Path planning optimization in industrial AGVs: A hybrid decentralized architecture

Panagiotis G. Giannopoulos
School of Social Sciences
Hellenic Open University
Patras, Greece
0000-0001-7530-2458

Vangelis Malamas
Department of Informatics
University of Piraeus
Piraeus, Greece
0000-0002-9154-8340

Dimitris Koutras
Department of Informatics
University of Piraeus
Piraeus, Greece
0000-0002-9154-8340

Thomas K. Dasaklis
School of Social Sciences
Hellenic Open University
Patras, Greece
0000-0003-1240-4822

***Abstract*—The emergence of Industry 4.0 technologies has significantly transformed supply chain operations, particularly through the deployment of Autonomous Guided Vehicles (AGVs) in logistics and manufacturing settings. Integrating Industrial Internet of Things (IIoT) devices with Artificial Intelligence (AI) has enhanced AGV autonomy by enabling real-time data-driven decision-making. However, challenges related to cybersecurity, data synchronization, and scalability, still persist in cyber-physical (CPS) manufacturing systems. Blockchain technology offers a prominent pathway towards ensuring data integrity, decentralization, and security, but its adoption in AGV applications remains limited due to scalability, latency, and computational constraints. To address this gap, this paper proposes a hybrid, blockchain-centric architecture that leverages the synergistic potential of Reinforcement Learning (RL) methods in multi-agent, collaborative AGVs. The architecture is designed to be scalable, interoperable, and resilient to cyber threats, making it suitable for a wide range of industrial applications.**

I. Introduction

The advent of Industry 4.0 technologies has transformed operations in the supply chain (SC), with mobile robotic systems—especially Autonomous Guided Vehicles (AGVs)—automating tasks from in-house logistics to urban delivery coordination. Integrating various Industry 4.0 technologies is crucial for enhancing AGVs' real-time autonomy, prompting the development of synergistic frameworks that boost different operational functions.

The integration of AGVs with Industrial Internet of Things (IIoT)-enabled devices in the hardware domain combined with the capabilities of AI methods in the software domain, has emerged as a key approach for enhancing AGV intelligence and operational efficiency [1]. By leveraging vast datasets generated through real-time sensor inputs and synchronization, AI-driven learning enables AGVs to continuously adapt to dynamic external conditions [2]. In this regard, the synergy between AI and IIoT is shaping a prominent pathway for intelligent robotics, particularly through the lens of Machine Learning (ML)-centric techniques, which are capable of analyzing diverse data types, including LiDAR, temperature, and location sensors.

979-8-3503-5348-8/24/$31.00 ©2024 IEEE

Towards their industrialization, applications in AGVs must safeguard the human and infrastructure safety. In this regard, several security concerns are raised in this domain, which mainly encompass algorithmic processes that ensure both appropriate mobility and data security. On the mobility side, main efforts are centered around training AI models to ensure human safety and optimizing performance for efficient operations. Toward this direction, significant strides have been made, due to the wealth of learning methods, which are feasible to be applied in cyber-physical systems (CPSs), which integrate AI and IIoT technologies to enhance autonomous decision-making. However, as manufacturing ecosystems become increasingly connected, cyber-security risks in IIoT networks, and AGVs operations continue to escalate, as industrial robots are increasingly exposed to cyber-attacks [3]. Recent cyber-attacks on critical infrastructure have exposed centralized data architectures' vulnerabilities, raising concerns about data integrity, unauthorized access, and operational disruptions [4], [5]. Beyond cyber-threats, challenges in data synchronization also persist, given the difficulties in coordinating real-time, AI-driven schemes in legacy industrial systems.

Blockchain, holds significant promise towards safeguarding immutability and data integrity, especially when integrated with IIoT-enabled devices. Such synergistic schemes have recently gained traction in industrial applications due to their safety-aware capabilities, in terms of providing tamper-resistant and transparent data management [6], [7]. Notwithstanding the significance of this technology in AI-centric applications, blockchain-based architectures remain relatively scarce in the field of AGVs. This may be attributed to the scalability and latency limitations inherent in blockchain technology in high-throughput manufacturing environments, computational overheads impacting real-time data exchange and the absence of standardized integration frameworks with existing industrial automation systems [8].

