entropy

Article

Survey of Reinforcement-Learning-Based MAC Protocols for Wireless Ad Hoc Networks with a MAC Reference Model

Zhichao Zheng *, Shengming Jiang, Ruoyu Feng, Lige Ge and Chongchong Gu

College of Information Engineering, Shanghai Maritime University, Shanghai 201306, China
* Correspondence: zhengzhichao0003@stu.shmtu.edu.cn

Abstract: In this paper, we conduct a survey of the literature about reinforcement learning (RL)-based medium access control (MAC) protocols. As the scale of the wireless ad hoc network (WANET) increases, traditional MAC solutions are becoming obsolete. Dynamic topology, resource allocation, interference management, limited bandwidth and energy constraint are crucial problems needing resolution for designing modern WANET architectures. In order for future MAC protocols to overcome the current limitations in frequently changing WANETs, more intelligence need to be deployed to maintain efficient communications. After introducing some classic RL schemes, we investigate the existing state-of-the-art MAC protocols and related solutions for WANETs according to the MAC reference model and discuss how each proposed protocol works and the challenging issues on the related MAC model components. Finally, this paper discusses future research directions on how RL can be used to enable MAC protocols for high performance.

Keywords: MAC reference model; wireless ad hoc network; medium access control protocols; reinforcement learning

check for updates

Citation: Zheng, Z.; Jiang, S.; Feng, R.; Ge, L.; Gu, C. Survey of Reinforcement-Learning-Based MAC Protocols for Wireless Ad Hoc Networks with a MAC Reference Model. *Entropy* **2023**, *25*, 101. https://doi.org/10.3390/e25010101

Academic Editors: Andrea Prati, Luis Javier García Villalba and Vincent A. Cicirello

Received: 23 November 2022
Revised: 19 December 2022
Accepted: 28 December 2022
Published: 3 January 2023

1. Introduction

The wireless ad hoc network (WANET), as a wireless communication network with frequently changing topology [1], has attracted extensive attention among scholars. A large number of mobile nodes (MNs) with functions such as information perception, data processing and wireless communication constitute a WANET environment. Therefore, the communication medium is often shared by multiple nodes at the same time and controlled with medium access control (MAC) protocols. A WANET is a decentralized type of wireless network not relying on a preexisting infrastructure [2,3]. Maintaining the reliability and security of wireless networks (WNs) is especially difficult, compared to the infrastructured device because of the mobility of nodes. Moreover, due to the hidden node problem, signal collision occurs frequently when two nodes do not know the existence of each other and need to access the shared media at the same time. The wireless network topology is highly dynamic because wireless links are very fragile due to path loss, poor interference immunity and the mobility of MNs. Moreover, whenever congestion occurs at intermediate nodes, packet loss and delays accumulate, which makes it impossible to obtain satisfactory network performance. It is a pressing demand for the development of efficient intelligent schemes to improve the performance of WANETs.

In highly dynamic WNs, MAC protocols are also required to be designed to adapt to changing conditions [4,5]. In the last several years, machine learning (ML) has been widely applied in solving different application-oriented problems. It allows the agent to learn different rules through statistics and different data sets by applying artificial intelligence (AI) to handle the design and development of algorithms and models [6,7]. ML is suitable for solving network problems due to the following two reasons: its capability of interacting with the environment from input information and its capability of making control decisions, i.e., wireless devices collect large quantities of observations, then feed them into ML algorithms to help make better decisions. For example, a series of techniques are used to design and optimize nodes in dynamic scenes to adapt to constantly changing environments. Some solutions are dedicated to rescue operations [8], forest fire monitoring and prevention [9]. Similarly, there are some works for underwater acoustic networks (UANs) [10–12], the Internet of things (IoT) [13–16], unmanned aerial vehicle (UAV) formations [17–21], vehicular ad hoc networks [22], flying ad hoc networks (FANET) [23–25], heterogeneous networks (HetNet) [26–29] and next-generation wireless communication [30–32]. In addition, there are some works aimed at security [33], robustness [34], energy saving [35,36], adaptability [37–40,40–42] and stability [43–51]. As recent WNs become more complex, more demands are placed on learning systems [52]. In such dynamic networks, intelligent MAC schemes should be designed to adapt to changing topologies. Instead of switching between different solutions, reinforcement learning (RL) is used to continuously evaluate channel conditions by inquiring and responding to a channel state in order to maintain a conflict-free schedule.

Considering the important role of intelligent model in addressing the above challenges, this paper presents an overview of intelligent applications at the MAC layer with a comprehensive understanding of ML-based MAC protocols under a MAC reference model. We focus on RL [53] and deep reinforcement learning (DRL) [54] approaches in the context of WANETs. The main contributions of this paper are summarized as follows:

- We describe how to use RL to classify observations into positive/negative types, and how to integrate it with different MAC components, to make a better decision.
- We review different RL algorithms used in various MAC components.
- We summarize open research issues and future directions in this field.

As shown in Figure 1, the rest of the paper is organized as follows. Section 2 discusses existing survey articles in this area, with some classic RL schemes introduced in Section 3. We classify the reviewed RL-based MAC protocols according to the MAC reference model [55] and summarize them in Section 4. Then, we present research challenges to improve the MAC design and discuss further research issues in Section 5. Finally, we conclude this study in Section 6.

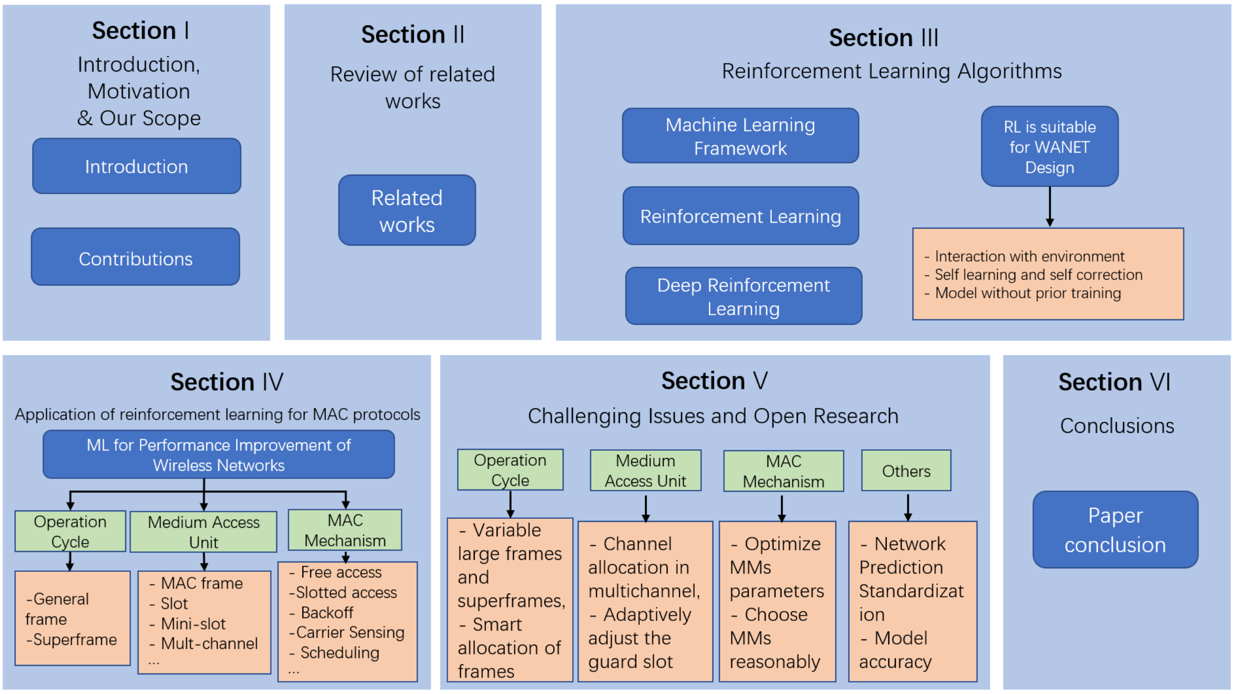
Section I
Introduction, Motivation & Our Scope
Introduction
Contributions
Section II
Review of related works
Related works
Section III
Reinforcement Learning Algorithms
Machine Learning Framework
Reinforcement Learning
Deep Reinforcement Learning
RL is suitable for WANET Design
- Interaction with environment
- Self learning and self correction
- Model without prior training
Section IV
Application of reinforcement learning for MAC protocols
ML for Performance Improvement of Wireless Networks
Operation Cycle
Medium Access Unit
MAC Mechanism
-General frame
-Superframe
- MAC frame
- Slot
- Mini-slot
- Mult-channel
...
- Free access
-Slotted access
- Backoff
-Carrier Sensing
- Scheduling
...
Section V
Challenging Issues and Open Research
Operation Cycle
Medium Access Unit
MAC Mechanism
Others
- Variable large frames and superframes,
- Smart allocation of frames
- Channel allocation in multichannel,
- Adaptively adjust the guard slot
- Optimize MMs parameters
- Choose MMs reasonably
- Network Prediction Standardization
- Model accuracy
Section VI
Conclusions
Paper conclusion

Figure 1. Paper outline.

