# NOMA-based Random Access: Multi-Agent Reinforcement Learning Method

Yu Zhao, Jun-Bae Seo, Joohyun Lee

Hanyang University, Gyeongsang National University, Hanyang University

zhaoyu0112@hanyang.ac.kr, jbseo@gnu.ac.kr, joohyunlee@hanyang.ac.kr

# NOMA 기반 랜덤 엑세스: 다중 에이전트 강화 학습 방법으로

자오유, 서준배, 이주현

한양대학교, 경상국립대학교, 한양대학교

## Abstract

Non-orthogonal Multiple Access (NOMA) is a promising technology for improving spectral efficiency in wireless communication systems. In this paper, we employ the NOMA method to the random access (RA) communication system to enhance the sum throughput. To this end, we propose a reinforcement learning-based (RL) algorithm to obtain an optimal transmission action for each user. The simulation results show that the proposed method can significantly improve the system throughput and is much higher than the traditional methods.

## Ⅰ. Introduction

The power-domain NOMA technique has been widely used in mobile communication systems to improve spectral efficiency [1]. In [2], an uplink NOMA RA protocol was proposed for unmanned aerial vehicle communications and the throughput of the proposed algorithm can achieve 0.5869 (packets/slot). In [3], two Non-Orthogonal RA techniques were proposed for 5G mobile communication networks and the maximum throughput can exceed 0.7 (packets/slot). However, [2] and [3] only consider two power levels. To this end, multilevel target powers were considered for NOMA uplink RA systems [4]. This is a significant throughput improvement compared to the traditional RA protocol such as Slotted-ALOHA 0.368 (packets/slot).

The existing works do not solve the collision problem altogether. In this paper, we proposed a novel RL-based NOMA RA protocol, where each time slot can be fully utilized. Moreover, the maximum throughput can be up to $L$ (packets/slot) due to the NOMA technique, where $L$ denotes the number of power levels.

## Ⅱ. System Model

We consider a multi-user uplink RA system, where the time domain is divided into frames. The number of time slots in each frame is defined as the least common multiple (LCM) of $N_{min}$, $N_{min}+1,\cdots$, $N_{max}$, where $N_{min}$ and $N_{max}$ represent the minimum and the maximum number of users, respectively. A frame is further divided into a number of subframes, each of which has k time slots for $\mathrm{k} \in \{N_{min}, N_{min}+1,\cdots,N_{max}\}$. Thus, the number of subframes in a frame is $\mathrm{LCM}(N_{min}, N_{min}+1,\cdots,N_{max})/\mathrm{k}$. Each user can transmit its data packet to the $l-th$ slot of every subframe, where $l \in \{1,2,\ldots,\mathrm{k}\}$.

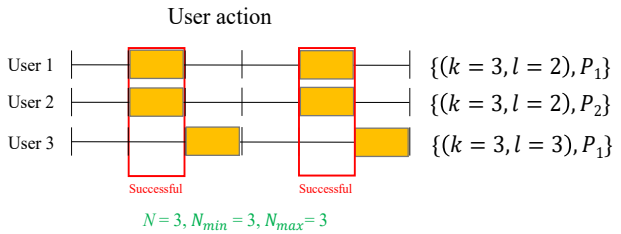


Fig. 1. Example of NOMA RA problem

In this paper, we apply the NOMA technique to the RA system. Let M denotes the number of power levels. Before transmitting the packet, each user selects a predefined power $P_i$, $i \in \{1,2,\ldots,M\}$. We assume that $P_1 > P_2 > \cdots > P_M$. Channel inversion is employed so that the received power can be the target power level and the base station (BS) decodes the received packets with the successive interference cancellation (SIC). The BS can successfully decode the received signals if the following conditions hold. First of all, a signal can be always successfully decoded if only one packet is received in a time slot with $P_i$. Secondly, if more than one packet is received in a time slot, a packet with $P_i$ can be successfully decoded if signal-to-interference noise ratio (SINR) satisfies

$$\frac{P_i}{\sum_{j=i+1}^{L} P_j \; + \; n0} \geq \gamma, \qquad (1)$$

where $n0$ and $\gamma$ represent noise power and the SINR threshold, respectively. Fig. 1 shows an example of the NOMA RA problem when $N = N_{min} = N_{max} = 3$, where $N$ denotes the number of users in the NOMA RA system. We assume that the BS decodes the received packets in decreasing order of power. In other words, the BS first decodes the packets with the highest power and the other signals treat as interference. After a packet is decoded, it is removed from the superimposed signal by SIC technology. Then, the BS decodes the packet with the second-highest power. In addition, if more than one packet is received with $P_i$, all of them cannot be decoded due to power collision, and the BS cannot decode the packet with $P_j$ for $j > i$ as well.

