

Model-Aware Nodes in Heterogeneous Networks:

A Supplementary Document to Paper “Deep-Reinforcement Learning Multiple Access for Heterogeneous Wireless Networks”

Yiding Yu, Taotao Wang, Soung Chang Liew
Department of Information Engineering, The Chinese University of Hong Kong
Email: {yy016, ttw, soung}@ie.cuhk.edu.hk

This document derives the optimal sum/individual throughputs that can be attained by a model-aware node when it coexists with TDMA networks, ALOHA networks and a mix of TDMA and ALOHA networks. This is a supplementary document to our journal paper: the paper puts forth a model-free wireless MAC protocol that leverages the techniques of deep reinforcement learning (DRL) for heterogeneous networking. The throughputs achievable by the model-aware protocols, derived here, serve as benchmarks for the DRL model-free protocol described in the paper.

Part 1 of this document derives the optimal sum and individual throughputs when the model-aware node aims to maximize the sum throughput of all nodes. The optimal results in Part 1 also serve as benchmarks for our conference paper. Part 2 of this document derives the individual throughputs of different nodes when the model-aware node aims to achieve proportional fairness among all nodes, as a representative example of the general α -fairness.

In addition, some extra discussions on our future work are provided in Part 3 of this document.

Part 1: Maximizing Sum Throughput

▪ Coexistence with a TDMA node

We first consider the coexistence of a model-aware node and a TDMA node. Thus, there are two nodes in the network. The TDMA node transmits in X specific slots within each frame of K slots in a repetitive manner. In this case, the model-aware node is a TDMA-aware node that has full knowledge of the X slots used by the TDMA node. To maximize the sum throughput, the TDMA-aware node will transmit in all the slots not used by the TDMA node and refrains from transmission in the slots occupied by the TDMA node. Thus, the optimal total throughput is 1.

▪ Coexistence with (N-1) q-ALOHA nodes

We next consider the coexistence of a model-aware node with $(N - 1)$ q -ALOHA nodes. The total number of nodes in the network is therefore N . The q -ALOHA node transmits with a fixed transmission probability q in each time slot. The transmission probabilities of different time slots are i.i.d. The model-aware node here is referred to as a q -aware node. It knows the value q as well as N . Given the q -ALOHA nodes transmit in time slots in an i.i.d. manner, the transmission strategy of the q -aware node should also be identical and independent in all the time slots. Consider a particular time slot, let p be the transmission probability for the q -aware node and $f(p)$ be the associated sum throughput. It is given by

$$f(p) = (N - 1)q(1 - q)^{N-2}(1 - p) + p(1 - q)^{N-1} \quad (1)$$

The derivative of $f(p)$ with respect to p is

$$f'(p) = -(N-1)q(1-q)^{N-2} + (1-q)^{N-1} \quad (2)$$

From the derivative $f'(p)$, we can see when $q > 1/N$, $f'(p) < 0$; when $q < 1/N$, $f'(p) > 0$. Therefore, to maximize $f(p)$, the optimal transmission probability for the q -aware node is

$$p^* = \begin{cases} 0 & \text{if } q > 1/N \\ 1 & \text{if } q \leq 1/N \end{cases} \quad (3)$$

or

$$p^* = \begin{cases} 0 & \text{if } q \geq 1/N \\ 1 & \text{if } q < 1/N \end{cases} \quad (4)$$

The corresponding optimal total throughput $f(p^*)$ is

$$f(p^*) = \begin{cases} (N-1)q(1-q)^{N-2} & \text{if } q \geq 1/N \\ (1-q)^{N-1} & \text{if } q < 1/N \end{cases} \quad (5)$$

■ Coexistence with a fixed-window ALOHA node

Here, we consider the coexistence of a model-aware node with a fixed-window ALOHA (FW-ALOHA) node. Thus, there are two nodes in the network. The FW-ALOHA node generates a random backoff counter value c in the range of $[0, W-1]$ after it transmits in a time slot. It then waits for c slots before its next transmission. The value of c is decremented at the end of each time slot. When c reaches 0, the FW-ALOHA node transmits in the next time slot. The model-aware node is named FW-aware node and it knows the transmission scheme of the FW-ALOHA node. Also, the FW-aware node knows the value of W and whether the FW-ALOHA node transmitted in past time slots by listening to the channel.

A subtlety about model-aware nodes needs to be pointed out here. Being aware of the MAC protocol used by the other node is different from knowing the instantaneous state of the MAC of the other node. In this case, although the FW-aware node knows the other node uses FW-ALOHA, it does not know the backoff counter value c of the FW-ALOHA node most of the time, except at the end of a time slot in which the FW-ALOHA node just transmitted (*the fact that $c=0$ in that time slot could be derived from listening to the channel (in case the FW-aware node did not transmit) or from the lack of an ACK from the receiver of the FW-aware node (in case the FW-aware transmitted and there was a collision)*). In particular, the FW-aware node does not know the exact c value at the beginning of a time slot, although it knows its range. For example, if the FW-aware node notices that the FW-ALOHA node did not transmit in the past n time slot, then it knows that the value of c in the next time slot is equally likely to be $0, 1, \dots, (W-1-n)$.

To maximize the sum throughput, the FW-aware node has two optimal strategies that draw on results (3) and (4) of the q -ALOHA case in the previous section:

❖ Optimal Strategy 1

In both optimal strategies, the time slots immediately following a transmission by the FW-ALOHA node serve as “renewal points”. In optimal strategy 1, after observing a transmission of the FW-ALOHA node in a particular time slot, the FW-aware node transmits in all the subsequent time slots except after observing consecutive $(W - 1)$ idle slots of the FW-ALOHA node (i.e., the FW-ALOHA node did not transmit in the next $(W - 1)$ slots after its previous transmission), in which case the FW-aware node refrains from transmission in the next time slot. If the FW-ALOHA node transmits again before $(W - 1)$ idle time slots (this leads to a collision which we assume to be known by the FW-aware node through feedback mechanisms from its receiver; i.e., the lack of an ACK), then the time slot following the new transmission is a renewal point for the above strategy of the FW-aware node.

The reason the above strategy is optimal is as follows. Immediately after a transmission by the FW-ALOHA node, the probability of the FW-ALOHA node transmitting in the next time slot is $1/W$ (i.e., it will transmit only when the counter value c is 0). This is equivalent to the single q -ALOHA case where $q = 1/W$ (with $N=2$ in (1)). Thus, if $W \geq 2$, the FW-aware node should transmit (see (3)). If the FW-ALOHA node does not transmit, the probability it will transmit in the time slot after next is $1/(W - 1)$. Again, according to (1), the FW-aware node should transmit unless $1/(W - 1) = 1$. Reasoning it this way, we see that, the FW-aware node will only transmit after it observes $(W - 1)$ idle time slots of the FW-ALOHA node.

The individual throughputs of the FW-ALOHA node and the FW-aware node, and the sum throughput of the FW-ALOHA node and the FW-aware node can be derived using a renewal process. We present the derivation as follows.

We define one round as the number of slots it takes for the FW-ALOHA node backoff counter to go to zero. We have

$$E[\text{number of slots in one round}] = \frac{1}{W} [1 + 2 + \dots + (W - 1) + W] = \frac{W + 1}{2} \quad (6)$$

$$E[\text{number of total success slots in one round}] = \frac{1}{W} [0 + 1 + 2 + \dots + (W - 2) + W] = \frac{W^2 - W + 2}{2W} \quad (7)$$

$$E[\text{number of FW-ALOHA success slots in one round}] = \frac{1}{W} \cdot 1 = \frac{1}{W} \quad (8)$$

$$E[\text{number of FW-aware success slots in one round}] = \frac{1}{W} [0 + 1 + 2 + \dots + (W - 2) + (W - 1)] = \frac{W - 1}{2} \quad (9)$$

The throughputs are therefore as follows:

$$\text{Sum throughput: } \frac{(7)}{(6)} = \frac{W^2 - W + 2}{W(W + 1)}$$

$$\text{FW-ALOHA throughput: } \frac{(8)}{(6)} = \frac{2}{W(W + 1)}$$

$$\text{FW-aware throughput: } \frac{(9)}{(6)} = \frac{W - 1}{W + 1}$$

❖ Optimal Strategy 2

In optimal strategy 2, after observing a transmission of the FW-ALOHA node, the FW-aware node transmits in all subsequent time slots except after observing consecutive $(W - 2)$ idle slots of the FW-ALOHA node, in which case the FW-aware node refrains from transmission in the next 2 time-slots. The difference between this strategy and the previous strategy is that the previous strategy follows (1) while this strategy follows (4).

