Proceedings Paper:
Lee, Haeyoung, Lee, Sunyoung and Ko, Youngwook (2023) Multichannel Relay assisted NOMA-ALOHA with Reinforcement Learning based Random Access. In: IEEE Vehicular Technology Conference. IEEE Vehicular Technology Conference, 20-23 Jun 2023 IEEE Vehicular technology conference (VTC), ITA.

https://doi.org/10.1109/VTC2023-Spring57618.2023.10200766

Multichannel Relay assisted NOMA-ALOHA with Reinforcement Learning based Random Access

Haeyoung Lee*, Sunyoung Lee†, Youngwook Ko‡
*School of Physics, Engineering and Computer Science, University of Hertfordshire, United Kingdom
†Entrust Microgrid Ltd., United Kingdom
‡School of Physics, Engineering and Technology, University of York, United Kingdom
Email: H.Lee@herts.ac.uk, firfler@gmail.com, youngwook.ko@york.ac.uk

***Abstract*—We investigate multichannel relay assisted non-orthogonal multiple access (NOMA) in slotted ALOHA systems, where each user randomly accesses one of different channel slots and different transmit power for uplink transmissions over two-hop links, to and from the relay. By using multi-agent reinforcement learning, we propose greedy and non-greedy random access methods so that each user can learn its best strategies of random access over multiple relay slots. Random collisions and fading over the relay slots are both considered. The behaviors of relay-aided NOMA-ALOHA strategies are evaluated with the simulation. It is shown that the greedy method outperforms the non-greedy method in terms of average success rate. For deployment of relay, the greedy method benefits in improving transmission reliability under the symmetric relay channels (between the two-hop links) compared to asymmetric channels. Thus, it is interpreted that the proposed greedy method is more promising to the NOMA-ALOHA systems under a symmetric multichannel relay.**

***Index Terms*—Non-orthogonal multiple access; random access; ALOHA; relay; reinforcement learning**

I. Introduction

Due to various emerging applications such as smart factory and smart home, significant increase of Internet of Thing (IoT) connections is anticipated [1]. To support a huge increase of connections, non-orthogonal multiple access (NOMA) has been considered [2]. In NOMA, while multiple users can share the radio resource by utilizing the different transmission powers with successive interference cancellation [3].

Considering IoT transmission characteristics [4], adopting NOMA into slotted-ALOHA systems has been recently studied [5]–[8]. As the first work that applied NOMA to slotted-ALOHA, [5] studies the throughput enhancement perspective, and the analysis of throughput bounds of NOMA-ALOHA is investigated with derivation of lower bound of throughput in [6]. Since users would contend for the shared channel slots in NOMA-ALOHA, multiple studies were conducted on the channel access scheme. In [7], a non-cooperative game theory based approach is proposed to decide the mixed strategy for transmissions and a reinforcement learning (RL) based approach [9] for NOMA-ALOHA is studied in [8]. In [10], the use of RL for ALOHA systems is studied under random collision.

When NOMA is adopted to IoT transmissions, users may not be able to reach the destination directly due to connectivity deterioration. The deployment of relays can help to improve the transmission reliability as well as outage probability [11], [12]. Thus, it is worth to study on how distributed users can benefit from both NOMA and multichannel relay slots for ALOHA systems, and how they can select the shared channel slots and power level the two-hop links.

In this paper, we study the use of a multichannel relay in the NOMA-ALOHA random access system and propose the RL-based random access strategies for NOMA-ALOHA-RELAY systems. Two different power levels (High, Low) are considered for users in NOMA. A multi-armed bandit is used to model random access of distributed users in a multi-agent learning framework. In this context, we develop two RL-based random access methods: *greedy* and ϵ-*greedy* action decision algorithm. By the simulation results, we demonstrate the superiority of the proposed system to the existing ALOHA systems. In addition, it is shown that the *greedy* selection outperform the ϵ-greedy selection in terms of average success rate, and the ϵ-*greedy* selection shows severely poor performance under a highly overloaded scenario in which users achieve successful transmission rates only below 0.1. Moreover, the impact of the relay node position between users and the destination on transmission success rate is evaluated. It is found that under the symmetric relay channels (the average channel gains between two-hop links are the same), the greedy method can improve transmission success rates at most.

