. First, the 7-node problem is used as an instance, and the solution results obtained using this algorithm are carefully analyzed. Then, to demonstrate the validity of the model, VaR and CVaR values are solved for four examples of different sizes with 7, 15, 30 and 50 nodes, and the results are analyzed.

The solution results of the VaR model and CVaR model for the 7-node problem with different values are shown in TABLE IV, where $\beta$ denotes the confidence level, i.e., the risk attitude of the customer, $T_0$ denotes the customer's latest acceptable delivery time, $C_0$ is the customer's latest acceptable delivery cost. The value of VaR is the optimal solution obtained from the VaR model, the value of CVaR is the optimal solution obtained from the CVaR model, the Best Path is the distribution path that corresponds to the optimal solution, and Best Rate is the probability that the total number of runs obtains the best solution when the algorithm is used to solve. At this time, the total number of runs is 100, Time/s indicates the time for the algorithm to run once.

TABLE III. Solutions and Parameter Settings for Different Instances

| Number ofNodes | $k_1$ | $k_2$ | *Episode* | *CVaR* | Best Path | Time/s |
|---|---|---|---|---|---|---|
| 7 | 0.6 | 0.3 | 100 | 32.0456 | $R=\{\upsilon_s,2,\upsilon_2,2,\upsilon_3,1,\upsilon_t\}$ | 0.9s |
| 15 | 0.1 | 0.7 | 200 | 3.8184 | $R=\{\upsilon_s,1,\upsilon_2,2,\upsilon_6,3,\upsilon_{13},2,\upsilon_t\}$ | 1s |
| 30 | 0.37 | 0.357 | 200 | 13.6412 | $R=\{\upsilon_s,1,\upsilon_4,2,\upsilon_8,1,\upsilon_{12},1,$ $\upsilon_{15},2,\upsilon_{18},4,\upsilon_{21},1,\upsilon_{25},1,\upsilon_t\}$ | 1.5s |
| 50 | 0.2 | 0.5 | 500 | 8.3652 | $R=\{\upsilon_s,3,\upsilon_{39},1,\upsilon_{28},1,\upsilon_{29},1,$ $\upsilon_{10},1,\upsilon_{42},3,\upsilon_{37},3,\upsilon_t\}$ | 9.2s |

TABLE IV. Solution Results of 7-node Problem when $T_0$=80 and $C_0$=73.

| $\beta$ | $T_0$ | $C_0$ | *VaR* | *CVaR* | Best Path | Best Rate | Time/s |
|---|---|---|---|---|---|---|---|
| 0.9 | 80 | 73 | 10.9730 | 22.0456 | $R=\{\upsilon_s,2,\upsilon_2,2,\upsilon_3,1,\upsilon_t\}$ | 1 | 0.9s |
| 0.95 | 80 | 73 | 19.4699 | 29.2428 | $R=\{\upsilon_s,2,\upsilon_2,2,\upsilon_3,1,\upsilon_t\}$ | 1 | 0.9s |
| 0.99 | 80 | 73 | 35.4087 | 43.3341 | $R=\{\upsilon_s,2,\upsilon_2,2,\upsilon_3,1,\upsilon_t\}$ | 1 | 0.9s |

From the data in the Table IV, it can be seen that the best path corresponding to the best VaR value and CVaR value is the same. When the confidence level is 0.9 and the VaR value that meets the cost constraint is 10.9730, it means that the 4PL supplier has a 90% probability of ensuring that the delay amount will not exceed 10.9730. The CVaR value that satisfies the cost constraint is 22.0456, indicating that the tardiness risk mean value when the delivery task's delay exceeds the VaR is 22.0456, The related delivery cost is 73, and the best delivery path is $R=\{\upsilon_s,2,\upsilon_2,2,\upsilon_3,1,\upsilon_t\}$, which refers to the selection of transit cities 2 and 3 for transportation from the source node city $s$ to the destination node city $t$, and the numbers of the 3PL supplier chosen between every two cities are 2, 2, and 1. When the confidence level is 0.95 and the VaR value that satisfies the cost constraint is 19.4699, it means that the 4PL supplier has a 95% probability of ensuring that the delay amount will not exceed 19.4699. The CVaR value that satisfies the cost constraint is 29.2428, indicating that the tardiness risk mean value when the delivery task's delay exceeds the VaR is 29.2428. The associated delivery cost is 73, and the best

