# PPO Framework for Efficient Freight Dispatch in Dynamic Logistics Environments

Ji-Hyeon Kim
*Dept. of Electrical and Electronics Engineering*
*Pusan National University*
Busan, Republic of Korea
kjihyeon@pusan.ac.kr

Soon-Young Kwon
*Dept. of Electrical and Electronics Engineering*
*Pusan National University*
Busan, Republic of Korea
ysk1680@pusan.ac.kr

Hyoung-Nam Kim
*School of Electrical and Electronics Engineering*
*Pusan National University*
Busan, Republic of Korea
hnkim@pusan.ac.kr

***Abstract*— When diverse transport scenarios are present in an automatic vehicle/cargo dispatch system, effective matching of vehicles and cargo is essential to optimize operational performance. In this paper, we propose a transportation dispatch optimization framework based on proximal policy optimization (PPO), a reinforcement learning algorithm, to handle various environmental factors such as transport distance, time constraints, and traffic conditions. Simulation results demonstrate that the proposed framework achieves higher aggregate performance measures and faster convergence compared to traditional methods. The PPO framework reliably generalizes to hybrid scenarios, ensuring consistent and efficient performance in dynamic logistics environments.**

***Keywords*— *automatic vehicle/cargo dispatch system , component, transportation optimization, proximal policy optimization (PPO)***

## I. Introduction

In the conventional port logistics industry, the dispatching of freight vehicles still heavily relies on manual processes, often resulting in frequent transportation delay [1-3]. Such issues result in inefficiencies in operations and fairness concerns in dispatching, especially when environmental factors such as traffic conditions, weather, and transportation distances are not adequately considered. To address these challenges, the digital transformation of logistics systems and the adoption of automated dispatch systems have been pursued [2-6].

Existing optimization models often prioritize specific objectives, such as minimizing empty travel distances or reducing costs, while reflecting only limited transportation conditions [4-6]. However, these models typically fail to address diverse scenarios or sufficiently account for dynamic environmental factors (e.g., traffic or vehicle conditions) [6].

In this paper, we propose a meta-reinforcement learning-based optimization algorithm utilizing proximal policy optimization (PPO) [7] to develop an automatic dispatch system for freight vehicles. The proposed algorithm is designed to generalize across diverse transportation scenarios, such as short, mid, and long-distance transport, while also considering various conditions like time of day and traffic conditions. By leveraging PPO, the optimization algorithm ensures stable policy updates while efficiently adapting to new scenarios, prioritizing equity in dispatching and achieving an optimal balance among time efficiency, vehicle utilization, and cargo throughput.

This paper organized as follows: Section II explains the various transportation scenarios considered in this study and briefly introduces the optimization framework using the PPO. Section III presents simulation results that evaluate the performance of the proposed algorithm across various transportation scenarios and dynamic conditions. Finally, we conclude this paper in Section IV.

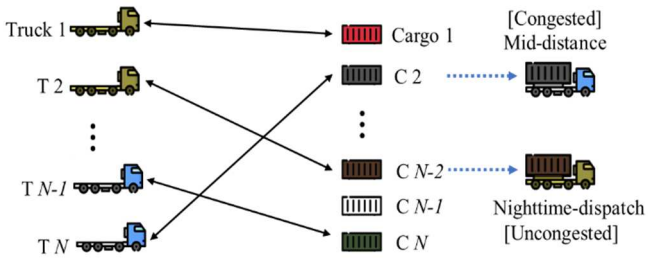


Fig. 1. The overview of transportation scenarios.

## II. Transportation Scenarios and Optimization Framework

### A. Transportation Scenarios

The efficient operation of a dispatch system requires adaptability to varying conditions, such as transportation distances, time constraints, and traffic patterns. These conditions are categorized into three primary scenarios: distance-based, time-based, and traffic condition scenarios, as shown in Fig. 1.

Transportation distance plays a crucial role in shaping the objectives of dispatch operations. For short-distance transport, defined as deliveries within a 50 km radius, the primary focus is minimizing travel time to enable frequent dispatches and improve time efficiency. Mid-distance transport, covering deliveries between 50 and 200 km, demands a balance between travel time and vehicle utilization, often constrained by moderate traffic conditions and scheduling requirements. Long-distance transport involves intercity deliveries exceeding 200 km, where maximizing cargo throughput and minimizing idle vehicle time are the main objectives. This scenario typically involves fewer but larger shipments.