The remaining part of this paper is organized as follows. Section II describes the model of relay aided MIMO-ALOHA system. In Section III, the modelling of the random access with RL is explained. Then, the proposed solutions are elaborated in IV. The performance validation are explained in Section V and this paper is concluded in Section VI.

II. System model

We consider a multi-channel relay aided time-slotted random access network. Since users are assumed not to be able to reach the destination directly, they are required to access the relay with one of M different channel slots. In our model, there are K users and M different time-slotted channels for uplink transmissions aided by a single relay node.

Each user can select an random access action (i,$m$sa) including the channel $m \in \{1, \cdots, M\}$ and the power level $i \in \{\mathsf{H}, \mathsf{L}\}$. H and L indicate the transmit power P_{H} and P_{L} ($P_{\mathsf{H}} > P_{\mathsf{L}} > 0$). Thus, a set of actions is denoted as

$\mathcal{A} = \{(\mathsf{H},1),\ldots,(\mathsf{H},M),(\mathsf{L},1),\ldots,(\mathsf{L},M),0\}$, and 0 stands for no transmission. Let $h_{k;m}$ denote the channel coefficient from user k on channel m to the relay. The received signal at the relay on channel m is given by

$$y_m = \sum_{k\in\mathcal{K}_{\mathsf{H},m}} \sqrt{P_{\mathsf{H}}} h_{k;m} s_k + \sum_{k\in\mathcal{K}_{\mathsf{L},m}} \sqrt{P_{\mathsf{L}}} h_{k;m} s_k + n_m, \quad (1)$$

where $\mathcal{K}_{\mathsf{H},m}$ and $\mathcal{K}_{\mathsf{L},m}$ are the index sets of users choosing channel m with power P_{H} and P_{L}, respectively, s_k is the transmitted signal, and $n_m \sim \mathcal{CN}(0,N_0)$ is the noise of channel m. Let $\mathbb{E}[s_k]=0$ and $\mathbb{E}[|s_k|^2]=1$ for normalization.

For the link from the relay to the destination, denote by $g_{k;m}$ the channel coefficient of the relayed channel m. The signal received at the BS on relayed channel m is

$$\tilde{y}_m = \sum_{k\in\mathcal{K}_{\mathsf{H},m}} \sqrt{P_{\mathsf{H}}} g_{k;m} s_k z_k + \sum_{k\in\mathcal{K}_{\mathsf{L},m}} \sqrt{P_{\mathsf{L}}} g_{k;m} s_k z_k + w_m, \quad (2)$$

where s_k is the signal relayed on behalf of user k, $z_k \in \{0,1\}$ is a successful decoding indicator at the relay for user k, and $w_m \sim \mathcal{CN}(0,N_0)$ is the background noise of the channel m.

We assume that users do not know the channel coefficients, $h_{k;m}$ and $g_{k;m}$, thus, no power control is employed. Additionally, since each user chooses only one action at a time, the index sets, $\mathcal{K}_{\mathsf{H},m}$ and $\mathcal{K}_{\mathsf{L},m}$, become disjoint.

It is noteworthy that the proposed relay aided NORA system exploits not only multiple slotted channels of a relay but also the power differences. Provided that users cannot directly communicate with the BS, users access a relay over one of time-slotted channels which are shared by others. Moreover, when two users compete for shared channels, the use of power difference can allow two users access the same channel. The multi-user superposed transmission in the 3GPP Release 13 was introduced to enable NOMA [13]. However, the use a small number of power levels would be desirable considering the tradeoff between the cost and power efficiency, the use of two power levels is considered. In view of this, for a given generalized number K of users, each user becomes to randomly access M channels with a random choice of one of power levels. This differs from the conventional two-user NOMA application, in which the users using two power levels are determined.

III. The Random Access Modelling in Relay-aided NOMA-ALOHA

A reinforcement learning model is considered for relay-aided NOMA-ALOHA. The players, actions, rewards are defined considering random collision and channel fading on channels over two-hop links (i.e., to and from the relay).

