A Fast Convergence ALOHA Based on Reinforcement Learning

Shuai Xiaoying, Yin Yuxia, and Zhang Bin

***Abstract*—To improve the throughput of ALOHA in Ad Hoc, we proposed a fast convergence framed ALOHA (F-ALOHA). F-ALOHA combines ALOHA with reinforcement learning to achieve an optimal way to select time slot. Q-learning is applied to update the Q-value of the slot by feedback and memory. The agents remember the number of consecutive conflicts or successes in the current slot and the idle slots in the last frame. The truncated binary exponential increase algorithm is adopted to update Q-value to accelerate convergence. The simulation results show that the average convergence time of this algorithm is significantly lower than other ALOHA algorithms, and the throughput is higher than others.**

***Index Terms*—Ad Hoc, ALOHA, reinforcement learning, time slot.**

I. Introduction

Wireless Ad Hoc is temporary dynamic topology structure using shared wireless channel to interconnect many nodes, and it doesn't rely on any fixed network infrastructure. Ad Hoc such as WSN (Wireless Sensor Network) has wide utility and has attracted the attention of researchers. MAC (Medium Access Control) is an import factor of Ad Hoc. It affects the performance of network, such as throughput, delay and so on.

Many MAC protocols for wireless networks have proposed by researchers. ALOHA is the simplest protocol but poor throughput [1]. The maximal throughput of slotted ALOHA cannot exceed 0.37 per slot [2]. CSMA (Carrier Sense Multiple Access) improves throughput by listening channel before transmission. CSMA/CA (Carrier Sense Multiple Access with Collision Avoidance) is widely used in WLAN (Wireless Local Area Network) [3]. To collision-free broadcast schedule in a fair way, TDMA (Time Division Multiple Access) is widely used in Ad Hoc [4]. However, TDMA needs to forward data such as topology, scheduling information and so on. In order to avoid possible collisions and reduce energy waste, many improvements have been proposed. RL (Reinforcement Learning) has been used for MAC protocols recently [5]-[7].

This paper proposed the F-ALOHA (Fast Convergence Framed ALOHA) combines ALOHA with RL to achieve an optimal way to select time slot. Q-Learning is integrated into FSA (Frame Slotted ALOHA) and each frame is composed of several time slots with Q-value. These are similar to Q-ALOHA (Q-learning ALOHA) [8]. Q-learning and ε-greedy are applied to framed ALOHA to intelligently select time slot. The agents remember the number of consecutive conflicts or successes in the current slot to achieve a binary exponential increase in rewards or penalties. The agent can also remember the idle slots in the current frame. If there is a conflict, these idle slots may be selected by the agent in the next frame. The simulation results show that the average convergence time of this algorithm is significantly lower than other ALOHA algorithms, and the throughput is higher than others.

Manuscript received November 30, 2020; revised March 1, 2021. This work was supported in part by the Taizhou University Foundation for the Talents QD2016035 (702065).

Shuai Xiaoying, Yin Yuxia, and Zhang Bin are with the Taizhou University, Taizhou, 225300, China (e-mail: xyshuai@163.com, shuaicz@yeah.net, 476382399@qq.com).

II. Related Work

ALOHA is widely applied to wireless networks since it was introduced. To reduce collision and improve performance, SA (Slotted ALOHA), FSA (Frame Slotted ALOHA) [9] and FLSA (Frameless Slotted ALOHA) [10], [11] have been proposed. SA is a random access technology where the time of transmission is divided into slots and the user transmits at the begin of each slot. FSA is variant of SA. FSA organizes slots into frames. Each user is allowed to randomly and independently choose only a slot to transmit packet per frame. FLSA differs from FSA in that the frame length is not set at the beginning. New slots are added until sufficiently high fraction of users has been resolved. ALOHA has the characteristics of simplicity and low overheads, but poor network performance due to transmit at any time.

For the special underwater environment, the L-ALOHA (Learning ALOHA) is proposed [12]. L-ALOHA adopts a new ACK (Acknowledgement) backoff scheme in an "infrastructure" network. Each node searches for the optimal transmission slot through continuous learning.

Reinforcement Learning is learning through trial-and-error search in a dynamic system to maximize reward [13]. The agent can discover the actions which yield the most reward by training them. In paper [14] RL is used to MAC in WSN for the optimization performance such as throughput and delay. DLMA (Deep-reinforcement Learning Multiple Access) to coexist with other protocol node in heterogeneous wireless networks is designed in [15]. Shangxing Wang *et al*. apply RL and DQN (Deep Q-Network) for dynamic multichannel access in wireless networks [16].