delivery path is $R=\{\upsilon_s,2,\upsilon_2,2,\upsilon_3,1,\upsilon_t\}$, which refers to the selection of transit cities 2 and 3 for transportation from the source node city $s$ to the destination node city $t$, and the 3PL supplier numbers selected between each two cities are 2, 2, and 1. When the confidence level is 0.99 and the VaR value that satisfies the time constraint is 35.4087, it means that the 4PL supplier has a 99% probability of ensuring that the delay amount will not exceed 35.4087. The CVaR value that satisfies the cost constraint is 43.3341, indicating that the tardiness risk mean value when the delivery task's delay exceeds VaR is 43.3341. The associated delivery cost is 73, and the best delivery path is $R=\{\upsilon_s,2,\upsilon_2,2,\upsilon_3,1,\upsilon_t\}$, which refers to the selection of transit cities 2 and 3 for transportation from the source node city $s$ to the destination node city $t$, and the 3PL supplier numbers selected between each two cities are 2, 2, and 1. The above data indicates that when other constraints are constant, as the confidence level grows, the best delivery path will not change. However, customers will face higher tardiness risks, and the corresponding VaR is often smaller than CVaR. This also verifies that the VaR model can only evaluate the probability of risk occurrence, while the CVaR model can effectively measure tail risk and estimate the tardiness risk mean value faced by delivery tasks in extreme situations. Compared to VaR models, it can better reflect potential tardiness risks.

The statistics in Table V show that the 7-node problem's solution is provided for various combinations of confidence level, delivery time, and delivery cost. We know from analyzing these data that while the confidence level and time constraints are consistent, as the delivery cost increases, the corresponding VaR and CVaR values will decrease, indicating that the delay risk faced by the delivery path will be reduced. When the delivery cost and time constraints remain unchanged and the confidence level increases, the corresponding VaR and CVaR values will increase, indicating an increase in the risk of tardiness faced by the delivery path. When the confidence level and delivery cost are constant, as the time constraint increases, the corresponding VaR and CVaR values will increase, and thus the delay risk faced by the distribution path will rise. In addition, by comparing the VaR value and the CVaR value, it can be seen that when other conditions are the same, the CVaR value obtained is always greater than the VaR value, which also verifies that CVaR is more able to reflect the potential value at risk than VaR. Therefore, when using the VaR model to measure tardiness risk failure, 4PL suppliers can use the CVaR model to make up for the shortcomings of the VaR model. Combined with the risk tolerance of customers, they can comprehensively consider the risk level and expected tardiness risk of the distribution scheme, monitoring of potential tardiness risks in real time, providing customers with the delivery path with the minimum tardiness risk mean value at a given confidence level, and estimating the tardiness risk generated when extreme events occur, making reasonable delivery service decisions, thereby improving customer satisfaction.

Tables VI to VIII provide the solution results of the VaR model and CVaR model for the 15 node problem with time constraints, cost constraints, 30 node problem with time constraints, and 50 node problem with time constraints and cost constraints, respectively. When using the Q-learning algorithm to solve the proposed model, the solution time for small and medium-sized examples is about 1 second, and for large-scale problems, the solution time does not exceed 10 seconds. This fully demonstrates that the Q-learning algorithm has a high solution speed and high stability. When solving a 15 node problem, the training number is 200, and the optimal solution rate of the algorithm is as high as 98%, that is, the algorithm runs 100 times, 98 times can obtain the optimal solution, and when solving the 50 node problem, the best rate also reaches 95%, that is, the algorithm runs 100 times, and 95 times can obtain the optimal solution.

### C. The Influence of Confidence Level

To investigate the influence of confidence level $\beta$ on the 4PLROP, we provide three values of $\beta$ with four different cases, which is customer's degree of risk appetite, shown in Tables V to VIII.