A. Formulation of Reinforcement Learning

K-user random relaying can be formulated with three elements: the set of users, $\mathcal{K}$, the set of actions, $\mathcal{A}$, and the rewards R_k (for user k). As for the rewards, user k is assumed to choose an action of (H,m) or (L,m), which indicates that the user chooses channel m with transmit power P_{H} or P_{L}. If the transmission from this user on channel m to the relay becomes successful, the relay will choose the same action of (H,m) or (L,m) to transmit to the BS over the relay channel m. The instantaneous rewards of user k, R_k, is denoted by $V_{k;m}$ and $W_{k;m}$ for selecting (H,m) and (L,m), respectively. In particular, the rewards are calculated by three cases depending on the chosen action, i.e., $R_k(\mathsf{H},m)=V_{k;m}, R_k(\mathsf{L},m)=W_{k;m}, R_k(0)=C_0$. C_0 denotes the cost of action $i=0$, i.e., no transmission. $V_{k;m}$ and $W_{k;m}$ becomes 1 if the transmissions to the BS are successful. Otherwise, it is equal to 0. Notice that the rewards R_k do not rely on only channel fading, but also others' actions.

B. SINR under channel fading and collision

When user k chooses an action (i,m) for $i\in\{\mathsf{H},\mathsf{L}\}$, the corresponding SINR can be expressed by

$$\mathrm{SINR}_k(i,m) = \alpha_{k;m}\frac{P_i}{I_m}, \quad \text{for } i\in\{\mathsf{H},\mathsf{L}\}, \quad (3)$$

where $\alpha_{k;m}=|h_{k;m}|^2$ for channel m to the relay, or $\alpha_{k;m}=|g_{k;m}|^2$ for relay channel m to the destination. Channel gains $|h_{k;m}|$ (and $|g_{k;m}|$) to the relay (and the destination) are assumed to follow independent Rayleigh fading. In particular, we have

$$\alpha_{k;m} \sim \mathrm{Exp}(\bar{\alpha}_{k;m}), \quad (4)$$

where $\bar{\alpha}_{k;m} = \mathbb{E}[\alpha_{k;m}]$. For the sake of simple notation let $\bar{\alpha}_{k;m}$ denote $\bar{\alpha}_{k;m}$ for relay channel m to destination. Interference I_m in (3) can be defined as

$$I_m = \sum_{k'\neq k} \alpha_{k';m}\left(P_{\mathsf{H}} Z_{k';\mathsf{H},m} + P_{\mathsf{L}} Z_{k';\mathsf{L},m}\right) + N_0, \quad (5)$$

where $Z_{k;i,m}$ for $i\in\{\mathsf{H},\mathsf{L}\}$ represent the user k's activity binary indicators. If user k chooses an action (i,m), $Z_{k;i,m}$ becomes 1, otherwise is equal to 0. It is noteworthy that $\mathbb{E}[Z_{k;i,m}] = x_{k;i,m}$ and $\sum_m Z_{k;\mathsf{H},m} + Z_{k;\mathsf{L},m} \in \{0,1\}$, for given k.

Notice that even with no use of a relay, the conventional approaches [5] [7] [14] consider that users know the CSI. The knowledge of the CSI at users may allow the power control so that a required SINR for successful decoding can be obtained if no collision occurs. Unlike this, as in (1) and (2), transmit nodes (i.e., users and relay) in this work do not know the CSI and thus no power control is utilized. Accordingly, the computation of the average rewards in this work requires a consideration not only on random collision but also on fading on both users-relay slots and relay-destination slots.

C. Relay Access Rewards

For action chosen by user k, the instantaneous reward R_k can be computed in relay assisted NORA.

1) Relay-Reward with (H,m): The signal transmitted from user k can be decoded successfully only when the following conditions are all satisfied and $R_k(\mathsf{H},m)=V_{k;m}=1$.

(H1) user k is only the user choosing (H,m);

(H2) the SINR on channel m to the relay is greater than or equal to Γ_{H};

(H3) and the SINR on channel m to the destination is greater than or equal to Γ_{H}.