Q-learning is a form of model-free RL, which is widely used to find a good policy [17], [18]. Q-learning is one of the most effective and popular algorithms for learning from delayed reinforcement to determine an optimal policy, in absence of the transition probability and reward function. A Q-learning agent with a pair (S, A), where S is the set of states and A is the set of actions, which learns an expected discounted reward when an action a is taken in the states from

DOI: 10.7763/IJCTE.2021.V13.1296

the policy π. The Q-learning update the Q-value according to [19]:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[R_{t+1} + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)] \quad (1)$$

where $a \in [0, 1]$ is the learning rate, γ is the discount rate. At each time t=1, 2, …, the agent observes states $s_t \in S$, takes an action at $\in$ A. It gets a reward R_{t+1} and then moves to next state s_{t+1}. The goal of agent is to gain maximal cumulative reward.

Selahattin Kosunalp and Yi Chu applied Q-Learning to slotted ALOHA [20]. There are several time slots in a frame. Each slot has a Q-value. If a collision occurs in a certain slot, the associated reward (R) is decreased by one, and if a packet is successfully transmitted in a slot, then that R is increased by one. The Q value is updated by:

$$Q_{t+1}(i, k) = Q_t(i, k) + \alpha(R - Q_t(i, k)) \quad (2)$$

where $\alpha \in (0,1]$ is the learning rate, $R \in \{-1,1\}$ is the current reward, i is the node ID (Identity), k notes the selected time slot. Q-learning is applied to select intelligently slot. Q-ALOHA adopts exponential backoff algorithm according to Ethernet. Q-ALOHA converges to steady state after certain time to improve throughput and network performance. These technologies based on RL require a length training period to achieve steady state solution.

III. Fast Convergence ALOHA Protocol

A. Baseline of F-ALOHA

Q-Learning is integrated into frame slotted ALOHA. Each frame is composed of several time slots with Q-value. Q-value of each time slot is initialized to 0. As depicted in Fig. 1, each agent selects one of the available slots by Q-value. The agent gets environment feedback S (success, collision, other) according to the outcome of this action A (select slot). At the same time, the agent will also obtain the feedback parameters related to the current action, such as slot ID i, Q-value, and the number of consecutive operations k. The agent remembers the feedback. The agent updates Q-value by the feedback. Through the iterative learning of agents, each node gets an optimal time slot and the network is in a steady state.

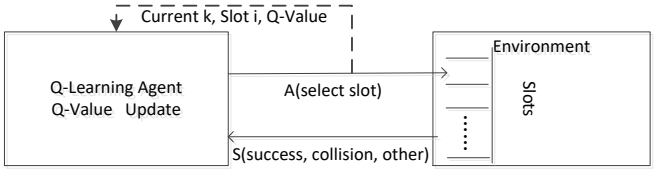


Fig. 1. Baseline of F-ALOHA.

B. Update Q-Value

If the transmission in the current slot is successful, the agent will get a reward R; if it fails, it will get a penalty R; otherwise, the Q value will remain unchanged. The Q-value update as (3):

$$Q_{t+1}(i, s) = \begin{cases} Q_t(i, s) + \gamma R & success \\ Q_t(i, s) - \gamma R & collision \\ Q_t(i, s) & otherwise \end{cases} \quad (3)$$

The R is the reward, γ is the learning rate. To accelerate convergence, R is increased by truncated binary exponential, as follow:

$$R = \begin{cases} 1 << threshold & k > threshold \\ 1 << k & otherwise \end{cases} \quad (4)$$

where k is the number of consecutive failures or successes of the current slot. If it can't continue, k is set to 0, as Fig. 2. At initialization, k and the Q-value of each slot are set to 0. In frame 1, node A and node B randomly select slot 1 and 2, while node C and node D get slot 5 and 6. Node A and node B transmit successfully. Node C and node D have failed. Node A and node B get reward 1 respectively. Node C and node D take punishment -1. The k is added by 1. In frame 2, Node B and node D have no data to send. After the frame, the k is reset to 0. Node A successfully transmitted in slot 1, so the Q-value is updated to 3 and k to 2. The transmission of node C is still conflict. The agent updated the Q-value to -3 and k to 2.