Similar to optimal Strategy 1, these throughputs of optimal strategy 2 are also derived using a renewal process as follows:

$$E[\text{number of slots in one round}] = \frac{1}{W} [1 + 2 + \dots + (W - 1) + W] = \frac{W + 1}{2} \quad (10)$$

$$E[\text{number of total success slots in one round}] = \frac{1}{W} [0 + 1 + 2 + \dots + (W - 3) + (W - 1) + (W - 1)] = \frac{W^2 - W + 2}{2W} \quad (11)$$

$$E[\text{number of FW-ALOHA success slots in one round}] = \frac{1}{W} \cdot 2 = \frac{2}{W} \quad (12)$$

$$E[\text{number of FW-aware success slots in one round}] = \frac{1}{W} [0 + 1 + 2 + \dots + (W - 2) + (W - 2)] = \frac{W^2 - W - 2}{2W} \quad (13)$$

The throughputs are therefore as follows:

$$\text{Sum throughput: } \frac{(11)}{(10)} = \frac{W^2 - W + 2}{W(W + 1)}$$

$$\text{FW-ALOHA throughput: } \frac{(12)}{(10)} = \frac{4}{W(W + 1)}$$

$$\text{FW-aware throughput: } \frac{(13)}{(10)} = \frac{W - 2}{W}$$

We note that although the two optimal strategies of the FW-aware node are different, they achieve the same sum throughputs.

▪ Coexistence with an exponential-backoff ALOHA node

Now, we consider the coexistence of a model-aware node and an exponential-backoff ALOHA (EB-ALOHA) node. Thus, there are two nodes in the network. EB-ALOHA is a variant of FW-ALOHA. Specifically, after each packet transmission without collision, an EB-ALOHA node randomly chooses a backoff counter value c in the range of $[0, W - 1]$. EB-ALOHA node doubles its window size each time when its transmission encounters a collision, until a maximum window size $2^m W$ is reached, where m is called the “maximum backoff stage”. Upon a successful transmission, the window size is reset to the initial value W .

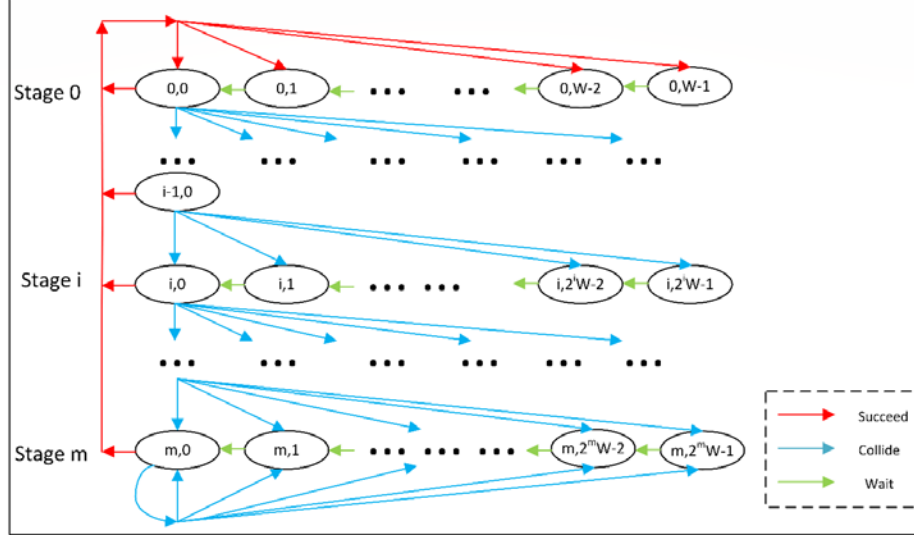


Fig. 1 Markov chain for backoff process of the EB-ALOHA node.

We define the *stage* of the EB-ALOHA node to be the number of previous collisions experienced by the current packet, i.e., stage 0, stage 1, ..., stage m . We define the *state* of the EB-ALOHA node to be (i, j_i) , where $i = 0, 1, 2, \dots, m$ is the backoff stage and $j_i \in [0, 2^i W - 1]$ is the backoff counter value. The backoff process of the EB-ALOHA node can be modelled by a two-dimensional Markov chain as in Fig. 1 (the transition probabilities are omitted in the figure). The red arrows correspond to the EB-ALOHA node transmitting successfully and its random selection of a new backoff counter value in the range of $[0, W - 1]$; the blue arrows correspond to the EB-ALOHA node encountering a collision and its random selection of a backoff counter value in the next stage (or in the last stage if the state number is already m); the green arrows correspond to the EB-ALOHA node not transmitting and decrementing its backoff counter by 1.

The mode-aware node in this case is called the EB-aware node and it knows the transmission scheme of the EB-ALOHA node, i.e., it knows that the EB-ALOHA node will double its window size upon a collision. Also, the initial window size W and the maximum backoff stage m are known to the EB-aware node. The EB-aware node knows whether the EB-ALOHA node transmitted in past time slots by listening to the channel (when the EB-aware node did not transmit) and from the lack of ACK from its receiver (when the EB-aware node transmitted and there was a collision). Thus, the EB-aware node knows the stage of the EB-ALOHA node by counting the number of collisions the current EB-ALOHA packet has encountered. Similar to the FW-aware node, the EB-aware node has no knowledge of the backoff counter value c of EB-ALOHA node at each stage, although the EB-aware node knows the stage value.

One possible strategy for the EB-aware node is to behave like the FW-aware node, i.e., after observing a transmission of the EB-ALOHA node in a particular time slot, the EB-aware node knows the stage of the EB-ALOHA node. If there was no collision, the EB-ALOHA node will be in stage 0; if there was a collision, the stage of the EB-aware node will increase by 1. If the EB-ALOHA is in stage i , then its counter value at that moment in time is equally likely to be $0, 1, \dots, 2^i W - 1$. The EB-aware node could behave like a FW-aware node when the FW-aware node interacts with a FW-ALOHA node with window size $2^i W$: i.e., the EB-aware node will transmit in all subsequent time slots except after $2^i W - 1$ idle time slots of the EB-ALOHA node.

It turns out that the optimal strategy is not always given by such dynamic emulation of the EB-aware node that depends on i . To see this, consider the case with a large m . A better strategy could be to cause lots of collisions with the EB-ALOHA node so that it seldom transmits. In particular, the EB-aware node could transmit greedily in all time slots, causing the EB-ALOHA nodes to get stuck in stage m . When m is large, the EB-ALOHA node then seldom transmits. The success probability of the EB-aware node will be close to 1 when m is large, allowing the throughput to approach the upper bound of 1. In general, the optimal strategy depends on m and W .

We now analyze the optimal strategy for the EB-aware node when $m = 2$ (used in our experiment). After a transmission of the EB-ALOHA node, the EB-aware node knows the EB-ALOHA node is at stage i . Let $g(i)$ be the number of observed consecutive idle slots of EB-ALOHA node at stage i ($g(i) \in [0, 2^i W - 1]$) after the EB-ALOHA node moves to stage i . When $g(i) < 2^i W - 1$, EB-aware node always transmits in the next time slot (as in optimal strategy 1 of the FW-aware node). When $g(i) = 2^i W - 1$ at the end of a time slot, EB-aware node needs to decide to transmit or not to transmit in the next slot. When $m = 2$, there will be 8 possible strategies for EB-aware node as summarized in Table 1 and the optimal strategy is one of the strategies. Next, we analyze the 8 strategies to find out the optimal strategy.

