IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 7, July 2026

DOI: 10.17148/IARJSET.2026.13712

Multi-Agent Reinforcement Learning (MARL) for "Open Autonomy" in Mixed-Fleet Underground Mining

Ashok Prajapat[1], Kratika Verma[2]

Department of Mining Engineering, PAHER University, Udaipur[1]

Department of Basic Sciences & Humanities, GITS Udaipur[2]

Abstract: Modern underground mining operations are progressively integrating autonomous haulage systems (AHS) and automated Load-Haul-Dump (LHD) vehicles. However, current commercial autonomy architectures are "closed systems" that rely on rigid, centralized dispatch servers and strict zone isolation, requiring human-driven utility vehicles to completely clear an area before automated fleets can operate. This paper reviews the emerging frontier of **Multi-Agent Reinforcement Learning (MARL)** designed to achieve **"Open Autonomy"**—a decentralized paradigm where heterogeneous, mixed fleets of autonomous and human-operated vehicles dynamically co-exist and cooperate. We evaluate MARL framework configurations (e.g., Centralized Training with Decentralized Execution [CTDE]), multi-agent communication topologies under intermittent subsurface wireless connectivity, and the integration of game-theoretic collision avoidance algorithms within tight, single-lane subterranean environments.

1. INTRODUCTION: THE RIGID PARADIGM OF CLOSED AUTONOMY

In underground mining excavations, efficiency depends entirely on the synchronized movement of material from the active blasting face (stope) to the underground crusher or vertical hoist shaft. While autonomous haulage has revolutionized surface mines, translating this success underground introduces a fundamental spatial constraint: **confined, single-lane geometries** with highly restricted lines of sight and no overtaking lanes except at designated passing bays (Langer et al., 2024).

```
[ COMMERCIALLY ISOLATED ZONE ]        [ THE "OPEN AUTONOMY" FRONTIER ]

┌──────────────────────────────────────┐
┌──────────────────────────────────────┐
│ 🤖 Automated LHD  (Stalled)       │     │ 🤖 Autonomous LHD (Agent A)      │
│                                       │ vs │        ▲ (Decentralized MARL)       │
│ 🛑 [ZONE LOCKED OUT]                 │    │      ▼ (Dynamic Yielding)           │
│ 🧑 Human Utility Jeep              │     │ 🧑 Human Utility Jeep (Agent B)   │
└──────────────────────────────────────┘
└──────────────────────────────────────┘
```

Current autonomous mining operations manage this constraint through **rigid spatial isolation**: if an autonomous LHD is assigned to a production drift, that entire zone is electronically locked down, preventing human-driven utility vehicles, scissor lifts, or maintenance jeeps from entering (Langer et al., 2024; ST Engineering, 2025).

- **The Productivity Bottleneck:** If a human operator needs to inspect a water line, production halts entirely while the autonomous fleet is cleared out.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ∺ Peer-reviewed & Refereed journal ∺ Vol. 13, Issue 7, July 2026

DOI: 10.17148/IARJSET.2026.13712

- **The Telemetry Failure Point:** Centralized dispatch algorithms (e.g., classic MILP or fleet management servers) require constant, unyielding Wi-Fi or LTE connectivity to every vehicle. In deep mines, where ongoing blasting frequently destroys fiber backhauls or creates wireless blind spots around corners, a centralized system immediately stalls the entire fleet for safety whenever communication drops (ST Engineering, 2025).

To overcome these limitations, research is pivoting toward **Open Autonomy** powered by MARL (Rizk et al., 2020; Oroojlooy & Hajinezhad, 2023). This architecture shifts decision-making directly onto the vehicles, allowing individual machines to behave as intelligent, adaptive agents that can negotiate right-of-way and coordinate schedules on the fly.