TABLE V. SOLUTION FOR 7-NODE PROBLEMS UNDER DIFFERENT CONFIDENCE LEVELS, TIME CONSTRAINTS, AND COST CONSTRAINTS

| $\beta$ | $T_0$ | $C_0$ | *Episode* | *VaR* | *CVaR* | Best Path | Time/s |
|---|---|---|---|---|---|---|---|
| 0.9 | 70 | 73 | 100 | 20.9730 | 32.0456 | $R=\{\upsilon_s,2,\upsilon_2,2,\upsilon_3,1,\upsilon_t\}$ | 0.9s |
| | | 75 | 100 | 17.2239 | 28.0198 | $R=\{\upsilon_s,3,\upsilon_2,2,\upsilon_3,1,\upsilon_t\}$ | 0.9s |
| | 80 | 73 | 100 | 10.9730 | 22.0456 | $R=\{\upsilon_s,2,\upsilon_2,2,\upsilon_3,1,\upsilon_t\}$ | 0.9s |
| | | 75 | 100 | 7.2239 | 18.0198 | $R=\{\upsilon_s,3,\upsilon_2,2,\upsilon_3,1,\upsilon_t\}$ | 0.9s |
| 0.95 | 70 | 73 | 100 | 29.4699 | 39.2428 | $R=\{\upsilon_s,2,\upsilon_2,2,\upsilon_3,1,\upsilon_t\}$ | 0.9s |
| | | 75 | 100 | 25.5084 | 35.0371 | $R=\{\upsilon_s,3,\upsilon_2,2,\upsilon_3,1,\upsilon_t\}$ | 0.9s |
| | 80 | 73 | 100 | 19.4699 | 29.2428 | $R=\{\upsilon_s,2,\upsilon_2,2,\upsilon_3,1,\upsilon_t\}$ | 0.9s |
| | | 75 | 100 | 15.5084 | 25.0371 | $R=\{\upsilon_s,3,\upsilon_2,2,\upsilon_3,1,\upsilon_t\}$ | 0.9s |
| 0.99 | 70 | 73 | 100 | 45.4087 | 53.3341 | $R=\{\upsilon_s,2,\upsilon_2,2,\upsilon_3,1,\upsilon_t\}$ | 0.9s |
| | | 75 | 100 | 41.0489 | 48.7762 | $R=\{\upsilon_s,3,\upsilon_2,2,\upsilon_3,1,\upsilon_t\}$ | 0.9s |
| | 80 | 73 | 100 | 35.4087 | 43.3341 | $R=\{\upsilon_s,2,\upsilon_2,2,\upsilon_3,1,\upsilon_t\}$ | 0.9s |
| | | 75 | 100 | 31.0489 | 38.7762 | $R=\{\upsilon_s,3,\upsilon_2,2,\upsilon_3,1,\upsilon_t\}$ | 0.9s |

TABLE VI. SOLUTION FOR THE 15-NODE PROBLEM WHEN $T_0$=70, $C_0$=115

| $\beta$ | *Eposide* | *VaR* | *CVaR* | Best Path | Best Rate | Time/s |
|---|---|---|---|---|---|---|
| 0.9 | 200 | 0.5969 | 3.7728 | $R=\{\upsilon_s,1,\upsilon_2,2,\upsilon_6,3,\upsilon_{13},2,\upsilon_t\}$ | 0.98 | 1s |
| 0.95 | 200 | 3.0340 | 5.8371 | $R=\{\upsilon_s,1,\upsilon_2,2,\upsilon_6,3,\upsilon_{13},2,\upsilon_t\}$ | 0.98 | 1s |
| 0.99 | 200 | 7.3406 | 9.4295 | $R=\{\upsilon_s,2,\upsilon_2,2,\upsilon_6,3,\upsilon_{13},2,\upsilon_t\}$ | 0.98 | 1s |

TABLE VII. SOLUTION FOR THE 30-NODE PROBLEM WHEN $T_0$=115, $C_0$=190

| $\beta$ | *Eposide* | *VaR* | *CVaR* | Best Path | Best Rate | Time/s |
|---|---|---|---|---|---|---|
| 0.9 | 200 | 10.5009 | 13.6412 | $R=\{\upsilon_s,1,\upsilon_4,2,\upsilon_8,1,\upsilon_{12},1,$ $\upsilon_{15},2,\upsilon_{18},4,\upsilon_{21},1,\upsilon_{25},1,\upsilon_t\}$ | 0.97 | 1.5s |
| 0.95 | 200 | 12.9107 | 15.6825 | $R=\{\upsilon_s,1,\upsilon_4,2,\upsilon_8,1,\upsilon_{12},1,$ $\upsilon_{15},2,\upsilon_{18},4,\upsilon_{21},1,\upsilon_{25},1,\upsilon_t\}$ | 0.97 | 1.5s |
| 0.99 | 200 | 17.4312 | 19.679 | $R=\{\upsilon_s,1,\upsilon_4,2,\upsilon_8,1,\upsilon_{12},1,$ $\upsilon_{15},2,\upsilon_{18},4,\upsilon_{21},1,\upsilon_{25},1,\upsilon_t\}$ | 0.97 | 1.5s |