Table 1 Strategies for the EB-aware node

Stage i	$g(i)$	After observing $g(i)$ idle slots, transmit or not (EB-aware node)?							
0	$W-1$	No	Yes	No	Yes	No	Yes	No	Yes
1	$2W-1$	No	No	Yes	Yes	No	No	Yes	Yes
2	$4W-1$	No	No	No	No	Yes	Yes	Yes	Yes
Strategy		NNN	YNN	NYN	YYN	NNY	YNY	NYY	YYY

First, we note that there is no difference between strategies NNY, YNY, NYY and YYY over the long term, because once the system moves to stage 2, it will never escape from stage 2 and this corresponds to the FW-ALOHA case with a window size of $4W$. We use Strategy xxY to represent these four strategies.

Next, we examine strategy xxY , together with the remaining four strategies: NNN, YNN, NYN and YYN.

❖ Strategy xxY

As analyzed above, Strategy xxY causes EB-ALOHA node to behave like an FW-ALOHA node with a window size of $4W$. Similarly, we can use a renewal process to analyze the throughputs as follows:

$$E[\text{number of slots in one round}] = \frac{1}{4W} (1 + 2 + \dots + (4W - 1) + 4W) = \frac{4W + 1}{2} \quad (14)$$

$$E[\text{number of EB-aware success slots in one round}] = \frac{1}{4W} (0 + 1 + 2 + \dots + (4W - 2) + (4W - 1)) = \frac{4W - 1}{2} \quad (15)$$

The throughputs are therefore as follows:

$$\text{Sum throughput: } \frac{4W - 1}{4W + 1}$$

EB-ALOHA throughput: 0

$$\text{EB-aware throughput: } \frac{(12)}{(11)} = \frac{4W - 1}{4W + 1}$$

❖ Strategy NNN

To analyze Strategy NNN, we construct a 3-state (state 0, state 1 and state 2) coarse Markov chain and “state” here corresponds to “stage” in Fig. 1. By saying “coarse”, we mean we do not care how many time slots are spent in each stage. The coarse Markov chain is illustrated in Fig. 2, where p_0 , p_1 and p_2 denote the stationary probability of EB-ALOHA node at state 0, state 1 and state 2, respectively. The transition probabilities (as a function of initial window size W) are given in Fig. 2.

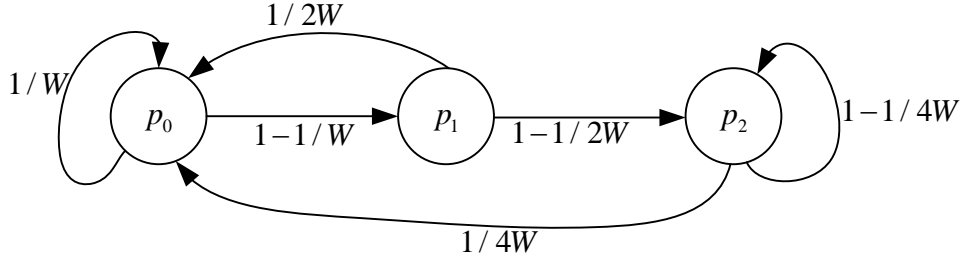


Fig. 2 Coarse Markov Chain for strategy NNN.

We can calculate p_0 , p_1 and p_2 as follows:

$$\left\{ \begin{array}{l} \frac{1}{W} p_0 + \frac{1}{2W} p_1 + \frac{1}{4W} p_2 = p_0 \\ \left(1 - \frac{1}{W}\right) p_0 = p_1 \\ \left(1 - \frac{1}{2W}\right) p_1 + \left(1 - \frac{1}{4W}\right) p_2 = p_2 \\ p_0 + p_1 + p_2 = 1 \end{array} \right. \Rightarrow p_0 = \frac{W}{(2W-1)^2}, p_1 = \frac{W-1}{(2W-1)^2}, p_2 = \frac{2(W-1)}{2W-1} \quad (16)$$

Then, the throughputs can also be derived using a renewal process (The results in (3)-(6) are used in the following derivation):

$$E[\text{number of slots in one round}] = p_0 \cdot \frac{W+1}{2} + p_1 \cdot \frac{2W+1}{2} + p_2 \cdot \frac{4W+1}{2} \quad (17)$$

$$E[\text{number of total success slots in one round}] = p_0 \cdot \frac{W^2 - W + 2}{2W} + p_1 \cdot \frac{(2W)^2 - (2W) + 2}{2(2W)} + p_2 \cdot \frac{(4W)^2 - (4W) + 2}{2(4W)} \quad (18)$$

$$E[\text{number of EB-ALOHA success slots in one round}] = p_0 \cdot \frac{1}{W} + p_1 \cdot \frac{1}{2W} + p_2 \cdot \frac{1}{4W} \quad (19)$$

$$E[\text{number of EB-aware success slots in one round}] = p_0 \cdot \frac{W-1}{2} + p_1 \cdot \frac{2W-1}{2} + p_2 \cdot \frac{4W-1}{2} \quad (20)$$

The throughputs are therefore as follows:

$$\text{Sum throughput: } \frac{(18)}{(17)}$$

$$\text{EB-ALOHA throughput: } \frac{(19)}{(17)}$$

$$\text{EB-aware throughput: } \frac{(20)}{(17)}$$

The derivations of strategies YNN, NYN and YYN are similar to strategy NNN. We only give the coarse Markov chain and the final results for each strategy below without detailed description.

❖ Strategy YNN

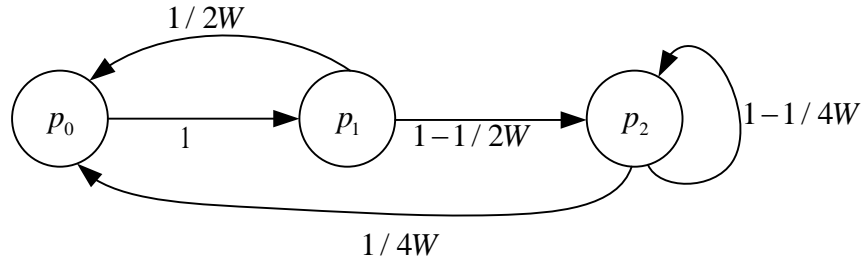


Fig. 3 Coarse Markov Chain for strategy YNN.

$$p_0 = \frac{1}{4W}, p_1 = \frac{1}{4W}, p_2 = \frac{2W-1}{2W}$$

$$E[\text{number of slots in one round}] = p_0 \cdot \frac{W+1}{2} + p_1 \cdot \frac{2W+1}{2} + p_2 \cdot \frac{4W+1}{2} \quad (21)$$

$$E[\text{number of total success slots in one round}] = p_0 \cdot \frac{W-1}{2} + p_1 \cdot \frac{(2W)^2 - (2W) + 2}{2(2W)} + p_2 \cdot \frac{(4W)^2 - (4W) + 2}{2(4W)} \quad (22)$$

$$E[\text{number EB-ALOHA success slots in one round}] = p_1 \cdot \frac{1}{2W} + p_2 \cdot \frac{1}{4W} \quad (23)$$

$$E[\text{number of EB-aware success slots in one round}] = p_0 \cdot \frac{W-1}{2} + p_1 \cdot \frac{2W-1}{2} + p_2 \cdot \frac{4W-1}{2} \quad (24)$$

The throughputs are therefore as follows:

$$\text{Sum throughput: } \frac{(22)}{(21)}$$

EB-ALOHA throughput: $\frac{(23)}{(21)}$

EB-aware throughput: $\frac{(24)}{(21)}$

❖ **Strategy NYN**

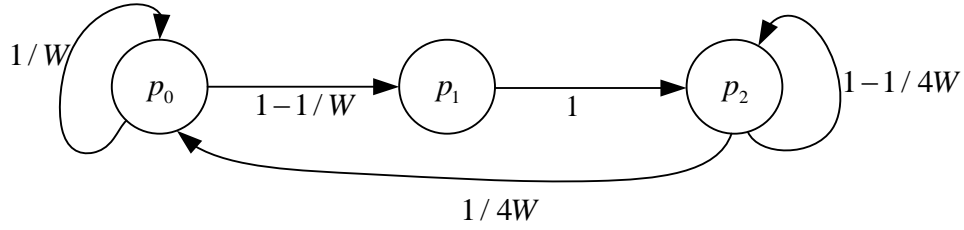


Fig. 4 Coarse Markov Chain for strategy NYN.