2. Review of Related Survey Articles

Through a comprehensive survey of the intersection of the two fields, Ref. [56] bridges the gap between deep learning (DL) and mobile wireless network research by introducing

DL methods and their potential applications in mobile networks. It also reviews various technologies that can be used to deploy DL on WNs, such as data analysis, network control, signal processing, etc. In addition, it also provides a comprehensive list of papers related to mobile/WNs and DLs and classifies them according to different application fields.

Ref. [57] introduces the latest application of ML in improving the performance of MAC throughput, energy efficiency and delay in wireless sensor networks. It comprehensively reviews the application of ML technology and supports the implementation of real-time-based Internet of things (IoT) applications by optimizing the performance of MAC protocols. A similar work was conducted in [58]. In order to reduce the energy consumption of IoT devices and achieve a higher throughput, this paper summarized the RL-based time-division medium access (TDMA) MAC protocol and compares them from several unified features.

Ref. [59] surveys ML models, radio and MAC analysis and network prediction approaches in WNs, considering all layers of the protocol stack: physical (PHY), MAC and network. A comprehensive review is conducted on the use of ML-based schemes to optimize wireless communication parameter settings to achieve an improved network quality of service (QoS) and quality of experience (QoE). Finally, research challenges and open issues are discussed. Ref. [32] also conducts an in-depth survey on recent network optimization AI methods to ensure end-to-end QoS and QoE in 6G. It introduces the ML-based works of end-to-end transport flows from network access, congestion control and adaptive flow control. Some outstanding problems and potential research directions are discussed. Refs. [30,31] also survey multiple access schemes for next-generation wireless systems and give a detailed comparisons of 1G to 6G.

Ref. [60] surveys the latest works of ML applications in WNs by classifying them into three types according to different protocol stacks: resource control in the MAC layer, routing and mobility management in the network layer and localization in the application layer. In addition, several conditions of applying ML to wireless communication are analyzed, and the performance comparison between the traditional methods and ML-based methods is summarized. Finally, challenges and unsolved problems are discussed.

The reconfigurable intelligent surface (RIS) technology is used to reconstruct the wireless propagation environment by integrating the characteristics of large arrays with inexpensive antennas based on metamaterials. However, the MAC of RIS channels supporting multiuser access is still in the preliminary stage of research. Ref. [61] elaborates the architecture of the integration of a RIS and a MAC protocol in a multiuser communication system, introducing four typical RIS-assisted multiuser scenarios. Finally, some challenges, prospects and potential applications of RIS technology related to MAC design are discussed.

Ref. [62] outlines various ML-based methods in wireless sensor networks (WSNs) for smart city applications. The ML methods in WSN-IoT was investigated in detail, including supervised and unsupervised learning algorithms and RL algorithms.

ML is expected to provide solutions to various problems of UAVs used for communication. All relevant research works have been investigated in detail in [63], which have been used in UAV-based communications from the aspects of channel control, resource allocation, positioning and security. Table 1 compares our survey with the discussed references.

Table 1. Comparison between our survey and the related existing ones.

Reference	Networking Schemes	Learning Model				Survey Contents
		RL	DL	DRL	Other	
[56] (2019)	Cellular, ad hoc, cognitive radio networks (CRNs), IoT		✓	✓		The crossovers between DL models and mobile/wireless networking
[57] (2017)	WSNs, IoT	✓			✓	ML algorithms for WSNs.
[58] (2022)	TDMA, IoT	✓			✓	RL-based schemes for TDMA-based MAC protocols.
[59] (2021)	IoT, mobile big data, CRNs, WSNs	✓		✓	✓	RL-based schemes for WNs of three layers: PHY, MAC and network.
[31] (2022)	1G to 6G				✓	Multiple access schemes for next-generation wireless.
[32] (2021)	WNs	✓	✓		✓	ML-based methods for ensuring end-to-end QoS and QoE.
[60] (2019)	Cellular, WNs	✓			✓	ML-based schemes for network access, routing for traffic control and streaming adaption in WSNs.
[61] (2021)	RIS-aided WNs				✓	AI-assisted MAC for RIS-aided WNs.
[62] (2021)	WSN, IoT	✓			✓	RL-based methods in WSNs for smart city applications.
[63] (2019)	UAV, WNs				✓	ML framework for UAV-based communications.
This paper	WSNs, IoT, cellular, ad hoc, CRNs, USNs, HetNet, UANs	✓	✓	✓		RL/DRL-based MAC protocols for WNs with a MAC reference model.

3. RL Algorithms

With ML, computers learn to perform tasks by learning a set of training examples training, then uses the new set of data to perform the same task and give the results. In this section, we give a brief introduction on the ML framework, and the differences and relationship between RL and DRL are discussed.

3.1. Machine Learning Framework

One of the most commonly used ML algorithms is a neural network (NN), which loosely models neurons in biological brains. A typical NN system consists of two parts: neuron and connection. Each connection has a weight, which determines how strongly a node at one end affects another [64].

3.2. Reinforcement Learning

Through continuous learning and self-correction, humans spontaneously interact with their surroundings to accomplish tasks, such as object recognition, adjustment and selection [65]. Inspired by human behavior, RL was proposed. According to the different operating mechanisms, RL can be divided into the following two types: model-based RL and model-free RL [65]. The model-based RL algorithm allows an agent to learn a model to describe how the environment works from its observation and then uses this model to make action decisions. When the model is very suitable for the environment, the agent can directly find the optimal strategy through learning algorithms. However, some tasks cannot be modeled. The model-free RL scheme is an effective solution. In this scheme, the agent does not try to model the environment, but updates its knowledge based on the value iteration method to estimate the future return, so as to understand the value of actions taken in a given state. According to the strategy updating method, the model-free RL algorithm can be divided into round-update algorithms and step-update algorithms. Policy optimization and $Q-$learning are widely used in step-update algorithms. An RL system consists of the following three parts: an agent which interacts with the environment, an environment from which can be extracted some features and an AI algorithm which processes data and makes decisions. The AI algorithm makes decisions through an agent to influence the state and makes decisions again through feedback given by the environment, as shown in Figure 2a.

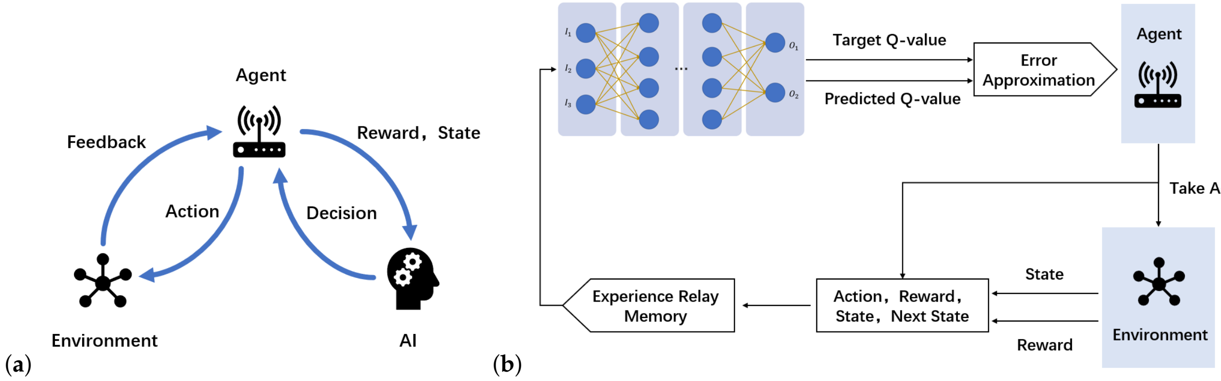

Figure 2. Schematic diagrams of RL and DRL. (**a**) Schematic diagram of reinforcement learning. (**b**) Schematic diagram of deep reinforcement learning [54].

