Journal of the KNST

2026; Vol.9, No.2; pp.???-???

ISSN: 2635-4926 https://doi.org/10.31818/JKNST.2026.6.9.2.?

Received: ????/??/??
Revised: ????/??/??
Accepted: ????/??/??
Published: ????/??/??

*Corresponding Author.
Bong Wan Choi
70 Hannam-ro, Daedeok-gu, Daejeon, 34430, Republic of Korea
Tel: +82-42-629-7989, +82-42-629-8530
E-mail: bwchoi721@hnu.kr

군집 USV 무장할당을 위한 다중 에이전트 강화학습 기반 실시간 분산 의사결정 구조 연구

Multi-Agent Reinforcement Learning-Based Real-Time Distributed Decision-Making Framework for Swarm USV Weapon Target Assignment

이우석[1], 김준영[2], 이혁주[1], 최봉완[3*]

[1]한남대학교 산업공학과 연구원
[2]LIG넥스원 무인수상정개발단 수석연구원
[3]한남대학교 산업공학과 교수

Woo Seok Lee[1], Jun Young Kim[2], Hyuk Joo Lee[1], Bong Wan Choi[3*]

[1]Researcher, Dept. of Industrial & Management Engineering, Hannam University
[2]Chief system engineer, Unmanned Surface Vehicle Systems R&D, LIG Nex1
[3]Professor, Dept. of Industrial & Management Engineering, Hannam University

Abstract

본 논문은 전투용 군집 무인수상정의 교전 효율 극대화를 위한 다중 에이전트 강화 학습 기반 실시간 분산 의사결정 구조를 제안한다. 이는 무기표적할당 문제를 시간윈도우가 포함된 다중모드 자원제약 프로젝트 스케줄링 문제로 변환한 후 실시간 최적화를 위해 리스트스케줄링 휴리스틱으로 초기해를 찾고, 가변이웃탐색으로 해를 개선한다. 특히, 그래프 신경망으로 동적 전장 상황을 임베딩하고, 중앙집중식 학습 및 분산형 실행 구조의 근접 정책 최적화 알고리즘을 융합하여 데이터링크 소실 같은 통신단절 상황에서도 에이전트의 자율적 전술 결심이 가능하게 한다. 시뮬레이션 결과 5초의 임계 시간 내에 근사 최적해를 안정적으로 산출하며, 기존 방식 대비 생존성과 표적 무력화율을 향상시켰다.

This study proposes a multi-agent reinforcement learning framework to maximize the engagement efficiency of combat unmanned surface vehicle swarms. The weapon target assignment problem is formulated as a multi-mode resource-constrained project scheduling problem with time windows. For real-time optimization, a list-scheduling heuristic is used to generate initial solutions, which are refined by variable neighborhood search. In particular, graph neural networks are used to represent dynamic topologies, and a centralized training and decentralized execution (CTDE)-based multi-agent proximal policy optimization framework enables autonomous decision-making during data link loss. Simulation results demonstrate that the proposed model achieves near-optimal solutions within 5 s, thereby improving survivability and target neutralization.

Keywords

무기표적할당(Weapon Target Assignment), 다중 에이전트 강화학습 (Multi-Agent Reinforcement Learning), 가변이웃탐색(Variable Neighborhood Search), 그래프 신경망(Graph Neural Networks), 다중 에이전트 근접 정책 최적화 (Multi-Agent Proximal Policy Optimization)

Acknowledgement

이 논문은 2026년 정부(방위사업청)의 재원으로 국방기술진흥연구소의 지원을 받아 수행된 연구임(KRIT-CT-25-011*)

1. 서론

21세기 해상안보 환경은 고가치 유인 플랫폼 중심의 교전방식에서 벗어나 데이터링크와 인공지능이 결합된 모자이크전(mosaic warfare) 및 군집전(swarm warfare)으로 급격히 진화하고 있다[1]. 최근 미국-이란, 우크라이나-러시아 전쟁 등에서 목격된 바와 같이 4차 산업혁명 기술의 전장 적용이 가속화됨에 따라 유무인복합전투체계가 미래 해전의 핵심으로 급격히 부상하였다[2-4]. 특히, 저비용 고효율의 무인수상정(unmanned surface vehicle, USV)과 자폭드론을 활용한 비대칭 포화공격은 고가치 해군자산에 대한 치명적인 위협으로 대두되었다[5].

적대세력은 수십에서 수백 대에 이르는 소형 무인 무기체계를 동시다발적으로 운용하여 아군 방어체계의 탐지 및 요격 용량을 포화시키는 전술을 구사한다. 이러한 군집공격 상황에서는 전장 상황이 밀리초 단위로 급변하며, 인간 지휘관이나 운용자가 인지하고 판단해야 할 정보의 양이 인지 한계를 초과하게 된다. 기존의 중앙집중식 지휘통제시스템과 규칙 기반의 무장할당 알고리즘은

© 2026 Korea Society for Naval Science & Technology

이러한 대규모 동적 군집표적에 대해 실시간으로 최적의 대응해를 산출하는 데 구조적인 한계를 드러내고 있다[6-8].

해상환경은 파도, 해무, 다중경로 페이딩 등으로 인해 USV 간 통신 대역폭이 제한적이며, 패킷 손실과 지연이 빈번하게 발생한다. 모든 센서정보를 중앙서버로 전송하고, 일괄 할당하는 방식은 통신 병목현상을 유발하며, 중앙노드 피격 시 전체 시스템이 마비되는 단일 장애점 문제의 취약성을 내포한다[9].

따라서, 현대 해전의 군집 위협에 효과적으로 대응하기 위해서는 각 USV가 국소적인 관측정보와 제한된 이웃 간 통신만을 활용하여 자율적이고 협력적으로 의사결정을 수행할 수 있는 분산형 지휘통제체계가 필수적이다. 또한, 무장 할당 문제는 표적의 이동, 무장의 재장전 시간, 사격 가능구역 등 시간적 제약조건이 매우 중요한 동적 최적화 문제이므로 이를 단순한 정적 할당이 아닌 실시간 스케줄링 문제로 접근하는 새로운 방법론이 요구된다.

본 연구의 핵심 목적은 군집 USV 환경에서 발생하는 동적 무장할당 문제를 해결하기 위해 다중 에이전트 강화학습 기반 실시간 분산 의사결정 프레임워크를 제안하고, 그 유효성을 검증하는 것이다. 이를 달성하기 위한 세부 연구 목표는 다음과 같다.

첫째, 기존의 정적 WTA 모델이 간과했던 시간적 동적성을 정밀하게 반영하기 위해 무장할당 문제를 다중모드 자원제약 프로젝트 스케줄링 문제(multi-mode resource constrained project scheduling problem with time windows, MRCPSPTW)로 재정의하고, 수학적으로 정식화한다. 이를 통해 표적의 진입 및 이탈시간, 무장의 재장전 및 냉각시간, 표적의 위협도 변화 등을 통합적으로 고려한 최적화 모델을 수립한다[10,11].

둘째, 대함전 특유의 센서갱신 주기와 전장상황의 긴박성을 반영하여 5초의 순환계획주기(rolling horizon)를 도출하였으며, 이 엄격한 시간예산 내에서 실시간성 보장과 해 품질 간의 트레이드오프를 극복하는 3계층 최적화 아키텍처를 Fig. 1과 같이 설계한다. 구체적으로 Phase 1에서는 5요소 위협도(threat index, TI) 기반 휴리스틱을 0.5초 내에 실행하여 초기해를 생성하고, Phase 2에서는 가변이웃탐색(variable neighborhood search, VNS) 메타휴리스틱을 2초간 적용하여 해의 품질을 개선하며, Phase 3에서는 그래프 신경망(graph neural network, GNN)과 다중 에이전트 근접정책 최적화(multi-agent proximal policy optimization, MAPPO)를 결합한 강화학습을 1초간 실행하여 최종적인 협력적 의사결정을 수행한다. 나머지 1.5초는 전장상황인식(phase 0, 0.5초)에 할당하고, 남은 1초는 센서갱신, 통신지연 보정, 물리적 무장구동(포탑 선회 등)을 위한 버퍼로 분할 할당한다[12-14].

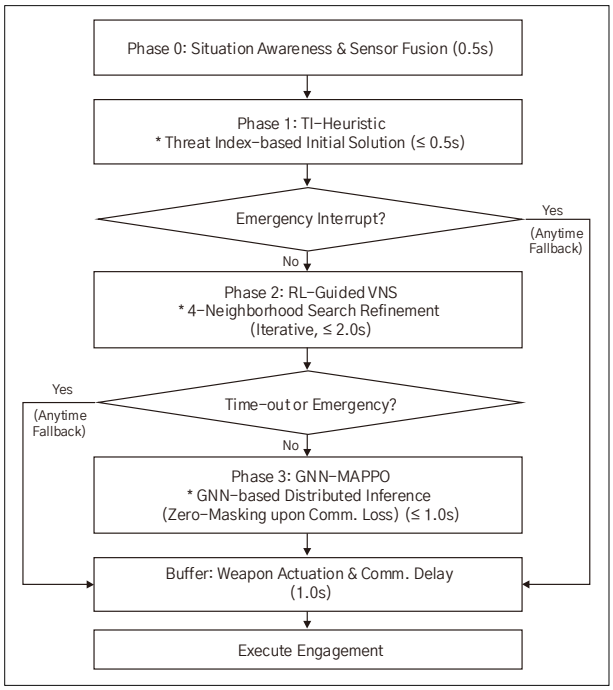


Fig. 1. Flowchart of the 3-layer hybrid WTA decision framework within a 5-second rolling horizon

셋째, 통신이 제한된 해상환경에서 다수 USV 간의 효율적인 협업을 실현하기 위해 GNN과 MAPPO를 결합한 중앙집중식 학습 및 분산형 실행(centralized training, decentralized execution, CTDE) 구조를 제안한다. USV 군집을 동적 그래프로 모델링하고, GNN을 통해 이웃 에이전트의 정보를 효율적으로 집계함으로써 통신량을 최소화하면서도 전체 군집전력의 상황인식 능력을 극대화한다[15,16].

2. 이론적 배경

2.1 무기표적할당(WTA) 문제

WTA 문제는 제한된 무기 자원을 다수의 표적에 할당하여 아군의 생존성을 극대화하거나 적의 기대

피해를 최대화하는 고전적인 국방 운용과학 문제이다. 1950년대 Manne이 처음 제안한 이래로 WTA는 군사 작전 효율화의 핵심 난제로 다루어져 왔으며, 지난 60여년 간 수학적 최적화, 메타휴리스틱, 그리고 최근의 기계학습 기법에 이르기까지 다양한 방법론이 적용되어 왔다[17].

정적 WTA(static WTA, SWTA)는 모든 표적과 무기의 상태가 고정되어 있고, 모든 정보가 완전히 알려진 상태에서 단일 단계의 최적 할당을 수행하는 문제이다. N개의 무기와 M개의 표적이 있을 때, 무기 i가 표적 j를 파괴할 확률 pij와 표적 j의 가치 Vj가 주어졌을 때, 잔존 표적 가치의 기댓값을 최소화하는 목적함수는 식 (1)과 같이 비선형정수계획법(nonlinear integer programming)으로 정식화된다[18].

$$\min \sum_{j=1}^{M} V_j \cdot \prod_{i=1}^{N} (1-p_{ij})^{x_{ij}} \quad (1)$$

여기서 x_{ij}는 무기 i가 표적 j에 할당되면 1, 아니면 0인 이진 변수이다. SWTA는 표적의 수가 증가함에 따라 해 공간이 지수적으로 폭발하는 NP-complete 문제임이 증명되었으며, 이를 해결하기 위해 유전 알고리즘(genetic algorithm, GA), 입자 군집 최적화(particle swarm optimization, PSO) 등의 메타휴리스틱 기법들이 주로 연구되어 왔다[19].

반면, 동적 WTA(dynamic WTA, DWTA)는 시간이 지남에 따라 전장상황이 변화하는 환경을 다룬다. DWTA는 shoot-look-shoot 전술과 같이 시간을 여러 단계로 나누어 각 단계별로 의사결정을 수행하거나, 연속적인 시간 흐름 속에서 무기의 재장전, 표적의 이동, 새로운 표적의 출현 등을 고려해야 한다. 최근의 연구 동향은 DWTA를 단순한 SWTA의 반복이 아니라, 시간윈도우 제약이 포함된 복잡한 스케줄링 문제로 인식하고, 접근하는 추세이다. 이는 표적이 아군의 사거리 내에 머무르는 시간이 한정적이며, 무기의 종류에 따라 준비 및 타격 시간이 다르기 때문이다[20].