The time of day also impacts transportation efficiency due to traffic congestion and operational priorities. Daytime dispatch, characterized by peak traffic hours, requires the algorithm to prioritize routes that minimize delays and fuel consumption. Conversely, nighttime dispatch benefits from reduced traffic, allowing the focus to shift toward maximizing cargo throughput and vehicle utilization through more efficient route planning.

Traffic conditions further influence dispatch system performance. Congested routes are defined as routes where the average vehicle speed falls below 30~40 km/h, reflecting high traffic density. In such scenarios, the optimization algorithm prioritizes strategies to minimize delays while maintaining operational efficiency. In contrast, uncongested routes, where the average vehicle speed is 60 km/h or higher, allow the algorithm to focus on improving cargo throughput and reducing transportation costs, as time delays are less critical.

The efficient operation of a dispatch system requires adaptability to varying conditions, such as transportation distances, time constraints, and traffic patterns. Transportation optimization is needed to address these operational requirements.

### B. Optimization Framework

#### 1) Proximal Policy Optimization (PPO) [7]

Proximal policy optimization (PPO) is a reinforcement learning algorithm designed to improve the stability and efficiency of policy training. Developed as an enhancement to trust region policy optimization (TRPO) [8], PPO simplifies the optimization process by avoiding the need for complex second-order approximations while retaining performance stability.

The critical feature of PPO is its use of a clipping mechanism that restricts policy updates to a specified range. This prevents drastic changes to the policy during learning, ensuring smooth and consistent updates. PPO achieves an effective balance between exploration and exploitation, making it particularly suitable for dynamic and complex environments like logistics and transportation systems.

#### 2) Integration of PPO with Dispatch Optimization

The proposed framework employs PPO to optimize dispatch decisions by training a policy $\pi_\theta$ that maps the current state of the system to optimal actions. The state space encompasses critical information such as vehicle locations, availability, cargo details (e.g., size, priority), and external factors like traffic conditions and time of day. This ensures that the policy dynamically adjusts to reflect real-time operational conditions. The action space involves assigning vehicles to specific cargo and determining optimal delivery routes. By focusing on these decisions, the framework ensures efficient resource utilization and timely deliveries under varying logistical constraints.

The reward function is designed to evaluate the effectiveness of each action based on key performance metrics:

$$R_t = w_1 \times \text{Time Efficiency} + w_2 \times \text{Vehicle Utilization} + w_3 \times \text{Cargo Throughput} + w_4 \times \text{Dispatch Equity} \quad (1)$$

Here, $w_1$, $w_2$, $w_3$, $w_4$ are scenario-specific weights that prioritize different objectives depending on the operational context. 'Time Efficiency' represents the average time required to complete the delivery of assigned cargo, normalized to reflect the reduction in delays. 'Vehicle Utilization' is defined as the ratio of active transportation time to the total available time of vehicles, measuring how effectively the fleet is utilized. 'Cargo Throughput' quantifies the total volume or number of cargoes successfully delivered within a given period, reflecting the system's capacity to handle transportation demands. 'Dispatch Equity' ensures a fair workload distribution among vehicles by minimizing variance in dispatch counts. By adjusting the weights, the reward function dynamically aligns with the priorities of different scenarios, such as emphasizing time efficiency in short-distance operations or cargo throughput in long-distance transport.

TABLE I. SIMULATION PARAMETERS

| Parameters | Values |
|---|---|
| Maximum transport distance | 400 km |
| Transportable distance (short/mid/long) | 50 km / 200 km / 400 km |
| Maximum driving time | 2 ~ 6 hours |
| Traffic congestion | congested / uncongested |
| Vehicle speed range | 30 ~ 100 km/h |

The overall effectiveness of the system is assessed using an aggregate performance measure [9], which averages rewards over the operational period:

$$\text{Aggregate Performance Measure} = \frac{1}{T}\sum_{t=1}^{T} R_t \quad (2)$$

This measure provides a comprehensive evaluation of the system's performance, ensuring that multiple objectives are balanced to achieve robust and adaptive optimization.

By integrating PPO with these components, the framework dynamically adapts to changing conditions, providing a scalable and efficient solution for dispatch optimization.

## III. SIMULATION RESULTS

In this section, simulations are conducted to evaluate the performance of the proposed PPO-based optimization framework under various transportation scenarios in the freight dispatch system. Simulations were designed to reflect realistic operational conditions, incorporating varying distances, time constraints, and traffic patterns.

