Scalable Strategies for Cooperative and Competitive Multi-Agent Systems

Grant Douglas
Adelaide University
Adelaide, Australia
grant.douglas@adelaide.edu.au

ABSTRACT

This research develops principled methods for enabling multi-agent systems to operate effectively at swarm scale, motivated by adversarial asset-protection scenarios. Existing studies either focus on small-team tactical coordination, where joint-state spaces remain tractable, or on large-swarm homogeneous behaviours like flocking. We bridge this gap by formalising structured coordination as a scalable alternative to end-to-end joint policy learning. The framework constructs heterogeneous, scenario-conditioned strategies using modular heuristics, evolutionary optimisation, and game-theoretic principles, producing decentralised policies that are robust to agent loss and environmental uncertainty. Offline design shifts complexity away from execution, minimising communication and computational overhead. The approach is validated in high-performance simulations of defending teams against large attacking swarms, demonstrating coordination at scales previously considered intractable. These results establish a generalisable methodology for designing, evaluating, and deploying large-scale adversarial multi-agent systems with interpretability, robustness, and resource-aware strategic reasoning.

KEYWORDS

Multi-Agent Systems; Scalable Coordination; Adversarial Swarms; Evolutionary Optimisation

ACM Reference Format:
Grant Douglas. 2026. Scalable Strategies for Cooperative and Competitive Multi-Agent Systems. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 3 pages. https://doi.org/10.65109/JCOC8756

1 INTRODUCTION

Coordinating large groups of autonomous agents remains a central challenge in multi-agent systems. While cooperative and competitive learning methods achieve notable success at small team sizes, scalability becomes increasingly difficult as the number of agents grows. Joint state–action spaces expand combinatorially [10, 21], rendering exhaustive planning and naive learning intractable and brittle under distributional shift or partial system failure [2, 22].

These limitations are particularly pronounced in adversarial and high-stakes settings, such as pursuit–evasion and asset protection, where the penetration of even a single attacker can lead to global task failure [10, 11, 13]. Although reinforcement learning and MARL show promise in small-scale cooperative and mixed environments [7, 14], their effectiveness degrades with scale due to non-stationarity, credit assignment challenges, sparse rewards, and high sample complexity [9, 15]. Many learning-based approaches remain practically limited to modest team sizes in continuous adversarial domains [3, 20].

More broadly, scalability in MAS is frequently approached through improved optimisation or architectural refinement, rather than as a primary structural design objective for heterogeneous coordination. Existing literature is effectively bifurcated: sophisticated tactical role differentiation is typically demonstrated in small teams via joint optimisation, whereas large-swarm research emphasises homogeneous navigation rules such as attraction–repulsion dynamics [17]. Approaches that maintain differentiated tactical roles at substantial scale remain comparatively underexplored.

This research addresses this gap by proposing a staged factorisation paradigm in which scale is treated as a first-class design constraint. Rather than relying on end-to-end joint policy learning, the proposed framework adopts a Centralised Design, Decentralised Execution (CDDE) methodology that decomposes coordination into modular, reusable components discovered through offline optimisation. Evolutionary search and structured heuristic representations provide interpretable, decentralised policies that are computationally lightweight and robust to agent loss, enabling heterogeneous tactical coordination at scales beyond those typically studied in adversarial MARL [8, 16].

2 CURRENT RESEARCH

The trajectory of this research is guided by three principles: (i) scalable systems favour simple, composable components over monolithic complexity; (ii) large coordination problems can be factorised into sub-problems that admit tractable solutions; and (iii) in adversarial, safety-critical domains, coordination mechanisms must remain transparent and auditable. These principles are consistent with observations from natural swarms and distributed systems, where complex global behaviour emerges from structured combinations of simple local rules rather than tightly coupled global control.

Together, these considerations motivate a staged factorisation of multi-agent coordination treating scale as a primary design constraint. Complexity is shifted to an offline optimisation phase, producing lightweight policies that execute in a decentralised manner. Communication is treated as an optional enhancement rather than a structural requirement, ensuring coordination remains robust under bandwidth limitations, agent loss, or degraded connectivity.

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/JCOC8756

All studies are evaluated in a high-performance JAX simulation environment enabling massively parallel adversarial swarm engagements. Here, a defending team must intercept an attacking swarm following nonlinear trajectories to protect a fixed asset.

2.1 Discovering Modular Building Blocks

The first study introduces *LLM-guided Hierarchical Evolutionary Learning with Permutation-invariant Surrogates (LLM-HELPS)* [5]. Rather than learning end-to-end joint policies, large language models are used offline to propose interpretable low-level heuristics, which are then composed hierarchically and optimised via evolutionary search. A permutation-invariant surrogate fitness model enables efficient team-level evaluation without full simulation for every candidate, substantially reducing computational cost and mitigating credit-assignment challenges common in MARL [9, 14]. This stage establishes a reusable library of behavioural primitives and demonstrates coordinated defence at moderate scales.