2) Relay-Reward with (L, m)*:* The transmission by user k can be successful only when the following conditions are all satisfied and $R_k(\mathsf{L}, m) = W_{k;m} = 1$.

(L1) user k is only the user choosing (L, m);
(L2) at most one another user $k' \neq k$, chooses (H, m);
(L3) and the signals from users k and k' (if exists) can be decodable with Γ_{L} and Γ_{H}.

IV. Proposed Machine Learning based algorithm

The reinforcement learning technique is adopted to help each user decide the best time slot and power to combat the random collisions caused by resource contending and channel fading. Each user observes the reward for the selected action at each time, and estimates the average reward value of the action indicating the probability of reliable transmission by the action. Then, the estimated reward values are exploited for action choice. We develop the action-value method based on multi-arm bandit problem which can be considered as MarKov Decision Process with a single state (or stateless). To present the action-value method, we focus on a certain user k and define the following elements:

1) the action $a = (i, m) \in \mathcal{A}$ where $\mathcal{A}$ is the set of actions;
2) the estimated mean reward of action a, $q_n(a)$, at time n;
3) the immediate reward of the action at time n, R_n;
4) the initial estimated rewards, $q_1(a) = 0$, for all $a \in \mathcal{A}$.

We consider the *sample-average method* to compute the estimated reward value. For a given arbitrary action a, the reward after at most the n-th selection of the action a can be denoted by $R_n(a)$ and its estimated mean reward after selecting $n-1$ times at most is denoted by $q_n(a)$. Then, the new average of all n rewards, $q_{n+1}(a)$ is expressed by

$$q_{n+1}(a) = \frac{1}{n}\sum_{i=1}^{n} R_i(a) = q_n(a) + \frac{1}{n}(R_n - q_n(a)), \quad (6)$$

where $\frac{1}{n}$ indicates the learning rate and $q_2(a) = R_1(a)$. It is worth to mention that $q_n(a)$ not only represents the sample average of R_n's over at most n selections of the action a, but also indicates the estimated probability of the successful transmissions with the action a. The term $R_n - q_n(a)$ in (6) expresses the estimation error. Notice that the factor scaling the gradient $\eta(n) = \frac{1}{n}$, as a step size of learning, varies from one step to another and control the learning rate. As n is larger, the impact of the error term $R_n - q_n(a)$ gets reduced, and the new estimate $q_{n+1}(a)$ becomes to rely more on the estimated mean of all $n-1$ rewards $q_n(a)$ than the immediate reward R_n. Therefore, the choice of the learning rate $\frac{1}{n}$ makes a variation of estimated rewards to be reduced over time, and guarantees a convergence to the true action value by the law of large numbers. Based on the above framework, we propose RL based random access algorithms for relay-assisted NORA systems: *greedy* and *ϵ-greedy action selection algorithm*.

Greedy algorithm is to select the action in a greedy manner, i.e., selecting the action producing the highest estimated reward at each time. For given $\mathbf{q}_n = \{q_n(a) | \ \forall a \in \mathcal{A}\}$, the action selection rule of the greedy algorithm is expressed as

$$a(n) = \underset{a \in \mathcal{A}}{\text{argmax}} \{q_n(a) |\ q_n(a) \in \mathbf{q}_n\}.$$

ϵ-greedy algorithm selects actions randomly in a while (with a probability $\epsilon \in [0, 1]$) regardless of the estimated rewards although it chooses actions greedily most of time. While the *greedy method* always selects the action to get the most reward, *ϵ-greedy method* intends to improve user's knowledge about each action by trying many actions. Two algorithms for NOMA-ALOHA-RELAY systems are explained in **Algorithm 1**. (Due to a lack of space, two algorithms are presented together). Notice that $bandit_H(a)$ and $bandit_L(a)$ (in line 8-9) will return $V_{k;i,m}$ and $W_{k;i,m}$, respectively, considering all conditions of (H1)-(H3) or (L1)-(L3) in the previous section.

