

계층 강화학습 기반 송신 전력 제어 슬롯형 ALOHA 프로토콜

김상호, 전용인, 유영빈, 안재하, 이호원, 유희정

고려대학교, 전자·정보공학과

kgn7392@korea.ac.kr, {yongin.jeon, youngbin.you}@lignex1.com, anjaha@add.re.kr, howon@ajou.ac.kr, heejungyu@korea.ac.kr

Hierarchical RL-Based Slotted ALOHA Protocol with Transmit Power Control

Sangho Kim, Yongin Jeon, Youngbin You, Jaha An, Howon Lee, Heejung Yu

Department of Electronics and Information Engineering, Korea University.

요약

본 논문은 Slotted ALOHA 프로토콜로 동작하는 무인항공기(Unmanned Aerial Vehicle: UAV) 군집비행 네트워크에서의 패킷 통신 충돌 확률을 줄이고, 각 UAV의 정보 도달 거리를 최대화하기 위해, 계층적 Q-learning(Hierarchical Q-learning)을 적용한 분산 제어 기반 통신 슬롯 및 송신전력 제어 알고리즘을 제안한다. 제안된 알고리즘은 외부 루프에서의 송신전력 제어와 내부 루프의 송신 슬롯 선택을 분리하여 학습함으로써, 각 UAV의 통신 환경에 따라 최적의 슬롯과 전송 반경을 선택할 수 있도록 설계되었다. 모의실험 결과 기존 방식 대비 효율적으로 전송 충돌률을 감소시켜, 슬롯 활용률과 통신 안정성이 개선됨을 확인했다.

I. 서론

무인 항공기(Unmanned Aerial Vehicle; UAV)는 초기 군사적 목적으로 개발되었으나, 최근에는 재난 감시, 물류 배송, 환경 관측 등 다양한 민간 분야에서 활용되고 있다. 특히, 복수의 UAV가 협력하여 임무를 수행하는 군집비행은 높은 기동성과 유연성을 바탕으로 효율성을 극대화할 수 있는 기술로 주목받고 있다[1], [2]. 그러나 UAV는 소형 기체의 특성상 제한된 배터리 용량과 환경 변화에 대한 민감도로 인해 통신 충돌이나 기체 간 간섭에 취약하며, 이는 군집 전체의 안정성과 임무 성공률에 부정적인 영향을 미칠 수 있다.

본 논문에서는 군집 UAV 네트워크에서 UAV간 충돌 방지를 위해서 자신의 위치 정보를 불특정 주변 UAV들에게 Slotted ALOHA 프로토콜로 전송(방송)하는 환경을 고려한다. 여기서, 각 UAV가 자율적으로 판단하여 통신 슬롯과 전송 반경, 즉 송신전력을 결정하는 분산 제어 방식을 기반으로 한다. 대표적인 분산 제어 방식 중 하나인 ALOHA는 간단한 구조를 가지고 있으나 슬롯 충돌 가능성이 높고 노드 수 증가에 따른 성공률 저하 문제가 존재한다. 이를 개선하기 위해 슬롯 기반 전송 구조인 Slotted ALOHA가 제안되었지만, 여전히 임의 슬롯 선택으로 인한 충돌 발생 가능성은 완전히 해소되지 않았다.

이에 본 논문에서는 전체 UAV의 수가 선택 가능한 slot의 수 보다 작은 환경도 송신 전력을 낮춰서 공간 재활용을 통하여 충동을 회피할 수 있는 Slotted ALOHA 학습 알고리즘을 제안하고자 한다. 즉, 다양한 네트워크 환경에서 UAV 간 통신 충돌을 최소화하고 슬롯 활용률을 개선하기 위해, UAV를 에이전트로 정의하는 계층적 강화학습 구조를 제안한다.

II. 시스템 모델

본 논문에서는 그림 1과 같이 군집비행 중인 다수의 UAV가 일정 지역 내에서 위치 정보를 주기적으로 교환하는 통신 환경을 가정한다. 각 UAV

는 비행 중 일정 시간 간격으로 자신의 위치 정보가 담긴 패킷을 주변 UAV들과 공유해야 한다. UAV 간 통신은 슬롯 기반 프레임 구조를 따르며, 프레임은 그림 2와 같이 lot 개의 슬롯으로 구성되어 있다. 두 개의 프레임이 하나의 수퍼프레임을 구성하게 된다. 각 UAV는 프레임마다 하나의 슬롯을 선택하여 자신의 패킷을 전송하며, 다른 UAV는 이 정보를 수신한다. 각 수퍼 프레임의 첫 프레임을 전송한 후 UAV는 둘째 프레임에서 첫 프레임과 동일한 슬롯을 선택하여 자신의 위치 정보와 첫 프레임의 정상 수신 여부를 알려주는 ACK 신호를 전송한다.