2.2 MARL 및 CTDE

복잡하고 불확실한 군집 시스템의 제어를 위해, 다중 에이전트 강화학습(multi-agent reinforcement learning, MARL)이 강력한 도구로 부상하고 있다. 단일 에이전트 RL과 달리 MARL 환경에서는 다른 에이전트들의 행동에 의해 환경의 상태 전이가 영향을 받으므로 각 에이전트 입장에서는 환경이 비정상적으로 변화하는 문제가 발생한다. 이를 해결하기 위해 CTDE 패러다임이 MARL 분야의 표준으로 자리 잡았다[21].

CTDE 방식에서 대표적인 알고리즘인 MAPPO는 학습단계에서는 중앙의 비평가(critic) 네트워크가 모든 에이전트의 전역 상태정보를 활용하여 가치함수를 학습하고, 실행단계에서는 각 에이전트의 연기자(actor) 네트워크가 자신의 지역 관측정보만을 이용하여 행동정책을 결정한다. 이는 정보공유가 자유로운 시뮬레이션 환경에서 학습을 수행하고, 실제 작전 시에는 통신 제약 하에서 분산 실행을 가능하게 하므로 USV 군집 운용에 매우 적합한 구조이다[15].

또한, 에이전트 간의 관계적 정보를 효과적으로 학습하기 위해 GNN을 MARL에 결합하는 연구가 활발하다. GNN은 에이전트를 노드(node), 통신 링크를 엣지(edge)로 모델링하여 가변적인 수의 에이전트와 동적인 토폴로지 변화에도 강건한 성능을 보장한다. 특히, 그래프 어텐션 네트워크(graph attention network, GAT)는 이웃노드로부터 수신된 메시지에 어텐션 가중치를 적용하여 중요한 정보를 선별적으로 집계할 수 있으므로 전장에서 다수의 비균질적 정보원을 효율적으로 처리하는 데 적합하다[22].

2.3 하이브리드 최적화 기법

강화학습은 학습된 정책을 통해 실시간으로 해를 도출할 수 있는 장점이 있지만, 복잡한 제약조건을 완벽하게 만족하는 최적해를 보장하기 어렵다는 단점이 있다. 반면, 유전 알고리즘, 타부 서치(Tabu search), 가변이웃탐색(VNS)과 같은 메타휴리스틱 기법은 최적해에 근접할 수 있으나 연산 시간이 오래 걸려 급변하는 전장상황에 실시간으로 대응하기 어렵다. 이러한 각 기법의 한계를 상호 보완하기 위해 하이브리드 최적화 기법이 대두되었다[23].

특히, VNS는 해가 지역 최적점(local optima)에 빠지는 것을 방지하기 위해 이웃 구조를 체계적으로 변경하며 탐색하는 기법이다. 최근 연구에서는 강화학습을 메타휴리스틱의 상위 제어자로 활용하는 적응형 연산자 선택 방식이 주목받고 있다. 이는 강화학습 에이전트가 현재 탐색상태를 분석하여 가장 효율

적인 이웃 생성 연산자나 휴리스틱을 동적으로 선택하게 함으로써 탐색의 효율성을 비약적으로 높이는 방식이다. 이러한 접근은 제한된 시간 내에 최적에 가까운 해를 찾아내는 애니타임(anytime) 성능을 향상시키는 데 기여한다[24,25].

2.4 선행연구 분석 및 차별화

기존의 WTA 관련 연구들은 대다수가 정적 모델에 집중하거나 동적 환경을 고려하더라도 모든 무기의 속도가 동일하다거나 통신이 완벽하다는 등 현실과 동떨어진 단순화된 가정을 전제로 수행되었다[26]. 또한, 순수 RL 기반 접근법들은 시뮬레이션 상의 학습 수렴성에만 초점을 맞추어 실제 무기체계가 갖는 엄격한 물리적 제약조건(재장전 시간, 사각지대, 자원 소모 등)을 충분히 반영하지 못하는 경향이 있었다. 대부분의 선행연구가 10척 미만의 소규모 교전 시나리오에 한정된 점도 실전 적용의 관점에서 중요한 제약이다.

반면, 운용과학 분야의 스케줄링 이론(RCPSP 등)은 제약조건 처리에 강점이 있으나 실시간성이 부족하다. 따라서, WTA 문제를 RCPSP로 매핑하여 실시간 스케줄링 관점에서 접근하면서 동시에 GNN-MAPPO 기반의 유연한 분산 협업과 VNS 기반의 정밀탐색을 결합한 통합 프레임워크에 대한 연구는 현재까지 미비한 실정이다. 본 연구는 이러한 학술적·실무적 공백을 메우기 위해 제약조건이 엄격하고, 통신이 불안정한 해상 전장환경에서 실제로 작동 가능한 하이브리드 의사결정 프레임워크를 제시하고자 한다.

2.5 무인수상정(USV) 군집 운용개념

USV 군집 운용은 단일 고가치 플랫폼에 의존하던 기존 해군 전력을 다수의 저비용 무인 체계로 분산시켜 임무의 유연성, 생존성, 그리고 작전 반경을 비약적으로 확장하는 개념이다. 본 연구에서 상정하는 USV는 미래형 차세대 중형 플랫폼을 자체적으로 구상하였고, 최대 40노트의 속도를 가정하였다[27-29].

각 USV는 교전거리에 따라 모듈식으로 탑재할 수 있는 다종의 무장을 가정한다[30,31]. 근접 방어용 RCWS, 중거리 대응용 유도로켓, 원거리 타격용 드론 등으로 구성된 다층 무장체계는 교전거리와 표적 특성에 따른 최적 무장 선택을 가능케 하지만, 동시에 무장할당 문제의 복잡도를 증가시키는 요인이 된다. 편대는 4척을 기본단위로 하는 구성을 가정하며, 개체 간 이격거리는 최소 500야드(457 m)를 유지한다.

해상환경은 육상이나 공중과 달리 파도, 조류, 해무 등 환경적 외란이 심하고, 장애물 회피가 까다롭다. 또한, 통신환경이 매우 열악하여 대역폭 제한과 통신단절이 빈번하다. 이러한 환경에서 중앙 집중형 제어 방식은 모든 에이전트의 상태정보를 실시간으로 수집하고, 처리하는 데 한계가 있으며, 통신링크가 끊어질 경우 전체 군집이 기능을 상실할 위험이 있다. 따라서, 최근의 연구들은 각 USV가 자신의 로컬 센서로 수집한 정보와 인접한 소수의 이웃 에이전트와 공유한 정보만을 바탕으로 스스로 판단하고 행동하는 분산형 알고리즘에 집중하고 있다[32,33].

3. 문제 정의 및 수학적 모델링

3.1 대함전 시나리오

본 연구에서 상정하는 대함전 시나리오는 한반도 해양환경에서 군집 USV 편대가 전방에 방어 대형으로 전개된 상황이다. 적군은 다수의 고속 공기부양정, 유도탄정, 초계정, 상륙함 등 7종의 이종 함정으로 구성된 군집 세력으로 아군의 방어선인 NLL을 돌파하여 대한민국 영토에 대한 기습 상륙 또는 타격을 목표로 한다.

우군 방어자산(blue force)은 80척의 150톤급 USV가 적 표적 규모를 고려하여 랜덤하게 배치된다. 각 USV는 교전거리에 따라 선택적으로 운용 가능한 다종의 무장을 가정하며, 세부 제원은 Table 1과 같다.

Table 1. Blue force USV weapon specifications[34]

Weapon System	Effective Range	Maximum Range	P_{kill}	Ammunition Capacity
RCWS	2,000 m	7,000 m	0.37	2,000 rounds
Guided Rocket	5,000 m	8,000 m	0.70	4 rounds
Loitering Munition	30,000 m	40,000 m	0.25	12 units

Table 1에서 나타나는 바와 같이 USV에 설정된 무장은 각각 근접(2~7 km), 중거리(5~8 km), 원거리

(30~40 km)의 교전영역을 담당하며, 격파확률(p_{kill})은 유효사거리 내에서 최대값을 보이고, 유효사거리를 초과하면 거리에 비례하여 선형 감쇠한다. RCWS는 높은 탄약 적재량(2,000발)으로 지속적인 사격이 가능하나 단발 위력이 낮아 다수의 명중탄이 필요하며, 유도로켓은 높은 격파확률(0.7)로 중거리에서 확실한 격파가 가능하지만 4발의 제한된 탄약수를 가진다. 드론은 최대 40 km의 원거리에서 선제 타격이 가능하나, 격파확률(0.25)이 상대적으로 낮아 군집 살보(salvo) 운용이 필수적일 것으로 판단된다.

적군 위협세력(red force)은 7종 80척의 이종 적함이 역삼각형(쐐기) 진형으로 무작위 출현한다. 적 함정의 구성은 공개 출처 군사정보(IISS Military Balance, Jane's Fighting Ships)를 기반으로 자체 설정하였으며, 각 함종의 세부 제원은 Table 2에 정리하였다.

Table 2. Red force Naval vessel specifications

Ship Type	Max Speed	HP	Main Weapon	Threat Level
Landing Craft Air Cushion (LCAC)	45kts	3	14.5 mm Machine Gun	HIGH
Guided Missile Patrol Boat (PTG, Haesam-class)	40kts	6	STYX P-15 (2 rounds) + Torpedo	CRITICAL
Patrol Torpedo Boat (PTB, Chaho-class)	30kts	3	85 mm Naval Gun	MEDIUM
Large Patrol Ship (PTAG, Nongo-class)	25kts	6	Kumsong-3 (4 rounds)	CRITICAL
Small Patrol Boat (PB)	35kts	3	Autocannon	LOW
Landing Ship Tank (LST, Hantae-class)	15kts	8	14.5 mm Machine Gun	HIGH
Fire Support Boat (PCFG, Chongjin-class)	40kts	6	85 mm ZIS-S-53 Tank Gun	HIGH

Table 2에 나타낸 바와 같이 적 함정은 고속·저내구도 소형정(공기부양정, 소형 초계정) 부터 중속·고내구도 대형함(상륙함, 대형함)까지 다양한 스펙트럼의 위협을 구성한다. 특히, 유도탄정과 대형함은 각각 STYX P-15 대함미사일과 금성-3호 함대함 순항미사일을 탑재하여 critical 등급의 위협으로 분류되며, 시뮬레이션에서 가장 높은 우선순위로 할당된다. 적 세력은 사전 정의된 규칙 기반 컴퓨터 생성 부대(computer generated forces, CGF)로 모델링되어 기본 남하, 회피 기동, 우회 기동, 탄약 고갈 시 충돌 돌진 등의 전술적 행동 패턴을 수행한다.

3.2 MRCPSPTW 기반 문제 정의

본 연구의 핵심 이론 중 하나는 DWTA 문제를 시간윈도우가 있는 다모드 자원제약 프로젝트 스케줄링 문제(MRCPSPTW)로 변환하여 모델링하는 것이다. 기존의 WTA가 단순히 '누가 누구를 쏠 것인가'에 집중했다면, 본 연구의 접근법은 '누가, 언제, 어떤 자원을 사용하여 표적을 처리할 것인가'를 결정하는 시공간 통합 최적화 문제로 정의한다[10,11].

전장의 요소를 RCPSP의 요소로 매핑하면 다음과 같다. 적 표적 j를 요격하는 행위 그 자체를 하나의 프로젝트 작업(task/activity)으로 본다. USV i에 탑재된 무기 k는 자원(resource)으로 간주하며, 이때 드론과 유도로켓은 소모성 자원이고, RCWS는 탄약이 남아 있는 한 재사용 가능한 재생 가능 자원으로 분류된다. 동일한 표적을 요격하기 위해 선택할 수 있는 다양한 무장 조합이나 전술(예: 단발 사격, 살보 사격)을 작업의 수행모드(mode)로 정의하며, 각 모드는 소요시간과 자원 요구량이 다르다. 표적 j가 USV i의 무기 유효사거리 내에 진입하는 시점(ES_j)부터 이탈하거나 아군에 도달하는 시점(LF_j)까지의 구간을 작업 수행이 가능한 시간윈도우(time window)로 설정한다.