### A. Experimental Setup

The simulation environment was constructed to model realistic freight dispatch conditions, incorporating a range of transportation scenarios and environmental variables. These scenarios included distance-based categories (short-distance, mid-distance, and long-distance), time-based operations (daytime and nighttime), and varying traffic conditions (congested and uncongested routes). The parameters used in the simulations are summarized in Table 1.

The evaluation of the dispatch system focused on four metrics: time efficiency, vehicle utilization, cargo throughput, and dispatch equity. Scenario-specific weights ($w_1$,$w_2$,$w_3$,$w_4$) were applied in the reward function to emphasize objectives tailored to each scenario, such as prioritizing time efficiency for short-distance operations or cargo throughput for long-distance deliveries.

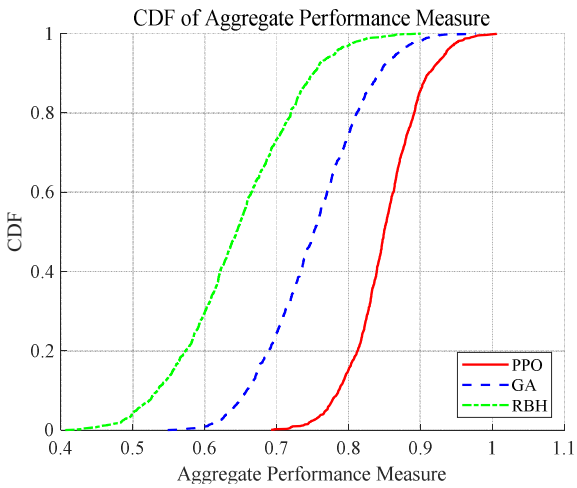


Fig. 2. Performance comparison of dispatch algorithms using CDF of Aggregate Performance (learning stage).

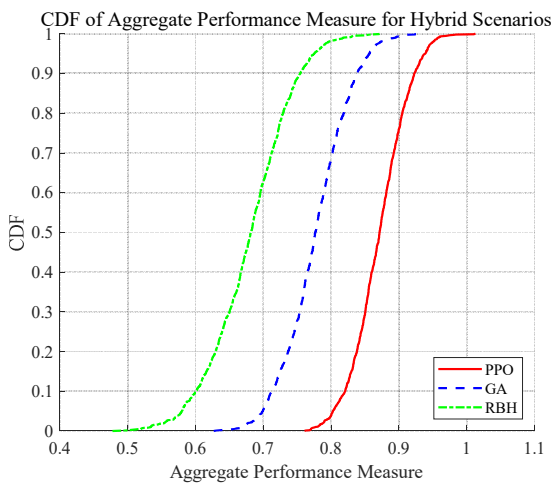


Fig. 3. Performance comparison of dispatch algorithms using CDF of Aggregate Performance for hybrid scenarios (testing stage).

The framework was implemented using PyTorch, and training was conducted over 10,000 iterations for each scenario. Testing involved 100 episodes per scenario, with randomized assignments of vehicles and cargoes. The system's overall effectiveness was measured using an aggregate performance measure, calculated by averaging rewards across the simulation period.

### *B. Performance Comparison*

This section evaluates the performance of the proposed PPO-based optimization framework. The evaluation is conducted in two stages: the learning stage and the testing stage. Both stages incorporate distance-based (short-, mid-, and long-distance), time-based (daytime and nighttime), and traffic condition (congested and uncongested) scenarios to ensure a comprehensive assessment. Additionally, the performance of PPO is compared with baseline algorithms, including genetic algorithm (GA) [5] and rule-based heuristic (RBH) [6].

Genetic algorithm (GA) [5] optimizes solutions using evolutionary techniques like mutation and crossover. While it can perform well in static or moderately dynamic environments, its slower convergence limits its effectiveness in highly dynamic scenarios. Rule-based heuristic (RBH) [6], on the other hand, relies on fixed rules for decision-making, offering simplicity but struggling to adapt to changing logistics conditions, resulting in suboptimal performance in complex scenarios.

Fig. 2 illustrates the cumulative distribution function (CDF) of aggregate performance measures for PPO, GA, and RBH algorithms during the learning stage. This stage quantifies how each algorithm improves its dispatch optimization policy through iterative training. The PPO algorithm achieves a faster ascent in the CDF curve compared to the other methods, reaching a normalized performance measure exceeding 0.9 within a small number of iterations. This indicates that PPO can effectively identify high-quality solutions with reduced training effort. In contrast, the GA algorithm shows a slower progression. While it achieves reasonable performance, its CDF curve ascends less steeply, indicating a need for additional training iterations to achieve stabilization. The RBH algorithm demonstrates the lowest performance across all scenarios, with its normalized performance measure rarely exceeding 0.75. The static nature of RBH limits its ability to adapt to the changing conditions of logistics operations, as reflected in its slower improvement and lower final performance levels.