2.2 Structural Scaling via Factorised Assembly

Building on these primitives, the second study [4] investigates how small-team coordination can be systematically scaled. A hybrid Genetic Algorithm–Dynamic Programming (GA–DP) framework evolves effective strategies for small sub-engagements and assembles them into large-scale formations through polynomial-time decomposition. By treating entire small-team chromosomes as compositional units, the approach preserves higher-order behavioural dependencies [8, 16]. This factorisation transforms an otherwise combinatorial search into tractable optimisation, enabling coordinated engagements at substantially larger scales than direct joint optimisation permits.

2.3 Scenario-Conditioned Composition

The third study [6] extends static assembly into adaptive deployment. Instead of applying a fixed global strategy, scenario-conditioned policy composition separates offline policy discovery from online allocation. A geometric dispersion metric characterises adversary configurations, allowing dynamic programming to partition swarms into spatially coherent sectors and assign appropriate sub-team strategies in polynomial time. Empirical results demonstrate that conditioning coordination on adversary structure improves robustness and scalability as swarm size increases.

2.4 Ongoing Work: Adaptive re-partitioning

A natural extension is to allow defensive formations to recompose when adversaries deliberately alter spatial configurations, similar to receding-horizon re-partitioning [1, 18]. Current work integrates these modular components into a unified neuro-symbolic controller based on graph state-space models. A graph neural backbone captures relational structure within the swarm, while a state-space module models temporal dynamics with linear scaling. Rather than outputting raw control actions, the network selects among previously discovered heuristics, preserving interpretability while leveraging deep representation learning. This approach aims to unify structured coordination and dynamic contextual recomposition without reverting to monolithic joint policy optimisation.

3 FUTURE DIRECTIONS

The next phase of this research extends scalable coordination toward higher-level strategic reasoning, building on the neuro-symbolic Graph-SSM framework and scenario-conditioned recomposition mechanisms. One focus is resource allocation across multiple assets, where heterogeneous defenders must be efficiently deployed under adversarial pressure. Game-theoretic principles will guide the assignment of limited defensive resources, drawing on established models of strategic security allocation such as Stackelberg security games used in real-world infrastructure protection contexts [12, 19].

A second direction addresses robustness under environmental and communication degradation. Future work will evaluate how decentralised heuristics and adaptive recomposition maintain performance under partial observability, low-bandwidth or delayed communication, and sim-to-real discrepancies. Understanding these dynamics is critical for deploying large-scale autonomous systems in realistic operational conditions.

Finally, the research will explore generalisability beyond adversarial defence. Planned applications include heterogeneous multi-target protection and distributed disaster response, testing scenario-conditioned decomposition and adaptive re-partitioning under dynamic, uncertain conditions. Together, these directions aim to advance adaptive, verifiable, and resource-aware multi-agent systems while preserving the principle that scalability must be an explicit design objective rather than an emergent by-product.

4 CONCLUSION AND IMPACT

This research addresses a central challenge in multi-agent systems: enabling cooperative and competitive coordination at scales where joint optimisation becomes impractical. Rather than extending small-team methods beyond their intended regime, the work treats scale as a primary architectural constraint, shaping how strategies are represented, constructed, and deployed.

The completed studies demonstrate that staged factorisation - combining modular heuristic discovery, structural assembly, and scenario-conditioned composition - can support heterogeneous coordination at substantially larger scales than conventional monolithic learning approaches. By shifting complexity to offline design and preserving lightweight decentralised execution, the framework achieves scalability while maintaining interpretability and bounded computational requirements.

Beyond adversarial defence, the broader contribution lies in articulating design principles for large-scale MAS: favouring compositional simplicity over monolithic complexity, decomposing global coordination into tractable substructures, and preserving transparency in safety-critical settings. These principles contribute not only algorithms, but an architectural perspective on how large-scale multi-agent systems can remain scalable, auditable, and robust under realistic operational constraints.

ACKNOWLEDGMENTS

G. Douglas has received support through the Australian Government Research Training Program Scholarship. This research is also supported in part by Saab Australia Pty Ltd through the Defence Trailblazer program and with super-computing resources provided by the Phoenix HPC service at Adelaide University.

REFERENCES

[1] J. Bellingham, A. Richards, and J.P. How. 2002. Receding horizon control of autonomous aerial vehicles. In *Proceedings of the 2002 American Control Conference*, Vol. 5. IEEE, Anchorage, AK, USA, 3741–3746 vol.5. https://doi.org/10.1109/ACC.2002.1024509 ISSN: 0743-1619.

[2] Daniel S Bernstein, Robert Givan, Neil Immerman, and Shlomo Zilberstein. 2002. The Complexity of Decentralized Control of Markov Decision Processes. *Mathematics of Operations Research* 27, 4 (2002), 819–840.

[3] Pei Chi, Jiahong Wei, Kun Wu, Bin Di, and Yingxun Wang. 2023. A Bio-Inspired Decision-Making Method of UAV Swarm for Attack-Defense Confrontation via Multi-Agent Reinforcement Learning. *Biomimetics* 8, 2 (2023), 222. https://doi.org/10.3390/biomimetics8020222 ObjectType-Article-1.