Algorithm 1 RL for NOMA-ALOHA-RELAY systems

1: User k, $\forall k$, independently run the following steps.
2: Initialization, for $a = 1$ to $2M + 1$,
3: $Z(a) = V(a) = W(a) = 0$, $q(a) = 0$, where $a \in A$ and $A(a)$ denotes $A(\cdot)$ of action a
4: **procedure** RANDOM-ACCESS$(M, K, P_H, P_L, RL, \epsilon)$
5: **while** 1 **do** ▷ A loop until a convergence
6: $a \leftarrow$ SEL-ACTION(RL, ϵ)
7: user k transmits by action a, ▷ NOMA process
8: $V \leftarrow bandit_H(a)$ if $a \in \{(H, 1), \cdots, (H, M)\}$,
9: $W \leftarrow bandit_L(a)$ if $a \in \{(L, 1), \cdots, (L, M)\}$,
10: $R \leftarrow V + W$,
11: $Z(a) \leftarrow Z(a) + 1$,
12: $q(a) \leftarrow q(a) + \frac{1}{Z(a)}(R - q(a))$,
13: **procedure** SEL-ACTION(RL, ϵ)
14: if $(RL ==$ greedy$)$ $a \leftarrow \text{argmax}_{a \in \mathcal{A}}\, q(a)$
15: else if$(RL == \epsilon -$ greedy$)$
16: if$(1 - rand() < \epsilon)$ $a \leftarrow \text{argmax}_{a \in \mathcal{A}}\, q(a)$
17: else $a \leftarrow$ a random action in $\mathcal{A}$
18: Learning outcomes: $\mathbf{q_k}$

V. Simulations and Discussions

To evaluate the performance of proposed reinforcement learning driven NORA-RELAY systems, two cases of high traffic are considered: the *double-distribution* case of M channels and $K(= 2M)$ users, and the *over-distribution* case of $K(> 2M)$ users. Two power levels are exploited for NOMA users. As a performance indicator, we measure the average success rate (ASR) of actions selected by each agent indicating the transmission reliability of actions. In addition, we illustrate the sum of estimated rewards (ERs) of action at each action to address the efficacy of accumulated action decisions. The simulation parameters are summarized in Table I.

We first consider the *double-distribution* case of 8 users and 4 channels with two power level. In Fig. 1, the ASRs of two systems, the proposed RL-NOMA-ALOHA-RELAY and RL-ALOHA-RELAY, are illustrated. To focus on comparison of two systems' performance, the greedy algorithm is used

TABLE I: Simulation Parameters

Parameter	Value
Number of users and channels, $\{K, N\}$	$\{8, 4\}$ and $\{12, 48\}$
Transmit power level, $\{P_{\mathsf{H}}, P_{\mathsf{L}}\}$	$\{0.8, 0.2\}$
Channel gain, $\{\bar{\alpha}_{k;m}, \tilde{\alpha}_{k;m}\}$	$\{10, 10\}$ dB
ϵ	0.2
Simulation time (transmissions)	$\{3500, 5000\}$

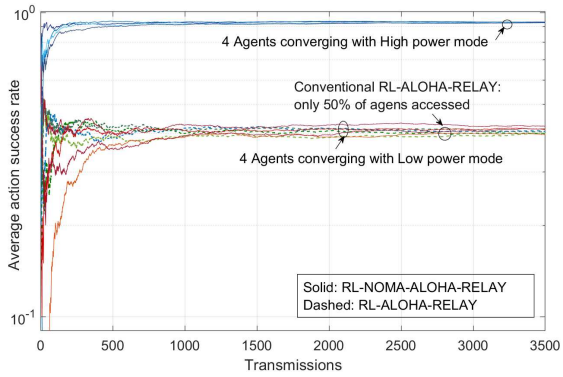


Fig. 1: The effect of NOMA in double-distribution case

for both systems. In the proposed system, users randomly select one of M channel slots and one of two power levels while users in RL-ALOHA-RELAY can randomly access one channel slot only with the high power. The ASRs of the proposed system are split with two user groups depending on the selected power level. While high power users (blue solid lines) are shown to achieve around 0.93 success rates after 1000 transmissions, low power users (red solid lines) achieve only around 0.4 success rates after 2000 transmissions. For the relaying ALOHA, only 4 users (dashed lines) are able to transmit with 0.4 success rates after 2000 transmissions.