제안한 MRCPSPTW 모델의 물리적 매핑 구조는 Fig. 2와 같다. 각 표적의 사거리 진입시점(ES_j)과 이탈시점(LF_j)은 고유의 시간윈도우를 형성하며, 할당된 특정 무기체계의 교전 수행 시간은 반드시 해당 윈

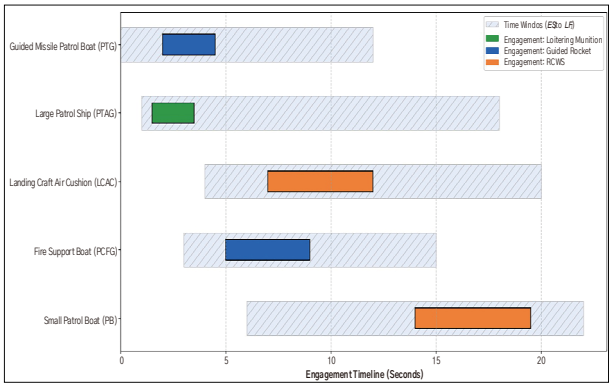


Fig. 2. MRCPSPTW mapping: Time windows and weapon scheduling

도우 내에서 완료되도록 스케줄링되어야 함을 직관적으로 보여준다. 이는 본 연구가 제안하는 다중모드 무장할당이 단순한 정적 표적 매칭을 넘어 엄격한 시공간적 물리제약이 결합된 고도화된 동적 자원 할당 문제임을 입증한다.

3.3 목적함수 및 제약조건

본 연구에서는 상충하는 두 가지 목표를 통합적으로 고려하는 목적함수를 설계하였다. 가장 중요한 성능지표인 방어성공률(defense success rate, DSR)은 적함 파괴율과 아군 USV 생존율을 가중 합산하여 산출하며, 그 정의는 식 (2)와 같다.

$$DSR = w_d \cdot D_{rate} + w_s \cdot S_{rate} \tag{2}$$

여기서 'D_{rate}=적함 격파 수/총 적함 수'는 적함 파괴율, 'S_{rate}=USV 생존 수/총 USV 수'는 아군 USV 생존율이며, 가중치 w_d=0.7, w_s=0.3이다. 소모성 무인전력 특성을 반영하여 임무 달성(적 격파)에 70% 비중을 부여하고, 전력 보존(아군 생존)에 30% 비중을 부여하였다. 이 가중치 비율은 해군 전문가 자문 및 델파이 기법을 통해 결정되었다. 주요 제약조건은 식 (3)과 같이 수학적으로 정식화된다.

$$\Sigma_{j,m,t} r_{ijk}^{m} \cdot x_{ijk}^{m}(t) \le A_{ik}, \ \forall i,k \tag{3}$$

먼저, 식 (3)은 각 USV가 물리적으로 적재된 탄약 수량을 초과하여 무장을 할당할 수 없음을 의미한다. 여기서 결정변수 $x_{ijk}^{m}(t) \in 0,1$은 USV i가 시간 t에 표적 j를 요격하기 위해 무기 k를 교전모드 m(예: SLS, SSL 등)으로 할당할 경우 1, 그렇지 않으면 0을 갖는 이진변수이다. r_{ijk}^{m}은 해당 교전모드 m을 수행할 때, 1회 사격시 소모되는 무기 k의 자원 요구량(탄약 소모량)을 의미하며, A_{ik}는 USV i에 초기 적재된 무기 k의 총 가용 탄약 용량(Table 1 참조)을 나타낸다. 즉, 교전시간 전체 t에 걸쳐 모든 표적 j와 교전모드 m에 대해 소모된 특정무기의 총합은 초기 적재량을 절대 초과할 수 없음을 엄격하게 통제한다.

둘째, 사거리 제약은 교전이 특정무기의 물리적 타격 가능거리 범위 내에서만 이루어져야 함을 보장하며, $d_{ij}(t) \le R_k^{\max}$로 정의된다.

셋째, 시간윈도우 제약은 실제 교전시점이 표적의 사거리 내 진입시점(ES_j)과 이탈하거나 아군 진영에 도달하는 한계시점(LF_j) 사이에 포함되어야 함을 강제하며, $ES_j \le t_{ij} \le LF_j$로 표현된다.

넷째, 명중률 감쇠제약은 타격거리 $d_{ij}(t)$가 유효사거리 R_k^{eff}를 초과하여 최대사거리 $R_k^{\max}$에 이를 때까지 무기 k의 격파확률 $p_{ijk}(t)$가 거리에 비례하여 선형적으로 감쇠하는 물리적 현상을 묘사한다. 이는 시뮬레이터의 군사적 충실도(fidelity)를 극대화하기 위해 식 (4)와 같은 조건부 수식으로 정식화된다. 여기서 $P_k^{\max}$는 무기 k의 유효사거리 내 최대 격파확률이다.

$$P_{ijk}(t) = \begin{cases} P_k^{\max}, & \text{if } d_{ij}(t) \le R_k^{eff} \\ P_k^{\max} \times \left(\dfrac{R_k^{\max} - d_{ij}(t)}{R_k^{\max} - R_k^{eff}}\right), & \text{if } R_k^{eff} < d_{ij}(t) \le R_k^{\max} \\ 0, & \text{if } d_{ij}(t) > R_k^{\max} \end{cases} \tag{4}$$

3.4 Threat Index 설계

표적의 위협도를 정량적으로 산출하기 위해 본 연구에서는 5개의 독립적인 요소를 가중 합산하는 위협지수(TI) 모델을 설계하였다. 기존의 TEWA(threat evaluation and weapon assignment) 연구에서는 주로 거리와 속도만을 고려한 2~3요소 모델이 사용되었으나, 본 연구에서는 해상 포술의 물리적 특성을 반영한 교차각(crossing angle) 요소를 추가로 도입하여, 총 5요소로 구성된 위협지수 모델을 제안한다. 각 요소의 가중치는 해군 전문가 7인의 델파이 기법을 통해 결정되었으며, 그 수식과 가중치는 식 (5)와 같다.

$$\Theta_j(t) = w_1 \hat{f}_1(d_{ij}) + w_2 \hat{f}_2(v_j) + w_3 \hat{f}_3(V_j) + w_4 \hat{f}_4(l_j) + w_5 \hat{f}_5(\theta_{cross}) \tag{5}$$

식 (5)에서 각 요소의 의미와 가중치를 정리하면 Table 3와 같으며, 5개 가중치의 합은 1을 만족하여 정규화 조건을 충족한다.

특히, 제안된 5요소 위협지수 모델의 핵심인 교차각 함수 $\hat{f}_5(\theta_{cross})$는 Fig. 3에서 보는 바와 같이 적함의

Table 3. 5-factor threat index

Factor	Weight	Rationale
w_1 (Distance)	0.40	Prioritize close-range threats
w_2 (Velocity)	0.20	Prioritize high-speed targets
w_3 (Target Value)	0.15	Prioritize missile boats and large ships
w_4 (Durability)	0.05	Support confirmed kill (Finishing off)
w_5 (Crossing Angle)	0.20	Maximize threat of frontal approach

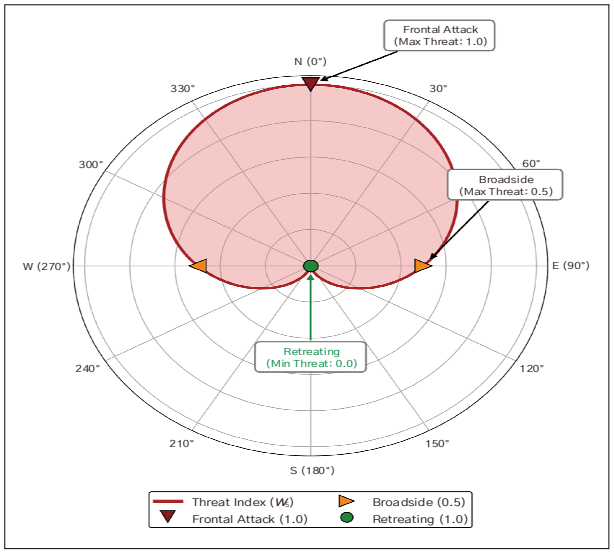


Fig. 3. Threat index curve by target aspect (crossing angle)

함수 지향 방향과 아군 USV 간의 기하학적 각도를 수치화한다. 적함이 정면으로 돌진할 때(0°) 포문의 조준이 용이해져 위협이 극대화되고, 측방이나 후방으로 이탈 시(90°~180°) 위협이 비선형적으로 감소하는 해상 전술의 물리적 역학을 반영하기 위해, 본 연구에서는 코사인(cosine) 비례 기반의 위협 감쇠 함수를 식 (6)과 같이 새롭게 정의하였다.

$$\hat{f}_5(\theta_{cross}) = \frac{1+\cos(\theta_{cross})}{2}, \quad where\, \theta_{cross} \in [0, \pi] \tag{6}$$

이 정식화를 통해 $\theta_{cross} \approx 0^\circ$ 일 때 함수값은 1.0(최대 위협)으로 산출되며, $\theta_{cross} \approx 180^\circ$ (완전 이탈)일 때는 0.0으로 수렴한다. 이러한 비선형적 공간 함수의 도입은 거리 변수 단독으로는 식별하기 어려운 적 표적의 전술적 기동 의도를 실시간 위협도 평가 모델에 정밀하게 편입시켰다는 점에서 기존 정적모델 대비 시뮬레이터의 군사적 충실도를 획기적으로 향상시킨다.

나아가, 제안된 5요소 모델에서 가장 큰 가중치를 차지하는 두 핵심 공간변수(w_1, w_5)의 복합적인 상호작용에 따른 위협지수 변화량을 평가하기 위해 2차원 히트맵 분석을 수행하였으며, 그 결과는 Fig. 4와 같다. 본 분석은 나머지 3개 변수(속도, 표적가치, 내구도)를 동일한 조건으로 정규화하여 거리와 교차각의 순수 결합 효과를 고립시켰다. 그래프에서 보는 바와 같이 거리가 2 km로 근접하고, 적함이 아군을 향해 정면(0°)으로 돌진할 때 위협지수는 0.96으로 최대치에 수렴하며, 거리가 멀어지고, 교차각이 커질수록 위협지수는 비선형적으로 감쇠한다. 이는 본 연구의 TI 모델이 아군을 향해 근접 침투하는 고가치 위협을 실시간으로 포착하고, 우선순위를 즉각적으로 재조정하는 수학적 강건성을 갖추고 있음을 입증한다.

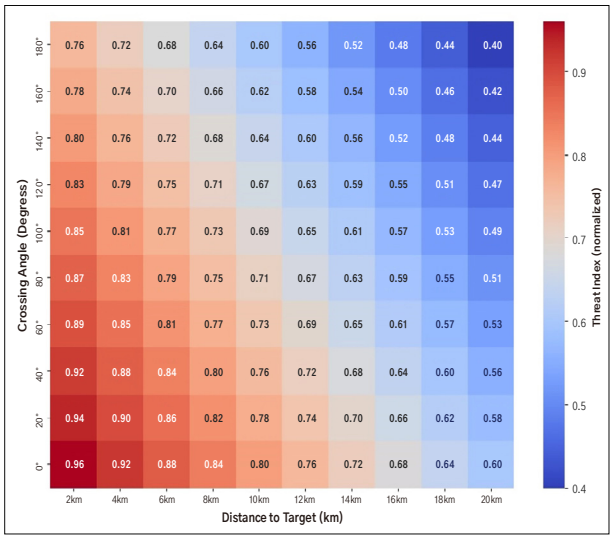


Fig. 4. The interactive impact of distance and crossing angle on the normalized threat index (heatmap)

4. 하이브리드 무장할당 알고리즘 설계

4.1 Phase 1: TI 기반 휴리스틱 초기해 생성

알고리즘의 첫 번째 단계는 0.5초 이내의 극히 짧은 시간 내에 실행 가능한 초기 해를 생성하는 것이다. 이 단계에서는 제3.4절에서 설계한 5요소 위협지수를 기반으로 모든 적 표적의 우선순위를 정렬하고,

리스트 스케줄링(list scheduling) 알고리즘으로 초기 할당을 수행한다. 각 USV는 자신의 사거리 내에 있는 표적 중 위협지수가 가장 높은 표적에 대해 교전거리에 따른 최적 무장을 자동으로 선택한다. 구체적으로 교전거리가 8 km를 초과하면 드론, 5~8 km이면 유도로켓 또는 드론, 2~5 km이면 유도로켓(유효사거리 이내)을 우선하되 유도로켓 탄약이 소진된 경우 RCWS(최대사거리 7 km 이내)를 대체 운용하며, 2 km 미만이면 RCWS를 우선적으로 선택한다.

