The calculation of the upper bound for the paper:
Deep Reinforcement Learning based MAC Protocol for Underwater Acoustic Networks

Xiaowen Ye
Xiamen University
Xiamen, China
xiaowen@stu.xum.edu.cn

Liqun Fu
Xiamen University
Xiamen, China
liqun@xum.edu.cn

I. Introduction

This is a supplementary document to the paper: Deep Reinforcement Learning based MAC Protocol for Underwater Acoustic Networks. In this paper, we give the upper bound on the network throughput when the DR-DLMA protocol coexists with other protocols, and we then use these conclusions as a benchmark for our paper. The method of proof is to refer to [1]. **We have delivered the paper, allowing reference to the following, but eliminating plagiarism.** If you want to make some valuable suggestions after reading the following derivation process, you are welcome to contact us by email.

II. coexistence with one kind of protocol

A. coexistence with TDMA

1. Coexistence with one TDMA node

Suppose there are B time-slots, TDMA node uses A time slots within B time-slots. In order to maximize the network throughput, the DR-DLMA node needs to take advantage of the available time-slots not used by the TDMA node. As a result, the optimal network throughput is 1.

2. Coexistence with N TDMA nodes

We extend the number of TDMA nodes to N in the network model. The collision among all TDMA nodes depends on their long propagation delays and transmission time-slots. The worst-case scenario is that TDMA nodes collide with each other in all transmission time-slots, i.e., their packets arrive at the AP at the same time. The best-case scenario is that they don't interfere with each other in all transmission time-slots. Therefore, the optimal network throughput is

$$f = 1 - C/B, \tag{1}$$

where C is all time-slots where collision occur, which are caused by N TDMA nodes.

B. coexistence with q-ALOHA

1. Coexistence with one q-ALOHA node

Consider that the case where one q-ALOHA node coexists with one DR-DLMA node. In a certain time slot, the average network throughput is

$$f(p) = p(1-q) + q(1-p), \tag{2}$$

where q and p is the transmission probability of the q-ALOHA node and the DR-DLMA node, respectively. Then we can know that when $q < 1/2$, the DR-DLMA node chooses to transmit in this time slot, and then the optimal network throughput is $1-q$; Otherwise the DR-DLMA node chooses to wait, and then the optimal network throughput is q.

2. Coexistence with N q-ALOHA nodes

Consider that there are $K \in \{1, 2, ..., N-1\}$ q-ALOHA nodes in the network, the transmission probability of i-th q-ALOHA node is q_i. For convenience, we use $\mathbf{q} = \{q_i, i \in 1, 2, ..., k\}$ to denote the set of transmission probabilities of N q-ALOHA nodes, which are not changed in different time slots. The average network throughput is a function of p, as follow:

$$\begin{aligned} f(p) =\ & p(1-q_1)(1-q_2)...(1-q_k)+ \\ & (1-p)\sum_{i=1}^{k}\Big[q_i \bigcap_{j=1_{j\neq i}}^{k}(1-q_j)\Big]. \end{aligned} \tag{3}$$

By deriving, we can get that when $A \geq 0$, the optimal network throughput is $(1-q_1)(1-q_2)...(1-q_k)$; Otherwise the optimal network throughput is $(\sum_{i=1}^{k} q_i) - (2\sum_{i,j=1_{i\neq j}}^{k} q_i q_j) + (3\sum_{i,j,l=1_{i\neq j\neq l}}^{k} q_i q_j q_l) - ... + [(-1)^{k+1} k q_1 q_2 ... q_k]$, where the value of A is

$$\begin{aligned} A =\ & 1 - (2\sum_{i=1}^{k} q_i) + (3\sum_{i,j=1_{i\neq j}}^{k} q_i q_j) - \\ & (4\sum_{i,j,l=1_{i\neq j\neq l}}^{k} q_i q_j q_l) - ... + [(-1)^{k+1}(k+1) q_1 q_2 ... q_k]. \end{aligned} \tag{4}$$

C. coexistence with FW-ALOHA

1. Coexistence with one FW-ALOHA node

From Fig. 1, we can deduce the stationary probability of continuous i idle-slots of one FW-ALOHA node as follow:

$$p_i = \frac{2(W-i)}{W(W+1)}, \tag{5}$$

where the W is the contention window size of the FW-ALOHA node. Particularly, $i = 0$ represents the FW-ALOHA node transmits a packet in the last time slot. Assume that when it is observed that the FW-ALOHA node stays in state i, the DR-DLMA node takes the action a_i, then the network