2. MARL ARCHITECTURAL DESIGN FOR SUBSURFACE FLEETS

Underground MARL applications cannot rely on fully centralized execution due to the reality of underground signal loss. Instead, modern implementations utilize a **Centralized Training with Decentralized Execution (CTDE)** framework (Oroojlooy & Hajinezhad, 2023).

```
[ STAGE 1: CENTRALIZED TRAINING (Surface Simulator) ]

- Global Value Function Approximation (Q-Mix / MAPPO)

- Simulates millions of tunnel-conflict iterations

                 |

                 ▼ (Deploying Trained Actor Networks)

[ STAGE 2: DECENTRALIZED EXECUTION (Underground Active Fleet) ]

 ┌──────────────────────────────────────┐
┌────────────────────────────────┐

 | 🤖 LHD Agent A (Onboard AI)       | ◄────► | 🤖 Haul Truck B (Onboard AI) |

 | - Evaluates local LiDAR           | Ad-Hoc |  - Evaluates local LiDAR     |

 | - Executes local Policy π_A       | V2V Co | - Executes local Policy π_B  |

 └──────────────────────────────────────┘
└────────────────────────────────┘
```

2.1. The Actor-Critic Framework Setup

In a mixed-fleet mine, each autonomous machine runs its own onboard deep neural network policy (π_{θ_i}), mapping its local sensor inputs (LiDAR, radar, articulated frame angle) directly to localized mechanical outputs (steering angle, throttle, braking, bucket hydraulics).

- **The State Space (S):** Contains local velocity, proximity point-clouds, relative distance to the closest passing bay, and basic velocity/heading data shared by neighboring vehicles via short-range **Vehicle-to-Vehicle (V2V)** wireless loops.

- **The Action Space (A):** Discrete or continuous control over machine velocity vectors and path selection at tunnel junctions.

2.2. Value-Factorization Networks (QMIX & MAPPO)

To ensure multiple self-driving vehicles cooperate instead of acting greedily (which would result in two trucks meeting nose-to-nose in a single-lane tunnel and blocking each other indefinitely), the training loop uses a value-factorization algorithm like **QMIX** or Multi-Agent PPO (MAPPO) (Oroojlooy & Hajinezhad, 2023).

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
Impact Factor 8.066 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 13, Issue 7, July 2026
DOI: 10.17148/IARJSET.2026.13712

During surface simulation training, the system evaluates a global reward function (R_{global}) based on total tons of ore moved and zero vehicle collisions. QMIX breaks this global reward down into individual, decentralized action-value choices (Q_i) for each machine. This ensures that even when executing decisions completely offline underground, an individual truck's local choice inherently maximizes the performance of the entire fleet.

3. ALGORITHMIC COORDINATION STRATEGIES IN MIXED TUNNELS

3.1. Game-Theoretic Right-of-Way Negotiation

When two automated machines—or an automated LHD and a human-driven jeep—approach each other from opposite ends of a single-lane ramp, the MARL policy must resolve the conflict. By applying a **Markov Game** formulation, the autonomous agent continuously estimates the human driver's intent based on their observed deceleration rate and distance to the nearest passing bay (Langer et al., 2024).

[SINGLE-LANE SUBTERRANEAN DRIFT]

🤖 Autonomous LHD (Agent A) ──► ◄── 🧑 Human Jeep

[Action: Calculate Target Bay] [Observed Deceleration]

| Agent A Yields & Pulls Into Passing Bay | | Human Proceeds Safely Down Main Ramp |

If the model calculates that the human-driven vehicle has passed the last available passing bay, the autonomous agent's policy executes a cooperative yielding maneuver. It autonomously reverses or pulls into its own adjacent passing bay, clearing the main haulage route without requiring any human dispatch intervention.

3.2. Robustness Under Intermittent Ad-Hoc Communication

Unlike surface robots that can query global GPS coordinates, underground agents rely on unstable, ad-hoc wireless communication. Modern MARL frameworks incorporate **Attention-Based Communication Mechanisms (e.g., ATOC)** (Rizk et al., 2020).

Instead of continuously broadcasting large data streams, the vehicles' neural networks utilize an attention gate to decide exactly *when* communication is physically necessary. When driving down an empty, straight tunnel, communication drops to zero. As a vehicle approaches a blind 90-degree intersection, the attention mechanism spikes, establishing a localized, high-priority ad-hoc mesh network with any active agent within a 100-meter radius to safely coordinate the crossing.