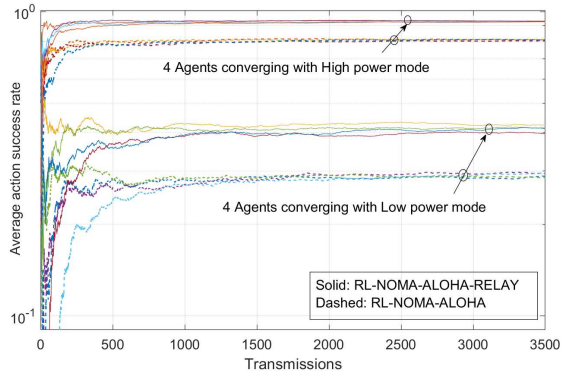


Fig. 2: The effect of relay assist in double-distribution case

When users cannot reach the destination directly, deploying a relay can extend the users' connectivity. In Fig. 2, the relay effect is illustrated by comparing the ASRs of two system, a relay assisted system and a non-relay system. It is observed that ASRs of two user groups are increased by using a relay (dashed to solid lines). Such gain is explained that employing a relay node is beneficial to combat the channel fading for both high and low power users. Actually, the relay node position can affect the system performance. In Fig. 3, the ASRs of symmetric and asymmetric relay channels are compared. For symmetric channels, the average SNR level of two-hop links (to and from the relay) is set to the same, $\{\bar{\alpha}_{k;m}, \tilde{\alpha}_{k;m}\} = \{10, 10\}$. For asymmetric channels, the average SNR of two-hop links is set differently, $\{\bar{\alpha}_{k;m}, \tilde{\alpha}_{k;m}\} = \{20, 0\}$ dB. It is observed that both high and low power users can obtain higher ASRs under the symmetric relay channels. With the assumption that the relay node is deployed much closer to users, we use the asymmetric channel setting of $\{20, 0\}$ dB. It is analyzed, by locating the relay closer to users, the impact of fading of the link from relay makes ASRs to be degraded. When relay is located closer to the destination with the setting $\{\bar{\alpha}_{k;m}, \tilde{\alpha}_{k;m}\} = \{0, 20\}$ dB, the similar result is observed (no graph due to a lack of space). Thus, it is found that the effect of relay can be the most under symmetric relay channels.

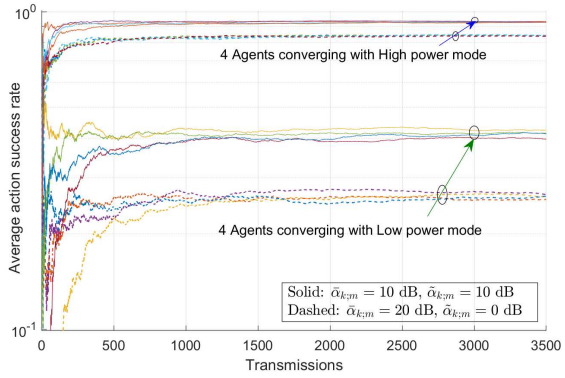


Fig. 3: The effect of a relay position with *greedy selection*

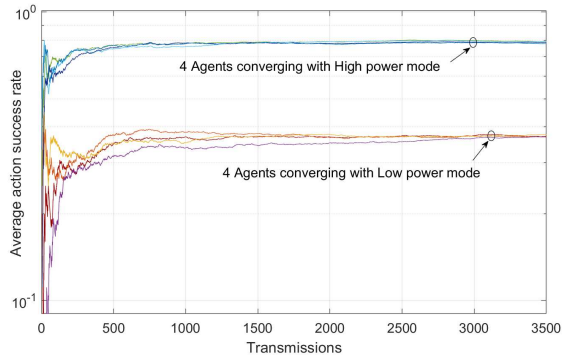
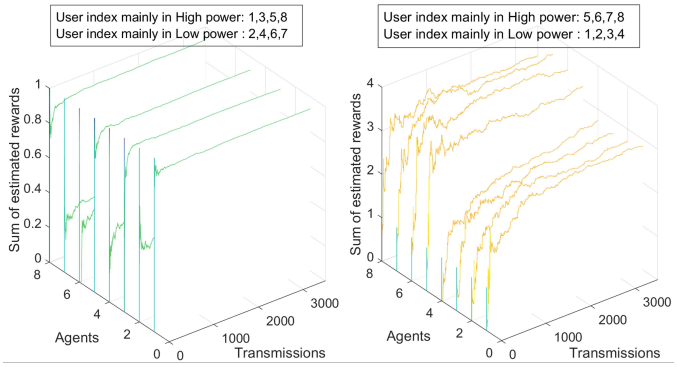