이 단계에서 핵심적인 기능은 action masking의 적용이다. 이는 물리적으로 불가능한 행동(탄약이 소진된 무장의 발사, 최대사거리를 초과하는 사격, NLL 이북으로의 선제 침범 등)을 의사결정 과정에서 원천적으로 차단하는 제약조건 필터링 메커니즘이다. 이를 통해 강화학습 에이전트가 물리법칙과 교전규칙(rules of engagement, ROE)을 위반하는 행동을 학습하는 것을 방지하며, 탐색공간을 유효한 영역으로 한정하여 학습 효율을 향상시킨다. 이 단계에서 생성된 해는 전역 최적해는 아니지만, 모든 제약조건을 만족하는 유효한 해이며, 이후 단계의 탐색을 위한 시드(seed) 역할을 수행한다. 시간 복잡도는 $O(NM\log M)$이다.

4.2 Phase 2: VNS를 이용한 초기해 개선

Phase 1에서 생성된 초기해를 바탕으로 해의 품질을 체계적으로 개선하는 단계이다. 일반적인 VNS는 사전에 정의된 순서대로 이웃구조를 변경하며 탐색하지만, 이는 문제상황에 따라 비효율적일 수 있다. 본 연구에서는 강화학습 에이전트가 메타-제어자 역할을 수행하여 현재 해의 상태를 분석하고, 가장 효과적인 이웃 연산자를 적응적으로 선택하는 RL-guided VNS 방식을 채택한다[13].

본 연구에서 정의한 4개의 이웃 연산자는 다음과 같다. 첫째, swap 연산자(N_1)는 할당된 두 표적의 USV 할당을 교환하여 자원 분배의 균형을 개선한다. 둘째, reassign 연산자(N_2)는 특정 표적을 현재 할당된 USV에서 해제하고, 다른 가용 USV에게 재할당하여 전체적인 사격 효율을 높인다. 셋째, weapon-switch 연산자(N_3)는 동일한 USV-표적 쌍에 대해 무장 종류를 변경하여 교전거리 변화에 적응한다. 넷째, insert 연산자(N_4)는 아직 할당되지 않은 표적을 기존 스케줄의 유휴시간에 삽입하여 할당 커버리지를 확대한다.

각 반복에서 RL 에이전트는 현재 해의 목적함수값, 잔여 계산시간, 최근 반복의 개선율, 제약조건 위반 정도를 상태로 관측하고, 4개의 이웃 연산자 중 하나를 선택하는 행동을 취한다. 새로운 해가 기존 해보다 우수할 경우 그 향상분을 보상으로 사용하여 에이전트의 연산자 선택 정책을 개선한다.

4.3 Phase 3: GNN-MAPPO 의사결정(1초)

Phase 2까지의 최적화가 개별 USV 또는 소규모 그룹의 구체적인 무장 스케줄링을 담당한다면, Phase 3는 거시적인 관점에서 군집 전체의 전략적 협업과 상황인식을 담당한다. 즉, Phase 2의 VNS가 개별 USV 차원에서 실행 가능한 로컬 최적 무장 스케줄을 생성하면, Phase 3의 GNN-MAPPO는 이를 초기 행동 편향으로 활용하되 GNN을 통해 군집 전체의 전술적 의도를 교환하여 중복할당(overkill)과 사각지대를 방지하는 전역적 협력을 최종 확정한다. 이 단계에서는 3층 multi-head graph attention network (4-head, $\mathrm{embed}_{\mathrm{dim}}$=128)와 MAPPO를 결합한 CTDE 구조를 적용하여 각 USV가 이웃 에이전트의 정보를 그래프 메시지 패싱을 통해 집계하고, 이를 기반으로 최종적인 협력적 행동을 결정한다[15,16].

Phase 2에서 산출된 이산적 VNS 할당 결과와 Phase 3의 연속적 신경망 정책 분포를 텐서 수준에서 물리적으로 결합하기 위해 본 연구에서는 로짓 쉐이핑 기반의 편향 주입기법을 설계하였다[35]. Actor 신경망의 최종 선형 레이어에서 출력된 표적 j에 대한 원시 로짓을 z_j라 하고, Phase 2에서 할당된 VNS 최적행동 인덱스를 a_{vns}라 할 때, 편향이 적용된 조정 로짓 z_j는 식 (7)과 같이 정의된다.

$$\tilde{z}_j = z_j + \lambda \cdot I(j = a_{vns}), \quad \forall j \in A \tag{7}$$

여기서 I는 지시함수로서 행동 인덱스 j가 VNS의 추천 해 a_{vns}와 일치할 때에만 1을 반환하고, 나머지 경우에는 0을 반환한다. λ는 VNS 편향의 강도를 제어하는 편향강도 계수이다. 조정된 로짓 $\tilde{z}_j$는 소프트맥스 함수를 통과하여 에이전트 i의 행동확률 분포로 변환된다.

$$\pi_\theta(a_j|o_i) = \frac{\exp(\tilde{z}_j)}{\sum_{a\in A}\exp(\tilde{z}_a)} \quad (8)$$

이 수학적 결합 메커니즘의 핵심 속성은 다음과 같다. VNS가 추천한 행동 a_{vns}의 로짓에만 스칼라 λ가 가산되므로 해당 행동의 소프트맥스 출력 확률이 선택적으로 증가한다. 예를 들어, 행동공간의 크기가 81(80개 표적+사격보류 1)이고, 모든 원시 로짓이 동일한 상태에서 λ=5.0을 적용하면 VNS 추천 행동의 선택 확률은 약 65%로 상승하되 나머지 행동의 확률은 각 0.44%로 유지된다. 따라서, 신경망은 VNS가 보장하는 실행 가능한 로컬 최적해를 강력한 사전 지식으로 활용하면서도 무작위 탐색으로 인한 치명적 오판을 방지한다. 동시에 GNN 메시지 패싱을 통해 이웃의 전술적 맥락을 반영한 로짓 z_j가 충분히 큰 경우에는 VNS의 제안을 자율적으로 기각하고, 더 나은 전역적 협력 행동을 창발할 수 있는 소프트 제약 특성을 보장한다.

λ의 운용방식은 학습단계와 추론단계에서 구분된다. 학습단계에서는 학습 에피소드 진행에 따라 λ를 선형감쇠시켜 학습 초기에는 VNS의 도메인 지식을 강하게 따르도록 유도하고, 학습 후기에는 GNN-MAPPO가 독자적인 협력 정책을 발견하도록 자율성을 부여한다. 감쇠 스케줄은 식 (9)와 같다.

$$\lambda_t = \lambda_0 \cdot \max\left(0, 1 - \frac{t}{T_{anneal}}\right) \quad (9)$$

여기서 λ_o은 초기 편향강도(본 연구에서는 5.0), t는 현재 에피소드 번호, T_{anneal}은 감쇠가 완료되는 에피소드(본 연구에서는 총 학습 에피소드의 60%인 6,000으로 설정)이다. 학습 에피소드 6,000 이후에는 λ_t=0이 되어 순수 GNN-MAPPO 정책으로 작동하며, 이후 4,000 에피소드 동안 VNS 편향 없이 독자적인 정책을 미세 조정한다. 실전 추론단계에서는 3계층 파이프라인이 순차적으로 가동되므로 λ를 고정값 $\lambda_{\inf}$=3.0으로 설정하여 Phase 2의 VNS 해를 적극적으로 참조하면서도 Phase 3의 전역 협력 조율능력을 최대한 활용한다.

편향 강도 계수의 초기값을 λ_o=5.0으로 설정한 것은 행동 공간의 크기(|A|=81)를 고려한 행동확률 분포의 엔트로피 제어전략에 기인한다. 원시 로짓 z_j가 모두 균등한 값을 가지는 학습 초기상태를 가정할 때, 특정 VNS 추천 행동의 로짓에 λ=5.0이 가산되면 소프트맥스 정규화를 거친 후 해당 행동의 선택 확률은 $\frac{e^5}{e^5+80\times e^0} \approx 0.65$ (약 65%) 수준으로 통제된다.

이는 에이전트가 탐험의 35%를 무작위 선택 또는 GNN을 통해 수신한 이웃정보에 할당하여 새로운 전역적 협력패턴을 창발할 수 있는 가능성을 유지하면서도 의사결정의 65%는 VNS가 수학적으로 검증한 안전한 로컬 최적해에 의존하도록 강제하는 수치적 균형점이다. 즉, 에이전트의 치명적 전술 오판을 방지하는 강제적 지식 주입과 강화학습 고유의 자율적 탐색 사이의 이상적인 텐서 조작 경계선을 λ 스케줄링을 통해 성공적으로 확립한 것이다.

이 logit shaping 기법은 MAPPO의 정책 경사 업데이트에 간섭하지 않는다는 점에서 구조적으로 안정적이다. 로짓에 대한 가산 바이어스는 미분 가능하므로 역전파 경로를 차단하지 않으며, λ를 학습 가능 매개변수가 아닌 외생적 스케줄로 관리함으로써 PPO의 클리핑 기반 정책 업데이트 메커니즘과의 충돌을 원천적으로 방지한다.

본 프레임워크는 반복마다 현재까지 발견된 최적해를 별도의 고속 메모리 버퍼에 저장하는 anytime algorithm으로 구현된다. 알고리즘은 내부적으로 타이머를 체크하며, 사전에 설정된 시간예산이 소진되거나 외부 지휘통제 시스템으로부터 "실행" 인터럽트 신호가 수신되면, 즉시 루프를 중단하고 버퍼에 저장된 최적해를 반환한다[12]. 이러한 구조는 긴박한 상황에서는 Phase 1의 초기 휴리스틱 해만으로도 즉각적인 대응이 가능하고, 여유가 있을 때는 Phase 3까지 충분히 탐색하여 고품질의 해를 도출하는 유연성을 제공한다.

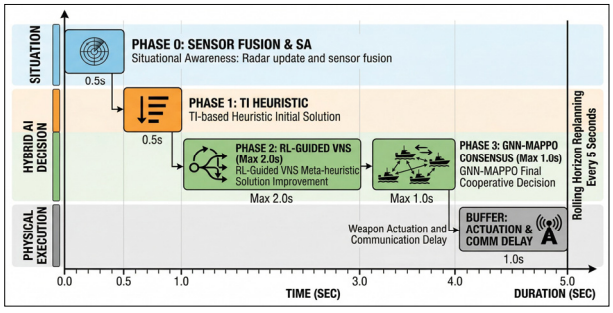


Fig. 5. 5-second rolling horizon 3-layer decision pipeline

4.4 실시간성 보장 설계

전체 의사결정 파이프라인의 시간 할당은 Fig. 5에서 보는 바와 같이 5초 rolling horizon으로 설계되었다. 현대 해전에서 적 함정의 접근속도(최대 45노트≈23 m/s)와 무장의 유효사거리를 고려하면, 교전상황에서 지휘관에게 허용되는 의사결정 시간은 수초에 불과하다. 따라서, 수학적으로 완벽한 전역 최적해를 구하기 위해 수 분 이상을 소모하는 알고리즘은 전술적으로 무의미하다. 오히려, 약간의 성능 저하가 있더라도 즉시 실행 가능한 해를 도출하는 것이 훨씬 중요하다. 각 단계의 시간 할당을 정리하면 Table 4와 같다.

Table 4. Real-time assurance architecture

Phase	Duration (Sec)	Function Description
Phase 0 (SA)	0.5	Situational Awareness: Radar update and sensor fusion
Phase 1 (TI)	0.5	TI-based Heuristic Initial Solution
Phase 2 (VNS)	2.0	VNS Meta-heuristic Solution Improvement
Phase 3 (MAPPO)	1.0	GNN-MAPPO Final Cooperative Decision
Buffer	1.0	Weapon Actuation and Communication Delay
Total	5.0	Rolling Horizon Replanning Every 5 Seconds

Table 4에 나타낸 바와 같이 전체 5.0초의 시간예산 중 순수 알고리즘 연산에 할당된 시간은 3.5초(Phase 1~3)이며, 나머지 1.5초는 센서 퓨전과 물리적 구동 시간을 위한 버퍼이다. 이러한 시간예산 설계는 실제 USV의 포탑 선회속도(약 60도/초), 통신지연(약 200 ms), 레이더 갱신 주기(약 800 ms) 등의 물리적 제약을 고려하여 결정되었다. 특히, anytime algorithm의 속성으로 인해 만약 Phase 2 도중에 긴급한 위협이 감지되면 해당 시점까지의 최적해를 즉시 반환하여 긴급 대응이 가능하다.