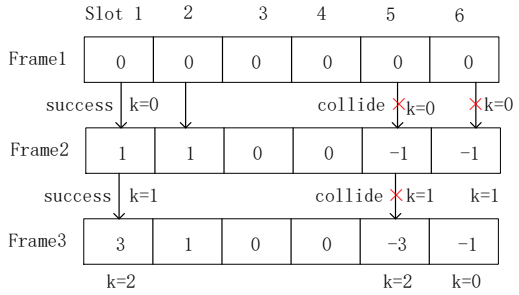
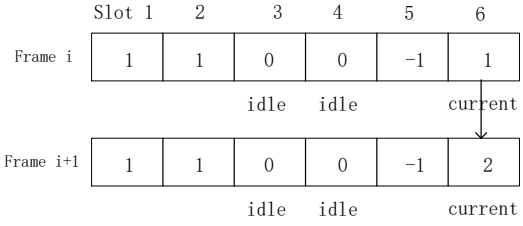


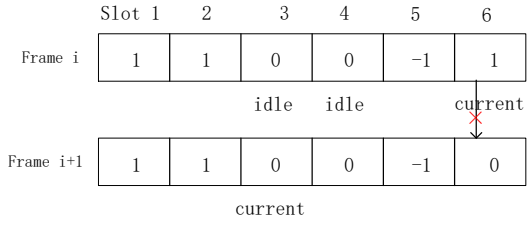
Fig. 2. Q-value update, γ=1.

C. Select Slot

F-ALOHA adopts exploration and exploitation to search the best action. If the transmission is successful, the node selects the slot with the maximum Q-value. Otherwise, using the ε-greedy policy, nodes select a random idle slot in the last frame with probability ε, select one slot with the highest Q-value with probability 1- ε. Nodes select slot as follow Fig. 3. At first, node randomly select slot 6 and transmit successfully in Fig. 3 (a). Then the node still selects the slot 6 with the maximum 2 in the frame i+1. As shown in Fig. 2 (b), the node selects the slot 6 in frame i, and a collision occurs. So, the node selects the idle slot 3 of the last frame by ε-greedy algorithm in frame i+1.



(a) Select one slot with the highest Q-value



(b) Select a random idle slot
Fig. 3. Select slot.

When each node is assigned to a different time slot, the network is in a convergent state. In this case, the algorithm is equivalent to TDMA.

D. Algorithm of F-ALOHA

Algorithm 1 Fast Convergence Aloha (F-ALOHA)

Initialize γ, ε, *threshold*, Q, k
Output: *slot*
1. select a random *slot*;
2.while not converge do
3. update Q-value (γ, *threshold*, Q, k, *slot*);
4. select slot (ε, Q, *slot*); //select slot
5. end while

Procedure update Q-value (γ, *threshold*, Q, k, *slot*)
1. if success then
2. R=1<<k; // binary exponential increase
3. Q=Q+γR;
4. compute k;
5. else
6. R=1<<k;
7. Q=Q-γR;
8. compute k;
9. end if
10. return Q

Procedure select slot (ε, Q, *slot*)
1. if collision then
2. random p
3. if p<ε then
4. select a random idle slot
5. else
6. select a random slot with highest Q-value
7. end if
8. else
9. select a random slot with highest Q-value
10. end if
11. return slot

IV. Simulation

We develop C program to run on Intel Core™ 2.6 GHz computer with 8 GB (Gigabyte) RAM (Random Access Memory) to compare the throughput and converge time of ALOHA algorithms. The number of nodes increased from 8 to 96, and the number of frames sent by each node increased from 10 to 100. We set γ to 1, ε to 0.01 and *threshold* to 3. Firstly, we test the average throughput of ALOHA, F-ALOHA and Q-ALOHA in different networks. The number of nodes in network changes from 8 to 96. As show in Fig. 4, with the increase in frame, throughput of F-ALOAH and Q-ALOHA increases respectively, but the change of ALOHA is very small. When the system is in stable state, the throughput of F-ALOAH and Q-ALOHA reaches the maximum value.

Wireless Ad Hoc has the characteristics of dynamic topology. So, the convergence rate is very important to network. Fig. 5 compares the convergence rate of F-ALOHA and Q-ALOHA. After iterative learning, agents get the best time slot and the network is steady. When the number of nodes in the network increases, iteration steps of F-ALOHA from the beginning to stability do not change significantly. However, the Q-ALOHA algorithm grows rapidly.