DOI: 10.17148/IARJSET.2026.13712

4. TECHNICAL COMPARISON: CLOSED VS. OPEN AUTONOMY

Architectural Metric	Legacy Closed Autonomy (Current Industry Standard)	MARL-Driven Open Autonomy (Future Vision)
Fleet Composition	Homo-homogeneous (Automated units only; zero human co-habitation).	Heterogeneous (Mixed autonomous haulers and human utility vehicles).
Decision Making	Centralized (Surface server dictates precise route schedules via strict time slots).	Decentralized (Onboard edge chips compute immediate velocity and routing profiles).
Network Reliance	High Vulnerability (A local Wi-Fi drop safely triggers an immediate emergency stop across the entire zone).	Fault-Tolerant (Vehicles utilize local V2V mesh networks and LiDAR to maintain operation during network outages).
Infrastructure Overhead	High (Requires electronic gates, laser barriers, and extensive zone lockout infrastructure).	Minimal (Relies on machine-integrated sensors and onboard edge computing power).

5. CRITICAL ENGINEERING AND SAFETY GAPS

5.1. The Sim-to-Real Transfer and Geomechanical Variance

MARL networks require millions of training iterations to learn safe driving behaviors, a process that must take place within high-fidelity 3D simulators. However, transferring a policy trained in a smooth, virtual simulator into a real underground mine creates a severe **Sim-to-Real gap**. Real tunnel floors are highly irregular, rutted, frequently flooded with muddy water, and change friction coefficients daily. A policy that calculates a precise stopping distance in a simulator may slip on a slick incline in a real mine, making robust domain randomization during training an absolute necessity.

5.2. Safety Verification and Non-Deterministic Failure Modes

For mine managers, the primary barrier to adopting MARL is the **lack of deterministic guarantees**. Because deep neural networks evaluate continuous sensor streams on the fly, it is mathematically difficult to prove that a reinforcement learning model will never choose an unhandled, dangerous action when encountering an entirely new scenario underground. To mitigate this risk, current research focuses on building **Control Barrier Functions (CBFs)** that act as hardcoded safety envelopes wrapped around the AI's choices. If the MARL policy outputs a steering angle or acceleration vector that violates physical safety limits, the CBF instantly overrides the command to prevent a collision.

6. CONCLUSION

Multi-Agent Reinforcement Learning represents the definitive pathway toward flexible, high-yield "Open Autonomy" in subsurface engineering. Moving past the era of strict zone lockouts allows mines to seamlessly integrate automated productivity with necessary human maintenance workflows. By leveraging decentralized CTDE architectures, autonomous fleets can maintain stable operations through localized communication dropouts, mitigating the infrastructure vulnerabilities that challenge current autonomous mine layouts. Resolving the remaining challenges in sim-to-real transfer and non-deterministic safety verification will unlock fully adaptive, self-organizing underground logistics ecosystems.

REFERENCES

[1]. Langer, J., et al. (2024). Operational challenges and safety validation framework for mixed-fleet autonomous operations in underground hard-rock mines. *International Journal of Mining Science and Technology*, 34(2), 201–215.

[2]. Oroojlooy, A., & Hajinezhad, D. (2023). A review of cooperative multi-agent deep reinforcement learning. *Applied Intelligence*, 53(11), 13620–13645. https://doi.org/10.1007/s10489-022-04105-y

[3]. Rizk, Y., Awad, M., & Tunstel, E. (2020). Decision making in multi-agent systems: A review. *Robotica*, 38(8), 1357–1384.

[4]. ST Engineering Technology. (2025). *De-risking autonomous haulage systems for deep subsurface mining applications*. Industry Insights Whitepaper, Vol. 12.

[5]. Tanimoto, J. (2021). *Fundamentals of Evolutionary Game Theory and Multi-Agent Systems*. Springer Nature. https://doi.org/10.1007/978-981-15-9643-8