Q-learning [66] is a widely used RL algorithm. A *Q*-table $Q(s, a)$ is the core and the expectation that obtain a reward by taking some action a in the state s over a period of time. The corresponding reward r is obtained from the feedback of the environment according to agent action a. Therefore, the key idea of *Q*-learning is to construct a table composed of states and corresponding actions to store the *Q*-value, and then selects one action that can obtain the maximum reward. Through continuous trial and error, an RL system aims to find a law from the mistakes and learn to achieve the goal without any idea at first. In addition, the actor–critic algorithm combines two kinds of reinforcement learning algorithms based on a value (such as Q learning) and policy (such as policy gradients).

A learning automaton (LA) is an RL algorithm that operates in a probability space and learns the optimal value by continuously interacting with an unknown environment. LA algorithms do not need to know their usage scenarios or any knowledge about the target to be optimized is a major advantage [67]. In recent years, some works [21,67,68] have attempted to apply LA to solve various practical problems, e.g., optimizing congestion-aware scheme and configuring the LA in the routing node.

3.3. From RL to DRL

As modern WNs become more and more complex, learning systems are required to have better performance, such as more powerful computing power, faster learning, better robustness and more extensible input and output mechanisms [52]. Therefore, DL applications for WNs have aroused widespread interest. Compared to a traditional RL algorithm, DL offers more improvements in WANET applications with more complex and large input data, by using multiple hidden layers between the input and output layers. Due to the increase of the action space and the state space from the input data, the *Q*-table is too huge to calculate. To address the above problems, a deep *Q*-learning (DQN) algorithm combines DL with *Q*-learning and uses the function instead of *Q*-tables to fit the relationship between the action and the state. A DQN algorithm extracts the features of the input state as the input, then calculates the value function through an NN as the output and trains the function parameters until convergence. The schematic structure of a DQN is shown in Figure 2b. The input data are accepted by the input neurons and transmitted to the hidden layer neurons. An agent obtains output data through the output layer after the training phase. The agent takes the result of the algorithm and computes the error by comparing it to the target model [69]. The policy is also updated using backpropagation by the agent. After a certain time, the agent adopts various actions and saves the current status, rewards obtained, next status and actions adopted into the experience replay memory (ERM), which is employed into the DQN framework to avoid potential nonconvergence of the gradient calculation. This makes the algorithm converge faster and more stably, and the agent can make decisions efficiently using the learned model. A taxonomy of RL is shown in Figure 3.

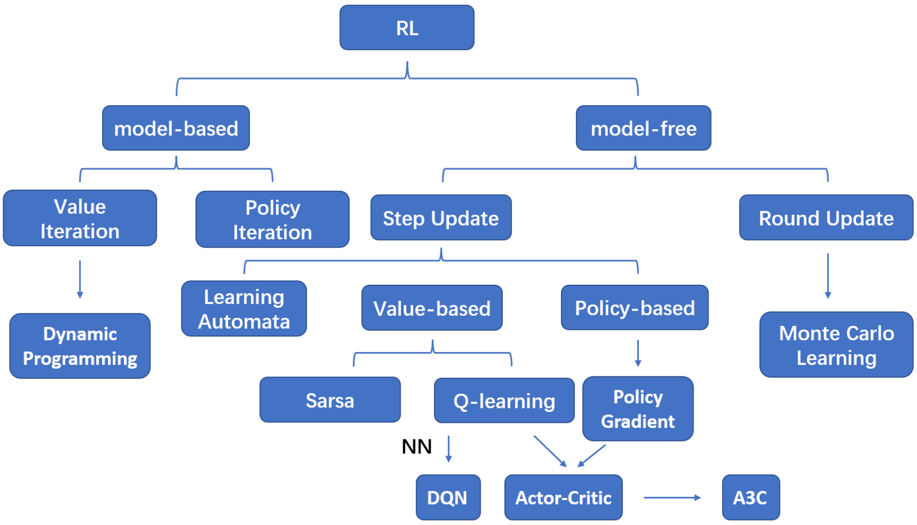

Figure 3. A taxonomy of reinforcement learning.

3.4. Optimal Design with RL

WNs have complex characteristics in their communication signals, channel quality, node topology, link congestion, etc. [70]. Moreover, the communication performance is significantly affected by MAC protocol targets, such as access delay, fairness, energy efficiency, congestion control, etc. The relationship between each state in the network environment satisfies the Markov property, and the stochastic process can be regarded as a Markov process (MP). When RL schemes are used to make decisions for the protocol, the reward and the action are introduced into the system, thus turning the MP problem into a Markov decision process (MDP). RL schemes have been widely developed to handle complicated wireless network situations with superiority over conventional schemes as listed below.

- Interaction with the environment: Due to node mobility, the network topology is constantly changing so that protocols are also required to be dynamically and automatically configured to achieve better network performance. RL can interact with the environment and feed back network changes to the protocol in time.
- Self-learning and self-correction: The characteristics of self-learning and interaction with the surroundings make systems adapt to the frequent changes of the network in time. Thus, the optimal scheme should be more adaptive, highly scalable and energy-efficient.
- Model without prior training: WNs are generated at any time and easy to construct; they do not need to be supported by preexisting networks. Compared with other learning schemes using trained models, RL schemes train the model by interacting with the networks in real time to make the trained model more suitable for the current environment.

4. Application of Reinforcement Learning for MAC Protocols

Several papers have investigated the applications of RL in MAC protocol design to provide accelerated data processing while maintaining a high throughput, with a minimization of collision and time saving. The review here adopts the MAC reference model proposed in [55], which compartmentalizes a MAC protocol into the following three components: operation cycle (OC), medium access unit (MAU) and MAC mechanism (MM). At the beginning of an OC, some MMs start running, which can be used to compete for the channel, and then, node accesses the medium with certain MAUs for transmission. Various MMs have different running sequences, the exchanged message content, the capacity and number of MAUs available in each OC constituting the MAC protocol process, which vary according to the function and design of different protocols [55]. The architecture diagram of MAC protocol components is illustrated in Figure 4. These components form the general structure of various MAC schemes for easier understanding.

Following this model, how RL schemes are used to improve the performance and effectiveness of these components are discussed below.

4.1. Operation Cycle

A MAC operation cycle is a repeated time epoch with either a fixed format or a random interval, which includes general frames, large frames and superframes [55]. One approach of using RL in MAC protocol design is modifying the OC, such as the length of the general frame and the structure of the frame. Modifying the superframe has been used in [46] for the purpose of protocol design and optimization.

The time slot is divided into a normal time slot and an emergency time slot, such as the superframe. Traditional MAC protocols lack efficient transmission rate control and backoff strategies in emergency transmission. This is a challenge for reliable emergency data transmission. To overcome these problems, an enhancement objective *Q*-Learning MAC (eOQ-MAC) protocol for emergency transmission was proposed in [46]. A novel superframe structure was designed for a slotted frame structure in eOQ-MAC, which modified the MAC frame by adding a new type of slot in it and selected a time slot to transmit data through a distributed slot selection algorithm with *Q*-Learning. If an emergency message needed to be sent, it automatically used the emergency time slot. If there was a data collision in the emergency time slot, the node without emergency data left its emergency time slot. The simulations showed that this packet loss rate of modification was lower than CSMA/CA's 79.6% in emergency transmissions.