[4] Grant Douglas, Stephen Franklin, Claudia Szabo, and Mingyu Guo. 2026. Building Large-Scale Drone Defenses from Small-Team Strategies. (2026). Accepted for publication.

[5] Grant Douglas, Stephen Franklin, Claudia Szabo, and Mingyu Guo. 2026. LLM-guided Hierarchical Evolutionary Learning with Permutation-invariant Surrogate. (2026). Manuscript under review.

[6] Grant Douglas, Stephen Franklin, Claudia Szabo, and Mingyu Guo. 2026. Scenario-Conditioned Policy Composition for Large-Scale Multi-Agent Coordination. (2026). Manuscript under review.

[7] Jakob N. Foerster, Gregory Farquhar, Triantafyllos Afouras, Nantas Nardelli, and Shimon Whiteson. 2018. Counterfactual multi-agent policy gradients. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence (AAAI'18/IAAI'18/EAAI'18)*. AAAI Press, New Orleans, Louisiana, USA, Article 363, 9 pages.

[8] David E. Goldberg. 1989. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley Pub. Co., Reading, Mass.

[9] Pablo Hernandez-Leal, Michael Kaisers, Tim Baarslag, and Enrique Munoz de Cote. 2019. A Survey of Learning in Multiagent Environments: Dealing with Non-Stationarity. https://doi.org/10.48550/arXiv.1707.09183 arXiv:1707.09183 [cs].

[10] Rufus Isaacs. 1951. *Games of Pursuit*. RAND Corporation, Santa Monica, CA.

[11] Rufus Isaacs. 1965. *Differential games : a mathematical theory with applications to warfare and pursuit, control and optimization*. Wiley, New York. Bibliography: p. 379-380.

[12] Y. Jiang, Huamin Yang, Jianping Zhao, Huan Wang, Jinqing Li, Xiaoqiang Di, Xu Liu, Ligang Cong, and Neale R. Smith. 2019. Allocating Limited Resources to Protect a Massive Number of Targets Using a Game Theoretic Model. *Mathematical problems in engineering* 2019, 2019 (2019), 1–16.

[13] Arthur Louette, Pascal Leroy, Yanis Geurts, and Damien Ernst. 2025. Existing Gaps In Reinforcement Learning For Drone Warfare. (2025). Author preprint.

[14] Ryan Lowe, Yi Wu, Aviv Tamar, Jean Harb, Pieter Abbeel, and Igor Mordatch. 2017. Multi-agent actor-critic for mixed cooperative-competitive environments. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (Long Beach, California, USA) *(NIPS'17)*. Curran Associates Inc., Red Hook, NY, USA, 6382–6393.

[15] Georgios Papoudakis, Filippos Christianos, Lukas Schäfer, and Stefano V. Albrecht. 2021. Benchmarking Multi-Agent Deep Reinforcement Learning Algorithms in Cooperative Tasks. https://doi.org/10.48550/arXiv.2006.07869 arXiv:2006.07869 [cs] version: 4.

[16] Michael A. Potter and Kenneth A. De Jong. 2000. Cooperative Coevolution: An Architecture for Evolving Coadapted Subcomponents. *Evolutionary Computation* 8, 1 (2000), 1–29.

[17] Craig W. Reynolds. 1987. Flocks, herds and schools: A distributed behavioral model. In *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '87)*. Association for Computing Machinery, New York, NY, USA, 25–34. https://doi.org/10.1145/37401.37406

[18] Sharad Kumar Singh, Puduru Viswanadha Reddy, and Bhaskar Vundurthy. 2022. Study of Multiple Target Defense Differential Games Using Receding Horizon-Based Switching Strategies. *IEEE Transactions on Control Systems Technology* 30, 4 (July 2022), 1403–1419. https://doi.org/10.1109/TCST.2021.3104857

[19] M. Tambe, M. Jain, J. A. Pita, and A. X. Jiang. 2012. Game theory for security: Key algorithmic principles, deployed systems, lessons learned. In *2012 50th Annual Allerton Conference on Communication, Control, and Computing*. IEEE, Monticello, Illinois, 1822–1829.

[20] Baolai Wang, Shengang Li, Xianzhong Gao, and Tao Xie. 2021. UAV Swarm Confrontation Using Hierarchical Multiagent Reinforcement Learning. *International Journal of Aerospace Engineering* 2021 (2021), 1–12. https://doi.org/10.1155/2021/3360116

[21] Huan Wang and Jintao Wang. 2024. Enhancing multi-UAV air combat decision making via hierarchical reinforcement learning. *Scientific reports* 14, 1 (2024), 4458–4458. https://doi.org/10.1038/s41598-024-54938-5 ObjectType-Article-1.

[22] Michael J. Wooldridge. 2009. *An introduction to multiagent systems* (2nd ed. ed.). John Wiley & Sons, Chichester, U.K.