$$p_0 = \frac{W}{4W^2 - 2W - 1}, p_1 = \frac{W - 1}{4W^2 - 2W - 1}, p_2 = \frac{4W^2 - 4W}{4W^2 - 2W - 1}$$

$$E[\text{number of slots in one round}] = p_0 \cdot \frac{W + 1}{2} + p_1 \cdot \frac{2W + 1}{2} + p_2 \cdot \frac{4W + 1}{2} \quad (25)$$

$$E[\text{number of total success slots in one round}] = p_0 \cdot \frac{W^2 - W + 2}{2W} + p_1 \cdot \frac{2W - 1}{2} + p_2 \cdot \frac{(4W)^2 - (4W) + 2}{2(4W)} \quad (26)$$

$$E[\text{number EB-ALOHA success slots in one round}] = p_0 \cdot \frac{1}{W} + p_2 \cdot \frac{1}{4W} \quad (27)$$

$$E[\text{number of EB-aware success slots in one round}] = p_0 \cdot \frac{W - 1}{2} + p_1 \cdot \frac{2W - 1}{2} + p_2 \cdot \frac{4W - 1}{2} \quad (28)$$

The throughputs are therefore as follows:

Sum throughput: $\frac{(26)}{(25)}$

EB-ALOHA throughput: $\frac{(27)}{(25)}$

EB-aware throughput: $\frac{(28)}{(25)}$

❖ **Strategy YYN**

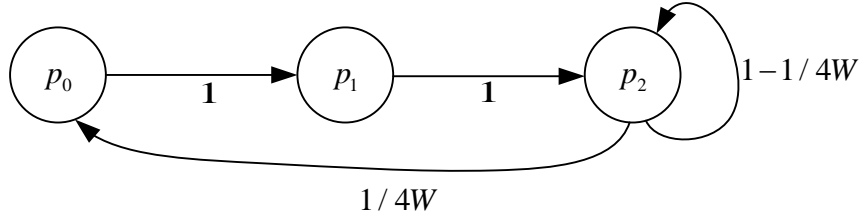


Fig. 5 Coarse Markov Chain for strategy YYN.

$$p_0 = \frac{1}{4W+2}, p_1 = \frac{1}{4W+2}, p_2 = \frac{4W}{4W+2}$$

$$E[\text{number of slots in one round}] = p_0 \cdot \frac{W+1}{2} + p_1 \cdot \frac{2W+1}{2} + p_2 \cdot \frac{4W+1}{2} \quad (29)$$

$$E[\text{number of total success slots in one round}] = p_0 \cdot \frac{W-1}{2} + p_1 \cdot \frac{2W-1}{2} + p_2 \cdot \frac{(4W)^2 - (4W) + 2}{2(4W)} \quad (30)$$

$$E[\text{number EB-ALOHA success slots in one round}] = p_2 \cdot \frac{1}{4W} \quad (31)$$

$$E[\text{number of EB-aware success slots in one round}] = p_0 \cdot \frac{W-1}{2} + p_1 \cdot \frac{2W-1}{2} + p_2 \cdot \frac{4W-1}{2} \quad (32)$$

The throughputs are therefore as follows:

$$\text{Sum throughput: } \frac{(30)}{(29)}$$

$$\text{EB-ALOHA throughput: } \frac{(31)}{(29)}$$

$$\text{EB-aware throughput: } \frac{(32)}{(29)}$$

Now we can calculate the throughputs for strategies xxY, NNN, YNN, NYN and YYN numerically for different initial window sizes and the results are given in Table 2. The optimal strategy can be concluded as follows:

When $W = 2$, the optimal strategy is strategy NNN;

When $W = 3$, the optimal strategies are strategy xxY, strategy NNN and strategy NYN;

When $W \geq 4$, the optimal strategy is strategy xxY.

Table 2 Throughputs for different strategies of EB-aware node.

W	strategy xxY			strategy NNN			strategy NYN		
	EB-aware	EB-ALOHA	Total	EB-aware	EB-ALOHA	Total	EB-aware	EB-ALOHA	Total
2	0.777777778	0	0.777777778	0.7230769231	0.0615384615	0.7846153846	0.7349397590	0.0481927711	0.7831325301

3	0.8461538462	0	0.8461538462	0.8251748252	0.0209790210	0.8461538462	0.8284023669	0.0177514793	0.8461538462
4	0.8823529412	0	0.8823529412	0.8712220762	0.0105124836	0.8817345598	0.8725376593	0.0092699884	0.8818076477
5	0.9047619048	0	0.9047619048	0.8978562421	0.0063051702	0.9041614123	0.8985176739	0.0057012543	0.9042189282
6	0.9200000000	0	0.9200000000	0.9152957648	0.0042002100	0.9194959748	0.9156742839	0.0038622465	0.9195365304
7	0.9310344828	0	0.9310344828	0.9276231263	0.0029978587	0.9306209850	0.9278597051	0.0027899562	0.9306496613
8	0.9393939394	0	0.9393939394	0.9368066283	0.0022468754	0.9390535037	0.9369642622	0.0021099829	0.9390742451
9	0.9459459459	0	0.9459459459	0.9439161653	0.0017465554	0.9456627207	0.9440264269	0.0016516792	0.9456781061

W	strategy YNN			strategy YN		
	EB-aware	EB-ALOHA	Total	EB-aware	EB-ALOHA	Total
2	0.7419354839	0.0322580645	0.7741935484	0.7500000000	0.0250000000	0.7750000000
3	0.8297872340	0.0141843972	0.8439716312	0.8323353293	0.0119760479	0.8443113772
4	0.8730158730	0.0079365079	0.8809523809	0.8741258741	0.0069930070	0.8811188811
5	0.8987341772	0.0050632911	0.9037974683	0.8993135011	0.0045766590	0.9038901601
6	0.9157894737	0.0035087719	0.9192982456	0.9161290323	0.0032258065	0.9193548388
7	0.9279279279	0.0025740026	0.9305019305	0.9281437126	0.0023952096	0.9305389222
8	0.9370078740	0.0019685039	0.9389763779	0.9371534196	0.0018484288	0.9390018484
9	0.9440559441	0.0015540016	0.9456099457	0.9441587068	0.0014695077	0.9456282145

■ Coexistence with a mix of TDMA and q -ALOHA networks

Now we consider the coexistence of one model-aware node with one TDMA node and $(N - 1)$ q -ALOHA nodes. Altogether, there are $(N + 1)$ nodes in the network. The optimal strategy for the model-aware node is given as follows. The model-aware node refrains from transmission in the slots used by the TDMA node. For the slots not used by the TDMA node, the model-aware node decides whether to transmit or not according to the value of N and q , as per (3) and (4).

Suppose that TDMA is assigned X slots of 10 slots in a frame, the optimal throughputs are given in the following table ((4) is used):

Table 3 Throughputs for the coexistence of a model-aware node and a mix of a TDMA and $(N-1)$ q -ALOHA nodes.

	TDMA	Each q -ALOHA	Mode-aware	Sum
$q \geq 1/N$	$\frac{X}{10}(1-q)^{N-1}$	$\left(1 - \frac{X}{10}\right)q(1-q)^{N-2}$	0	$\frac{X}{10}(1-q)^{N-1} + \left(1 - \frac{X}{10}\right)q(1-q)^{N-2}$
$q < 1/N$	$\frac{X}{10}(1-q)^{N-1}$	0	$\left(1 - \frac{X}{10}\right)(1-q)^{N-1}$	$(1-q)^{N-1}$

For the coexistence of multiple model-aware nodes with a mix of TDMA and q -ALOHA nodes, the optimal results can also be derived. Consider a system with L model-aware nodes, one TDMA node and $(N - L)$ q -ALOHA nodes. Altogether, there are $(N + 1)$ nodes in the network. Suppose the L model-aware nodes are aware of each other and they transmit in a round-robin manner (other transmission strategies are also possible, e.g., one model-aware node transmits, and others do not transmit). Thus, the L model-aware nodes can be regarded as one integrated model-aware node. The optimal strategy for the integrated model-aware node is to refrain from transmission in the slots used by the TDMA node. For the slots not used by the TDMA node, the integrated model-aware nodes decide whether to transmit or not according to the value of N and q , as per (3) and (4).