Notwithstanding the constraints imposed by the above factors, designing architectures that seamlessly coordinate industrial infrastructures while facilitating security and intelligence aspects, holds significant promise towards achieving the smart manufacturing paradigm. Recent approaches have highlighted that building synergistic AI-blockchain schemes are vital in vehicular settings [9], like those describing the

AGVs' operations. Aligning with this perspective, we propose a novel, blockchain-centric architecture that bridges industrialization-oriented gaps and exemplifies the Security-as-a-Service paradigm in field of collaborative AGVs. The theoretical contributions of the proposed architecture, are summarized in Table I. Technically, our approach leverages multi-agent reinforcement learning (RL) techniques to enhance both the navigation capabilities and the real-time synergies in the operation of AGVs.

TABLE I
SECURITY-ORIENTED CONTRIBUTIONS OF THE PROPOSED ARCHITECTURE

Challenges	**Description**	**Proposed solution**
Decentralized log and verification	Transmitting all the data logs in the blockchain is computationally intensive.	A hybrid architecture that ensures both verifiability (on-chain) and low latency (off-chain) by incorporating the IIoT-fused logs in an IPFS module.
Secure communication	Challenges related to communication of AGVs, include interception and authorization issues.	The architecture acts on top of devices enhanced with end-to-end encryption protocols, that are coupled with verification mechanisms through on-chain processes.
Tamper-proof task allocation	AI engine-related vulnerabilities, including potential manipulation of AGV task assignments.	The architecture accommodates hash-locked, AI-driven decisions by incorporating smart contracts in its on-chain module.

II. LITERATURE REVIEW

The groundwork for implementing RL applications in multi-robot environments was laid by [10], who first proposed a methodology to minimize the learning space through the use of behaviors and conditions, and deal with the credit assignment problem in the form of heterogeneous reinforcement functions and progress estimators. Beyond simple RL forms, cooperative path planning is also the central in studies like [11], which elaborates upon integrating fuzzy logic and RL approaches, particularly by applying a fuzzy Q-learning form with a modified Wolf-PH algorithm to address the challenges posed by simple RL methods that typically proved effective when applied in small and discrete state spaces.

Safety is a critical factor in autonomous motion planning. For instance, [12] develops an approach tailored to mitigating the lack of safety and uncontrollability challenges, often faced by mobile robots when trained under simplistic RL schemes. By composing multi-agent RL with control barrier function-based shields, the authors develop a scheme capable of advancing both planning effectiveness and scalability in ensuring motion safety. In a similar vein, [13] proposes a multi-mode filtering target tracking method for mobile robots using multi-agent RL to enhance path planning, motion control and target tracking in unknown environments. Particularly, the paper incorporates extended Kalman filter and probability data association to continuously update the observations and estimate the state of each robot in real time.

A wealth of research contributions has been directed towards warehouse and production automation. In this research stream, path finding is a central theme, as illustrated in [14], which introduces a priority-aware communication framework using graph attention networks and a deadlock-centric RL mechanism to dynamically adjust multi-agent path finding in warehouse automation. In-house logistics is addressed by [15], who incorporate multi-layer perceptrons to facilitate the development of robust attention mechanisms that fosters the applicability of actor-critic algorithms, by improving collision avoidance and target accuracy in path planning. Lifelong multi-agent path finding, crucial in large-scale warehouse robotics, is explored by [16] who introduce a decentralized multi-agent training framework for optimal real-time path planning in partially observable environments. Studies modeling the interplays between dispatching, path planning and route execution are also prominent in this stream of applications, as evidenced in [17], which proposes a decentralized training scheme in which using RL agents bid on orders based on individual observations, enhancing adaptability in dynamic environments. The reward sparsity problem is another area of study in multi-agent RL systems, as discussed in [18], which aims to enhance agent coordination by introducing a novel rewarding mechanism. This mechanism is experimentally developed through the analysis of vast state spaces, incorporating complex actions and diverse routes to improve learning efficiency.