TABLE VIII. SOLUTION FOR THE 50-NODE PROBLEM WHEN $T_0$=150, $C_0$=165

| $\beta$ | *Eposide* | *VaR* | *CVaR* | Best Path | Best Rate | Time/s |
|---|---|---|---|---|---|---|
| 0.9 | 500 | 4.4900 | 8.3652 | $R=\{\upsilon_s,3,\upsilon_{39},1,\upsilon_{28},1,\upsilon_{29},1,$ $\upsilon_{10},1,v_{42},3,\upsilon_{37},3,\upsilon_t\}$ | 0.95 | 1.5s |
| 0.95 | 500 | 7.4637 | 10.4637 | $R=\{\upsilon_s,3,\upsilon_{39},1,\upsilon_{28},1,\upsilon_{29},1,$ $\upsilon_{10},1,\upsilon_{42},3,\upsilon_{37},3,\upsilon_t\}$ | 0.95 | 1.5s |
| 0.99 | 500 | 13.042 | 15.8157 | $R=\{\upsilon_s,3,\upsilon_{39},1,\upsilon_{28},1,\upsilon_{29},1,$ $\upsilon_{10},1,\upsilon_{42},3,\upsilon_{37},3,\upsilon_t\}$ | 0.95 | 1.5s |

The result in these four tables shows that the CVaR, the tardiness risk, increases with the confidence level $\beta$ increasing. Because with a smaller $\beta$, the 3PL can make more effort and the tardiness risk is smaller. 4PL can select the proper delivery solution for the investor to control the tardiness risk according to the customer's degree of risk aversion.

### D. The Influence of Cost and Due Date

To investigate the influence of cost $C_0$ and due date $T_0$ on the 4PLROP, we provide two values of $C_0$ and $T_0$ with 7-node problem, shown in Table V.

The result in Table V shows that the CVaR, the tardiness risk, decreases with the $C_0$ and $T_0$ increasing when $\beta$ in fixed. Because with a fixed $\beta$, if the budget or time is enough, the 3PL can make more effort and the tardiness risk is smaller. 4PL can select the proper delivery solution for the investor to control the tardiness risk according to the customer's budget and due date.

### E. Algorithm Comparison

This section uses the improved Q-learning and GA embedded with Dijkstra algorithm to solve three different scale examples, and the comparative data is shown in Table IX.

TABLE IX. COMPARISON OF RESULTS OF DIFFERENT ALGORITHMS

| Number of Nodes | Algorithm | *CVaR* | Best Path | Best Rate | Time/s |
|---|---|---|---|---|---|
| 7 | Improved Q-learning | 22.0456 | $R=\{\upsilon_s,2,\upsilon_2,2,\upsilon_3,1,\upsilon_t\}$ | 1 | 0.9s |
| | GA embedded with Dijkstra | 22.0456 | $R=\{\upsilon_s,2,\upsilon_2,2,\upsilon_3,1,\upsilon_t\}$ | 0.95 | 19.4s |
| 15 | Improved Q-learning | 3.7728 | $R=\{\upsilon_s,1,\upsilon_2,2,\upsilon_6,3,\upsilon_{13},2,\upsilon_t\}$ | 0.98 | 1s |
| | GA embedded with Dijkstra | 3.7728 | $R=\{\upsilon_s,1,\upsilon_2,2,\upsilon_6,3,\upsilon_{13},2,\upsilon_t\}$ | 0.94 | 24.5s |
| 30 | Improved Q-learning | 13.641 | $R=\{\upsilon_s,1,\upsilon_4,2,\upsilon_8,1,\upsilon_{12},1,$ $\upsilon_{15},2,\upsilon_{18},4,\upsilon_{21},1,\upsilon_{25},1,\upsilon_t\}$ | 0.95 | 1.5s |
| | GA embedded with Dijkstra | 13.641 | $R=\{\upsilon_s,1,\upsilon_4,2,\upsilon_8,1,\upsilon_{12},1,$ $\upsilon_{15},2,\upsilon_{18},4,\upsilon_{21},1,\upsilon_{25},1,\upsilon_t\}$ | 0.9 | 28.5s |

It can be seen from the Table IX that both methods can find the optimal solution when solving small-scale problems with 7 nodes. However, the latter performs poorly in solving speed because that GA embedded with Dijkstra algorithm is used to generate the simple graph based on the multi-graph shown in Fig. 1 and the Dijkstra algorithm is used to generate the shortest path on the generated simple graph, but the shortest path may be not met the constraints. Thus, it needs a lot of time to find the feasible solution of the problem. However, the improved Q-learning algorithm directly solves the problem on

the multi-graph shown in Fig. 1, which saves much computational time. Furthermore, as the solution size increases, the improved Q-learning algorithm exhibits higher solving speed and quality.