4.2. Medium Access Unit

A medium access unit is the fundamental unit for nodes to access the shared medium. It can be determined not only by the MAC scheme, e.g., MAC frames, time slots and minislots, but also by the multiplexing solution in the physical layer, e.g., time slots, code and sub-bands [55]. The second approach of using RL in MAC protocol design is modifying a MAC unit or multiplexing unit in the MAU, such as the channel access method [10,45].

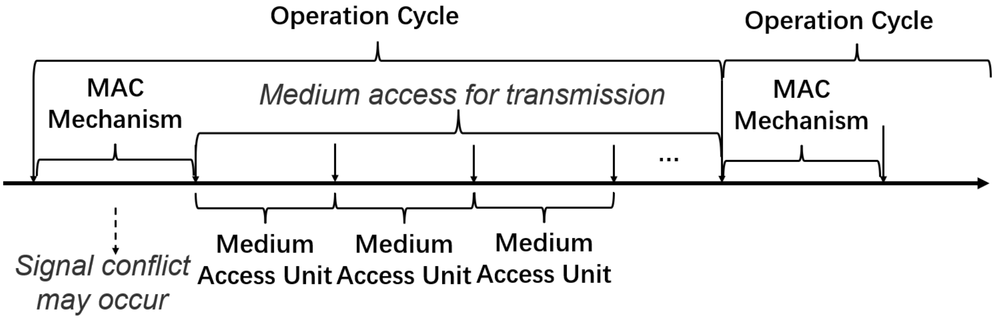

Figure 4. The architecture diagram of MAC protocol components [55].

In [10], the author proposed a slotted carrier sensing multiple access (CSMA) protocol, using RL to extend the life cycle of nodes in underwater acoustic wireless sensor networks (UA-WSNs), which have energy constraints due to the complex environment, where power replacement and charging for underwater devices are difficult and expensive. Meanwhile, the channels are vulnerable and changing with a poor delay and limited available bandwidth [55]. The underwater acoustic channel was divided into several subchannels. The design of the protocol mainly considered three factors: the transmission power, the number of shared channel neighbors and the selection of subchannels. These factors were considered to design the reward function in the protocol, using *Q*-learning to adjust the time slot length and choose a suitable subchannel to ensure the success of the transfer. It was shown that the proposed scheme was able to extend the lifetime of the network by adjusting the parameters of the algorithm to adapt to the underwater environment and reduce energy as much as possible.

Ref. [45] designed an adaptive MAC scheme for WSNs with the irregular repetition slotted ALOHA (IRSA) protocol. Due to uncertain and changing channel conditions in WSNs, especially the waterfall effect in the traditional IRSA protocol, *Q*-learning was used there to allocate a limited frame size and improve the common channel throughput. It tried to find the optimal policy that optimized expected rewards for all actions by learning the input–output relationship. RL-IRSA was optimized for small frame sizes, and the global throughput for high channel loads was effectively improved.

HetNet is a large-scale network with more access competition. Ref. [28] proposed a novel DRL-based MAC scheme for HetNet, named deep-reinforcement learning multiple access (DLMA). By correctly defining the state space, action space and rewards in a novel multidimensional DRL system, this scheme could maximize the total throughput and ensure the fairness of agents by selecting a specific time slot for transmission. The work in [71] also proposed an RL-based MAC protocol to select the appropriate time slot to access the network for slotted CSMA. By collecting collision conditions among slots, the RL system could effectively reduce collisions and packet delay. The above-discussed MAC protocols are summarized in Table 2.

Table 2. Comparison of the RL techniques used in OC and MAU.

Protocols (Reference)	MAC Components	Network	Algorithm	Data Used for Learning	Performance
eOQ-MAC [46] (2019)	Superframe (OC)	WNs	*Q*-learning	Packet loss rate, emergency time slot, throughput	Lowered packet loss rate for emergency data transmission.
UA-Slotted CSMA [10] (2013)	Slot (MAU)	UA-WSNs	*Q*-learning	Lifetime, channel status, energy consumption, bandwidth	Minimized power consumption and extended the life of the network.
RL-IRSA [45] (2020)	MAC frame, slot (MAU)	CRNs	*Q*-learning	Convergence time, channel status, throughput	Significantly reduced convergence time with optimized degree distributions for small frame sizes.
DLMA [28] (2019)	Slot (MAU)	HetNet	DRL	Convergence time, throughput	Maximized the total throughput with faster learning speed.
Q-CSMA [71] (2021)	Slot (MAU), scheduling (MM)	WNs	*Q*-learning	Channel status, packet loss rate, delay	Reduced the number of collisions and packet delay.

4.3. MAC Mechanism

A MAC mechanism is the competitive action taken by nodes expecting access to the medium, which can be divided into basic mechanisms and complex mechanisms (combination of basic mechanisms) [55]. Some significant basic mechanisms are listed below: free access, slot access, backoff, signaling, carrier sensing, messaging, scheduling, prioritization, etc. The third approach of using RL in MAC protocol design is modifying the MM for collision avoidance. MMs are modified and controlled by RL schemes to access the channel and find approximate solutions for optimization.

4.3.1. Backoff

In [8], the author proposed a fully distributed rate adaptive CSMA/CA protocol, which was an energy-saving multiuser scheduling MAC protocol based on RL for delay-sensitive data. Enabling the system to effectively trade off consumption and delay by using traditional CSMA/CA schemes is challenging. In order to balance consumption and delay, the energy-saving and delay-sensitive multiuser resource allocation problem was formulated as an MDP. Then, an RL algorithm was used to solve the agent's access task, and an optimal policy for each agent was learned through interaction with the environment. Unlike traditional CSMA/CA, this scheme adaptively changed the congestion windows (CWs) so that a node could increase the rate of transmission if the queue buffer was insufficient. Additionally, the proposed MAC protocol assumed that the backoff counters would be updated to fit its current situation instead of freezing it when multiple nodes grabbed the channel. Compared with the traditional CSMA/CA, this solution enabled nodes to achieve tighter latency limits with an energy consumption of 53%. Moreover, the proposed learning algorithm could be updated multiple times in each time slot to speed up the convergence speed and improve runtime efficiency.

In [72–76], a CW size adjustment was also used to improve network quality. Ref. [72] obtained a more efficient data exchange process by adjusting the size of CWs for vehicular ad hoc networks. The proposal could extract channel characteristics and predict the transmission success rate of packets. The simulations showed that the proposed scheme performed better than the existing backoff schemes. Ref. [73] tried to use the RL model to adjust the CW size by counting the current network throughput, while [74] tracked the wireless traffic load to promote the RL model for a novel CSMA/CA-based MAC. In [75], the authors proposed a novel RL model based on channel observation to predict the future channel state by observing whether the channel state was busy or idle and reduced the collision of access channels in dense WNs by controlling the size of the CW. A wireless smart utility network (Wi-SUN) is a large-scale smart city network that uses unslotted CSMA/CA. With the increase of the network scale, the competition becomes more serious, and the performance of a wireless MAC protocol will deteriorate rapidly. Ref. [76] proposed a Q-learning-based unslotted CSMA/CA (QUC) scheme to select the best backoff delay. The performance of the packet delivery ratio was improved by 20%.

Due to the poor performance of the latest CSMA scheme in multihop scenarios, Ref. [37] proposed a scheme to adopt an experience-driven approach and train a CSMA-based MAC with a DRL algorithm named Neuro-DCF. The solution adopted a multiagent RL framework for stable training with a distributed execution and introduced a new graph neural network (GNN)-based training construction for training a unified approach that incorporated various perturbation patterns and configurations. A GNN is a network that is used to process graph-type data. The representation of each node in a graph is calculated from the characteristics of the node, the characteristics of the edges connected to the node, the neighbor representation of the node and the characteristics of its neighbor nodes. That scheme comprehensively investigated multiple factors in the network, such as throughput, delay and channel utilization, which can accelerate the speed of decision convergence. Simulation showed that the scheme could improve the delay performance while preserving optimal utility.