Beyond the above approaches, hybrid deep RL methods have also been explored. For instance, [19] proposes a hybrid, master agent-like framework that integrates multi-agent and multi-task learning, where a reflective agent assists the master agent by utilizing second-order historical reward information to enhance decision-making. This approach combines value-based deep RL techniques with policy gradients, enabling agents to reflect on past decisions and identify optimal actions for real-time path planning in unknown environments. Policy-based, hybrid schemes are also evident, including the hybrid, deep RL-centric approach introduced in [20]. This approach integrates proximal policy optimization with the dynamic window approach; and incorporates a new cost function with two sub-functions to model agent-goal interactions towards optimizing decentralized multi-agent path planning.

The later research pertains to the social-awareness of AGVs' operations. In the field of warehouse logistics, the authors in [21] address cross-contamination risks in logistics by integrating conservative deep Q-learning into distributed multi-robot control frameworks to optimize task allocation and minimize human workload through optimal human-robot collaborations. In the field of external-type, pedestrian operations, where direct path planning in unfamiliar, collision-prone environments poses challenges, some authors [22] employ neural networks and kernel smoothing to approximate greedy actions and improve navigation efficiency. Expanding on these methods, the authors in [23] introduce a hybrid decentralized deep RL framework for multi-agent path planning, where agents independently learn using feature fusion techniques to prioritize vision-based information. In social-aware multi-agent RL, the authors in [24] leverage temporal-spatial graph encoders for enhanced field-of-view awareness, alongside a K-step look-ahead reward system and a multi-head global attention module, improving policy updates and adaptive multi-robot coordination.

III. Proposed Architecture

For safeguarding that the path planning decisions made by the AGVs are immutable, verifiable and autonomously generated by the AI module, we propose a hybrid, decentralized architecture. This approach leverages the verifiability of blockchain, while mitigating the scalability issues, often related to the on-chain applications in the industrial settings. A high-level overview of the proposed architecture, is depicted in Figure 1. Evidently, our architecture comprises: (i) the industrial infrastructure, (ii) the centralized AI module; and (iii) the Blockchain layer, each playing a decisive role in achieving an optimally-performed, security-aware functionality for the collaborative AGVs.

A. Industrial infrastructure

The cyber layer is designed to enhance the functionality of the physical layer, which comprises both static elements (machinery) and dynamic entities (humans, robots). Towards achieving an optimal performance, the AGVs learn to collaborate to minimize traveled distances while maximizing load capacity, ensuring that all allocated delivery tasks are completed. In this regard, we assume that all AGVs are capable of real-time mapping of the industrial environment; and that their communication and data-sharing devices comply with the ISO/IEC 27001 standard, ensuring end-to-end encryption between industrial systems. On the infrastructure side, the AGVs are assumed to be infused with LiDAR sensors for high-precision environment perception and obstacle detection, RFID scanners for automated inventory package identification; and ultrasonic sensors to facilitate close-range collision avoidance and maneuverability.

Operation-wise the AGVs handle: (i) ready-to-store and (ii) work-in-progress (WIP) products. The first class pertains to products that have already been manufactured, which need to be transferred from the production station either to an intermediate point, where another machine will transport them to the warehouse, or to an in-house delivery point, from which another vehicle will take over their transportation to the final user. On the other hand, WIP products refer to partially completed items that must be transported from one workstation to another, where their manufacturing process will be finalized.

Towards better coordination, the AGVs exchange sensor-informed logs, including congestion updates, location of obstacles, and delivery progress, which collectively facilitate the real-time collaboration by leveraging peer-to-peer (P2P), encrypted protocols. These logs are incorporated into the learning process, specifically serving as state inputs, to allow dynamic adjustments for each agent's policy based on real-time system conditions, fostering decentralized collaboration while maintaining efficiency in navigation and task execution.