Suppose that TDMA is assigned X slots of 10 slots in a frame, the optimal throughputs are given in the following table ((4) is used):

	TDMA	Each q -ALOHA	Sum of Mode-aware	Sum
$q \geq 1/N$	$\frac{X}{10}(1-q)^{N-L}$	$\left(1 - \frac{X}{10}\right)q(1-q)^{N-L-1}$	0	$\frac{X}{10}(1-q)^{N-L} + \left(1 - \frac{X}{10}\right)q(1-q)^{N-L-1}$
$q < 1/N$	$\frac{X}{10}(1-q)^{N-L}$	0	$\left(1 - \frac{X}{10}\right)(1-q)^{N-L}$	$(1-q)^{N-L}$

Part 2: Achieving Proportional Fairness

Before deriving the individual throughputs achieved when the model-aware node aims to achieve proportional fairness, we first introduce α -fairness local utility function and the system objective. Consider a system with N nodes. For a particular node i , its throughput is denoted by $x^{(i)}$; its α -fairness local utility function is given by

$$f_{\alpha}^{(i)}(x^{(i)}) = \begin{cases} \log(x^{(i)}) & \text{if } \alpha=1 \\ (1-\alpha)^{-1} (x^{(i)})^{1-\alpha} & \text{if } \alpha \neq 1 \end{cases} \quad (33)$$

The objective of the overall system is to maximize the sum of all the local utility functions:

$$\begin{aligned} &\text{maximize } F(x^{(1)}, x^{(2)}, \dots, x^{(N)}) = \sum_{i=1}^N f_{\alpha}^{(i)}(x^{(i)}) \\ &\text{subject to } \sum_{i=1}^N x^{(i)} \leq 1. \\ &\quad x^{(i)} \geq 0, \quad \forall i \end{aligned} \quad (34)$$

When $\alpha=0$, maximizing the α -fairness objective corresponds to maximizing the sum throughput (Part 1); when $\alpha=1$, maximizing the α -fairness objective corresponds to achieving proportional fairness.

The MAC strategies of TDMA, q -ALOHA, FW-ALOHA and EB-ALOHA considered here are the same as those in Part 1. We derive optimal results without repeating each node's MAC strategy in this part.

■ Coexistence with a TDMA node

We first consider the coexistence of a model-aware node and a TDMA node. Thus, there are two nodes in the network. To achieve proportional fairness, the optimal strategy for the model-aware node is to transmit in the slots not occupied by TDMA node and not transmit in the slots occupied by TDMA node.

■ Coexistence with (N-1) q -ALOHA nodes

We next consider the coexistence of a model-aware node with $(N-1)$ q -ALOHA nodes. The total number of nodes in the network is therefore N . The model-aware node here is referred to as a q -aware node. It

knows the value q as well as N . We focus on a time period of T time slots. Suppose the q -aware node transmits in M slots out of the T time slots. Then the throughputs for q -aware node and each q -ALOHA node are $(1-q)^{N-1} M / T$ and $q(1-q)^{N-2} (T-M) / T$, respectively. To achieve proportional fairness, the q -aware node should maximize

$$F(M) = \log\left(\frac{M}{T}(1-q)^{N-1}\right) + (N-1)\log\left(\frac{T-M}{T}q(1-q)^{N-2}\right) \quad (35)$$

Taking the derivative, the maximum of $F(M)$ can be achieved when $M = T / N$. Therefore, the throughputs for q -aware node and each q -ALOHA node are $(1-q)^{N-1} / N$ and $(1-1/N)q(1-q)^{N-2}$, respectively.

■ Coexistence with a fixed-window ALOHA node

Here, we consider the coexistence of a model-aware node with a FW-ALOHA node. Thus, there are two nodes in the network. The model-aware node here is referred to as a FW-aware node. As shown in Fig. X, we can construct a Markov chain to analyze the individual throughputs. We define state i ($i = 0, 1, 2, \dots, W-1$) of the Markov chain to be the number of idle slots of FW-ALOHA node that has been observed by FW-aware node. State transition probabilities are given in Fig. X. We can derive the stationary probability p_i of state i using the balance equations below:

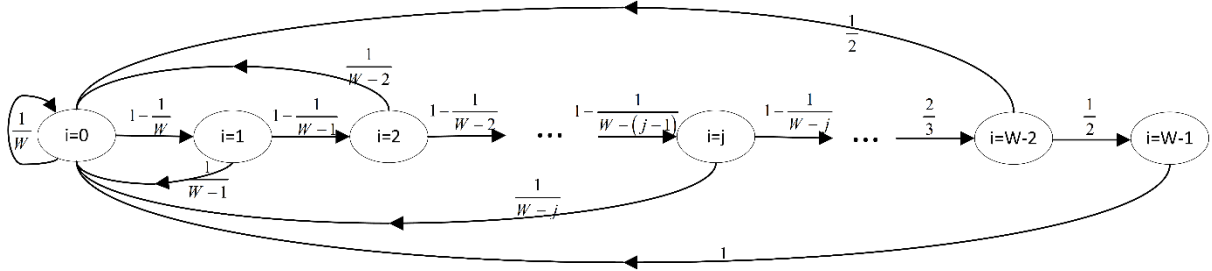


Fig. 6 Markov chain of the FW-aware node.

$$\left\{ \begin{array}{l} p_0 = \frac{1}{W} p_0 + \frac{1}{W-1} p_1 + \frac{1}{W-2} p_2 + \dots + \frac{1}{2} p_{W-2} + p_{W-1} \\ p_1 = \left(1 - \frac{1}{W}\right) p_0 \\ p_2 = \left(1 - \frac{1}{W-1}\right) p_1 \\ \vdots \\ p_{W-2} = \frac{2}{3} p_{W-3} \\ p_{W-1} = \frac{1}{2} p_{W-2} \\ p_0 + p_1 + \dots + p_{W-1} = 1 \end{array} \right. \quad (36)$$

Thus, the stationary probability p_i can be calculated as $p_i = \frac{2(W-i)}{W(W+1)}$.

We use $a_i \in \{0,1\}, i=0,1,2,\dots,W-1$ to represent the FW-aware node's decision at state i , i.e., if transmits, $a_i = 1$; else $a_i = 0$. Thus, the objective of the FW-aware node is to maximize

$$\begin{aligned} F(a_0, a_1, \dots, a_{W-1}) &= \log \left(\sum_{i=0}^{W-1} p_i a_i \left(1 - \frac{1}{W-i}\right) \right) + \log \left(\sum_{i=0}^{W-1} p_i (1-a_i) \frac{1}{W-i} \right) \\ &= \log \left(\sum_{i=0}^{W-1} a_i \left(1 - \frac{1}{W-i}\right) \cdot \frac{2(W-i)}{W(W+1)} \right) + \log \left(\sum_{i=0}^{W-1} (1-a_i) \frac{1}{W-i} \cdot \frac{2(W-i)}{W(W+1)} \right) \\ &= 2 \log \frac{2}{W(W+1)} + \log \left(\left(\sum_{i=0}^{W-1} (W-i-1) a_i \right) \left(\sum_{i=0}^{W-1} (1-a_i) \right) \right) \\ &= 2 \log \frac{2}{W(W+1)} + \log \left(\left((W-1) \sum_{i=0}^{W-1} a_i - \sum_{i=0}^{W-1} i \cdot a_i \right) \left(W - \sum_{i=0}^{W-1} a_i \right) \right) \end{aligned} \quad (37)$$

To maximize (37), we only need to maximize:

$$f(a_0, a_1, \dots, a_{W-1}) = \left((W-1) \sum_{i=0}^{W-1} a_i - \sum_{i=0}^{W-1} i \cdot a_i \right) \left(W - \sum_{i=0}^{W-1} a_i \right) \quad (38)$$

Without loss of generality, let $\sum_{i=0}^{W-1} a_i = j$, $j \in \{0,1,2,\dots,W-1\}$, then we can determine that there are j elements in $\{a_0, a_1, \dots, a_{W-1}\}$ equaling to 1. Here, we prove that $a_i = 1$ if $i < j$ and $a_i = 0$ if $i \geq j$.