Besides the channel allocation schemes, energy consumption is also considered as a solution for network performance improvement. Ref. [77] proposed an RL-based CSMA scheme for energy saving, which took into account the input and output of a power control method and multiple variables at the PHY layer. A similar study was proposed in [78], which introduced a deep-$Q-$learning-based scheme named ambient energy harvesting CSMA/CA (AEH-CSMA/CA) for the IoT. This scheme obtained the energy of the environment to intelligently adjust the size of its initial backoff window during the backoff process. The simulation results showed that the throughput of this scheme was 20% higher than IEEE 802.11 with a low energy supply. The above-discussed MAC protocols are summarized in Table 3.

Table 3. Comparison of the RL techniques used for the backoff mechanism.

Protocols (Reference)	MAC Components	Network	Algorithm	Data Used for Learning	Performance
Distributed rate adaptive CSMA/CA [8] (2016)	Backoff (MM)	WNs	*Q*-learning	Channel status, CW, energy consumption	Enabled users to reduce energy consumption based on their latency-limited needs with faster convergence.
Backoff Improvement [72] (2020)	Backoff (MM)	Vehicular networks	*Q*-learning	Channel status, transmission success rate	Had a more efficient data exchange process and ensured fairness.
CW adjustment scheme [73] (2021)	Backoff (MM)	WNs	*Q*-learning, supervised learning	CW, throughput	Effectively reduced the collision and improved the system throughput.
Performance enhancement CSMA [74] (2021)	Backoff, Scheduling (MM)	WNs	Reinforcement learning	Channel status, energy consumption, traffic load	Had a stable throughput.
Channel-observation-based MAC [75] (2018)	Backoff, scheduling (MM)	Dense WLANs	DRL	CW, throughput, channel status	Efficiently predicted the future channel state.
QUC [76] (2022)	Backoff (MM)	Wireless smart utility networks	*Q*-learning	Throughput, delay	The performance of the MAC layer was improved by 20%.
Neuro-DCF [37] (2021)	Backoff (MM), Slot(MAU)	WNs	DRL + GNN	Throughput, delay, channel utilization	Reduced the end-to-end delay while preserving optimal utility.
AEH-CSMA/CA [78] (2020)	Backoff (MM) (2020)	IoTs	Deep *Q*-learning	Throughput, CW, energy	Ensured a high throughput and low energy supply.

4.3.2. Scheduling

In cognitive ad hoc networks lacking a central controller, the secondary users (SUs) and the primary users (PUs) interfere with each other and synchronization is difficult. To overcome the above problems, Ref. [33] proposed a multichannel cognitive MAC protocol based on distributed RL for opportunistic spectrum access scheduling. This learning mechanism selected channels based on SUs' detection of PU traffic, avoided collisions with the PUs and kept SUs synchronized. Due to the burstiness of PUs' traffic, SUs needed to detect transmission opportunities in channels, and each SU employed an LA mechanism whose actions were updated based on environmental feedback from each channel. Having the same PU between clusters of nodes consisting of SUs, they could exchange their PU's existence experience within the cluster. Meanwhile, the protocol employed a control channel for resource reservation to handle interference and hidden terminal issues between SUs. Compared with the transmission opportunity through the statistical mechanism, the channel utilization of the proposed protocol was improved by 40% and the collision with the PUs was reduced.

In [43], the author proposed a novel p-persistent CSMA protocol based on RL by optimizing channel access scheduling. The channel allocation for SUs is a challenging issue due to packet collisions, while SUs may degrade the performance of PUs. This scheme achieved better channel utilization with reduced collisions for SUs by decreasing p in the presence of collisions and increasing p for successful transmissions. Meanwhile, this method maintained the throughput of the PUs and improved the channel utilization rate through sharing the PUs and SUs' traffic. The results showed that when the PUs increased the delay, the SUs could effectively use the available channels. Considering the transmission probability greatly affects the network throughput in WNs using p-persistent CSMA schemes, Ref. [79] proposed a multistate RL scheme. This scheme learned its optimal strategy by sensing the channel history information, especially the number of collisions or successful transmissions. Simulation showed that the average transmission success rate of this scheme was 45% higher than the traditional schemes.

The performance of the IEEE 802.15.4 MAC depends on the correct configuration of MAC parameters. In [80], an optimization method for adaptively configuring IEEE 802.15.4 MAC parameters was proposed. This scheme determined the MAC parameters based on the channel traffic status and channel characteristics with the predictive feature of RL algorithms and hence supported the dynamic nature of communications. The proposed solution built a learning model that could obtain optimized parameters, including sending rate frequency and packet interarrival time by considering a multilayer perceptron and random forests. Compared to the IEEE 802.15.4 standard with default parameters, this scheme reduced the end-to-end delay and ensured a stable throughput.

The performance of RL schemes relying on timely feedback rewards is severely limited because of the poor delay in UANs. An improved two-stage *Q*-learning scheme was proposed to obtain hidden reward for UANs, named the packet flow ALOHA with *Q*-learning (ALOHA-QUPAF) MAC protocol [11]. The two stages in that scheme were denoted as slot selection and flow harmony. They could not transmit and receive simultaneously in a UAN with half-duplex communication. Therefore, this scheme penalized these received slots to avoid collisions by reducing their *Q*-values. In a healthy channel, there is a continuous flow of packets on the chain. Thus, when the first packet is received, the receiver expects a packet in every frame that follows. As long as the packet flow is interrupted, a packet collision is inferred. The scheme isolated the receiving slot from the transmitting slot to avoid collisions by modifying the scheduling in the MM.

An RL-based MAC protocol for multimedia sensing in UANs was proposed in [12]. It improved the efficiency of mobile relay nodes by using transmission opportunity (TXOP) for relay nodes in multihop UANs. Transceiver nodes and relay nodes were allocated traffic demands based on the competition of sensors using the initial phase of *Q*-learning. Moreover, this solution allocated TXOP resources for the uplink devices based on the traffic demands. The simulations showed that the scheme had an efficient packet delivery rate and throughput.

To control the data transmission rate of nodes in CSMA/CA systems, Ref. [81] designed an RL scheme. By learning the timeout events of packets in the environment, the agent selected actions to control the data transmission rate of the node and adjusted the modulation and coding scheme (MCS) level of data packets to effectively use the available bandwidth in the dynamic channel conditions. Similarly, Ref. [82] also proposed a *Q*-learning system to control the data transmission rates for CSMA/CA-based systems. The RL agent tried to collect channel information and control the MCS level of data packets to obtain better network throughput.

When nodes become active and inactive in the WSN as nodes randomly join and leave the network, the long-term average network age of information (AoI) of their respective processes can be collectively minimized at the remote monitors. An optimized modification of the ALOHA-QT algorithm was proposed in [35], which employed a policy tree (PT) and RL to achieve high throughput. A PT is a prediction model, which represents a mapping relationship between attributes and values. The proposed algorithm was provided with an upper bound on the average AoI and a pointer to select its key parameters. The simulations showed that the proposed algorithm outperformed ALOHA-QT in terms of AoI while requiring less consumption and computation. The above schemes based on scheduling changes are summarized in the Table 4.

Table 4. Comparison of the RL techniques used for the scheduling mechanism.

Protocols (Reference)	MAC Components	Network	Algorithm	Data Used for Learning	Performance
MCC-MAC [33] (2015)	Scheduling (MM)	CRNs	*Q*-learning + LA	Network traffic, channel status	Avoided the conflict between SUs and PUs.
p-persistent CSMA [43] (2011)	Scheduling (MM)	CRNs	*Q*-learning	Throughput, channel utilization	Had good robustness and the SU efficiently utilized the available channel at the expense of an extra delay of the PU.
Optimal parameters for IEEE 802.15.4 [80] (2020)	Scheduling (MM)	WNs	Supervised learning	MAC parameters, delay, channel status	Increased the dynamic adaptability of nodes.
ALOHA-QUPAF [11] (2021)	Scheduling (MM)	UA-WSNs	Modified *Q*-learning	Throughput, channel status	Isolated the receiving slot from the transmitting slot to avoid collisions.
RL-MAC [12] (2022)	Scheduling (MM)	UA-WSNs	*Q*-learning	Channel traffic, energy consumption	Improved throughput with limited energy.
Improved ALOHA-QT [35] (2022)	Scheduling (MM)	WNs	RL + PT	Throughput, AoI, energy consumption	Adapted to the changing number of active agents with less energy.
Rate adaptation scheme [81] (2020)	Scheduling (MM)	WNs	*Q*-learning	MCS channel utilization, bandwidth	Obtained better network throughput.
Improved *p*-persistent CSMA [79] (2018)	Scheduling (MM)	WNs	Multi-state *Q*-learning	Channel status	Investigated the application of multistate RL algorithm.