## Ⅲ. Reinforcement Learning Method

In this paper, we define the action $\mathrm{a}_n$ of user $\mathrm{n}$ as the slot index in a subframe and the power selected, i.e., $\{(\mathrm{k}, l), P_i\}$. Since our goal is to achieve maximum throughput, we define the instant reward function of the user $\mathrm{n}$ as

$$\mathrm{r}_n = x\mathrm{R} - y\mathrm{C}, \qquad (2)$$

where $x$ and $y$ denote the number of successes and failures in a frame, respectively. A positive reward $\mathrm{R}$ is given if a successful transmission occurs, and a negative reward (or cost) $-\mathrm{C}$ is given, otherwise.

The Upper Confidence Bound (UCB) algorithm is used in this problem [5], in which each action is assigned a confidence bonus term, as below:

$$a_n(t) = argmax_a \left\{ \bar{a}_{n,a}(t) \; + \; \delta \sqrt{\frac{\ln t}{c_{n,a}(t)}} \right\}, \qquad (3)$$

where $\bar{a}_{n,a}(t)$ and $c_{n,a}(t)$ denote the user $\mathrm{n}$'s average reward of action $\mathrm{a}$ and the number of times that action $\mathrm{a}$ is visited by user $\mathrm{n}$ up to frame $\mathrm{t}$, respectively. $\delta > 0$ represents the degree of exploration.

## Ⅳ. Simulation Results

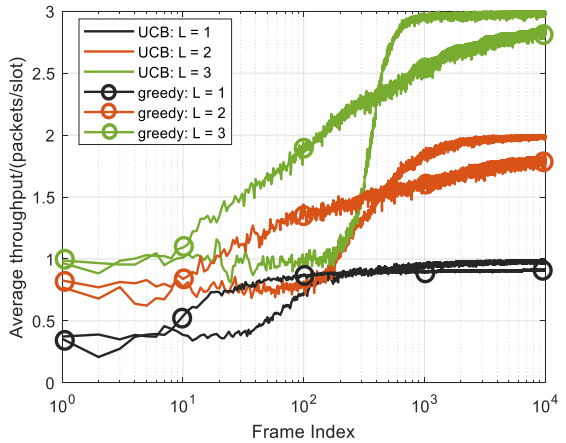


Fig. 2. Average throughput of UCB and ε-greedy algorithms.

Throughout the simulation, we set $\mathrm{R} = 1$, $\mathrm{C} = 2$, $\delta = 1$, $\gamma = 0.2$, $N = 3$, $N_{min} = 1$ and $N_{max} = 3$. In Fig. 2, we show the impact of the different power levels on the average throughput (packets/slot) of the proposed method during the learning phase. After sufficient learning, it can be seen that the throughput of the UCB algorithm is 2 (packets/slot) when $\mathrm{L} = 2$, 3 (packets/slot) if $\mathrm{L} = 3$. Therefore, we can also infer that the maximum throughput of the proposed method can be up to $\mathrm{L}$ (packets/slot). In contrast to the Slotted-ALOHA and [3], our method can significantly improve the system throughput. Also, we can see that the UCB algorithm outperforms the ε-greedy algorithm because it achieves the best trade-off between exploration and exploitation. The ε-greedy shows a low throughput because when an optimal action is found, the exploration of other sub-optimal or non-optimal actions does not stop and is still explored with probability ε.

## Ⅴ. Conclusion

In this paper, we proposed a NOMA RA protocol to improve the system sum throughput. To obtain an optimal transmission policy, we proposed the UCB-based multi-agent RL algorithm for the users to exercise. The simulation results showed that more than one packet can be transmitted per slot due to the NOMA.

## ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2021R1C1C1005126). This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.RS-2022-00155885, Artificial Intelligence Convergence Innovation Human Resources Development (Hanyang University ERICA)).

## REFERENCES

[1] Ghafoor, U., Ali, M., Khan, H. Z., Siddiqui, A. M., & Naeem, M. (2022). NOMA and future 5G & B5G wireless networks: A paradigm. *Journal of Network and Computer Applications*, 103413.

[2] Seo, J.B., S. Pack, and H. Jin. "UplinkNOMA random access for UAV-assisted communications." *IEEE Transactions on Vehicular Technology* 68.8 (2019): 8289-8293.

[3] Seo, J.B., B.C. Jung, and H. Jin. "Nonorthogonal random access for 5G mobile communication systems." *IEEE Transactions on Vehicular Technology* 67.8 (2018): 7867-7871.

[4] Seo, J.B, B. C Jung, and H. Jin. "Performance analysis of NOMA random access." *IEEE Communications Letters* 22.11 (2018): 2242-2245.

[5] Lattimore, Tor, and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.