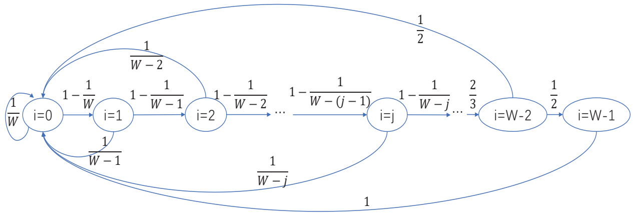

Fig. 1. Markov chain of the continuous idle slot of the FW-ALOHA node

throughput is

$$F(a_i) = \sum_{i=0}^{W-1}\left(p_i a_i (1-\frac{1}{W-i})\right) + \sum_{i=0}^{W-1}\left(p_i (1-a_i)\frac{1}{W-i}\right)$$
$$= \sum_{i=0}^{W-1}\left[\frac{2a_i(W-i-1)}{W(W+1)} + \frac{2(1-a_i)}{W(W+1)}\right] \quad (6)$$
$$= 2\sum_{i=0}^{W-1}\frac{(W-2)a_i - ia_i + 1}{W(W+1)}.$$

Now, the current goal is:

$$\begin{gathered} maximize\ F(a_i) \\ subject\ to\ 0 \le \sum_{i=0}^{W-1} a_i \le W-1, \\ a_i = 0\ or\ 1. \end{gathered} \quad (7)$$

We set $\sum_{i=0}^{W-1} a_i = j$, then

$$\sum_{i=0}^{W-1} i a_i \ge \frac{j(j-1)}{2}. \quad (8)$$

So, the objection (7) is replaced by (9).

$$\begin{gathered} maximize\ \frac{-j^2+(2W-3)j+2W}{W(W+1)} \\ subject\ to\ j \in \{0,\ 1,\ ...,\ W-2,\ W-1\}. \end{gathered} \quad (9)$$

It is a univariate quadratic function of j, and we can get the maximum value of (9) at $j = round[W-1.5]$ by derivation. That's to say, when $j = round[W-1.5]$, the network throughput will be optimal.

2. Coexistence with two FW-ALOHA nodes

Similar to only one FW-ALOHA node, when there are two FW-ALOHA nodes in the network, the Markov chain as shown in Fig. 2. We can obtain the stationary probability $p_{(i,j)}$ based on the Markov chain as follow:

$$p_{(i,j)} = \bigcap_{k=1}^{i}\left[1-\frac{1}{W-(k-1)}\right]\bigcap_{l=1}^{j}\left[1-\frac{1}{W-(l-1)}\right]p_{(0,0)},$$
$$\sum_{i=0}^{W-1}\sum_{j=0}^{W-1} p_{(i,j)} = 1. \quad (10)$$

where i and j are the number of consecutive idle time-slots for two FW-ALOHA nodes, respectively. The average network throughput is

$$F(a_{(i,j)}) = p_{(i,j)}a_{(i,j)}(1-\frac{1}{W-i})(1-\frac{1}{W-j}) + p_{(i,j)}\cdot$$
$$(1-a_{(i,j)})[(1-\frac{1}{W-i})\frac{1}{W-j} + (1-\frac{1}{W-j})\frac{1}{W-i}]$$
$$,(i,j = 0,1,...,W-1). \quad (11)$$

Further simplification, we can get

$$F(a_{(i,j)}) = p_{(i,j)}[\frac{W^2-(i+j+4)W+ij+2i+2j+3}{(W-i)(W-j)}]\cdot$$
$$a_{(i,j)} + p_{(i,j)}[\frac{1}{W-i} + \frac{1}{W-j} - \frac{2}{(W-i)(W-j)}],$$
$$i,j = 0,1,...,W-1. \quad (12)$$

where $a_{(i,j)}$ denotes the action taken by the DR-DLMA node when the DR-DLMA node observes the continuous idle time-slots of two FW-ALOHA nodes are i and j, respectively.

When $W^2-(i+j+4)W+ij+2i+2j+3 \ge 0$, the DR-DLMA node chooses to $Transmit$, i.e., $a_{(i,j)} = 1$; Otherwise $a_{(i,j)} = 0$. And then the network throughput will be optimal.