4.3.3. Other MAC Mechanisms

A smart-CSMA (SCSMA) MAC scheme with control messages was proposed for low-power and low-bandwidth WSNs in [44]. The 802.15.4 MAC does not use control messages for transmission due to the small size of the data, and the cost of control messages is relatively high. The proposal assigned different start delay slot numbers (SDNs) to neighboring nodes by making each node rebroadcast a message. A node was designed to collect the above data to reduce collisions from neighbor nodes and hidden terminals. For two competing hidden users, their waiting time difference was adjusted to be greater than the sum of the transmission times of RTS and CTS. Therefore, the busy channel status was detected by users with a longer waiting time. The blind learning (BL) algorithm was proposed to optimize the exchange of messages. By listening the control messages, BL-based agents adjusted their SDN intelligently with hidden terminals or topology changes. The result showed that this scheme appropriately adjusted its SDN to improve the throughput by 42.1% and reduced the energy consumption by 15.7%.

In [39], a distributed-RL MAC protocol with localized information in WSNs was proposed. In order to maximize the network throughput, the solution adopted distributed learning of the optimal MAC transmission strategy among nodes. Moreover, nodes also obtained channel information from the closest two-hop neighbor nodes and adaptively adjusted their transmission strategy to maintain maximum throughput under dynamic topology loads. This solution was also designed to parametrically change the access priority of nodes while learning, which is very practical for WANET applications with different priorities. The results showed that the performance of this scheme was improved by 20%.

Many RL solutions proposed so far assume real-time feedback of the results of packet transmission. In [83], a novel RL-based MAC protocol ALOHA-dQT was proposed, which improved the channel utilization by letting nodes periodically broadcast short summaries of their known channel history. The channel history for the last N slot states was stored by each node, and it iteratively merged its information with the channel history based on the node broadcast. The RL system updated the relevant information to make the node adaptive.

Ref. [23] proposed a novel RL-based scheme for FANET, named position-prediction-based directional MAC (PPMAC). This scheme addressed the directional deafness problem by using position information, which is very practical for FANET applications in a dynamically changing environment. The results showed that the reward of this learning system was improved by 16%.

In addition, in order to achieve a better end-to-end QoS and ensure fairness in user-dense scenarios, work on the optimization of learning algorithms has also been proposed. In [84], a novel Monte Carlo (MC) reward update scheme for DRL training was proposed. This scheme used the access history of each site to derive the DRL-based MAC protocol. Compared with the traditional distributed coordination function (DCF), this scheme could effectively improve the network throughput. The above-discussed MAC protocols are summarized in Table 5.

Table 5. Comparison of the RL techniques used for other mechanisms.

Protocols (Reference)	MAC Components	Network	Algorithm	Data Used for Learning	Performance
SCSMA [44] (2020)	Messaging (MM)	WSNs	Blind learning [44]	SDN, throughput, channel status	Improved throughput, reduced energy consumption and could avoid hidden terminal problems.
Distributed ALOHA [39] (2022)	Prioritization (MM)	IoTs	*Q*-learning	Throughput, channel status	Maintained maximum throughput under dynamic topology loads.
ALOHA-dQT [83] (2020)	Messaging (MM)	WNs	Reinforcement learning	Channel history, throughput	Achieved a high channel utilization.
PPMAC [23] (2018)	Messaging (MM)	FANET	*Q*-learning	Channel status, position information	Provided an intelligent and highly adaptive communication solution.
SPCA [85] (2019)	Messaging (MM)	WSNs	DRL	Spectrograms of TDMA, channel utilization	Reduced the number of collisions with efficient channel utilization.

4.4. Summary

We can find that most of the proposals try to improve the MMs by obtaining medium access opportunity and then performing optimal scheduling. The traditional medium access schemes cannot well meet the rapid changes of the wireless network topologies, while RL solutions can learn what the regular nodes' behavior is by creating models from past scenarios. Modifying the MAU has been proposed, such as assigning subchannels to make networks have a better performance, and with scheduling, the sender makes decisions locally to reduce collision from different senders.

Meanwhile, RL schemes are often used to optimize the channels, as some of the channel variations occur in a changeable environment, to let the node learn and make real-time decisions tailored to each situation, especially with the advantage of enabling nodes to perceive and interact with the environment.

5. Open Research Issues

Here, we discuss some challenging issues following the MAC reference model and give some directions for future research in this section. A summary of the discussed is presented in Table 6.

5.1. Operation Cycle

Both large frame and superframe divide the time into several segments. The former allocates each segment equally to the nodes and uses the same access scheme, while the latter adopts different access schemes for each segment, which is mainly used in scenarios that need to run multiple transmission methods at the same time, such as prioritization and emergency transmission. In this regard, RL solutions can be used for adaptive selective access schemes. In addition, coordinated scheduling with neighbors can use RL schemes to adapt the changing networks and topologies in future research, e.g., for nodes to learn the parameters selection and the extent of dynamic network changes.

5.2. Medium Access Unit

Several schemes have been proposed in certain functions, such as adjusting the slot length to handle the channel access and channel allocation with multiple access schemes. However, there are still some issues to be discussed. For example, with TDM, time is divided into regular time units. Although collisions are avoided by inserting a guard time between two adjacent time slots, it reduces bandwidth utilization. RL schemes can be used by nodes to adaptively adjust the size of the guard time according to the distribution of active nodes. With learning individual node patterns and requirements, RL schemes can help agents to learn which slot should be accessed to maximize bandwidth utilization. Another issue in the MAU is data channel allocation in multichannel protocols. Since a node can only work either on the control channel or the data channel, but not both at the same time, RL schemes can be applied to learn individual node's channel conditions and requirements to allocate channels required by the node without the control channel.

5.3. MAC Mechanism

A MM can be used alone to prevent collisions, or multiple MMs can be combined to form a new complicated mechanism [55]. State-of-the-art RL-based works focus on tuning the MMs' parameters. An issue is how to choose those MMs reasonably to effectively improve the MAC protocol performance. RL schemes can be used to choose MMs and form them to design a MAC protocol for different network scenarios by creating RL models and learning from current and past network status.

5.4. Other Issues

In WANETs, there are usually no nodes dedicated to coordinate communication, and each node dynamically forms a temporary network due to mobility. Mobility is one of the main factors affecting network performance since it produces a frequently changing network topology. In the following topics, some applications of RL for adaptability are discussed.

(1) Environmental prediction: The wireless network environment is complex and changeable; RL schemes are used to discover the rules of potential changes between environmental characteristics and reduce the impact of uncertain factors on the protocol performance to make better real-time decisions. For example, a method called spatial reuse, which enables two adjacent transmitters to send signals simultaneously without affecting each other. It is feasible to exploit similar features to maximize channel utilization as much as possible by using RL.

(2) Resource allocation and task scheduling: Channel resources are usually limited, and actual power allocation and channel constraints are the main causes of message collisions. The latest mathematical optimization theory still lacks a systematic resource allocation method, and resource scheduling will increase the complexity of the protocol. The protocol design can further use RL schemes to improve the capabilities of nodes, by decreasing complexity and high overheads.

(3) Systematical cross-layer design: Some MAC protocols using a cross-layer design also focus solely on MAC performance and ignore the performance degradation of the network as a whole. For example, the physical layer data is employed to detect channel variation and quality; meanwhile, the network layer data, for instance congestion, is employed to determine the status of an entire link. Using the powerful data aggregation function of RL, a protocol design can comprehensively consider all available optimization options to enhance performance [55].

(4) Underwater acoustic networks: Many modifications have noticed the drawbacks of long propagation delays in UANs and tried to turn the drawbacks of long delays into advantages by setting relevant parameters [55]. However, some protocols do not thoroughly consider the peculiar features of UANs, such as an unstable channel, rapidly changing channel states and the leveraging of long propagation delay. With RL, these features can be fully considered as much as possible to reduce the probability of signal collision.