Fig. 3 shows the CDF of aggregate performance measures for PPO, GA, and RBH algorithms during the testing stage in hybrid scenarios. The testing stage evaluates the robustness of the optimization policies in unseen combinations of operational conditions, including hybrid scenarios that mix distance-, time-, and traffic-based elements. These scenarios simulate the variability and complexity typical of real-world logistics environments. In hybrid scenarios, the PPO algorithm consistently maintains higher performance levels, with the CDF curve reaching a normalized performance measure of 0.9 or higher in most cases. This suggests that PPO can reliably generalize its trained policy to new scenarios without significant degradation in performance. GA achieves moderate performance, with normalized performance measures ranging from 0.7 to 0.85. However, its progression shows more significant variability, reflecting its reduced stability under hybrid conditions. The RBH algorithm performs the poorest, with the majority of its results below 0.75, emphasizing its limited capability in dynamic environments.

These results highlight PPO's ability to provide faster learning and higher generalization performance compared to GA and RBH. The results suggest that PPO is well-suited for optimizing freight dispatch systems in complex and dynamic conditions, where adaptability and consistency are critical for operational success.

## IV. Conclusion

We proposed a transportation dispatch optimization framework using PPO, a reinforcement learning algorithm, to enhance the performance of automatic dispatch systems under various logistical scenarios. Simulation results demonstrated that the proposed framework achieved higher aggregate performance measures compared to traditional algorithms such as GA and RBH. Specifically, the PPO-based framework consistently maintained superior efficiency and adaptability, achieving a normalized performance measure exceeding 0.9 across diverse and hybrid scenarios. These results validate that the PPO framework not only accelerates convergence during the learning phase but also generalizes effectively in the

testing phase, ensuring reliable performance under dynamic and complex conditions. Future studies will focus on incorporating additional optimization objectives and evaluating the framework under larger-scale, real-world logistics environments to enhance its practicality and robustness further.

## ACKNOWLEDGMENT

This work was supported by the Korea Technology & Information Promotion Agency for SMEs (TIPA) grant funded by the Korea government (MSIT) (NTIS identification no. 1425178206).

## REFERENCES

[1] Busan Port Authority, "*Informatization Strategy Plan Establishment Service Report for Establishing Global Shipping and Port Logistics Integrated Platform*," 2019.

[2] J.-H. Kim, S.-Y. Kwon, H.-N. Kim, "Meta-Learning Based Transport Matching Algorithm for Automatic Vehicle/Cargo Dispatch System," *The 11th International Conference on Green and Human Information Technology (ICGHIT)*, Bangkok, Thailand, Feb. 2023.

[3] J.-H. Kim, D.-M. Shin, S.-E. Lee, and H.-N. Kim, "Performance Analysis of Vehicle/Cargo Transport Matching Algorithm based on MAML," *Journal of The Institute of Electronics and Information Engineers*, vol. 60, no. 9, Sep. 2023.

[4] H.-J. Lee, W.-S. Jang, S.-J. Lee, and D.-K. Kim, "Optimal Dispatch Modeling Method Based on Multi-agent Reinforcement Learning in Port Logistics Environment," *Proceedings of KIIT Conference*, Jeju, Korea, pp. 135-138, Dec. 2022.

[5] H.-J. Lee, W.-S. Jang, S.-J. Lee, and D.-K. Kim, "Optimal Dispatch Method Based on Genetic Algorithm in Port Logistics Environment," *Proceedings of KIIT Conference*, 2023, Jeju, Korea, pp. 236-239, Nov. 2023.

[6] D.-H. Hong and G.-J. Kim, "Optimizing the Vehicle Dispatching for Enhancing Operation Efficiency of Container Terminal," *Journal of the Korea Convergence Society*, vol. 8, no. 10, pp. 19–28, Oct. 2017.

[7] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "*Proximal Policy Optimization Algorithms*," arXiv:1707.06347v2 [cs.LG] Aug. 2017.

[8] J. Schulman, S. Levine, P. Moritz, M. Jordan, and P. Abbeel, "*Trust Region Policy Optimization*," arXiv:1502.05477v5 [cs.LG].

[9] Jordan, Scott, et al. "Evaluating the performance of reinforcement learning algorithms," *International Conference on Machine Learning. PMLR*, 2020.