Fig. 4: ASRs of ϵ-*greedy* algorithm

We also evaluate the ASR of ϵ-*greedy* algorithm depicted in Fig. 4. Compared to the performance of *greedy* selection illustrated in Fig. 1, lower ASRs of both group users are achieved with ϵ-*greedy* selection. Moreover, ϵ-*greedy* selection needs more transmissions to reach to convergence. For example, in *greedy* selection, convergence of the high power group's ASRs occurs after 1000 transmissions, but it happens after 2000 transmissions in ϵ-*greedy* selection. It is interpreted that

random selection in ϵ-*greedy* algorithm tries to explore different actions but it actually results in decrease of transmission success rate and increases of convergence time.

The rewards of actions can be visualised by using the sum of estimated rewards (ERs) from the individual user perspective. The sum of estimated rewards is calculated over all actions at a certain time for each user and is illustrated over the transmission trials in Fig. 5. While Fig. 5(a) shows the sum of ERs of *greedy* algorithm, the sum of ERs per each user is observed to converge over transmission trials. Similar to curves shown in Fig. 1-3, the sum of ERs is grouped depending on the chosen power level. In Fig. 5(b), the sum of ERs of the ϵ-Greedy action selection is depicted. The sum of ERs of users can be grouped but a small amount of ripples is observed. Such ripple is explainable that each user selects the non-greedy actions with the probability of ϵ.



(a) Greedy selection (b) ϵ-Greedy selection

Fig. 5: Sum of ERs across users over transmission

The *over-distribution case* of $K > 2M$ is considered with the setting of $K = 48$ and $M = 12$. First, we apply the *greedy* selection and the obtained ASRs in RL-NOMA-ALOHA-RELAY are illustrated in Fig. 6. Similar to the *double-distribution* case presented in Fig. 1, the ASRs are split with two group depending on the selected power level. However, both group users can achieve much lower ASRs (the ASRs of low power group reaches to only 0.06) and each 12 user (the same number of channels M) in two power groups could have data transmission opportunities. Moreover, this *over-distributed case* needs the longer convergence time. We also evaluated ASRs of ϵ-Greedy selection even though the result is not included. While all 48 users only achieve the very low ASRs below 0.1, ASRs are shown not to be converged. It can be interpreted that non-greedy action selection in ϵ-Greedy algorithm distributes the slot access chances to users but hinders convergence. Based on observation, it is found ϵ-*greedy* selection is not effective in highly overloaded scenarios.

VI. Conclusions

We studied the multichannel relay assisted NOMA-ALOHA system where each user can randomly select one of different time slots over multiple two-hop channels and exploit power differences for uplink transmissions. We proposed the