Table 6. A summary of open research issues.

MAC Components	Open Research Issues
Operation cycle	Variable large frames and superframes Smart frame allocation algorithms
Medium access unit	Channel allocation in multichannel Adaptively adjust the guard slot
MAC mechanism	Optimize MMs' parameters Choose MMs reasonably
Other issues	Network prediction standardization and model accuracy Resource allocation and task scheduling Systematical cross-layer design Fully consider the features of UANs

6. Conclusions

This paper reviewed the methodologies of applying an RL scheme in WNs within the context of the MAC reference model. Although some works have been conducted in developing RL in MAC protocol design, there are still some open issues that urgently need to be addressed in this field, covering some directions according to different MAC components, especially the MMs. We hope that this article will help the reader to understand the novel RL-based MAC schemes and provide a guide for researchers to optimize WNs using RL techniques.

Author Contributions: Conceptualization, Z.Z., S.J., R.F., L.G. and C.G.; investigation, Z.Z.; writing—original draft, Z.Z.; writing—review and editing, Z.Z., S.J., R.F., L.G. and C.G.; project administration, S.J.; funding acquisition, S.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by The Innovation Program of Shanghai Municipal Education Commission of China, grant number 2021-01-07-00-10-E00121.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kou, K.; Lei, L.; Zhang, L.; Cai, S.; Shen, G. Intelligent Selection: A Neural Network-Based MAC Protocol-Selection Mechanism for Wireless Ad hoc Networks. In Proceedings of the IEEE 19th International Conference on Communication Technology (ICCT), Xián, China, 16–19 October 2019; pp. 424–429.
2. Toh, C.K. *Wireless ATM and Ad-Hoc Networks: Protocols and Architectures*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012.
3. Hu, X.; Liu, S.; Chen, R.; Wang, W.; Wang, C. A Deep Reinforcement Learning-Based Framework for Dynamic Resource Allocation in Multibeam Satellite Systems. *IEEE Commun. Lett.* **2018**, *22*, 1612–1615. [CrossRef]
4. Tilwari, V.; Bani-Bakr, A.; Qamar, F.; Hindia, M.N.; Jayakody, D.N.K.; Hassan, R. Mobility and queue length aware routing approach for network stability and load balancing in MANET. In Proceedings of the 2021 International Conference on Electrical Engineering and Informatics (ICEEI), Kuala Terengganu, Malaysia, 12–13 October 2021; pp. 1–5.
5. Abbas, T.; Qamar, F.; Hindia, M.N.; Hassan, R.; Ahmed, I.; Aslam, M.I. Performance analysis of ad hoc on-demand distance vector routing protocol for MANET. In Proceedings of the 2020 IEEE Student Conference on Research and Development (SCOReD), Johor, Malaysia, 27–29 September 2020; pp. 194–199.

6. Ahmadi, H.; Bouallegue, R. Exploiting machine learning strategies and RSSI for localization in wireless sensor networks: A survey. In Proceedings of the 13th International Wireless Communications and Mobile Computing Conference (IWCMC), Valencia, Spain, 26–30 June 2017; pp. 1150–1154.
7. Bangotra, D.K.; Singh, Y.; Selwal, A. Machine Learning in Wireless Sensor Networks: Challenges and Opportunities. In Proceedings of the Fifth International Conference on Parallel, Distributed and Grid Computing, Solan, Himachal Pradesh, India, 20–22 December 2018.
8. Mastronarde, N.; Modares, J.; Wu, C.; Chakareski, J. Reinforcement Learning for Energy-Efficient Delay-Sensitive CSMA/CA Scheduling. In Proceedings of the IEEE Global Communications Conference (GLOBECOM), Washington, DC, USA, 4–8 December 2016; pp. 1–7.
9. Khamukhin, A.A.; Bertoldo, S. Spectral analysis of forest fire noise for early detection using wireless sensor networks. In Proceedings of the International Siberian Conference on Control and Communications (SIBCON), Moscow, Russia, 12–14 May 2016; pp. 1–4.
10. Lu, J.; Huang, D.D. A slotted CSMA based reinforcement learning approach for extending the lifetime of underwater acoustic wireless sensor networks. *Comput. Commun.* **2013**, *36*, 1094–1099.
11. Alhassan, I.B.; Mitchell, P.D. Packet flow based reinforcement learning MAC protocol for underwater acoustic sensor networks. *Sensors* **2021**, *21*, 2284. [CrossRef] [PubMed]
12. Gazi, F.; Ahmed, N.; Misra, S.; Wei, W. Reinforcement Learning-Based MAC Protocol for Underwater Multimedia Sensor Networks. *ACM Trans. Sens. Netw. (TOSN)* **2022**. [CrossRef]
13. Kassab, R.; Destounis, A.; Tsilimantos, D.; Debbah, M. Multi-agent deep stochastic policy gradient for event based dynamic spectrum access. In Proceedings of the 2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications, London, UK, 31 August–3 September 2020; pp. 1–6.
14. Zhu, J.; Song, Y.; Jiang, D.; Song, H. A new deep-Q-learning-based transmission scheduling mechanism for the cognitive Internet of Things. *IEEE Internet Things J.* **2017**, *5*, 2375–2385. [CrossRef]
15. Tang, F.; Mao, B.; Fadlullah, Z.M.; Kato, N. On a novel deep-learning-based intelligent partially overlapping channel assignment in SDN-IoT. *IEEE Commun. Mag.* **2018**, *56*, 80–86. [CrossRef]
16. Jiang, N.; Deng, Y.; Nallanathan, A.; Chambers, J.A. Reinforcement learning for real-time optimization in NB-IoT networks. *IEEE J. Sel. Areas Commun.* **2019**, *37*, 1424–1440. [CrossRef]
17. Ye, H.; Li, G.Y.; Juang, B.H.F. Deep Reinforcement Learning Based Resource Allocation for V2V Communications. *IEEE Trans. Veh. Technol.* **2019**, *68*, 3163–3173. [CrossRef]
18. Chen, M.; Saad, W.; Yin, C. Liquid state machine learning for resource and cache management in LTE-U unmanned aerial vehicle (UAV) networks. *IEEE Trans. Wirel. Commun.* **2019**, *18*, 1504–1517. [CrossRef]
19. Chen, M.; Wang, T.; Ota, K.; Dong, M.; Zhao, M.; Liu, A. Intelligent resource allocation management for vehicles network: An A3C learning approach. *Comput. Commun.* **2020**, *151*, 485–494. [CrossRef]
20. Hoel, C.J.; Driggs-Campbell, K.; Wolff, K.; Laine, L.; Kochenderfer, M.J. Combining planning and deep reinforcement learning in tactical decision making for autonomous driving. *IEEE Trans. Intell. Veh.* **2019**, *5*, 294–305. [CrossRef]
21. Daknou, E.; Tabbane, N.; Thaalbi, M. A MAC multi-channel scheme based on learning-automata for clustered VANETs. In Proceedings of the 2018 IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA), Krakow, Poland, 16–18 May 2018; pp. 71–78.
22. Shah, A.S.; Ilhan, H.; Tureli, U. Designing and Analysis of IEEE 802.11 MAC for UAVs Ad Hoc Networks. In Proceedings of the 2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York, NY, USA, 10–12 October 2019; pp. 934–939.
23. Zheng, Z.; Sangaiah, A.K.; Wang, T. Adaptive Communication Protocols in Flying Ad Hoc Network. *IEEE Commun. Mag.* **2018**, *56*, 136–142. [CrossRef]
24. Rezwan, S.; Choi, W. A survey on applications of reinforcement learning in flying ad-hoc networks. *Electronics* **2021**, *10*, 449. [CrossRef]
25. Karabulut, M.A.; Shah, A.S.; Ilhan, H. A Novel MIMO-OFDM Based MAC Protocol for VANETs. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 20255–20267. [CrossRef]
26. Ding, H.; Zhao, F.; Tian, J.; Li, D.; Zhang, H. A deep reinforcement learning for user association and power control in heterogeneous networks. *Ad Hoc Netw.* **2020**, *102*, 102069. [CrossRef]
27. Zhao, N.; Liang, Y.C.; Niyato, D.; Pei, Y.; Wu, M.; Jiang, Y. Deep Reinforcement Learning for User Association and Resource Allocation in Heterogeneous Cellular Networks. *IEEE Trans. Wirel. Commun.* **2019**, *18*, 5141–5152. [CrossRef]
28. Yu, Y.; Wang, T.; Liew, S.C. Deep-Reinforcement Learning Multiple Access for Heterogeneous Wireless Networks. *IEEE J. Sel. Areas Commun.* **2019**, *37*, 1277–1290. [CrossRef]
29. Zhang, L.; Tan, J.; Liang, Y.C.; Feng, G.; Niyato, D. Deep reinforcement learning-based modulation and coding scheme selection in cognitive heterogeneous networks. *IEEE Trans. Wirel. Commun.* **2019**, *18*, 3281–3294. [CrossRef]
30. Shah, A.S.; Qasim, A.N.; Karabulut, M.A.; Ilhan, H.; Islam, M.B. Survey and performance evaluation of multiple access schemes for next-generation wireless communication systems. *IEEE Access* **2021**, *9*, 113428–113442. [CrossRef]