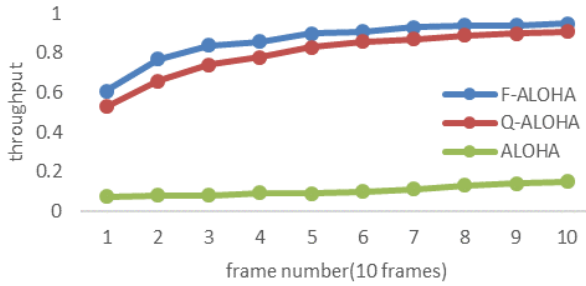


Fig. 4. Throughput comparisons.

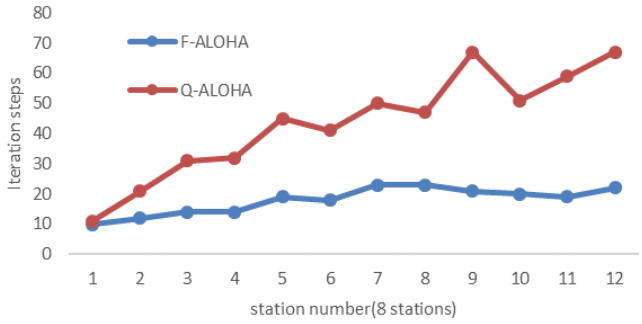


Fig. 5. Converge time comparisons.

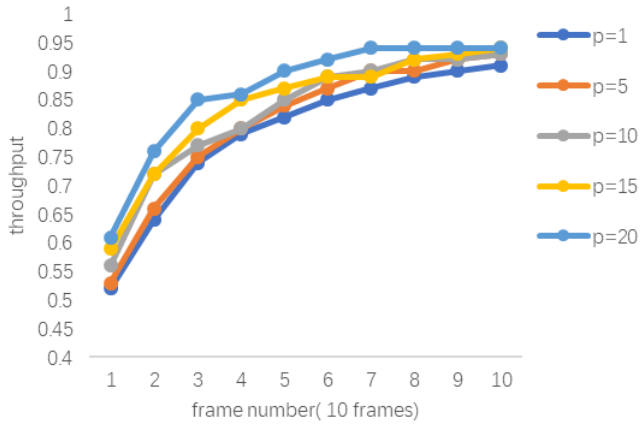


Fig. 6. Throughput with different p.

Finally, we discuss the impact of different probabilities p on F-ALOHA throughput shown in Fig. 6. In the early stage, different probabilities have a slightly greater impact on throughput, but not in the later stage. We apply ε-greedy policy to select slot. We only simulate the throughput of the algorithm under different probabilities, without considering the automatic adjustment of the probability during the operation. Ad Hoc may change dynamically with time (e.g., nodes are leaving and joining the network). Having an

adaptive ε at all time allows the decision policy to adapt to future changes.

The simulations show that the average converge time of F-ALOHA is significantly lower than that of the Q-ALOHA, and throughput of F-ALOHA is higher than of the Q-ALOHA and ALOHA.

V. Conclusion

This paper proposed a slot scheduling algorithm based on ALOHA and reinforcement learning for wireless ad hoc. Nodes intelligently select time slots through Q-learning. Agents adopt truncated binary exponential increase algorithm to compute immediate reward and adopt exploration and exploitation to search the best action. Simulations show that this algorithm is better than others aloha in throughput and convergence time. In the future, we will further study the problem of slot allocation in multi hop networks using RL to improve network performance.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Shuai Xiaoying performed the research involving existing protocols for RL for Ad Hoc, designed the algorithm of F-ALOHA, and wrote the manuscript; Yin Yuxia and Zhang Bin developed simulation programs and analyzed the results; all authors had approved the final version.

References

[1] L. G. Robert, "Aloha packet system with and without slots and capture," *ACM SIGCOMM Computer Communication Review*, vol. 5, no. 2, pp. 28–42, 1975.

[2] H. Okada, Y. Igarashi, and Y. Nakanishi, "Analysis and application of framed ALOHA channel in satellite packet switching networks-FADRA method," *Electronics and Communications in Japan*, vol. 60, pp. 72–80, 1977.

[3] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 535-547, 2000.