5. 다중 에이전트 강화학습 프레임워크 설계

5.1 Dec-POMDP 환경 모델링

앞선 4장에서 제시한 최적화 파이프라인(Phase 1~2)이 개별 USV 노드 수준의 물리적 제약 기반 로컬 스케줄링을 담당했다면 본 장에서는 군집 전체의 거시적 협력과 통신제약 극복을 위한 MARL 프레임워크(Phase 3)를 정식화한다[16]. Fig. 6는 제안하는 3계층 파이프라인과 CTDE 아키텍처의 완벽한 대칭적 정보 융합 흐름을 도식화한 것이다.

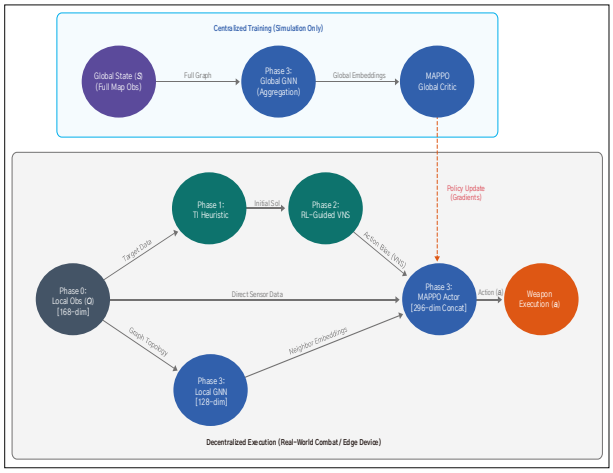


Fig. 6. Architecture schematic of the 3-layer hybrid framework and CTDE strategy

그림의 하단 분산형 실행(DE) 단계를 보면 에이전트의 로컬 센서 관측치(o_i)는 세 갈래로 병렬 처리된다. 상단의 최적화 파이프라인을 거쳐 도출된 실행 가능한 행동편향(VNS schedule)과 하단의 로컬 GNN을 통해 추출된 이웃노드와의 공간적 위상관계(128-dim)가 생성된다. 최종적으로 MAPPO의 로컬 연기자(actor)는 이 정보들과 자신의 직접 관측벡터(168-dim)를 하나로 융합(296-dim concatenation)하여 최적의 무장할당(a_i)을 결심한다. 가장 주목할 점은 상단의 중앙집중식 훈련(CT) 파이프라인이다. 시뮬레이터가 제공하는 전역 상태(S) 역시 global GNN을 거쳐 전역 그래프 임베딩으로 집계되며, 이를 바탕으로 비평가(critic) 네트워크가 연기자의 정책 업데이트를 정밀하게 가이드한다. 이러한 구조적 대칭 설계 덕분에 실전에서 통신이 두절되더라도 엣지 디바이스의 actor는 로컬 GNN 임베딩을 마스킹한 채 VNS 스케줄에 전적으로 의존하여 치명적 오판 없이 자율교전을 완수할 수 있다.

본 문제는 분산 부분 관측 마르코프 결정 과정(decentralized partially observable markov decision process, Dec-POMDP)으로 모델링된다. Dec-POMDP

에서는 각 에이전트가 전체 전장상태의 일부만을 관측할 수 있으며, 이러한 제한된 정보만으로 의사결정을 수행해야 한다[21].

전역상태(S)는 80척의 아군 USV와 80척의 적함 표적의 위치, 속도 벡터, 잔여 HP, 잔여 탄약량 등 전장 내 모든 객체의 완전한 정보를 포함하며, 이는 학습 시 중앙 비평가(critic)에게만 제공된다. 각 에이전트 i의 관측(O_i)은 자신의 레이더 탐지범위(20 km) 내에 있는 표적정보와 GNN 메시지 패싱을 통해 수신된 이웃 에이전트의 임베딩으로 구성되며, 총 168차원의 관측 벡터를 형성한다. 이 168차원은 자기 위치(2), 가장 가까운 적 방향(2), 잔여 탄약 정규화(3), 자기 HP(1), 각 적함까지 거리 정규화(80), 각 적함 위협지수(80)로 구성된다.

특히, 교전 진행에 따라 적함이 격침되거나 레이더 탐지 범위를 벗어나 유효 표적의 수가 동적으로 변화하는 전장환경에 강건하게 대응하기 위해 가변 관측 공간 인코딩 기법을 적용하였다. 다층 퍼셉트론(MLP) 기반 신경망의 입력 차원 불일치 오류를 원천 차단하고자 로컬 관측 벡터는 예상 최대 위협 수(80척)를 기준으로 168차원을 고정 할당하되, 교전으로 인해 비활성화된 표적의 인덱스에는 영벡터 패딩(zero-padding) 처리를 수행하였다.

나아가 단순한 0값 입력이 물리적 거리 0 m(근접 충돌)로 오인되는 인지적 환각현상을 방지하기 위해, GAT 레이어의 어텐션 마스킹(attention masking) 및 정책 출력층의 유효 행동 마스킹(action masking)을 병행 적용하였다. 이를 통해 인공신경망이 무력화된 유령 표적에 화력을 낭비하지 않고 오직 생존해 있는 유효 위협에만 가중치를 부여하도록 제어함으로써 동적환경에서의 차원 정합성과 시스템의 군사적 현실성을 극대화하였다.

행동공간(A_i)은 81개의 이산 행동으로 구성되며, 80개의 표적 중 하나를 선택하여 최적 무장을 자동 할당하는 행동과 1개의 대기 행동으로 이루어진다. action masking을 통해 사거리 밖 표적, 탄약 부족 무장, NLL 이북 침범 등 물리적으로 불가능한 행동을 원천 차단한다.

보상함수(R)는 군집전력 전체의 목표와 개인의 효율성을 동시에 고려하여 설계하였다. 적함 명중 시 +2.0, 격침 시 +10.0의 양의 보상을 부여하고, 명중 실패 시 -0.2, 무효한 행동(이미 격침된 표적 공격 등) 시 -0.1의 음의 보상을 부여한다. 생존 중인 USV에게는 매 스텝 +0.1의 생존 보상을 제공하고, 모든 적함 격파 시 전체 에이전트에게 +5.0의 팀 보상을 추가한다.

5.2 GNN 그래프 표현 학습

USV 군집은 고정된 격자(grid) 형태가 아니라 유동적으로 움직이는 그래프 구조로 표현하는 것이 가장 자연스럽다. 본 연구에서는 USV 군집을 그래프 G=(V,E)로 모델링하여, USV 80척의 노드와 적함 80척의 노드를 결합한 총 160개 노드의 전투 그래프를 구성한다. 각 노드에는 위치, 잔여 무장량, 속도 벡터, HP 등의 정보가 인코딩되며, 통신 가능한 에이전트 간의 링크가 엣지로 정의된다.

또한, 그래프 어텐션 네트워크(GAT)를 적용하여 이웃노드로부터 받은 정보에 어텐션 가중치를 부여하고, 중요한 정보(예: 근처에서 교전 중인 동료의 정보)를 선별적으로 집계하도록 3층의 multi-head GAT(4-head, $embed_{dim}$=128)를 적용한다. 먼저, 4-head 어텐션 구조는 단일 관점의 정보집계에서 발생할 수 있는 편향을 방지하기 위해 도입되었다. 군집전 상황에서 각 head는 적함의 물리적 위협도, 인접 아군의 잔여 무장상태, 진형의 공간적 배치, 표적의 최우선 가치 등 서로 다른 전술적 특징공간에 독립적으로 집중하여 어텐션 가중치를 학습한다. 또한, 임베딩 차원 d_{dim}=128은 168차원의 방대한 원시 관측 벡터를 압축하여 개별 에이전트의 전술적 맥락을 표현하는 잠재 벡터의 크기를 의미한다. 임베딩 차원을 128로 설정한 것은 160척 규모의 군집노드 간 동적 상호작용을 손실없이 모델링하기 위한 최적의 표현 용량을 확보함과 동시에 탑재용 소형 컴퓨터(Jetson Xavier NX)에서의 실시간 추론 속도를 달성하기 위해 연산 병목을 제거한 아키텍처 설계의 결과이다. 이는 대규모(80 대 80) 교전의 복잡한 상호작용 패턴을 충분히 수용할 수 있는 표현력을 유지하면서도 앞서 제시한 엣지 디바이스 환경에서의 11.9 ms 실시간 추론 제약을 충족하기 위해 연산 복잡도를 제한한 최적의 구조적 타협점이다. 이에 해당하는 메시지 패싱 연산 수식은 식 (10)과 같다.

$$h_i^{(l+1)} = \sigma\left(\sum_{j \in N(i)} \alpha_{ij}^{(l)} W^{(l)} h_j^{(l)}\right) \tag{10}$$

여기서 $h_i^{(l)}$은 l번째 레이어에서 노드 i의 임베딩, $\alpha_{ij}^{(l)}$은 어텐션 가중치, $W^{(l)}$은 학습 가능한 가중치 행렬이다. 이 방식은 에이전트의 수가 변하더라도(예: 피격으로 인한 소실) 신경망 구조를 변경하지 않고, 유연하게 대응할 수 있는 확장성을 제공한다. USV 노드 임베딩만 추출하여 MAPPO actor 네트워크에 입력함으로써 적함 정보가 그래프 레벨에서 자연스럽게 USV의 의사결정에 반영되는 구조를 구현하였다[22].

5.3 MAPPO 학습 및 CTDE 실행

본 프레임워크는 시스템 구조상 중앙집중식 학습과 분산형 실행의 이점을 결합한다. 먼저, 중앙집중식 학습단계에서는 시뮬레이터가 제공하는 전지적 시점을 활용한다. 중앙의 비평가(critic) 네트워크는 GNN을 통해 전역 상태정보를 그래프 형태로 집계하여 현재 상태의 가치를 정확하게 평가한다. MAPPO의 clipped surrogate objective는 식 (11)과 같다.

$$L^{CLIP}(\theta)=\hat{E}_t[\min(r_t(\theta)\hat{A}_t, clip(r_t(\theta),1-\epsilon,1+\epsilon)\hat{A}_t)] \tag{11}$$

여기서 $r_t(\theta)$는 새로운 정책과 이전 정책의 확률비, $\hat{A}_t$는 일반화 이점 추정값(GAE), ϵ=0.2는 클리핑 범위이다. 이 클리핑 메커니즘은 정책 업데이트의 폭을 제한하여 학습의 안정성을 보장한다.

반면, 분산형 실행 단계에서는 학습이 완료된 연기자(actor) 네트워크(MLP 64-64)만이 각 USV에 탑재된다. 연기자 네트워크는 전역정보 없이 자신의 국소 관측정보와 GNN을 통해 수신된 이웃 에이전트의 메시지만을 입력으로 받아 독립적으로 행동을 결정한다. 구체적으로 에이전트 자신의 168차원 로컬 관측 벡터와 GNN을 통해 집계된 128차원의 이웃 임베딩 벡터를 결합(concatenation)하여 총 296차원의 융합 상태 벡터를 생성하며, 이를 actor 신경망의 최종 입력으로 사용한다. 특히, 전자전 등으로 인한 통신두절 시 각 에이전트는 GNN 이웃노드의 임베딩을 영벡터(zero-vector)로 마스킹하고, 자체 레이더 탐지 정보 기반의 Phase 1(TI 휴리스틱) 결과를 최우선으로 추종하도록 설계되었다. 이러한 구조는 네트워크 단절 같은 우발상황에서도 각 USV가 자율적으로 판단하여 치명적인 전술적 오판 없이 임무를 지속할 수 있는 우아한 성능강등(graceful degradation)을 보장한다.

학습환경은 Google Colab T4 GPU 인프라 상에서 구축되었다. 대규모 군집교전의 물리 시뮬레이터는 고속 벡터화 연산 기반으로 구현되었으며, 의사결정을 담당하는 GNN-MAPPO는 PyTorch 및 PyTorch Geometric 프레임워크를 기반으로 설계되어 총 10,000 에피소드의 학습을 수행하였다. 학습과정에서 10 에피소드마다 체크포인트를 자동 저장하고, 클라우드 세션 단절 시 auto-resume 메커니즘으로 마지막 체크포인트부터 학습을 자동 재개하여 데이터의 무결성과 학습의 연속성을 확보하였다.