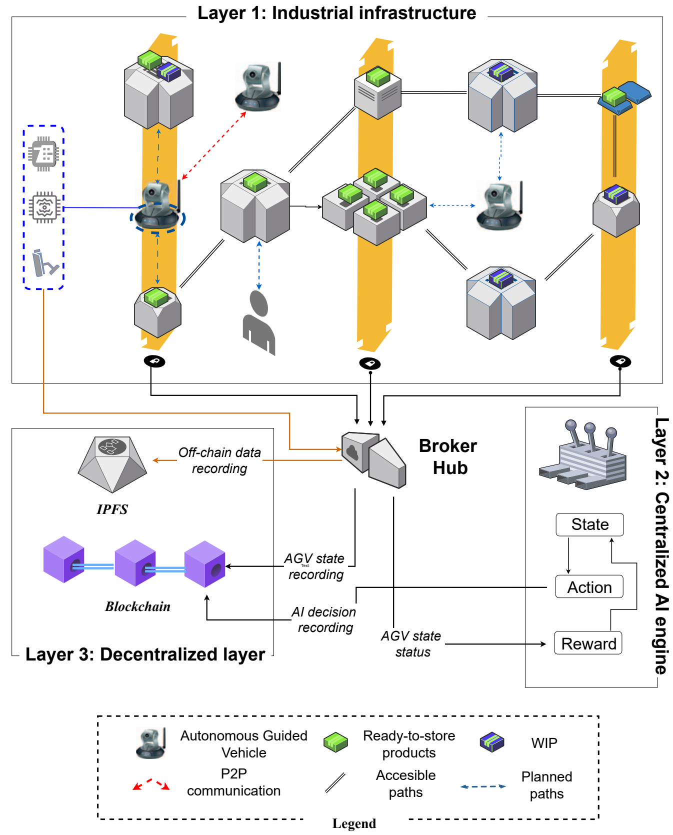

Fig. 1. High level architecture

B. AI module

An AI engine is incorporated into the proposed architecture to manage task allocation and fleet optimization in distributed AGV networks. This engine leverages multi-agent RL with value-based methods to train AGVs in learning a collaborative policy that minimizes congestion and resource conflicts, while ensuring that all the allocated tasks are met. Operating at a centralized level, the AI engine continuously receives sensor logs through a vertical broker; and is equipped with decryption mechanisms that allow for reward calculations based on real-time evidence. Task assignments are event-driven, meaning they are updated when a delivery is successfully completed. After task assignment completion, each AGV executes its own movements and local path planning autonomously, using learned policies to navigate, avoid obstacles, and adapt to dynamic conditions without requiring continuous communication with the central system.

Evidently, the proposed AI engine pertains to the Centralized Learning – Decentralized Execution (CL-DE) paradigm, which is designed to facilitate a robust integration with the Blockchain layer, to develop a seamless, yet efficient workflow. By implementing this architecture, the only on-chain logs that should be secured through the use of Smart Contracts (SMs) pertain to the actions undertaken by the agents, rather than all the corresponding sensor logs. In this regard, our approach is lightweight and secured, since it prevents tampering and unauthorized modifications, for any other reader except for the AI engine.

Technically, the CL-DE is implemented by incorporating two distinct reward mechanisms, as presented in Equation 1. To ensure a fair and unbiased reward distribution among the AGVs, the reward mechanism corresponding to the central engine (R_c) is designed in an integral form, capturing the cumulative impact of system-wide performance between two allocation plans. Specifically, we propose the integration of the following key performance indicators: (i) congestion levels (C_i), (ii) energy consumption (E_i); (iii) and task completion rates (S_i) over the interval $[T_1, T_2]$, to ensure that the assessment of AGV effectiveness is based on a global evaluation rather than isolated performance snapshots.

The second reward mechanism (R_i) is of a static nature, pertaining to the overall efficiency achieved by each AGV at a specific time (i). This mechanism is tailored to both incentivizing collaborations ($C_{\text{collab.}}$), ensuring accurate pick-up (S_{pickup}) and delivery executions ($S_{\text{del.}}$), while avoiding collisions ($C_{\text{coll.}}$), minimizing unnecessary traveled distances ($d_{\text{tr.}}$), thus optimizing local energy consumption. Both of these mechanisms are novel and tailored to facilitating the complexities of path planning in the context of industrial AGVs.

$$\begin{cases} R_{\text{c}}(T_1, T_2) = -\displaystyle\int_{T_1}^{T_2} C_i(t)dt - \int_{T_1}^{T_2} E_i(t)dt + \sum_{i=1}^{N} S_i \\ R_i = -d_{\text{tr.}} - C_{\text{coll.},i} + S_{\text{pickup}} + S_{\text{del.}} + C_{\text{collab.}} \end{cases} \tag{1}$$