[4] A. Sgora, Dimitrios, J. Vergados, and D. Vergados, "A survey of TDMA scheduling schemes in wireless multihop networks," *ACM Computing Surveys*, vol. 5, pp. 1-41, 2013.

[5] N. Z. Zubir, A. F. Ramli, and H. Basarudin, "Optimization of wireless sensor networks MAC protocols using machine learning: A survey," in *Proc. International Conference on Engineering Technologies and Technopreneurship*, 2017.

[6] I. Kakalou, G. I. Papadimitriou, P. Nicopolitidis, P. G. Sarigiannidis, and M. S. Obaidat, "A reinforcement learning-based cognitive MAC protocol," in *Proc. IEEE International Conference on Communications*, 2015, pp. 7230-7234.

[7] C. Zhang, P. Patras, and H. Haddadi, "Deep learning in mobile and wireless networking: A survey," *IEEE Communications Surveys & Tutorials*, vol. 21, pp. 2224–2287, 2019.

[8] Y. Chu, "Application of reinforcement learning on medium access control for wireless sensor networks," Ph.D. dissertation, University of York, 2013.

[9] L. Zhu and T. P. Yum, "Optimal frame aloha-based anti-collision algorithm for RfiD systems," *IEEE Transactions on Communications*, vol. 58, pp. 3583–3592, 2010.

[10] C. Stefanovic, P. Popovski, and D. Vukobratovic, "Frameless ALOHA protocol for wireless networks," *IEEE Commun. Lett.*, vol. 16, pp. 2087–2090, 2012.

[11] D. Jia, Z. Fei, and M. Xiao, "Enhanced frameless slotted ALOHA protocol with Markov chains analysis," *Science China Information Sciences*, vol. 61, pp. 1-11, 2018.

[12] C. Lin, K. Chen, and E. Cheng, "A learning ALOHA protocol for underwater acoustic sensor networks," *Journal of Harbin Engineering University*, vol. 16, pp. 1-8, 2019.

[13] R. S. Sutton and A. G. Barto, *Reinforcement Learning -An Introduction*, MIT Press, 2017.

[14] Z. Liu and I. Elhanany, "RL-MAC: A QoS-aware reinforcement learning based MAC protocol for wireless sensor networks," *International Journal of Sensor Networks*, vol. 1, pp. 1-6, 2006.

[15] Y. Yu, T. Wang, and S. C. Liew, "Deep-reinforcement learning multiple access for heterogeneous wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 37, pp. 1277-1290, 2019.

[16] S. Wang, H. Liu, P. H. Gomes, and B. Krishnamachari, "Deep reinforcement learning for dynamic multichannel access in wireless networks," *IEEE Transactions on Cognitive Communications and Networking*, vol. 4, pp. 257-265, 2018.

[17] A. Pottier, P. D. Mitchell, F.-X. Socheleau, and C. Laot, "Q-learning based adaptive channel selection for underwater sensor networks," in *Proc. Underwater Communications and Networking Conference*, 2018.

[18] R. Ali, N. Shahin, Y. B. Zikria, B.-S. Kim, and S. W. Kim, "Deep reinforcement learning paradigm for performance optimization of channel observation–based MAC protocols in dense WLANs," *IEEE Access*, vol. 7, pp. 3500-3511, 2019.

[19] Y. Li. (2018). Deep reinforcement learning. [Online]. Available: https://arxiv.org/abs/1810.06339

[20] S. Kosunalp, Y. Chu, and P. D. Mitchell, "Use of Q-learning approaches for practical medium access control in wireless sensor networks," *Engineering Applications of Artificial Intelligence*, vol. 6, pp. 146-154, 2016.

Shuai Xiaoying received his BS degree in computer science from Anhui Normal University, China, in 1996, and his MS and PhD degrees in computer application technology from University of Chinese Academy of Sciences, China, in 2005 and 2010, respectively. He is a professor of Computer Science and Technology Department, Taizhou University, Jiangsu, China. His research interest is wireless network.

Yin Yuxia was born in Anhui, China. She graduated from Hefei University of Technology, China, in 2009. She majored in computer application technology. She is currently a librarian at Taizhou University.

Zhang Bin received his BS degree in communication engineering from Jiangsu University of Technology, China, in 2013; and his MS degree in computer science from Liaoning Technical University, China, in 2015. He is currently a teaching assistant at Taizhou University. His major field of study is artificial intelligence.