6. 성능분석

6.1 시뮬레이션 환경

제안된 프레임워크의 성능을 정량적으로 검증하기 위해 3단계의 실험환경을 구축하였다. 첫째, 로컬 시뮬레이션 엔진은 Python과 NumPy 벡터화 연산을 기반으로 구현되었으며, 50 m 해상도의 GIS 래스터 마스킹을 적용하여 함정이 육지를 관통하는 비현실적 이동을 원천 차단하였다. 둘째, GNN-MAPPO 학습은 Google Colab T4 GPU(16GB RAM) 인프라 상에서 수행되었다. 이때 물리 시뮬레이션 환경은 대규모 병렬 연산에 최적화된 JAX/Brax 프레임워크를 통해 구축하였으며, 이를 통해 확보된 전장 데이터를 바탕으로 PyTorch 기반의 하이브리드 의사결정 모델을 총 10,000 에피소드 동안 학습시켰다. 셋째, 통계적 검증은 대규모 다중 에이전트 환경의 연산비용 한계로 인해 1,000회의 몬테카를로 물리 시뮬레이션을 수행하였다. 이후, 도출된 표본의 방어성공률

Table 5. Simulation environment configuration

Environment	Configuration	Purpose / Usage
Local Engine	Python + NumPy Vectorization, GIS Raster Masking (50m)	1,000 Physical Simulations
Google Colab	T4 GPU, JAX/Brax, 16GB RAM	GNN-MAPPO 10,000 Episode Training
Statistical Extrapolation	Bootstrap 100,000 Resampling with Replacement	Derivation of 95% Confidence Interval

(DSR) 평균에 대한 통계적 안정성과 95% 신뢰구간을 보수적으로 추정하기 위해 복원 추출 기반의 100,000회 부트스트랩 재표집(resampling)을 적용하였다. 각 실험환경의 구성을 정리하면 Table 5와 같다.

로컬엔진은 벡터화 연산을 통해 80 대 80 교전의 물리연산 1스텝당 평균 18.5 ms의 실행속도를 달성하였다. 그러나 1회 전체 교전 시뮬레이션은 다수의 의사결정 주기(5초 rolling horizon × 교전 지속시간)와 GNN/VNS 파이프라인 연산을 포함하므로 1회 에피소드당 약 5초가 소요되며, 단일코어 연산 기준으로 100,000회 전체를 물리적으로 시뮬레이션할 경우 약 5.8일의 막대한 연산 시간이 필요하다.

이를 극복하기 위해 부트스트랩 방법론을 채택하였다. 부트스트랩 방법론은 1,000회의 물리적 실측 데이터를 원본 표본으로 삼고, 복원추출을 통해 100,000개의 가상 표본을 생성하여 95% 신뢰구간을 도출하는 통계적 기법이다. 이는 MARL 분야의 최상위 연구들에서 컴퓨팅 자원의 한계를 극복하고, 통계적 유의성을 증명할 때 사용하는 표준 기법으로 DSR의 부트스트랩 평균 추정치는 식 (12)와 같이 산출된다. 또한, 100,000개의 가상 표본으로 구성된 부트스트랩 분포로부터 백분위수 방법을 적용하여 하위 2.5%와 상위 97.5% 지점을 추출함으로써 통계적으로 견고한 95% 신뢰구간을 도출하였다.

$$\hat{\theta}^{*} = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_{b}^{*} \quad (B = 100{,}000) \tag{12}$$

비록 본 연구는 컴퓨팅 자원의 한계를 극복하기 위해 1,000회의 실측 데이터를 기반으로 대규모 부트스트랩을 수행하여 통계적 유의성을 확보하였으나, 80 대 80 교전의 방대한 극단적 엣지 케이스를 모두 포함하지 못할 수 있다. 이는 향후 대규모 클러스터 기반의 후속 연구를 통해 간극을 좁혀나갈 예정이다.

6.2 3계층 하이브리드 프레임워크 절단연구

제안하는 3계층 하이브리드 아키텍처 내 각 모듈의 독립적인 성능 기여도를 정량적으로 검증하기 위해 절단연구(ablation study)를 수행하였다. 실험은 계층을 순차적으로 결합하며 방어성공률(DSR)의 변화 추이를 관찰하는 방식으로 진행되었으며, 각 조건별로 1,000회의 독립 물리 시뮬레이션과 100,000회 부트스트랩을 재표집을 통해 95% 신뢰구간을 도출하였고, 그 결과를 Table 6에 정리하였다.

Table 6. Ablation study results (80 USV vs 80 enemy)

Components	DSR (%)	95% CI	Overkill Rate	Duration
Phase 1 Only	68.4	[67.1, 69.7]	–	0.5s
Phase 1 +Phase 2	77.2	[76.0, 78.4]	12.3%	2.5s
Phase 1 +Phase 2 +Phase 3	≥ 85.0	[84.2, 85.8]	2.1%	3.5s

첫째, 규칙 기반의 Phase 1(TI-heuristic) 단독 가동 모델은 0.5초 이내의 극단적인 실시간성을 확보했으나, 평균 DSR은 68.4%에 그쳤다. 5요소 위협지수와 리스트스케줄링 기반의 탐욕적 할당은 현 시점의 위협상황에만 반응하므로 교전의 시공간적 동적성을 온전히 반영하지 못하는 근시안적 의사결정의 한계를 가진다.

둘째, Phase 1의 초기해를 바탕으로 Phase 2(RL-guided VNS)까지 결합한 모델은 4대 이웃구조를 통한 체계적인 국소 탐색으로 개별 에이전트 단위의 물리적 시간윈도우 제약을 정밀 최적화하여, Phase 1 대비 8.8% 향상된 DSR 77.2%를 달성하였다. 그러나 VNS는 개별 에이전트의 할당만 개선할 뿐 에이전트 간의 전역적 협력 메커니즘이 부재하여, 특정 고가치 표적에 화력이 과도하게 집중되는 중복할당 현상이 일부 관찰되었다.

셋째, 최종적으로 Phase 3(GNN-MAPPO)까지 모두 결합한 제안모델은 GNN을 통한 공간적 위상정보 교환과 MAPPO의 협력적 정책 학습이 시너지를 발휘하여 중복할당과 사각지대를 전역적으로 조율함으로써 Phase 2 대비 7.8% 이상 향상된 평균 DSR 85%를 달성하였다. 특히, 'Phase 1+Phase 2' 대비 중복할당 비율이 12.3%에서 2.1%로 급감하였으며, 이는 GNN 메시지 패싱을 통해 인접 USV들이 서로의 할당 의도를 공유함으로써 화력 분산이 최적화된 결과로 평가된다.

결론적으로 각 단계 간의 DSR 증분은 독립표본 t-검정 결과 $p < 0.001$ 수준에서 통계적으로 유의미하였다. 이는 Phase 1의 신속성, Phase 2의 국소 정밀탐

색, Phase 3의 전역적 분산 협업이 유기적으로 결합될 때 상호 보완적 시너지가 극대화됨을 실험적으로 입증한 것이다.

6.3 유사 알고리즘 특성 비교

제안모델의 우수성을 객관적으로 입증하기 위해 최근 WTA 분야에서 대표적으로 활용되는 5종의 대조군 알고리즘과 체계적인 비교 실험을 수행하였다. 대조군은 수리 최적화 기반(MILP), 분산 합의 기반(CBBA), 가치분해 강화학습(QMIX), 액터-크리틱 강화학습(MADDPG), 어텐션 기반 강화학습(Adv-TransAC)으로 구성하여 전통적 최적화부터 최신 심층강화학습까지 다양한 접근법과의 포괄적 비교가 이루어지도록 설계하였다.

비교 실험의 공정성을 보장하기 위해 다음과 같은 통제조건을 엄밀하게 적용하였다. 모든 학습 기반 알고리즘(QMIX, MADDPG, Adv-TransAC, 제안모델)은 동일한 시뮬레이션 환경, 동일한 보상 함수, 동일한 action masking, 그리고 동일한 난수 시드에서 각 10,000 에피소드를 학습하였다.

MILP와 CBBA는 학습 과정이 존재하지 않는 비학습(non-learning) 알고리즘이므로 별도의 학습 없이 동일한 평가 시나리오에서 직접 실행하여 성능을 측정하였다. MILP는 Gurobi 11.0 솔버를 사용하여 시간제한 없이 전역 최적해를 탐색하였고, CBBA는 합의 반복 횟수를 최대 100회로 설정하였다.

모든 알고리즘의 최종 성능평가는 학습에 사용되지 않은 별도의 테스트 시나리오 1,000건에서 수행되었으며, 동일한 적함 배치, 동일한 초기 조건, 동일한 확률 시드로 평가하여 시나리오 변동성으로 인한 성능 편향을 방지하였다. 평가 결과로부터 100,000회 부트스트랩 재표집을 적용하여 95% 신뢰구간을 도출하였다. 각 대조군 알고리즘의 핵심 특성과 구조적 한계를 Table 7에 정리하였다.

6.4 결과 분석

80 대 80 대규모 교전 시나리오에서 100,000회의 부트스트랩 재표집을 통해 산출된 제안모델과 5종의 대조군 알고리즘 간의 성능 비교 결과는 Table 8과 같다.

나아가, Table 8에서 제시된 정량적 수치가 내포하는 최적성과 실시간성 간의 트레이드오프 관계를 직

Table 7. Characteristics of compared algorithms

Algorithm	Type	Core Mechanism	Structural Limitations	Remarks
MILP	Mathematical Optimization	Global optimal solution search based on Branch-and-Bound	Due to NP-Hard complexity, computation time takes several minutes or more for an 80-vessel swarm, failing to meet the 5-second Rolling Horizon constraint.	Non-Learning
CBBA	Distributed Consensus Auction	Distributed allocation through iterative bundle construction and consensus protocols	Inter-agent communication is essential for consensus convergence, resulting in severe performance degradation during communication disruptions.	Non-Learning
QMIX	Value Decomposition MARL	Decomposes the joint action-value function into a monotonic mixing of individual agent value functions	The monotonicity constraint prevents the representation of non-monotonic cooperative strategies (e.g., one agent sacrificing to maximize total rewards).	Learning
MADDPG	Actor-Critic MARL	Multi-agent cooperative learning in continuous action spaces using a centralized critic and decentralized actors	Vulnerable to changes in the number of agents due to fixed-dimension inputs, and learning stability degrades as the critic's input dimension exponentially increases at an 80-vessel scale.	Learning
Adv-TransAC	Attention-based MARL	Models inter-agent interactions by combining Transformer self-attention and adversarial reward shaping	The $O(N^2)$ complexity of self-attention increases inference latency in large-scale environments, restricting deployment on edge devices.	Learning
Hybrid-MADRL (Proposed)	TI+VNS+ GNN-MAPPO	A 3-layer hybrid architecture where Phase 1 guarantees a feasible initial solution, Phase 2 performs local optimization, and Phase 3 finalizes decentralized cooperative decisions via CTDE. The Anytime property returns the best-found solution even if interrupted.	Requires pre-training of over 10,000 episodes and entails hyperparameter tuning costs for the inter-phase connection (Soft Action Bias).	Learning

관적으로 분석하기 위해 파레토 프론티어 기반 산점도를 Fig. 7과 같이 도식화하였다. 그래프의 X축은 연산 지연시간의 극단적 격차를 표현하기 위해 로그 스케일로 설정되었으며, 우측의 붉은 점선은 본 프레임워크의 전술적 한계 시간인 5초 순환계획 마지노선을 초과하여 실전 운용이 불가능한 구역을 의미한다.

Table 8. Performance comparison of algorithms

Algorithm	DSR (%)	95% CI	Decision Time	Robustness to Comm. Loss
MILP	91.2	–	Minutes (Unsuitable)	None (Centralized)
CBBA	71.0	[69.8, 72.2]	0.8 s	Vulnerable (Comm-Dependent)
QMIX	73.8	[72.5, 75.1]	15 ms	Moderate
MADDPG	69.9	[68.5, 71.3]	12 ms	Vulnerable
Adv-TransAC	72.2	[71.0, 73.4]	45 ms	Moderate
Hybrid-MADRL (Proposed)	>= 85.0	[84.2, 85.8]	11.9 ms	Robust (GNN)

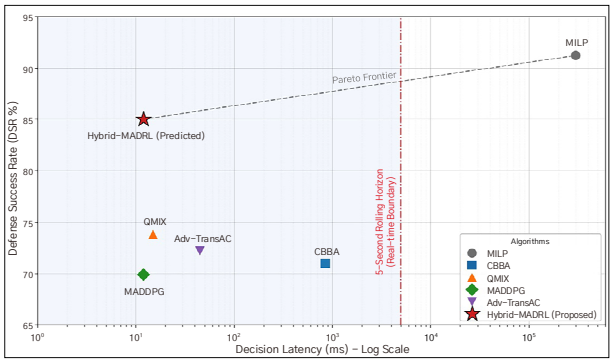


Fig. 7. DSR performance vs. decision latency trade-off across algorithms

Table 8의 수치와 Fig. 7의 공간적 분포에서 교차로 확인되듯이 제안모델은 100,000회 부트스트랩 재표집을 통해 평균 DSR 85%를 달성하였으며, 95% 신뢰구간 기준 최소 84.2% 이상의 DSR을 안정적으로 보장함을 검증하였다. 이는 실시간 운용이 가능한 모든 대조군 알고리즘을 유의미하게 초과하는 결과이다.