Both of these mechanisms should be incorporated in a specific RL algorithm to exemplify the multi-agent, CL-DE paradigm. A key challenge in simplified CL modules, especially matrix-based, pertains to their scalability limits due to high-dimensional state representations. Notwithstanding the various, greedy techniques that have been proposed towards overcoming such challenges, we argue that they hinder the potential of the RL methods by over-simplifying them. On the contrary, we believe that replacing extensive Q-tables with Deep Q-Network (DQN), paves the way for robust approximations, thus enabling efficient learning in high-dimensional spaces. In this regard, we propose incorporating the (R_c) into a DQN form for centralized task allocation, in a manner that both minimizes computational overhead due to its off-policy nature, while allowing for fast and adaptive decisions without exhaustive look-ups.

On the other hand, DE execution could be feasibly implemented under a Q-learning scheme, as the actions that could be undertaken by each agent are typically discrete. However, this may not be the most suitable approach for modern AGVs, whose capabilities extend beyond simple discrete movement decisions to include continuous control tasks such as adaptive speed adjustments and dynamic path planning. Given this complexity, we propose a policy-based RL approach, specifically the computationally efficient Soft Actor-Critic (SAC), which excels in handling continuous action spaces and enables real-time adaptability.

SAC is also an off-policy, RL technique designed for continuous action spaces. Particularly, this technique aims at maximizing the entropy of action spaces to ensure a balance between exploration and exploitation. The optimization objective consists of three key components. First, the policy is trained to maximize the expected cumulative reward while maintaining high entropy, specifically by implementing the Equation 2, in which $R(S_t, A_t)$ represents the reward function, $\mathcal{H}(\pi(\cdot|S_t))$ is the entropy term promoting stochasticity in action selection, and α is a controlling parameter that models the trade-offs between exploration and exploitation.

$$J(\pi) = \sum_{t=0}^{\infty} \mathbb{E}\left[R(S_t, A_t) + \alpha \mathcal{H}(\pi(\cdot|S_t))\right] \tag{2}$$

The second component comprises of a soft Q-function that estimates action values while accounting for entropy. In Equation 3, $Q_\theta(S, A)$ refers to the expected return of taking action A in state S and $V_\psi(S)$ provides a stabilized estimate of the value function. The final component (see Eq. 4) is used as a update mechanism, that facilitates the identification of optimal, entropy-regularized policy at a specific time-frame.

$$\begin{cases} Q_\theta(S, A) = \mathbb{E}\left[R(S, A) + \gamma \mathbb{E}_{S' \sim p}\left[V_\psi(S')\right]\right], \\ V_\psi(S) = \mathbb{E}_{A \sim \pi}\left[Q_\theta(S, A) - \alpha \log \pi(A|S)\right]. \end{cases} \tag{3}$$

$$J_\pi = \mathbb{E}_{S \sim D}\left[\mathbb{E}_{A \sim \pi}\left[\alpha \log \pi(A|S) - Q_\theta(S, A)\right]\right]. \tag{4}$$

The functionality that makes the SAC technique particularly suitable for AGVs' DE, stems from its dual Q-network architecture, which has been shown to effectively mitigate overestimation bias and prevent the convergence to over-simplified policies [24]. This ensures that AGVs make more reliable action-value estimations, leading to improved decision-making in dynamic environments.

C. Decentralized layer

The decentralized layer in the proposed architecture is designed in order to enhance data integrity and efficiency in industrial AGV operations by integrating both on-chain and off-chain mechanisms. These mechanisms can achieve verifiability of AGV decisions along with optimized computational efficiency and the decentralized layer in the proposed architecture is directed towards introducing enhanced data integrity, security, and efficiency in industrial AGV operations by introducing both on-chain and off-chain mechanisms.

The on-chain mechanism employs a permissioned blockchain to immutably record major operational decisions made by the AI module. All the actions taken by every AGV, for example, path planning decisions, task execution, and collision avoidance maneuvers, are immutably recorded through the use of smart contracts. The decisions are, therefore, tamper-proof and can be verified in real-time with transparency maintained in fleet coordination. Recording all the sensor logs and AI-generated data on-chain directly would, however, result in significant latency and storage overhead. To avert this, only high-level operational logs and final decision records are kept on the blockchain for the sake of computational feasibility without compromising security. To supplement this, the off-chain mechanism utilizes the InterPlanetary File System (IPFS) for the decentralized storage of bulk sensor data and intermediate AI processing results. Under this mechanism, AGVs are able to share and retrieve real-time environmental data, congestion reports, and obstacle detections without loading the blockchain.