Note that because the coexistence of N FW-ALOHA nodes is difficult to discuss, we only consider the coexistence of two FW-ALOHA nodes here. If there are more FW-ALOHA nodes in the network, the formula of the optimal throughput can be derived in a similar way.

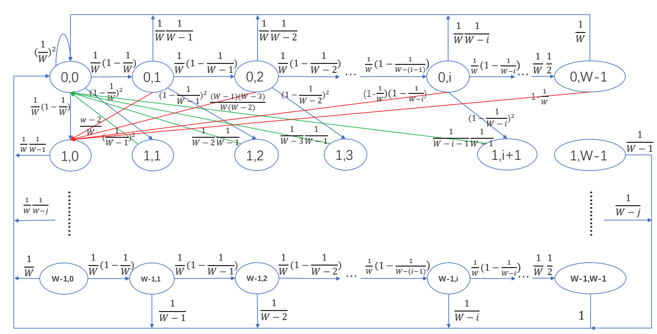

Fig. 2. Markov chain of the continuous idle slot of two FW-ALOHA nodes

III. Coexistence with two different protocols

A. coexistence with TDMA and q-ALOHA

Consider that there are M TDMA nodes, N q-ALOHA nodes, and L RD-DQN nodes in the network. Suppose the L DR-DLMA nodes are aware of each other and can be regarded as one integrated DR-DLMA node (we call it a big agent), i.e., the L DR-DLMA nodes can learn the optimal policy and then allocate the time-slots not used by M TDMA nodes and N q-ALOHA nodes internally, such as they can transmit in a round-robin manner. In the tth slot time, the total transmission probability of M TDMA nodes is p, the transmission probability of i-th q-ALOHA is q_i, and the action taken by a big agent is a_t. We can get the average network throughput is

$$F(a_t) \;=\; p\bigcap_{i=1}^{N}(1-q_i)(1-a_t)+(1-p)\bigcap_{i=1}^{N}(1-q_i)a_t+$$
$$(1-p)\sum_{i=1}^{N}\big[q_i\bigcap_{j=1_{j\neq i}}^{N}(1-q_j)\big](1-a_t). \tag{13}$$

Then, we can know the optimal network throughput is

$$(1-p)\bigcap_{i=1}^{N}(1-q_i),\; if\;\; A\geq 0;$$
$$p\bigcap_{i=1}^{N}(1-q_i)+(1-p)\sum_{i=1}^{N}\big[q_i\bigcap_{j=1_{j\neq i}}^{N}(1-q_j)\big],\; otherwise, \tag{14}$$

where the value of A is

$$A \;=\; (1-p)\bigcap_{i=1}^{N}(1-q_i)-$$
$$(1-p)\big[\sum_{i=1}^{N}q_i-2\sum_{i=1}^{N}\sum_{j=i+1}^{N}q_iq_j+...+(-1)^{N+1}\bigcap_{i=1}^{N}q_i\big]. \tag{15}$$

That's to say, when $A\geq 0$, the DR-DLMA node chooses to transmit in each slot time; Otherwise, the DRL node should stay in waiting. And then the network throughput will be optimal.

B. coexistence with FW-ALOHA and TDMA

Assume that there are one FW-ALOHA node, M TDMA nodes, and L DR-DLMA nodes (we think of them a big agent as in III-A). Based on (13), we can get the average network throughput as follow:

$$\begin{aligned} F(a_i) = \sum_{i=0}^{W-1}\big[\frac{2a_i(W-i-1)}{W(W+1)}(1-p)+\\ \frac{2(1-a_i)}{W(W+1)}(1-p)+\frac{2(1-a_i)(W-i-1)}{W(W+1)}p\big], \end{aligned} \tag{16}$$

where p is the total transmission probability of M TDMA nodes. Using (8), then our objection becomes

$$\begin{aligned} maximize\;\Bigg(\frac{(2p-1)j^2+[2W-3-4p(W-1)]j}{W(W+1)}+\\ \frac{2W+pW^2-3pW}{W(W+1)}\Bigg)\\ subject\;to\; j\in\{0,\;1,\;...,\;W-2,\;W-1\}. \end{aligned} \tag{17}$$

When $0\leq p\leq 1/2$, we can get the optimal network throughput as in (18), otherwise the optimal network throughput can be obtained at $j=0$.

$$j = round[\frac{2W-3-4p(W-1)}{2(1-2p)}]. \tag{18}$$

References

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