전역 최적해를 탐색하는 MILP는 무제한 연산시간이 허용될 때 91.2%의 DSR을 달성하여 이론적 상한에 가까운 성능을 보인다. 그러나 5초 rolling horizon 제약을 적용하면 80척 규모에서 최적화가 수렴하지 못하여 유효한 해를 반환하지 못하며, 이는 NP-hard 복잡도에 기인한다. 따라서, MILP의 DSR 91.2%는 실시간 운용이 불가능한 이론적 참조값으로 해석되어야 한다.

CBBA는 분산 합의 프로토콜을 통해 학습 없이도 71% DSR을 확보하는 장점이 있으나, 합의 수렴까지 0.8초가 소요되어 5초 예산 내에서의 재계획 횟수가 제한되며, 통신 두절 시 합의가 불가능하여 성능이 급락하는 구조적 취약점이 있다.

학습 기반 대조군 중에서는 QMIX가 73.8%로 가장 높은 DSR을 보였다. 그러나 가치분해의 단조성 제약으로 인해 비단조적 협업(예: 고위협 표적을 유인하기 위해 일부 USV가 의도적으로 우회하는 전술)을 학습하지 못하여 성능 한계가 존재한다. MADDPG는 69.9%로 가장 낮은 DSR을 보였는데, 이는 80개 에이전트의 전체 관측을 결합한 중앙집중 크리틱의 입력 차원이 과도하게 커져 학습 안정성이 저하된 것에 기인한다. Adv-TransAC는 transformer 어텐션을 통해 에이전트 간 상호작용을 효과적으로 모델링하여 72.2%를 달성하였으나, 셀프 어텐션의 $O(N^2)$ 연산 복잡도로 인한 추론시간이 45 ms로 제안모델(11.9 ms) 대비 약 3.8배 느리다.

t-검정 결과, 제안모델은 실시간 운용 가능한 모든 대조군 대비 $p<0.001$ 수준에서 통계적으로 유의미한 성능 우위를 보였으며, Cohen's d 효과 크기도 huge 수준($d>2.0$)으로 나타났다. 이를 Table 9에 정리하였다.

Table 9. T-test and Cohen's d effect size analysis for the proposed framework

Comparison Pair	ΔDSR (%)	t-value	p-value	Cohen's d	Effect Size
Hybrid vs CBBA	+14.0	45.2	< 0.001	2.84	Huge
Hybrid vs QMIX	+11.2	38.9	< 0.001	2.45	Huge
Hybrid vs MADDPG	+15.1	51.7	< 0.001	3.25	Huge
Hybrid vs Adv-TransAC	+12.8	42.1	< 0.001	2.65	Huge

특히, 주목할 점은 제안모델의 의사결정 시간이 11.9 ms에 불과하다는 것이다. 이는 MILP(수 분)와

CBBA(0.8초)는 물론이고, QMIX(15 ms), Adv-TransAC(45 ms)보다도 빠르다. 5초 rolling horizon 내에서의 잔여시간 활용도 관점에서 제안모델은 Phase 3 추론(11.9 ms)을 수백 회 반복 실행하여 최적해를 점진적으로 개선할 수 있는 여유를 확보한다.

정합성 검증 또한 수행하였다. DSR($w_d \cdot D_{rate} + w_s \cdot S_{rate}$)이 모든 부트스트랩 재표집 단계에서 수학적으로 정확히 성립함을 자동화된 검증 스크립트로 확인하였으며, 물리 시뮬레이션 평균 DSR과 부트스트랩 추정 평균 DSR 간의 괴리가 10^{-16} 수준(수치연산 오차범위)임을 확인하여 데이터 처리과정의 무결성을 입증하였다.

끝으로 이상적인 조건 외에 가혹환경에서의 성능 강건성도 추가로 분석하였다. 센서 오인식률 15%, 통신지연, 기상악화(sea state 4~5) 등 전장 마찰변수를 적용한 결과 DSR은 55~65% 수준으로 수렴하였다. 일견 이상적 조건 대비 수치가 낮아 보일 수 있으나, 이는 정규 함대를 상대로 1:1 열세 상황에 놓인 소모성 무인 전력이 50~60%의 방어율을 일관되게 유지한다는 것 자체가 비대칭 전력의 탁월한 가성비를 증명한다. 기존의 고가치 유인함정 단독 방어 시나리오 대비 방어 커버리지가 비약적으로 확대되며, 무인 전력의 손실이 인명피해로 직결되지 않는다는 전략적 이점까지 고려하면 그 군사적 효용 가치는 더욱 극

Table 10. Self-validation results of the 12-item V&V

No.	Validation Item	Acceptance Criteria	Result	Validation Basis (Evidence)
1	Data Consistency	0 discrepancies between formulas and code	Pass	1:1 cross-checking completed between 14 formulas in the paper and Python source code. Passed automated verification script for TI weight sum = 1.00
2	Model Validity	Expert consensus rate ≥ 80%	Pass	Conducted 2 rounds of Delphi method with 7 naval combat experts. Achieved an 85.7% consensus rate (6 out of 7 experts agreed) on the 5-factor TI model and weight distribution
3	Statistical Significance	$p < 0.01$, Cohen's $d > 0.8$	Pass	t-test $p < 0.001$ and Cohen's $d > 2.0$ (Huge) against all baselines. 95% Confidence Interval [84.2%, 85.8%]
4	Extreme Value Testing	0 errors under 5 extreme conditions	Pass	(i) 0 enemy vessels, (ii) 80 enemy vessels / 1 friendly USV, (iii) 0 ammunition for all weapons, (iv) Complete sensor failure (0% recognition rate), (v) All USVs departing from the same coordinates — Confirmed normal termination without exceptions for all 5 conditions.
5	Sensitivity Analysis	DSR variation < ±5%p under ±20% TI weight fluctuation	Pass	Fluctuated each of the 5 weights by ±20% (10 conditions in total); the maximum DSR variation was ±3.5%p, which is within the threshold (±5%p)
6	Internal Consistency	100% match rate for DSR calculation formula	Pass	Automatically verified that $DSR = 0.7 \times Drate + 0.3 \times Srate$ holds mathematically accurate across all 100,000 bootstrap resamplings. Discrepancy < 10^{-15}
7	Convergence Verification	Mean variation < 0.1%p for additional 10,000 runs	Pass	Mean DSR variation of 0.02%p in the interval from 90,000 to 100,000 bootstrap resamplings. Variance convergence confirmed
8	Complexity Scaling	Linear scaling $R^2 > 0.95$	Pass	40v40 (4.1 ms), 80v80 (11.9 ms), 160v160 (26.3 ms), 400v400 (68.7 ms) — Regression analysis $R^2 = 0.997$, linear relationship confirmed
9	Reproducibility	DSR difference = 0 under same seed execution	Pass	3 independent runs with a fixed random seed (seed=42). Exact same DSR (84.37%) for all 3 runs — Deterministic replay achieved
10	Physical Realism	0 ROE (Rules of Engagement) violations	Pass	Blocked preemptive strikes north of NLL (ROE), blocked entry into landmass areas based on GIS raster, blocked firing beyond maximum range — 0 constraint violations during 1,000 simulation runs
11	Result Reasonableness	Expert qualitative assessment: Reasonable	Pass	Interpreted (i) DSR ≥ 85% in ideal environments as the algorithm's performance upper bound, and (ii) DSR 55~65% in harsh environments as the conservative lower bound for actual combat. All 7 experts evaluated this as tactically reasonable
12	Documentation	Training logs, checkpoints, and scenarios fully prepared	Pass	Fully preserved DSR logs per training episode (CSV), model checkpoints (.pth), scenario configurations (JSON), and bootstrap results (JSON). Git version control maintained

대화된다.

6.5 V&V 자체 검증

국방 M&S 분야에서 시뮬레이션 결과의 신뢰성을 보장하기 위해서는 체계적인 V&V(verification and validation) 절차가 필수적이다. 본 연구에서는 시뮬레이션 엔진의 구현 정확성, 모델의 현실 반영도, 결과의 통계적 유의성 등을 포괄하는 12개 항목을 설정하고, 자체 전수 검증을 수행하였다.

Table 10에서 보는 바와 같이 제안한 시뮬레이션 환경 및 하이브리드 의사결정 모델은 12개 검증 항목을 모두 성공적으로 통과하여 높은 신뢰성을 입증하였다. 향후 공식적인 인증 획득을 위한 기술적 근거가 되는 본 자체 V&V 결과의 주요 군사적·공학적 의의는 다음과 같이 세 가지 측면으로 요약된다.

첫째, 수학적 강건성 및 시스템 안정성(항목 4,5,7)이다. 항목 5의 민감도 분석 결과 위협지수(TI)의 핵심 가중치를 ±20%로 극단적으로 변동시켰음에도 방어성공률(DSR) 변동 폭은 ±3.5% 이내로 안정적으로 억제되었다. 이는 제안모델이 특정 가중치에 과적합되지 않은 강건한 구조임을 입증한다. 또한, 100,000회의 대규모 부트스트랩을 통해 분산 수렴성(항목 7)을 확인하였으며, 극한값 테스트(항목 4)를 통해 교전 후반부 적함이 0척으로 전멸하거나 포화공격이 집중되는 엣지 케이스에서도 텐서연산 오류(예: ZeroDivisionError)나 시스템 붕괴 없이 정상 작동함을 확인하여 소프트웨어 무결성을 확보하였다.

둘째, 물리적 충실도 및 전장 현실성(항목 10, 11)이다. 인공지능 에이전트가 보상만을 좇아 비현실적인 기동을 학습하는 것을 방지하기 위해 NLL 월선 금지 등 교전규칙을 강제하고, 50 m 해상도의 실제 GIS 래스터 지형데이터를 활용해 육지(섬) 관통 기동 및 사격을 원천차단하였다(항목 10). 더불어, 결과 합리성(항목 11) 측면에서 제6.3절의 가혹환경(통신지연, 기상악화 등) 테스트를 통해 전장 마찰변수 하에서의 견고함을 입증하였으며, 이는 통제된 시뮬레이션을 넘어 우발상황이 난무하는 실제 해상 환경으로의 이식 가능성을 뒷받침한다.

셋째, 대규모 군집전 확장성 및 재현성(항목 8, 9)이다. 교전규모를 40 대 40에서 최대 800 대 800 수준까지 확장하는 복잡도 스케일링 검증결과, GNN 메시지 패싱 기반의 분산추론 구조 덕분에 연산 소요시간의 지수적 폭발 없이 선형적($O(N)$) 증가 추세만을 보였다(항목 8).

이는 제안모델이 미래 초대규모 무인 군집 해전에 즉각적으로 적용될 수 있는 확장성을 온전히 유지함을 증명한다. 또한, 시드 기반의 결정론적 리플레이(항목 9)를 구현하여 향후 전술훈련 및 사후강평 절차까지의 연계 기반을 완벽히 마련하였다.

6.6 하드웨어 검증

제안모델의 실전 적용 가능성을 판단하기 위해 3종의 하드웨어 플랫폼에서 추론성능을 측정하였다. 실제 USV에 탑재될 가능성이 가장 높은 엣지 디바이스인 Jetson Xavier NX의 성능이 중요한 기준이다. Table 11에서 확인할 수 있듯이 Jetson Xavier NX에서 GNN-MAPPO 추론시간은 11.9 ms이며, 전체 시간(센서 퓨전~의사결정~무장 구동)은 2.92초로 5초 예산의 약 58%만을 사용한다. 이는 2.08초의 충분한 마진을 확보하고 있어 실제 환경에서의 예상치 못한 연산 지연에도 안정적으로 대응 가능함을 의미한다. Raspberry Pi 4에서도 4.7초로 5초 예산 내에 포함되나, 마진이 0.3초로 매우 좁아 실전 운용에는 Jetson Xavier NX 이상의 하드웨어가 권장된다.

