

# 수중 센서 네트워크에서 AoI 향상을 위한 강화학습 기반 작업 할당 기법 연구

이상화, 설승환, 김민호, 정재학\*  
인하대학교

sanghwa0224@naver.com, koe3400@naver.com, ho2617@naver.com  
jchung@inha.ac.kr\*

## A Study on a Reinforcement Learning-Based Task allocation Scheme for Improving Age of Information in Underwater Sensor Networks

Lee Sang-Hwa, Seol Seung-Hwan, Kim Min-Ho, Chung Jae-Hak\*  
Inha Univ

### 요약

본 논문은 AUV가 제한된 시간, 배터리 조건에서 여러 subregion을 순회하며 데이터를 수집하는 환경에서 AoI 최소화 문제를 해결하기 위해 심층 강화학습 기반 작업 할당 기법을 제안한다. 최소 보상 문제를 완화하기 위해 보상형성을 적용하였으며 시뮬레이션을 통해 제안 기법이 라운드 로빈 대비 잔여 배터리 비율을 유지하면서 평균, 최대 AoI를 개선함을 확인하였다.

### I. 서 론

해양환경 모니터링, 해저 지진 감시 등과 같은 수중 Internet of things, 수중 센서 네트워크 응용에서 관측 정보의 최신성은 시스템 성능을 좌우한다. 따라서 정보의 최신성을 정량화하는 지표인 Age of information(AoI) 기반의 운용이 요구된다. 특히 Autonomous underwater vehicle(AUV)는 능동적으로 분산된 관측 지점을 순회하며 데이터를 수집·운반함으로써 기지국에 최신 정보를 제공하고 AoI를 효율적으로 관리할 수 있다[1]. 그러나 제한된 운용시간, 에너지로 인해 고정 규칙만으로 최적의 방문 및 전달 전략을 결정하기 어렵다. 이에 본 연구에서는 다수 subregion의 평균 AoI 최소화를 목표로 하는 심층 강화학습 기반 작업 할당 기법을 제안한다.

### II. 제안 기법

본 연구에서는 단일 AUV가 제한된 시간 내  $S$  개의 subregion을 순회하며 데이터를 수집하고 이를 정지되어 있는 base station에 전달하는 시나리오를 고려한다. Subregion의 데이터는 확률적으로 생성되며, AoI는 다음과 같이 정의한다.

$$A_i(t) = t - t_i^{base} \quad (1)$$

여기서  $t_i^{base}$ 는 subregion  $i$  ( $i = 1 \sim S$ )의 데이터가 base station에 마지막으로 전달된 시각이며 AUV가 base station에 전달한 시점에서만 AoI가 생성된다. 해당 의사결정 문제를 Markov decision process로 정의한다. AUV를 에이전트로 하여 각 step에서 상태  $s_t$ 는 AoI, 배터리 잔량( $B(t)$ ), 현재 위치, 이동 시 배터리 소모량으로 구성되며, 행동  $a_t \in \{0, 1, \dots, S\}$ 은 다음 목적지를 선택하는 것으로  $a_t = 0$ 은 base station 방문,  $a_t = i$ 는  $i$  번째 subregion 방문을 의미한다. AoI가 base station에서만 생성되는 특성상 base station 방문에만 보상을 부여할 경우 보상 신호가 희소하기 때문에 보상 형성을 적용하여 효율적인 정책 학습을 유도한다. 그러므로 보상은 아래와 같이 정의한다.

$$\begin{cases} r_t = -\alpha \overline{A(t)} - \beta B(t) - \emptyset(A_{max}(t)) + \gamma I_t & (a_t = 0) \\ r_t = -\alpha \overline{A(t)} - \beta B(t) - \emptyset(A_{max}(t)) & (a_t = i) \end{cases} \quad (2)$$

여기서  $\emptyset(A_{max}(t))$ 는  $A_{max}$ 가 임계값  $A_{th}$ 를 초과할 때 페널티를 부여하는 함수이며 base station 방문시에 AoI 전달 개선 보너스( $I_t$ )를 추가한다.

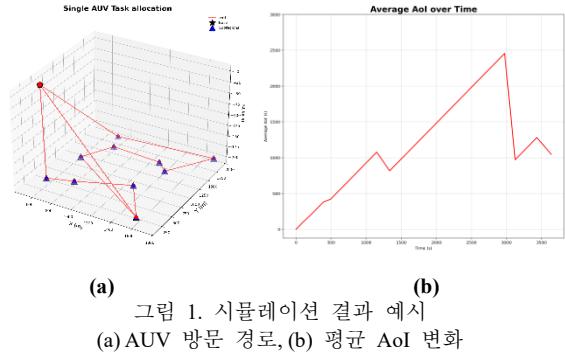


그림 1. 시뮬레이션 결과 예시  
(a) AUV 방문 경로, (b) 평균 AoI 변화

### III. 시뮬레이션 결과

본 연구에서는 각 step마다 이동시간에 다른 time step이 가변적인 환경에서 학습을 15,000 에피소드 진행하였고 100에피소드를 통해 검증을 진행하여 평균 AoI, 최대 AoI, 배터리 잔여량을 평가하였다. 그 결과, 제안 기법은 라운드 로빈 방식 대비 평균 AoI를 약 44.8%, 최대 AoI를 27.8% 개선을 하면서도 잔여 배터리 비율을 평균 약 25% 유지할 수 있음을 확인하였다.

### IV. 결론

이를 통해 동일한 에너지 소모 속에서도 정보의 적시성을 높이고, subregion의 장기 방치를 완화하는 데 효과적임을 확인하였다. 이러한 결과를 바탕으로 향후 연구에서 다중 협력 AUV 시나리오로 확장하고, 해류 예측 오차 및 통신 지연·손실을 반영한 다목적 최적화를 통해 정보의 적시성 및 에너지 효율성 향상에 기여할 수 있을 것으로 기대한다.

### ACKNOWLEDGMENT

이 논문은 2025년도 정부(방위사업청)의 재원으로 국방기술진흥연구소의 지원을 받아 수행된 연구임(KRIT-CT-23-026, 미래 기술 적응형 통합수중감시)

### 참 고 문 현

- [1] Z. Fang et al., "AoI-Inspired Collaborative Information Collection for AUV-Assisted Internet of Underwater Things", IEEE Internet of Things Journal, Vol.8, no.19, pp. 14559-14571, Oct. 2021