By storing hashed metadata pointers to on-chain smart contracts, the design ensures that all data logged is verifiable without subjecting the blockchain to undue storage. The hybrid design reduces latency while maintaining a secure, distributed AGV coordination architecture. The integration of the two mechanisms enhances the security of the architecture from cyber attacks and enables AGVs to operate free from centralized dependencies. On-chain verification and off-chain scalability enable the system to find a balance among security, efficiency, and flexibility under practical industrial environments. By decentralizing data storage and decision-making, the architecture enables an efficient and scalable system with real-time, AI-based AGV coordination and provable security guarantees.

IV. Functionality Description

The goal of the proposed architecture is to facilitate decentralized, AI-based path planning for AGVs with data integrity and security guaranteed through the application of blockchain technology. The system possesses clearly defined steps of operation, each of which plays a critical role in guaranteeing smooth and secure AGV coordination (see Figure 2).

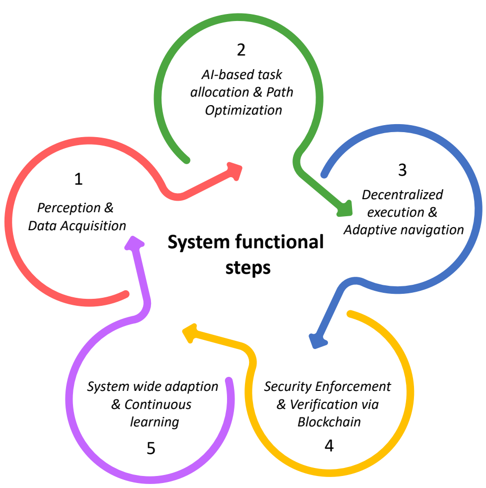

Fig. 2. System functional steps

Step 1: Perception and Data Acquisition. It begins with AGVs continuously gathering multi-modal sensory inputs to perceive the environment. With LiDAR for spatial mapping with high accuracy, ultrasonic sensors for collision detection in close proximity, and RFID scanners for inventory management automatically, AGVs generate vast amounts of real-time data. Such data streams provide critical inputs for AI-driven decision-making, logging environmental dynamics such as dynamic obstacles, congestion levels, and workspace availability. For secure transmission, all data exchange between AGVs and infrastructure components adheres to ISO/IEC 27001 encryption standards, protecting against unauthorized access and cyber threats.

Step 2: AI-Based Task Allocation and Path Optimization. Once sensor data is collected, an AI module interprets real-time input to achieve task assignment and path optimization. The AI module operates under a CL-DE paradigm with centralized training and decentralized autonomous execution. A DQN algorithm implements centralized decision-making with the utilization of historical AGV performance records in allocating optimum routes. Then, the AI engine operates through multi-agent RL for imparting collaborative fleet optimization, load balancing, energy minimization, and responsiveness to environmental uncertainties. Path optimization is designed through a reward function as a composite of efficiency of task completion, congestion minimization, and power conservation with guaranteed globally optimum navigation strategies.

Step 3: Decentralized Execution and Adaptive Navigation. Following task assignment, AGVs execute their travels independently, with SAC RL for continuous motion adaptation. In contrast to traditional discrete-action methods, SAC enables dynamic speed control and real-time obstacle avoidance by optimizing movement in continuous state-action spaces. Each AGV learns to adapt its navigation policy by peer-to-peer encrypted communication, reporting congestion and recommending detours without reliance on a central server. Decentralized operation in this way significantly reduces latency and allows adaptive decision-making, enabling AGVs to maintain efficiency even in dynamic industrial conditions.