### F. Discussion

In summary, the results show that when the delivery costs and time constraints remain unchanged, the higher the confidence level, the higher the corresponding values of VaR and CVaR, which means that the risk of delivery tardiness faced by customers will increase. When other conditions such as confidence level, time, and cost constraints are the same, the obtained CVaR value is always greater than the VaR value. The proposed CVaR model can reflect the average delay risk of delayed delivery exceeding the VaR value due to various factors, better compensating for the shortcomings of the VaR model in measuring tardiness risk, and real-time monitoring of potential tardiness risks that may occur during the delivery process. 4PL can adjust the customer's aversion to risk, use this model to calculate the tardiness risk mean value and provide a reliable delivery plan.

## VI. Conclusion

This article fully considers the tardiness risk caused by the uncertainty of transit time and transportation time in the actual delivery process in complex and uncertain environments. A risk measurement tool CVaR is introduced to measure and control the risk, and a mathematical model with CVaR minimization as the optimization objective and distribution cost as the constraint is established. Meanwhile, the proposed algorithm is compared with GA embedded with Dijkstra. The results demonstrate that the proposed model is effective for the 4PLROP and improved Q-learning algorithm can solve the large-scale 4PL path problem rapidly and with excellent stability. 4PL can adjust the customer's aversion to risk, use this model to calculate the tardiness risk mean value and provide a reliable delivery plan. Customers can obtain multiple schemes according to their risk preferences and take corresponding measures. This article provides scientific decision making basis and efficient and safe distribution plans for the 4PL, which can improve the level of logistics service.

Meanwhile, the stochastic variables' probability distributions may follow other distributions, such as the exponential distribution, uniform distribution, etc. Therefore, our research can be extended to a robust 4PLROP considering delay risk or multiple risks.

## Acknowledgment

This research was funded by the National Natural Science Foundation of China, grant number 62203202; Natural Science Fund Project of Liaoning Province, grant number 2022-BS-295; Youth Project of the Educational Department of Liaoning Province, Grant number LJKQZ20222432.

## References

[1] Y. X. Zhang, Z. M. Gao, M. Huang, S. C. Jiang, M. Q. Yin, and S. C. Fang, "Multi-period distribution network design with boundedly rational customers for the service-oriented manufacturing supply chain: a 4PL perspective,"Int. J. Prod. Res., vol. 62, pp. 7412-7431, 2022.

[2] E. Lee, "Alibaba's cainiao logistics conffrms first financing at $7.7B valuation," The technode, March 15. 2016. https://technode.com/2016/03/15/alibaba-cainiaofunding/.

[3] F. Q. Lu, W. D. Chen, W. J. Feng, and H. L. Bi, "4PL routing problem using hybrid beetle swarm optimization," Soft Comput., vol. 27, pp. 17011, 2023.

[4] M. Huang, J. Tu, X. Chao, and D. Jin, "Quality risk in logistics outsourcing: A fourth party logistics perspective," Eur. J. Oper. Res., vol. 276, pp. 855-879, 2019.

[5] M. Huang, L. W. Dong, H. B. Kuang, Z. Z. Jiang, L. H. Lee, X.W. Wang, "Supply chain network design considering customer psychological behavior−a 4PL perspective," Comput. Ind. Eng. pp. 159, 2021.

[6] M. Q. Yin, M. Huang, X. H. Qian, D. Z. Wang, X. W. Wang, L. H. Lee, "Fourth-party logistics network design with service time constraint under stochastic demand," J. Intell. Manuf. vol. 34, pp. 1203‐1227, 2023.

[7] F. Q. Lu, H. L. Bi, W. J. Feng, Y. L. Hu, S. X. Wang, and X. Zhang, "A two-stage auction mechanism for 3pl supplier selection under risk aversion," Sustainability, vol. 13, pp. 9745, 2021.

[8] H. Y. Wang, M. Huang, H. F. Wang, and Y. J. Zhou, "Fourth party logistics service quality management with logistics audit," J. Indus. Manag. Optim., Vol. 19, pp. 7105-7129, 2023.