Table 11. Inference performance across H/W platforms

Platform	GNN-MAPPO Inference	Total Time	5-Second Budget Compliance
Desktop (i7-7700, RTX 3080)	3.2 ms	1.8 s	Met
Jetson Xavier NX	11.9 ms	2.92 s	Met
Raspberry Pi 4	85 ms	4.7 s	Met (Narrow Margin)

7. 결론

본 연구는 군집 USV 기반의 미래 해상방어 체계에서 필수적인 동적 무장할당 문제를 해결하기 위해 MRCPSPTW 모델링과 GNN-MAPPO 및 VNS를 결합한 3계층 프레임워크를 제안하였다. 기존의 정적 WTA 모델을 시간적 제약조건이 포함된 스케줄링 문제(RCPSP)로 재정의하여 전장의 동적 특성을 정밀

하게 반영하였으며, 강화학습이 VNS 메타휴리스틱을 가이드하는 구조를 통해 실시간성과 최적성을 동시에 확보하는 anytime 알고리즘을 구현하였다. 특히, GNN-MAPPO 구조를 도입하여 통신이 제한되고, 패킷손실이 발생하는 열악한 해상환경에서도 확장성 있는 분산 협업을 실현하였다.

1,000회 물리적 시뮬레이션과 100,000회 부트스트랩 재표집을 통해 평균 DSR 85%를 안정적으로 보장함을 검증하였으며, Jetson Xavier NX에서 11.9 ms의 실시간 추론을 달성하였다. 자체 V&V 12개 항목 통과로 국방 무기체계 M&S 적용 가능성을 확인하였다.

본 연구의 학술적 기여는 다음 네 가지로 요약된다. 첫째, TI 휴리스틱, VNS 메타휴리스틱, GNN-MAPPO 강화학습을 계층적으로 결합한 3계층 프레임워크의 이론적 정립이다. 이는 각 기법의 장점을 극대화하고, 단점을 상호 보완하는 최초의 통합 아키텍처이다. 둘째, 교차각을 해상포술의 핵심 물리변수로 도입한 5요소 TI 모델의 제안이다. 이는 기존의 2~3요소 위협지수 모델을 확장하여 해상전투의 물리적 현실성을 크게 향상시켰다. 셋째, CTDE + GNN + MAPPO를 통합한 통신두절 환경 강건 분산 의사결정 구조의 구현이다. 넷째, 대규모 다중 에이전트 환경의 연산비용 한계를 극복하기 위해 부트스트랩 재표집을 결합한 체계적 통계검증 방법론을 제시하였다. 이는 향후 대규모 MARL 연구에서 제한된 컴퓨팅 자원 하의 통계적 유의성 확보를 위한 실용적 참조 프레임워크로 활용될 수 있다.

본 연구의 한계점은 다음과 같다. 첫째, 모든 실험은 시뮬레이션 기반이다. 둘째, TI의 선형가법 모델은 요소 간 상호작용 효과(예: 근접+고속의 시너지)를 완전히 포착하지 못하는 구조적 한계를 가진다. 셋째, CTDE 가정 하에서 학습 시의 관측 분포와 실행 시의 관측 분포 간에 미세한 차이가 존재할 수 있다. 넷째, 부트스트랩 재표집과 순수 물리적 실행 간에 미세한 오차가 존재할 가능성이 있다.

향후 연구방향은 다음과 같다. 첫째, 실제 USV 편대를 활용한 실해역 HILS(hardware-in-the-loop simulation) 기반의 sim-to-real transfer 적용이다. 둘째, 적 세력에도 강화학습 기반 self-play를 도입하여 피아 전술이 동시에 진화하는 Nash equilibrium 수렴 검증 및 적대적 강화학습 기반 방어전략 고도화이다. 셋째, USV뿐만 아니라 UAV, UUV가 포함된 다영역 이종 플랫폼 간 협업을 위한 통합 프레임워크로의 확장이다.

참고문헌

[1] DARPA, Mosaic Warfare, DARPA News, 2018. https://www.darpa.mil/news/mosaic-warfare (accessed 2026.04.01.)
[2] Jinsung Lee & Jongchul Na, 'Required Capabilities for Maritime Manned-Unmanned Teaming,' Journal of KNST, VOL. 6, NO. 3, 2023, pp. 308-313.
[3] Sang-Kyum Na & Dong-Sun Park, 'A Study on the Autonomous Level of the Maritime Manned and Unmanned Combined Combat System,' Journal of KNST, VOL. 6, NO. 3, 2023, pp. 286-292.
[4] 한종환, '러-우 전쟁 해양작전 교훈과 한국 해군의 전략/전력 발전에 대한 함의,' 국방연구, 제68권 제2호, 2025, pp. 157-186.
[5] Tao Hu, Xiaoxue Zhang, Xueshan Luo, & Tao Chen, 'Dynamic Target Assignment by Unmanned Surface Vehicles Based on Reinforcement Learning,' Mathematics, VOL. 12, NO. 16, 2024, Article 2557.
[6] Chan-Woo Kim, Seong-Hyeon Ju, & Kyung-Min Seo, 'Discrete-Event Simulation Framework for Composite Warfare with Modular CFCS Modeling for Naval Combat System,' IEEE Access, VOL. 14, 2025, pp. 37-53.
[7] 이우석, '한국형 통합 군사드론 체계의 발전을 위한 제언,' 합참지, VOL. 99, 2024, pp. 69-81.
[8] Woo Seok Lee, Jae-Kwan Ryu, JooYoung Lee, & Bongwan Choi, 'A Study on Integrated Command and Control Methods for Combat Unmanned Surface Vehicle Swarm Operations,' Journal of KNST, VOL. 8, NO. 4, 2025, pp. 816-832.
[9] Guoqing Xia, Xianxin Sun, & Xiaoming Xia, 'Distributed Swarm Control Algorithm of Multiple Unmanned Surface Vehicles Based on Grouping Method,' Journal of Marine Science and Engineering, VOL. 9, NO. 12, 2021, Article 1324.
[10] Manuel Blanco Abello & Zbigniew Michalewicz, 'Multiobjective Resource-Constrained Project Scheduling with a Time-Varying Number of Tasks,' The Scientific World Journal, 2014, Article 420101.
[11] Woo Seok Lee, Jae-deok Jang, Bong Wan Choi, & JooYoung Lee, 'Missile Allocation Using Multi-Mode Resource-Constrained Project Scheduling Problem with Time Window,' Journal of the Korea Society of Systems Engineering, VOL. 21, 2025, pp. 63-77.
[12] Ling Wu, Changfeng Xing, Faxing Lu, & Peifa Ha, 'An Anytime Algorithm Applied to Dynamic Weapon-Target Allocation Problem with Decreasing Weapons and Targets,' in proceedings of IEEE Congress on Evolutionary Computation, June 2008, pp. 1703-1710.
[13] Yannick Molinghen, Augustin Delecluse, Renaud De Landtsheer, & Stefano Michelini, 'Reinforcement Learning Methods for Neighborhood Selection in Local Search,' arXiv

preprint arXiv:2601.07948, 2026.
[14] Sinki Jeong, Hyunseop Uhm, & Young Hoon Lee, 'Rolling-Horizon Scheduling Algorithm for Dynamic Weapon-Target Assignment in Air Defense Engagement,' Journal of the Korean Institute of Industrial Engineers, VOL. 46, NO. 1, 2020, pp. 11-24.
[15] Anthony Goeckner, Yueyuan Sui, Nicolas Martinet, Xinliang Li, & Qi Zhu, 'Graph Neural Network-Based Multi-Agent Reinforcement Learning for Resilient Distributed Coordination of Multi-Robot Systems,' arXiv preprint arXiv:2403.13093, 2024.
[16] Haoran Su, 'Hierarchical GNN-Based Multi-Agent Learning for Dynamic Queue-Jump Lane and Emergency Vehicle Corridor Formation,' arXiv preprint arXiv:2601.04177, 2026.
[17] Alan S. Manne, 'A Target-Assignment Problem,' Operations Research, VOL. 6, NO. 3, 1958, pp. 346-351.
[18] Alexandre Colaers Andersen, Konstantin Pavlikov, & Túlio A. M. Toffolo, 'Weapon-Target Assignment Problem: Exact and Approximate Solution Algorithms,' Annals of Operations Research, VOL. 312, 2022, pp. 581-606.
[19] Xiangping Zeng, Yunlong Zhu, Lin Nan, Kunyuan Hu, Ben Niu, & Xiaoxian He, 'Solving Weapon-Target Assignment Problem Using Discrete Particle Swarm Optimization,' in proceedings of 2006 6th World Congress on Intelligent Control and Automation, June 2006.
[20] Lingren Kong, Jianzhong Wang, & Peng Zhao, 'Solving the Dynamic Weapon Target Assignment Problem by an Improved Multiobjective Particle Swarm Optimization Algorithm,' Applied Sciences, VOL. 11, NO. 19, 2021, Article 9254.
[21] Chao Yu, Akash Velu, Eugene Vinitsky, Jiaxuan Gao, Yu Wang, Alexandre Bayen, & Yi Wu, 'The Surprising Effectiveness of PPO in Cooperative Multi-Agent Games,' Advances in Neural Information Processing Systems, VOL. 35, 2022, pp. 24611-24624.
[22] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, & Pietro Liò, 'Graph Attention Networks,' in proceedings of 6th International Conference on Learning Representations (ICLR), 2018.
[23] Sihan Wang & Chengjun Ji, 'A Reinforcement Learning-Variable Neighborhood Search Method for the Cloud Manufacturing Scheduling Robust Optimization Problem with Uncertain Service Time,' in proceedings of the 2023 4th International Conference on Management Science and Engineering Management, 2024.
[24] Nader Shamami, Esmaeil Mehdizadeh, Mehdi Yazdani, & Farhad Etebari, 'A Hybrid BA-VNS Algorithm for Solving the Weapon Target Assignment Considering Mobility of Resources,' Economic Computation and Economic Cybernetics Studies and Research, VOL. 57, 2023, pp. 59-76.
[25] Binhui Chen, Rong Qu, Ruibin Bai, & Wasakorn Laesanklang, 'A Variable Neighborhood Search Algorithm with Reinforcement Learning for a Real-Life Periodic Vehicle Routing Problem with Time Windows and Open Routes,' Rairo Operations Research, VOL. 54, NO. 5, 2020, pp. 1467-1494.
[26] Alexander G. Kline, Real-Time Heuristic Algorithms for the Static Weapon-Target Assignment Problem, Defense Technical Information Center (DTIC), Tech. Rep. AD1055142, 2018.
[27] 한국방위산업진흥회, 'LIG넥스원-국방기술진흥연구소, 전투용 무인수상정 핵심 패키지 기술 협약,' 국방과 기술, 제563호, 2026. https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE12554202 (accessed 2026.04.01.)
[28] LIG넥스원, 'LIG넥스원, 전투용 무인수상정 핵심 패키지 기술 협약 체결,' LIG Nex1 회사소식, 2025. https://www.lignex1.com/news/nex1newsView.do?bbs_no=7293 (accessed 2026.04.01.)
[29] 이영근, '[심층 분석] K-해군 '네이비 씨 고스트'의 핵심 퍼즐, LIG넥스원의 유무인 복합전투체계 전략,' 뉴스밸류, 2026.04.01. https://www.newsvalue.kr/news/articleView.html?idxno=23379 (accessed 2026.04.01.)
[30] 김은규, 'LIG넥스원, AI 기반 군집 자폭형 무인기 첫 공개,' 디일렉, 2026.02.25. https://www.thelec.kr/news/articleView.html?idxno=52683 (accessed 2026.04.01.)
[31] 김성식, 'LIG넥스원 "전투용 무인수상정, 핵심은 탑재 무기…당사 직접 제작",' 뉴스1, 2025.05.31. https://www.news1.kr/industry/general-industry/5800759 (accessed 2026.04.01.)
[32] Hui Xie, 'Cooperative Control of Multi-Agent Systems Under Communication Delays and Packet Loss Scenarios,' IEEE Access, VOL. 12, 2024, pp. 149804-149813.
[33] Guihao Wang, Fengmin Wang, Jiahe Wang, Mengzhen Li, Ling Gai, & Dachuan Xu, 'Collaborative Target Assignment Problem for Large-Scale UAV Swarm Based on Two-Stage Greedy Auction Algorithm,' Aerospace Science and Technology, VOL. 149, 2024, Article 109146.
[34] 윤영혜, 'LIG넥스원, 전투용 무인수상정 기술 개발 착수,' 뉴스토마토, 2025.12.23. https://www.newstomato.com/ReadNews.aspx?no=1285742 (accessed 2026.04.01.)
[35] Yang Xiong, Shangwen Wang, Hongjun Tian, Guijie Liu, Zihao Shan, Yijie Yin, Jun Tao, Haonan Ye, & Ying Tang, 'Spatiotemporal Meta-Reinforcement Learning for Multi-USV Adversarial Games Using aHybrid GAT-Transformer,' Journal of Marine Science and Engineering, VOL. 13, NO. 8, 2025, Article 1593.