첫째 프레임 구간 동안 UAV i 는 자신이 성공적으로 수신한 패킷을 기반으로, 패킷 송신한 UAV의 인덱스와 해당 UAV가 사용한 슬롯 정보를 포함한 집합 $P_{rx}(i)$ 과 집합 $S_{rx}(i)$ 을 생성한다.

둘째 프레임 구간에서 각 UAV는 첫프레임과 동일한 슬롯 및 전송 반경을 갖고 자신의 위치 정보와 함께 $P_{rx}(i)$ 을 같이 전송한다. 그리고, UAV i 는 다른 UAV들로부터 수신한 $k \in i$ 인 $P_{rx}(k)$ 에 자신의 인덱스인 “ i ”가 포함된 k 들로 구성된 $ACK_{tx}(i)$ 를 구성한다. 즉, $ACK_{tx}(i) = \{k \mid i \in P_{rx}(k)\}$ 이다. 이 경우, 한 수퍼 프레임을 마친 후에 UAV i 는 자신이 전송한 패킷이 다른 수신 UAV에서 충돌이 생겼는지를 $P_{rx}(i)$ 와 $ACK_{tx}(i)$ 를 비교하여 파악할 수 있다. 즉, $P_{rx}(i) = ACK_{tx}(i)$ 이면, $P_{rx}(i)$ 의 모든 UAV들이 자신의 정보를 잘 수신했다는 것을 의미한다. 반면에, $P_{rx}(i) \neq ACK_{tx}(i)$ 이면, $P_{rx}(i)$ 에는 포함되지만, $ACK_{tx}(i)$ 에 포함되지 않은 UAV는 패킷 충돌로 인해서 UAV i 의 위치 정보를 수신하지 못한 것을 의미한다.

또한, 이와 같은 과정을 통하여 다른 UAV들이 충돌없이 사용 중인 슬롯을 확인할 수 있으며, 이 경우 학습 과정에서 해당 슬롯을 피하는 방식으로 학습의 수렴 속도를 향상 시킬 수 있을 것이다. 그리고 충돌을 완벽하게 회피하였다면 다음 에피소드에 자신이 이전 수퍼프레임에서 선택했던

슬롯을 동일하게 선택하는 정책 역시 수렴 속도 향상에 도움이 될 수 있다.

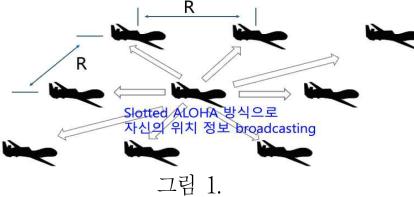


그림 1.

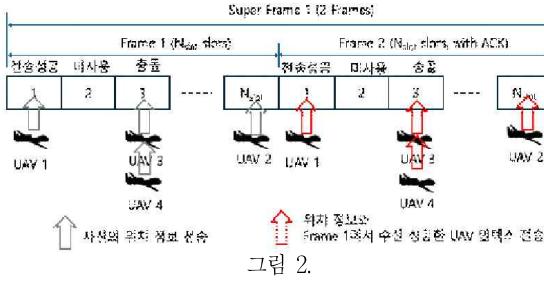


그림 2.

학습의 에이전트(agent), 상태(state), 행동(action), 정책(policy), 보상(reward)의 요소를 다음과 같이 정의했다. 여기서 t 는 현재 에피소드, 수퍼프레임을 의미하다.

-에이전트: 각의 UAV는 모두 독립된 에이전트로 동작한다.

-상태($s(t)$): 이전 에피소드에서 UAV가 선택한 행동

$$(t) = (a_{lot}(t-1), a_{range}(t-1)) \quad (1)$$

-행동($a(t)$)

·내부 루프 행동: UAV가 현재 에피소드에서 사용할 슬롯

$$a_{slot}(t) \in \{1, 2, \dots, N_{slot}\} \quad (2)$$

·외부 루프 행동: UAV가 패킷을 전송할 전송 반경

$$a_{range}(t) \in \{1R, 2R, \dots, R_{AX}\} \quad (3)$$

여기서, R_{AX} 를 최대 통신 거리로 가장 멀리 있는 UAV까지 정보를 전달할 수 있는 거리를 의미한다.

-정책: Decaying epsilon greedy 정책을 사용하였다. 해당 정책은 아래의 수식과 같이 동작한다.

$$a(t) = \begin{cases} \operatorname{argmax}_{a(t) \in A} (Q(s(t), a(t))), & \text{with probability } 1 - \epsilon(t) \\ \text{random action} & \text{, with probability } \epsilon(t) \end{cases} \quad (4)$$

$$\epsilon(t) = \epsilon_{init} d^{(t-1)} \quad (5)$$

여기서, ϵ_{init} 는 초기값을, d 는 감소 속도를 조절하는 파라미터이다.