[9] H. Y. Wang, M. Huang, H. F. Wang, X. H. Feng, and Y. J. Zhou, "Contract design for the fourth party logistics considering tardiness risk," Int. J. Indus. Eng. Comput., vol. 13, pp. 13-30, 2022.

[10] F. Q. Lu, W. J. Feng, M. Y. Gao, H. L. Bi, and S. X. Wang, "The fourth-party logistics routing problem using ant colony system-improved grey wolf optimization," J. Adv. Transp., vol. 2022, pp. 9864064, 2022.

[11] J. H. Wu, "The relationship between port logistics and international trade based on VAR model," J. Coastal Res., pp. 601-604, 2020.

[12] X.F. Lyu, Y. C. Song, C. Z. He, Q. Lei, and W. F. Guo, "Approach to integrated scheduling problems considering optimal number of automated guided vehicles and conflict-free routing in flexible manufacturing systems," IEEE Access, vol. 7, pp. 74909-74924, 2019

[13] G. Bo, M. Huang, "Model and Solution of Routing Optimization Problem in the Fourth Party Logistics with Tardiness Risk,"ComplexSyst. Complex. Sci, vol. 15, no. 03, pp. 66‐74, 2018.

[14] J. Li, Y. Liu, Y. Zhang, and S. Xu, "Algorithms for routing optimization in multipoint to multipoint 4PL system," Discr. Dyn. Nat. Soc., vol. 2015, pp. 426947, 2015.

[15] Y. Tao, E. P. Chew, L. H. Lee, and Y. Shi, "A column generation approach for the route planning problem in fourth party logistics," J. Oper. Res. Soc., vol. 68, pp. 165-181, 2017.

[16] W. Hong, Z. Xu, W. Liu, L. Wu, and X. Pu, "Queuing theory-based optimization research on the multi-objective transportation problem of fourth party logistics," Proc. Inst. Mech. Eng. B J. Eng. Manuf., vol. 235, pp. 1327-1337, 2021.

[17] M. Huang, L. Dong, H. Kuang, Z. Z. Jiang, L. H. Lee, and X. Wang, "Supply chain network design considering customer psychological behavior-a 4PL perspective," Comp. Ind. Eng., vol. 159, pp. 107484, 2021.

[18] D. Yue, M. Huang, M. Yin, "PSO algorithm for the fourth party logistics network design considering multi-customer behavior under stochastic demand," In 2017 29th Chinese Control And Decision Conference, Chongqing, China, 01 May 2017.

[19] F. Lu, W. Feng, M. Gao, H. Bi, and S. Wang, "The fourth-party logistics routing problem using ant colony system-improved grey wolf optimization," J. Adv. Transp., vol. 2020, pp. 1-15, 2020.

[20] M. Huang, L. Dong, H. Kuang, and X. Wang, "Reliable fourth party logistics location-routing problem under the risk of disruptions," IEEE Access, vol. 9, pp. 84857-84870, 2021.

[21] X. Liu, G. Bo, "Q-learning algorithm for fourth party logistics route optimization considering tardiness risk," Proceedings of the 2022 International Conference on Cyber-Physical Social Intelligence, 2022.

[22] R. T. Rockafellar and S. Uryasev, "Optimization of conditional value-at-risk," J. risk, vol. 2, pp. 21-42, 2000.

[23] W. Xue, L. Ma, and H. Shen, "Optimal inventory and hedging decisions with CVaR consideration," Int. J. Prod. Econ., vol. 162, pp. 70-82, 2015.

[24] V. Dixit, P. Verma, and M. K. Tiwari, "Assessment of pre and post-disaster supply chain resilience based on network structural parameters with CVaR as a risk measure," Int. J. Prod. Econ., vol. 227, pp. 107655, 2020.

[25] F. Ding, M. Liu, S. M. Hsiang, P. Hu, Y. Zhang, and K. Jiang, "Duration and labor resource optimization for construction projects-a conditional-value-at-risk-based analysis,"Buildings, vol. 14, pp.1-20, 2024.

[26] R. T. Rockafellar and S. Uryasev, "Conditional Value-at-Risk for general loss distributions," J. Bank. Fin., vol. 26, pp. 1443-1471, 2002.

[27] J. Li, Y. Liu, and Z. Hu, "Routing optimization of fourth party logistics with reliability constraints based on Messy GA," J. Ind. Eng. Manag., vol. 7, pp. 1097-1111, 2014.

[28] C. J. C. H. Watkins, "Learning from delayed rewards," Ph.D., University of Cambridge, Cambridgeshire, May 1989.