Proof:

Remember that the objective is to maximize (38). Substitute $\sum_{i=0}^{W-1} a_i = j$ into (38) we get

$$f(a_0, a_1, \dots, a_{W-1}) = \left((W-1)j - \sum_{i=0}^{W-1} i \cdot a_i \right) (W-j) \quad (39)$$

To maximize (39), we need to minimize $\sum_{i=0}^{W-1} i \cdot a_i$. Because $a_i \in \{0, 1\}$, we have

$$\sum_{i=0}^{W-1} i \cdot a_i \geq 0 \cdot 1 + 1 \cdot 1 + 2 \cdot 1 + \dots + (j-1) \cdot 1 = \frac{j(j-1)}{2} \quad (40)$$

where the equality holds when $a_i = 1$ for $i < j$ and $a_i = 0$ for $i \geq j$.

Now, the objective is to maximize

$$g(j) = \left((W-1)j - \frac{j(j-1)}{2} \right) (W-j) = \frac{1}{2}j^3 - \frac{1}{2}(3W-1)j^2 + \frac{1}{2}W(2W-1)j \quad (41)$$

Suppose j is continuous (the domain is $[0, W]$). Taking the derivative of $g(j)$ we get

$$g'(j) = \frac{3}{2}j^2 - (3W-1)j + \frac{1}{2}W(2W-1) \quad (42)$$

We can see that when $0 < j < (3W-1-\sqrt{3W^2-3W+1})/3$, $g'(j) > 0$; when $(3W-1-\sqrt{3W^2-3W+1})/3 < j < W$, $g'(j) < 0$. Therefore, the maximum of $g(j)$ can be achieved when $j = (3W-1-\sqrt{3W^2-3W+1})/3$. However, in our case j is an integer within $[0, W]$, the optimal \hat{j} should be the integer around $j = (3W-1-\sqrt{3W^2-3W+1})/3$ that maximizes (5) and \hat{j} can be derived accordingly.

Therefore, the optimal throughputs for the FW-aware node and the FW-ALOHA node are given as follows:

$$\text{FW-aware throughput: } \sum_{i=0}^{W-1} p_i a_i \left(1 - \frac{1}{W-i} \right) = \sum_{i=0}^{\hat{j}-1} \left(1 - \frac{1}{W-i} \right) \frac{2(W-i)}{W(W+1)} = \frac{-\hat{j}^2 + (2W-1)\hat{j}}{W(W+1)}$$

$$\text{FW-ALOHA throughput: } \sum_{i=0}^{W-1} p_i (1-a_i) \frac{1}{W-i} = \sum_{i=\hat{j}}^{W-1} \left(\frac{1}{W-i} \right) \frac{2(W-i)}{W(W+1)} = \frac{2(W-\hat{j})}{W(W+1)}$$

▪ Coexistence with an exponential-backoff ALOHA node

Now, we consider the coexistence of a model-aware node and an EB-ALOHA node. Thus, there are two nodes in the network. The model-aware node here is referred to as a EB-aware node. We define the *stage* of the EB-ALOHA node to be the number of previous collisions experienced by the current packet, i.e., stage 0, stage 1, ..., stage m . We define the *state* of the EB-ALOHA node to be (s, t_s) , where $s = 0, 1, 2, \dots, m$ is the backoff stage and $t_s \in [0, 2^s W - 1]$ is the number of idles slots of EB-ALOHA node has been observed by the EB-aware node. We use $a_{(s, t_s)} \in \{0, 1\}$ to represent the EB-aware node's decision at state (s, t_s) , i.e., if transmits, $a_{(s, t_s)} = 1$; else $a_{(s, t_s)} = 0$. The Markov chain and the state transition probabilities when $m = 2$ are illustrated in Fig. 2.

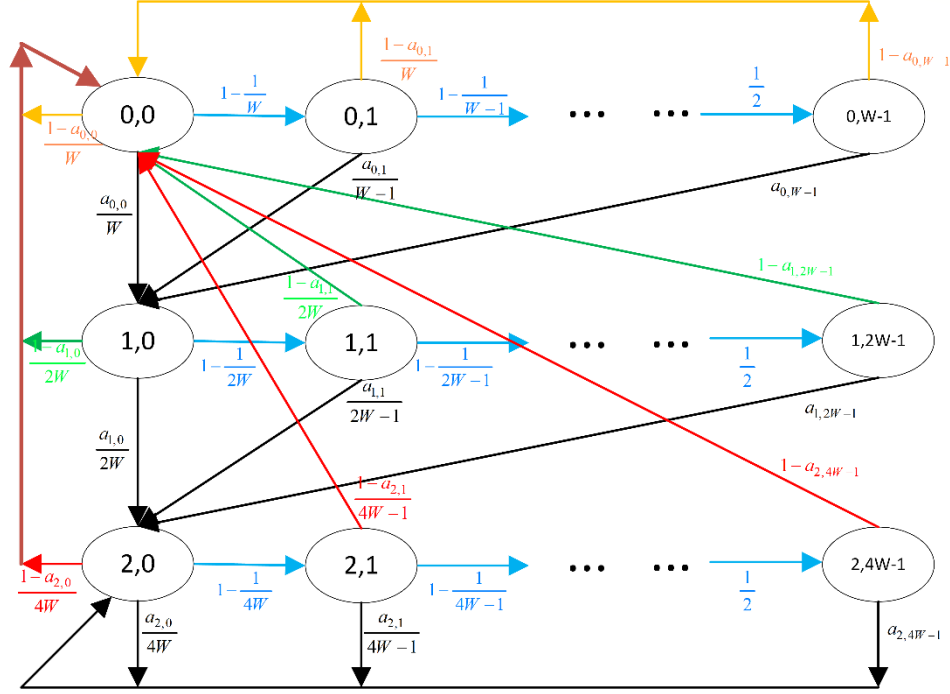


Fig. 2 Markov chain of the EB-aware node.

We can derive the stationary probability p_{s, t_s} of state (s, t_s) using the balance equations below:

$$\left\{ \begin{array}{l}
p_{0,0} = (1 - a_{0,0}) \frac{1}{W} p_{0,0} + (1 - a_{0,1}) \frac{1}{W-1} p_{0,1} + (1 - a_{0,2}) \frac{1}{W-2} p_{0,2} + \dots + (1 - a_{0,W-2}) \frac{1}{2} p_{0,W-2} + (1 - a_{0,W-1}) p_{0,W-1} \\
+ (1 - a_{1,0}) \frac{1}{2W} p_{1,0} + (1 - a_{1,1}) \frac{1}{2W-1} p_{1,1} + (1 - a_{1,2}) \frac{1}{2W-2} p_{1,2} + \dots + (1 - a_{1,2W-2}) \frac{1}{2} p_{1,2W-2} + (1 - a_{1,2W-1}) p_{1,2W-1} \\
+ (1 - a_{2,0}) \frac{1}{4W} p_{2,0} + (1 - a_{2,1}) \frac{1}{4W-1} p_{2,1} + (1 - a_{2,2}) \frac{1}{4W-2} p_{2,2} + \dots + (1 - a_{2,4W-2}) \frac{1}{2} p_{2,4W-2} + (1 - a_{2,4W-1}) p_{2,4W-1} \\
p_{0,1} = \left(1 - \frac{1}{W}\right) p_{0,0} \\
p_{0,2} = \left(1 - \frac{1}{W-1}\right) p_{0,1} \\
\vdots \\
p_{0,W-1} = \frac{1}{2} p_{0,W-2} \\
p_{1,0} = a_{0,0} \frac{1}{W} p_{0,0} + a_{0,1} \frac{1}{W-1} p_{0,1} + a_{0,2} \frac{1}{W-2} p_{0,2} + \dots + a_{0,W-2} \frac{1}{2} p_{0,W-2} + a_{0,W-1} p_{0,W-1} \\
p_{1,1} = \left(1 - \frac{1}{2W}\right) p_{1,0} \\
p_{1,2} = \left(1 - \frac{1}{2W-1}\right) p_{1,1} \\
\vdots \\
p_{1,2W-1} = \frac{1}{2} p_{1,2W-2} \\
p_{2,0} = a_{1,0} \frac{1}{2W} p_{1,0} + a_{1,1} \frac{1}{2W-1} p_{1,1} + a_{1,2} \frac{1}{2W-2} p_{1,2} + \dots + a_{1,2W-2} \frac{1}{2} p_{1,2W-2} + a_{1,2W-1} p_{1,2W-1} \\
+ a_{2,0} \frac{1}{4W} p_{2,0} + a_{2,1} \frac{1}{4W-1} p_{2,1} + a_{2,2} \frac{1}{4W-2} p_{2,2} + \dots + a_{2,4W-2} \frac{1}{2} p_{2,4W-2} + a_{2,4W-1} p_{2,4W-1} \\
p_{2,1} = \left(1 - \frac{1}{4W}\right) p_{2,0} \\
p_{2,2} = \left(1 - \frac{1}{4W-1}\right) p_{2,1} \\
\vdots \\
p_{2,4W-1} = \frac{1}{2} p_{2,4W-2}
\end{array} \right. \quad (43)$$