Step 4: Security Enforcement and Verification via Blockchain. To ensure the integrity of AGV decisions, blockchain-based verification is supported in this step. The permissioned blockchain, along with the smart contracts published on it, logs important operational events, including task completions, movement verification, and collaborative interactions. This ensures immutability, tamper-proof auditing, and traceability of AGV behavior, with no potential for unauthorized modification. Due to the vast volume of sensor data, raw data is not stored directly on-chain. For this purpose, a decentralized IPFS network is used, for off-chain storage, where cryptographic hashes are stored on the blockchain through the use of smart contracts for verifiable data retrieval.

Step 5: System-Wide Adaptation and Continuous Learning. In the final stage, there is ongoing learning and system adaptation based on logged AGV performance. The AI module ingests blockchain-verified operational logs to adjust training models, refining RL methods to future tasks. Blockchain ensures that only verified, tamper-resistant data is utilized for AI retraining, improving the reliability and resilience of the system to adversarial interference.

V. Discussion and Conclusions

The integration of AI-driven decision-making and blockchain-driven security in industrial AGVs is a novel approach to address the needs of autonomous navigation and task coordination. The proposed hybrid architecture effectively solves some key challenges in AGV deployment, including real-time path planning, cybersecurity risks, and decentralized coordination, using multi-agent RL and a permissioned blockchain. By structuring the system into distinct operational steps, the proposed architecture achieves seamless coordination among autonomous agents and also a verifiable, tamper-proof record of AGV operations.

A key contribution of this work is the integration of CL-DE with a permissioned blockchain infrastructure. The AI module, trained via multi-agent reinforcement learning (RL), learns dynamically to optimize route planning and fleet coordination based on congestion, obstacle distribution, and energy efficiency. The RL framework ensures fair task allocation among AGVs, minimizing workload imbalances and bottlenecks. SAC-based motion control enables precise, real-time navigation in continuous state-action spaces. From a security standpoint, blockchain adds a decentralized, immutable trust layer for validating AGV decisions. The hybrid on-chain/off-chain architecture addresses scalability by storing only essential AGV actions on-chain and managing large sensor data volumes via IPFS. A major advantage is scalability: the framework adapts to growing AGV fleets, complex logistics, and diverse cyber-physical infrastructures. Distributed, peer-to-peer communication enhances fault tolerance and decentralized coordination, reducing dependence on centralized servers as potential failure points.

Despite its advantages, the architecture has some challenges that need further research. The application of RL models involves considerable training time and computational overhead, which may hinder rapid deployment in real-world manufacturing environments. Transfer learning techniques must be explored in future work to facilitate quicker AGV adaptation in new factory settings. In addition, while the blockchain layer introduces enhanced security, it also introduces latency in the execution of smart contracts, which may impact real-time decision-making in high-speed operational settings. Another open problem is the need for optimization of consensus protocols and also the efficiency of smart contract execution.

Acknowledgment

This work has been supported by the Pandora European Horizon project (Grant Agreement No 101135775).

References

[1] F. Tao, M. Zhang, Y. Liu, and A. Y. Nee, "Digital twin in industry: State-of-the-art," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2405–2415, 2019.

[2] S. Levine, C. Finn, T. Darrell, and P. Abbeel, "End-to-end training of deep visuomotor policies," *Journal of Machine Learning Research*, vol. 17, pp. 1–40, 2021.

[3] F. Hu, R. Xie, and H. Wang, "Cybersecurity for robotic systems in smart manufacturing: Threats and countermeasures," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 5, pp. 3253–3263, 2020.

[4] D. Koutras, V. Malamas, P. Kotzanikolaou, and T. Dasaklis, "A risk assessment methodology for supply chain tracking services," in *2023 International Conference On Cyber Management And Engineering (CyMaEn).* IEEE, 2023, pp. 555–559.

[5] H. Boyes, B. Hallaq, J. Cunningham, and T. Watson, "The industrial internet of things (iiot): An analysis framework," *Computers in Industry*, vol. 101, pp. 1–12, 2018.

[6] V. Malamas, G. Palaiologos, P. Kotzanikolaou, M. Burmester, and D. Glynos, "Janus: Hierarchical multi-blockchain-based access control (hmbac) for multi-authority and multi-domain environments," *Applied Sciences*, vol. 13, no. 1, p. 566, 2022.