-보상($r(t)$)

·내부 루프

$$r(t) = \begin{cases} -25P_C(t), & \text{if } P_C(t) = 0 \\ 25P_S(t), & \text{, if } P_C(t) = 0 \end{cases} \quad (6)$$

·외부 루프

$$r(t) = (25P_S(t)) - (10 \times P_C(t)) \quad (7)$$

여기서, $P_C(t)$ 와 $P_S(t)$ 는 해당 UAV가 ACK을 통해 계산한 충돌 패킷의 수와 전송 성공 패킷의 수를 각각 나타낸다. 또한 외부 루프는 일정 주기(T_{outer})를 기준으로, $T_{outer} \bmod t = 0$ 을 만족할 때만 학습을 수행하도록 설정하였다. 제안한 계층 강화학습 기반 구조의 유효성을 검증하기 위해 MATLAB을 통해 모의실험을 진행하였으며 그림3과 같은 파라미터를 사용하였고 총 50회 모의실험을 반복했다. T_{outer} 의 경우 200에서 시작하여 에피소드가 진행됨에 따라 점차 감소하여 최종적으로 2에 수렴하게 조정하였다. 또한 Slotted ALOHA 및 Q-learning 방식과의 비교를 통해 제안한 방식의 유효성을 검증하고자 했다.

Parameter	value
수퍼프레임 수 (에피소드 수)	10000
프레임 내 Slot 수 (N_{slot})	25
군집 내 UAV 수	30
$\epsilon_{slot-init}$ $\epsilon_{range-init}$	1
d_{slot}	0.999
d_{range}	0.995
α	0.1
γ	0.9

그림 3.

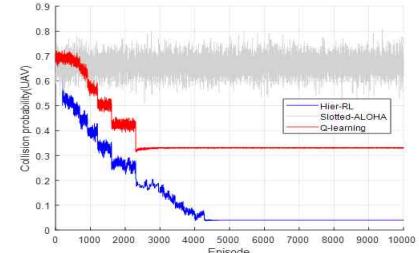


그림 4는 50회 반복 실험을 통해 얻은 평균 충돌 확률을 나타낸 것으로, 각 에피소드에서 충돌이 발생한 UAV 수를 전체 UAV 수로 나눈 값을 기반으로 충돌 확률을 정의하였다.

실험 결과 기존 Slotted ALOHA 방식은 일정 수준 이상의 높은 충돌 확률이 지속되었고, Q-learning 기법은 초기에는 충돌을 일부 완화하였으나 일정 수준 이하로 충돌을 회피하지 못했다. 그러나 제안한 알고리즘은 50회 반복에서 모든 경우 충돌을 회피하지는 못했으나, 전체적으로 가장 낮은 충돌 확률을 가지며 우수한 성능을 보였다.

III. 결론

본 논문에서는 UAV 군집비행의 안정성 향상을 위한 계층 강화학습을 적용한 Slotted ALOHA 기반 알고리즘을 제안하였다. 제안한 알고리즘은 노드 수가 많아질수록 충돌 확률이 급격히 증가하는 기존 Slotted ALOHA 방식의 한계를 개선하고자 하였으며, 모의실험 결과를 통해 해당 알고리즘이 효과적으로 동작함을 확인할 수 있었다. 그러나 실제 통신 환경에서는 다수의 UAV가 협력하여 임무를 수행하는 상황뿐만 아니라, 일부 UAV만이 독립적으로 동작하는 상황도 고려 되어야 한다. 또한, UAV의 수가 동적으로 변하거나 배치가 변화하는 등 유동적인 네트워크 환경에는 슬롯의 재할당 및 새로운 학습이 필수적이다. 향후 연구에서는 이러한 다양한 실제 적용 환경을 반영하여, 유연하고 환경 변화에 강인한 알고리즘을 제시할 계획이다.

ACKNOWLEDGMENT

본 연구는 정부(방위 사업청)의 지원으로 국방과학 연구소 연구비 지원을 받아 수행되었습니다. (UG223047VD)

This research was supported by Defense Acquisition Program Administration (DAPA) and Agency for Defense Development (ADD). (UG223047VD)

참 고 문 헌

- [1] Zhang, G., & Gu, X. (2023). "A survey on UAV-assisted wireless communications." Recent advances and future trends. Computer Communications, 208, 44-78.
- [2] Sharma, V., & Kumar, R. (2021). "Energy-efficient data collection in UAV-assisted wireless sensor networks." IEEE Transactions on Green Communications and Networking, 5(1), 25-36