The above equations can be simplified to

$$\left\{ \begin{array}{l}
p_{0,i} = \frac{W-i}{W} p_{0,0}, i = 0, 1, 2, \dots, W-1 \\
p_{1,j} = \frac{2W-j}{2W} p_{1,0}, j = 0, 1, 2, \dots, 2W-1 \\
p_{2,k} = \frac{4W-k}{4W} p_{2,0}, k = 0, 1, 2, \dots, 4W-1 \\
p_{0,0} = p_{0,0} - \frac{p_{0,0}}{W} \sum_{i=0}^{W-1} a_{0,i} + p_{1,0} - \frac{p_{1,0}}{2W} \sum_{j=0}^{2W-1} a_{1,j} + p_{2,0} - \frac{p_{2,0}}{4W} \sum_{k=0}^{4W-1} a_{2,k} \\
p_{1,0} = \frac{p_{0,0}}{W} \sum_{i=0}^{W-1} a_{0,i} \\
p_{2,0} = \frac{p_{1,0}}{2W} \sum_{j=0}^{2W-1} a_{1,j} + \frac{p_{2,0}}{4W} \sum_{k=0}^{4W-1} a_{2,k}
\end{array} \right. \quad (44)$$

Also, we have

$$\sum_{i=0}^{W-1} p_{0,i} + \sum_{j=0}^{2W-1} p_{1,j} + \sum_{k=0}^{4W-1} p_{2,k} = 1 \quad (45)$$

Let $\sum_{i=0}^{W-1} a_{0,i} = b_0$, $\sum_{j=0}^{2W-1} a_{1,j} = b_1$ and $\sum_{k=0}^{4W-1} a_{2,k} = b_2$, combining (44) with (45), we get:

1) When $b_2 \neq 4W$,

$$p_{0,0} = \frac{2}{(W+1) + \frac{b_0(2W+1)}{W} + \frac{2b_0b_1(4W+1)}{W(4W-b_2)}}, \quad p_{1,0} = \frac{b_0}{W} p_{0,0}, \quad p_{2,0} = \frac{2b_0b_1}{W(4W-b_2)} p_{0,0};$$

2) When $b_2 = 4W$,

$$p_{0,0} \rightarrow 0, \quad p_{1,0} \rightarrow 0, \quad p_{2,0} = \frac{1}{4W+1}.$$

The objective of the EB-aware node is to maximize

$$\begin{aligned} F = & \log \left[\sum_{i=0}^{W-1} p_{0,i} a_{0,i} \left(1 - \frac{1}{W-i} \right) + \sum_{j=0}^{2W-1} p_{1,j} a_{1,j} \left(1 - \frac{1}{2W-j} \right) + \sum_{k=0}^{4W-1} p_{2,k} a_{2,k} \left(1 - \frac{1}{4W-k} \right) \right] \\ & + \log \left[\sum_{i=0}^{W-1} p_{0,i} (1 - a_{0,i}) \frac{1}{W-i} + \sum_{j=0}^{2W-1} p_{1,j} (1 - a_{1,j}) \frac{1}{2W-j} + \sum_{k=0}^{4W-1} p_{2,k} (1 - a_{2,k}) \frac{1}{4W-k} \right] \end{aligned} \quad (46)$$

Equation (46) can be simplified to

$$\begin{aligned} F = & \log \left[\frac{p_{0,0}}{W} \left((W-1)b_0 - \sum_i^{W-1} i \cdot a_{0,i} \right) + \frac{p_{1,0}}{2W} \left((2W-1)b_1 - \sum_j^{2W-1} j \cdot a_{1,j} \right) + \frac{p_{2,0}}{4W} \left((4W-1)b_2 - \sum_k^{4W-1} k \cdot a_{2,k} \right) \right] \\ & + \log \left[\frac{p_{0,0}}{W} (W - b_0) + \frac{p_{1,0}}{2W} (2W - b_1) + \frac{p_{2,0}}{4W} (4W - b_2) \right] \end{aligned} \quad (47)$$

From (47), we have

$$\sum_i^{W-1} i \cdot a_{0,i} \geq \frac{b_0(b_0-1)}{2}, \quad \sum_j^{2W-1} j \cdot a_{1,j} \geq \frac{b_1(b_1-1)}{2}, \quad \sum_k^{4W-1} k \cdot a_{2,k} \geq \frac{b_2(b_2-1)}{2} \quad (48)$$

Combining (47) with (48), we get

$$F \leq \log \left[\frac{p_{0,0}}{W} \left(-\frac{b_0^2}{2} + \left(W - \frac{1}{2} \right) b_0 \right) + \frac{p_{1,0}}{2W} \left(-\frac{b_1^2}{2} + \left(2W - \frac{1}{2} \right) b_1 \right) + \frac{p_{2,0}}{4W} \left(-\frac{b_2^2}{2} + \left(4W - \frac{1}{2} \right) b_2 \right) \right] \quad (49)$$

$$+ \log \left[\frac{p_{0,0}}{W} (W - b_0) + \frac{p_{1,0}}{2W} (2W - b_1) + \frac{p_{2,0}}{4W} (4W - b_2) \right]$$

We can run simulations to find the maximum of F and the corresponding b_0 , b_1 and b_2 . Then we can derive the optimal throughputs for the EB-ALOHA node and the EB-aware node as follows:

$$\text{EB-aware throughput: } \frac{p_{0,0}}{W} \left(-\frac{b_0^2}{2} + \left(W - \frac{1}{2} \right) b_0 \right) + \frac{p_{1,0}}{2W} \left(-\frac{b_1^2}{2} + \left(2W - \frac{1}{2} \right) b_1 \right) + \frac{p_{2,0}}{4W} \left(-\frac{b_2^2}{2} + \left(4W - \frac{1}{2} \right) b_2 \right)$$

$$\text{EB-ALOHA throughput: } \frac{p_{0,0}}{W} (W - b_0) + \frac{p_{1,0}}{2W} (2W - b_1) + \frac{p_{2,0}}{4W} (4W - b_2)$$

■ Coexistence with a mix of TDMA and q -ALOHA networks

Now we consider the coexistence of one model-aware node with one TDMA node and $(N - 1)$ q -ALOHA nodes. The optimal strategy for the model-aware node is a combination of the optimal strategies when model-aware node coexists with TDMA node and when model-aware node coexists with $(N - 1)$ q -ALOHA nodes.

Suppose that TDMA is assigned X slots of 10 slots in a frame, the optimal throughputs are given in the following table:

Table 4 Throughputs for the coexistence of a model-aware node and a mix of a TDMA and $(N-1)$ q -ALOHA nodes.