[7] R. Xu, F. Yu, T. Li, Y. Peng, and B. Yang, "Blockchain: A new safeguard for cyber-physical systems," *Future Generation Computer Systems*, vol. 118, pp. 624–634, 2021.

[8] B. Guo, Z. Yu, X. Zhou, D. Zhang, and Y. Li, "Blockchain-enhanced cyber-physical systems for industry 4.0: A survey," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 51, no. 1, pp. 172–190, 2019.

[9] H. Chang, Y. Liu, and Z. Sheng, "Distributed multi-agent reinforcement learning for collaborative path planning and scheduling in blockchain-based cognitive internet of vehicles," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 5, p. 6301 – 6317, 2024.

[10] M. J. Matarić, "Reinforcement learning in the multi-robot domain," *Autonomous Robots*, vol. 4, no. 1, p. 73 – 83, 1997.

[11] D. Luviano and W. Yu, "Continuous-time path planning for multi-agents with fuzzy reinforcement learning," *Journal of Intelligent and Fuzzy Systems*, vol. 33, no. 1, p. 491 – 501, 2017.

[12] H. Cao, H. Xiong, W. Zeng, H. Jiang, Z. Cai, L. Hu, L. Zhang, and W. Lu, "Safe reinforcement learning-based motion planning for functional mobile robots suffering uncontrollable mobile robots," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 5, p. 4346 – 4363, 2024.

[13] X. Li, J. Ren, and Y. Li, "Multi-mode filter target tracking method for mobile robot using multi-agent reinforcement learning," *Engineering Applications of Artificial Intelligence*, vol. 127, 2024.

[14] J. Gao, Y. Li, Z. Ye, and X. Wu, "Pce: Multi-agent path finding via priority-aware communication & experience learning," *IEEE Transactions on Intelligent Vehicles*, pp. 1–13, 2024.

[15] C. Yang, B. Yuan, and P. Zhai, "Actor-hybrid-attention-critic for multi-logistic robots path planning," *IEEE Robotics and Automation Letters*, vol. 9, no. 6, pp. 5559–5566, 2024.

[16] M. Damani, Z. Luo, E. Wenzel, and G. Sartoretti, "Primal2: Pathfinding via reinforcement and imitation multi-agent learning-lifelong," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, p. 2666 – 2673, 2021.

[17] A. Malus, D. Kozjek, and R. Vrabič, "Real-time order dispatching for a fleet of autonomous mobile robots using multi-agent reinforcement learning," *CIRP Annals*, vol. 69, no. 1, p. 397 – 400, 2020.

[18] H. Lee, J. Hong, and J. Jeong, "Marl-based dual reward model on segmented actions for multiple mobile robots in automated warehouse environment," *Applied Sciences (Switzerland)*, vol. 12, no. 9, 2022.

[19] H. Meng and H. Zhang, "Mobile robot path planning method based on deep reinforcement learning algorithm," *Journal of Circuits, Systems and Computers*, vol. 31, no. 15, 2022.

[20] C.-T. Vu and Y.-C. Liu, "Deep reinforcement learning for multi-robot local path planning in dynamic environments," in *2024 13th International Workshop on Robot Motion and Control (RoMoCo)*, 2024, pp. 267–272.

[21] P. Wu, S. Luo, L. Tian, B. Mao, and W. Chen, "Collaborative scheduling task learning for affordance-heterogeneous multi-robot systems," *ICIC Express Letters*, vol. 17, no. 8, p. 881 – 887, 2023.

[22] D. L. Cruz and W. Yu, "Path planning of multi-agent systems in unknown environment with neural kernel smoothing and reinforcement learning," *Neurocomputing*, vol. 233, p. 34 – 42, 2017.

[23] M. Parooei, M. Tale Masouleh, and A. Kalhor, "Map3f: a decentralized approach to multi-agent pathfinding and collision avoidance with scalable 1d, 2d, and 3d feature fusion," *Intelligent Service Robotics*, vol. 17, no. 3, p. 401 – 418, 2024.

[24] L. Dong, Z. He, C. Song, X. Yuan, and H. Zhang, "Multi-robot social-aware cooperative planning in pedestrian environments using attention-based actor-critic," *Artificial Intelligence Review*, vol. 57, no. 4, 2024.