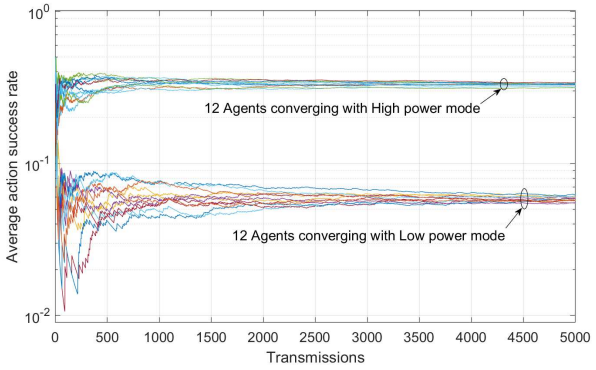


Fig. 6: ASRs for *greedy* algorithm in the over-distribution case

reinforcement learning based random access algorithms with no central cooperation, no channel information and no power control. In the multi-agent learning framework, each user can learn its own strategies to improve transmission success rates. Performance of the proposed *greedy* and ϵ-*greedy* algorithm was evaluated in terms of average success rates. *greedy* algorithm outperforms ϵ-*greedy* algorithm, and even in highly overloaded scenarios, *greedy* algorithm shows its effectiveness. For deployment of a relay, it was found that the condition of symmetric relay channels would be beneficial for the *greedy* method to maximise the effect of relay.

References

[1] S. Khairy, P. Balaprakash, L. X. Cai, and H. V. Poor, "Data-Driven Random Access Optimization in Multi-Cell IoT Networks Using NOMA," *IEEE Trans. Wirel. Commun.*, vol. 21, no. 7, pp. 4938–4953, 2022.

[2] Y. Yuan, S. Wang, Y. Wu, H. V. Poor, Z. Ding, X. You, and L. Hanzo, "NOMA for Next-Generation Massive IoT: Performance Potential and Technology Directions," *IEEE Commun. Mag.*, vol. 59, no. 7, 2021.

[3] F. Fang, J. Cheng, and Z. Ding, "Joint Energy Efficient Subchannel and Power Optimization for a Downlink NOMA Heterogeneous Network," *IEEE Trans. Veh.*, vol. 68, no. 2, pp. 1351–1364, 2019.

[4] M. Noor-A-Rahim, Z. Liu, H. Lee, M. O. Khyam, J. He, D. Pesch, K. Moessner, W. Saad, and H. V. Poor, "6G for Vehicle-to-Everything (V2X) Communications: Enabling Technologies, Challenges, and Opportunities," *Proc. IEEE*, vol. 110, no. 6, pp. 712–734, 2022.

[5] J. Choi, "NOMA-based random access with multichannel ALOHA," *J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2736–2743, Dec 2017.

[6] ——, "On Throughput Bounds of NOMA-ALOHA," *IEEE Wireless Commun. Lett.*, pp. 165–168, 2022.

[7] J. Choi, "Multichannel NOMA-ALOHA game with fading," *IEEE Trans. Commun.*, vol. 66, no. 10, pp. 4997–5007, 2018.

[8] Y. Ko and J. Choi, "Reinforcement learning for NOMA-ALOHA under fading," *IEEE Trans. Commun.*, vol. 70, no. 10, pp. 6861–6873, 2022.

[9] M. E. Morocho-Cayamcela, H. Lee, and W. Lim, "Machine Learning for 5G/B5G Mobile and Wireless Communications: Potential, Limitations, and Future Directions," *IEEE Access*, vol. 7, pp. 137 184–137 206, 2019.

[10] S. H. Park, P. D. Mitchell, and D. Grace, "Reinforcement learning based MAC protocol (UW-ALOHA-Q) for underwater acoustic sensor networks," *IEEE Access*, vol. 7, pp. 165 531–165 542, 2019.

[11] U. Uyoata, J. Mwangama, and R. Adeogun, "Relaying in the Internet of Things (IoT): A Survey," *IEEE Access*, vol. 9, 2021.

[12] A. Rauniyar, P. E. Engelstad, and O. N. Østerbø, "On the Performance of Bidirectional NOMA-SWIPT Enabled IoT Relay Networks," *IEEE Sens. J.*, vol. 21, no. 2, pp. 2299–2315, 2021.

[13] *Study on Downlink Multiuser Superposition Tramsmission (MUST) for LTE (Version 13.1.0 Release 13)*, 3GPP TR 36.869, 2015.

[14] W. Yu, C. H. Foh, A. u. Quddus, Y. Liu, and R. Tafazolli, "Throughput Analysis and User Barring Design for Uplink NOMA-Enabled Random Access," *IEEE Trans. Wirel. Commun.*, pp. 6298–6314, 2021.