TDMA	q -ALOHA (sum)	Mode-aware
$\frac{X}{10} (1-q)^{N-1}$	$\left(1 - \frac{X}{10} \right) \frac{(N-1)^2 q (1-q)^{N-2}}{N}$	$\left(1 - \frac{X}{10} \right) \frac{(1-q)^{N-1}}{N}$

When multiple model-aware nodes coexist with a mix of TDMA node q -ALOHA nodes, the optimal strategy is similar as the single model-aware node case, except we regard multiple model-aware nodes as one integrated model-aware node. Consider M model-aware nodes coexist with one TDMA node and $(N - M)$ q -ALOHA nodes. The TDMA node transmits in X slots out of 10 slots within a frame. The individual throughputs are given in table X.

Table 5 Throughputs for the coexistence of a M model-aware node and a mix of one TDMA and $(N - M)$ q -ALOHA nodes.

TDMA	q -ALOHA (sum)	Mode-aware (sum)
$\frac{X}{10} (1-q)^{N-M}$	$\left(1 - \frac{X}{10} \right) \frac{(N-M)^2 q (1-q)^{N-M-1}}{N}$	$\left(1 - \frac{X}{10} \right) \frac{M}{N} (1-q)^{N-M}$

Part 3: Future Work Discussions

▪ Coexistence with CSMA

[Comment] I expect that learning approaches would do well with regular MACs, how would this fare with something more random like CSMA/CA?

[Response] This question is interesting in that it touches on extension work in progress by us. The obvious difference between the MACs studied in this work and CSMA/CA is that CSMA/CA has the carrier sensing capability while the MACs in this paper do not. To coexist with CSMA/CA, we are currently exploring a carrier sensing based deep reinforcement learning MAC protocol. In this extension work, we assume a DRL agent uses a minislot to perform carrier sensing before transmitting a packet. Our preliminary results have been submitted to a conference, and its arXiv version can be found in [1]. In particular, in [1], we compared our carrier-sense based DLMA with a WiFi-like (a typical CSMA/CA protocol) protocol when coexisting with MACs studied in the current paper – i.e., we are interested in whether a carrier-sense DLMA can outperform a WiFi-like CSMA protocol when coexisting with other MACs. We found that carrier-sense DLMA outperforms WiFi-like protocol in maximizing the sum throughput of all nodes as well as individual throughputs of nodes. A detailed study of the coexistence of carrier-sense DLMA with CSMA/CA is still in progress.

▪ Large system size

[Comment] The experimental evaluation is limited to very small systems: two, three, or four nodes (or gateways). This does not seem realistic to me. Why did the authors not consider larger systems? Is it because the computational complexity of the learning algorithm becomes infeasible to compute?

[Response] The current work is the first step to explore the possibility of deep reinforcement learning based MAC protocol in heterogeneous wireless networks. To demonstrate the capability of such a MAC protocol in achieving near-optimal performance, we investigate small systems in the current paper. For large systems, e.g., dozens of nodes, scalability of our algorithm may or may not be an issue, depending on the situation.

For example, if the non-DRL nodes are TDMA nodes of a TDMA network. We can simply treat these TDMA nodes as one virtual TDMA node transmitting in the slots of all TDMA nodes. The goal of DLMA is to learn to avoid the slots used by the virtual TDMA node since the virtual TDMA node transmits in a regular manner and the transmission pattern is not influenced by the transmission pattern of DLMA.

If the non-DRL nodes are q -ALOHA nodes, then we can compute an effective q for an ensemble of q -ALOHA nodes. Specifically, if there are N q -ALOHA node, the probability of a slot being used by at least one q -ALOHA node $q' = 1 - (1 - q)^N$. The goal of DLMA is to be able to learn q' to make a decision to optimize certain objective. We do not see the problem of having a large number of q -ALOHA nodes either and we expect that DLMA will be able to scale in this case.

For other non-DRL nodes, e.g., FW-ALOHA or EB-ALOHA, to learn the random behavior of these non-DRL nodes may be more challenging. We conjecture that it will be more difficult for DLMA to learn the joint behavior of a large number of EB-ALOHA nodes, since the individual behavior of each EB-ALOHA node is affected by both the DRL nodes and other EB-ALOHA nodes. Whether our conjecture is correct, and if so, what are the possible solutions, is a future research direction of interest to us.

Last but not least, we would like to point out that the main issue may not be computation complexity, which can be solved by higher computation power or more complex neural networks, but the convergence issue. In practical wireless networks, each time slot is in the order of one millisecond. Thus, DLMA can have an observation of the environment only once in one millisecond, so that the “experiences” for the neural network are being gathered at the rate of one experience per millisecond. If the converged training of the neural network requires a lot more experiences in a large wireless network than in a small wireless network, then it will take a longer time for DLMA to achieve optimality or good performance. In other words, the system is “experience-bound” (limited by the number of experiences gathered per unit time) rather than “computation-bound” (limited by the computation capacity of the neural network).

A detailed and comprehensive study awaits future work.

▪ Saturated and unsaturated scenarios

[Comment] The authors assume that all nodes always have packets ready to send and construct their protocol accordingly. Assuming realistic traffic, what would change in protocol design and evaluation, and what would be the expected results?

[Response] This paper considers the saturated scenario in which all nodes always have packets to transmit. This will be the case, for example, when the nodes are transmitting large files containing many packets. In practice, some nodes may be unsaturated in that they only have packets to transmit intermittently. In the following, we sketch tentative ways to deal with scenarios in which the heterogeneous network has a mix of saturated and unsaturated nodes.

1) Saturated and Unsaturated DRL Nodes:

First, we need to redesign the DLMA algorithm to take into account the possibility of DRL nodes not having packets to transmit. For that, all DRL nodes can let the gateway know how many packets they currently have to transmit through a control channel. The gateway will simply not poll (instruct) unsaturated DRL nodes to transmit when they do not have any packet to transmit. In this case, the unsaturated nodes that do not previously have packets to transmit that suddenly have packets to transmit may be allocated more slots by the gateway. In this paper, the gateway first decides the slots to be used by DLMA, and then the gateway instructs the saturated DRL nodes to use the DLMA slots in a round-robin manner. In the new setup, with a mixture of saturated and unsaturated DRL nodes, instead of round-robin transmissions among the DRL nodes, the gateway can use the α -fairness mechanism to allocate DLMA slots among the DRL nodes.

2) Saturated and Unsaturated non-DRL Nodes:

Next, we need to consider the possibility of non-DRL nodes being unsaturated. Since our model-free approach is to assume we do not know the protocols of these non-DRL nodes, and we could only learn their behaviors by observations---our DLMA protocol remains largely unchanged. For example, consider the coexistence of q-ALOHA and DLMA. If some of the q-ALOHA nodes are unsaturated, then the value of their “effective q ” may vary over time depending on whether they have packets to transmit. When a q-ALOHA node does not have packets to transmit, its effective q is zero. With a number of unsaturated q-ALOHA nodes, their overall effective q may be changing over time: specifically, if among all q-ALOHA nodes, only N of them have packets to transmit at a particular moment. Then, at that moment, the effective q for the ensemble of q-ALOHA nodes is $q' = 1 - (1 - q)^N$, where q' is the probability that at least one of the q-ALOHA nodes transmits. The responsibility of DLMA is to learn and track the changing environment induced by the changing number of nodes with packets to transmit.

In the above, although we have sketched an approach to address the issue pointed out by the reviewer, we feel that the issue deserves a separate detailed study. We intend to study the issue together with the issue of large networks with large numbers of nodes, and the CSMA-based DLMA, in a separate paper – at the moment, our ongoing investigative effort is focusing on CSMA-based DLMA. In the revised manuscript script, we point out the possible extensions to our current manuscript along those lines – we did not, however, include the sketch above in the revised manuscript because the sketch is rather brief and does not do justice to various intricate nuances arising from further detailed considerations, such as speed at which our algorithm will converge under the new settings, etc. Including these detailed considerations, on one hand, will significantly increase the length of the paper; and, on the other hand, deserves a more careful and comprehensive studies.

References:

[1] Y. Yu, S.C. Liew, and T. Wang. "Carrier-Sense Multiple Access for Heterogeneous Wireless Networks Using Deep Reinforcement Learning." arXiv preprint arXiv:1810.06830 (2018).