31. Shah, A.S. A Survey From 1G to 5G Including the Advent of 6G: Architectures, Multiple Access Techniques, and Emerging Technologies. In Proceedings of the 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC), Virtual, 26–29 January 2022; pp. 1117–1123.
32. Tang, F.; Mao, B.; Kawamoto, Y.; Kato, N. Survey on Machine Learning for Intelligent End-to-End Communication Toward 6G: From Network Access, Routing to Traffic Control and Streaming Adaption. *IEEE Commun. Surv. Tutor.* **2021**, *23*, 1578–1598. [CrossRef]
33. Kakalou, I.; Papadimitriou, G.I.; Nicopolitidis, P.; Sarigiannidis, P.G.; Obaidat, M.S. A Reinforcement learning-based cognitive MAC protocol. In Proceedings of the IEEE International Conference on Communications (ICC), London, UK, 8–12 June 2015; pp. 5608–5613.
34. Nisioti, E.; Thomos, N. Robust Coordinated Reinforcement Learning for MAC Design in Sensor Networks. *IEEE J. Sel. Areas Commun.* **2019**, *37*, 2211–2224. [CrossRef]
35. Deshpande, Y.; Ayan, O.; Kellerer, W. Improving AoI via Learning-based Distributed MAC in Wireless Networks. *arXiv* **2022**, arXiv:2202.09222.
36. Yang, H.; Zhong, W.D.; Chen, C.; Alphones, A.; Xie, X. Deep-reinforcement-learning-based energy-efficient resource management for social and cognitive internet of things. *IEEE Internet Things J.* **2020**, *7*, 5677–5689. [CrossRef]
37. Moon, S.; Ahn, S.; Son, K.; Park, J.; Yi, Y. Neuro-DCF: Design of Wireless MAC via Multi-Agent Reinforcement Learning Approach. In Proceedings of the Twenty-Second International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing, Seoul, Republic of Korea, 26–29 July 2021 ; pp. 141–150. [CrossRef]
38. Chou, P.Y.; Chen, W.Y.; Wang, C.Y.; Hwang, R.H.; Chen, W.T. Deep Reinforcement Learning for MEC Streaming with Joint User Association and Resource Management. In Proceedings of the ICC 2020—2020 IEEE International Conference on Communications (ICC), Dublin, Ireland, 7–11 June 2020; pp. 1–7. [CrossRef]
39. Dutta, H.; Biswas, S. Distributed Reinforcement Learning for scalable wireless medium access in IoTs and sensor networks. *Comput. Netw.* **2022**, *202*, 108662. [CrossRef]
40. Wang, S.; Liu, H.; Gomes, P.H.; Krishnamachari, B. Deep reinforcement learning for dynamic multichannel access in wireless networks. *IEEE Trans. Cogn. Commun. Netw.* **2018**, *4*, 257–265. [CrossRef]
41. Li, W.; Zhou, F.; Chowdhury, K.R.; Meleis, W. QTCP: Adaptive congestion control with reinforcement learning. *IEEE Trans. Netw. Sci. Eng.* **2018**, *6*, 445–458. [CrossRef]
42. Zafaruddin, S.M.; Bistritz, I.; Leshem, A.; Niyato, D. Distributed learning for channel allocation over a shared spectrum. *IEEE J. Sel. Areas Commun.* **2019**, *37*, 2337–2349. [CrossRef]
43. Bao, S.; Fujii, T. Q-learning Based p-pesistent CSMA MAC Protcol for Secondary User of Cognitive Radio Networks. In Proceedings of the Third International Conference on Intelligent Networking and Collaborative Systems, Fukuoka, Japan, 30 November–2 December 2011; pp. 336–337.
44. Nguyen, T.T.; Oh, H. SCSMA: A Smart CSMA/CA Using Blind Learning for Wireless Sensor Networks. *IEEE Trans. Ind. Electron.* **2020**, *67*, 10981–10988. [CrossRef]
45. Nisioti, E.; Thomos, N. Fast Q-Learning for Improved Finite Length Performance of Irregular Repetition Slotted ALOHA. *IEEE Trans. Cogn. Commun. Netw.* **2020**, *6*, 844–857. [CrossRef]
46. Liu, Y.; Liu, H.; Fang, H.; Li, P.; Yu, C.; Zhang, H. Enhancement Objective Q-Learning MAC for Emergency Transmission. In Proceedings of the 6th International Symposium on Electromagnetic Compatibility (ISEMC), Milan, Italy, 23–27 June 2019; pp. 1–5.
47. Klaine, P.V.; Imran, M.A.; Onireti, O.; Souza, R.D. A Survey of Machine Learning Techniques Applied to Self-Organizing Cellular Networks. *IEEE Commun. Surv. Tutor.* **2017**, *19*, 2392–2431. [CrossRef]
48. Li, Z.; Guo, C. Multi-agent deep reinforcement learning based spectrum allocation for D2D underlay communications. *IEEE Trans. Veh. Technol.* **2019**, *69*, 1828–1840. [CrossRef]
49. Nakashima, K.; Kamiya, S.; Ohtsu, K.; Yamamoto, K.; Nishio, T.; Morikura, M. Deep reinforcement learning-based channel allocation for wireless lans with graph convolutional networks. *IEEE Access* **2020**, *8*, 31823–31834. [CrossRef]
50. Chen, S.; Chen, J.; Chen, J. A deep reinforcement learning based network management system in smart identifier network. In Proceedings of the 2020 4th International Conference on Digital Signal Processing, Chengdu, China, 19–21 June 2020; pp. 268–273.
51. Tang, F.; Mao, B.; Fadlullah, Z.M.; Liu, J.; Kato, N. ST-DeLTA: A novel spatial-temporal value network aided deep learning based intelligent network traffic control system. *IEEE Trans. Sustain. Comput.* **2019**, *5*, 568–580. [CrossRef]
52. Mao, Q.; Hu, F.; Hao, Q. Deep Learning for Intelligent Wireless Networks: A Comprehensive Survey. *IEEE Commun. Surv. Tutor.* **2018**, *20*, 2595–2621. [CrossRef]
53. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*; MIT Press: Cambridge, MA, USA, 2018.
54. Patterson, J.; Gibson, A. *Deep Learning: A Practitioner's Approach*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2017.
55. Jiang, S. State-of-the-Art Medium Access Control (MAC) Protocols for Underwater Acoustic Networks: A Survey Based on a MAC Reference Model. *IEEE Commun. Surv. Tutor.* **2017**, *20*, 96–131. [CrossRef]
56. Zhang, C.; Patras, P.; Haddadi, H. Deep Learning in Mobile and Wireless Networking: A Survey. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 2224–2287. [CrossRef]

57. Zuriatunadhirah binti Zubir, N.; Ramli, A.F.; Basarudin, H. Optimization of wireless sensor networks MAC protocols using machine learning; a survey. In Proceedings of the 2017 International Conference on Engineering Technology and Technopreneurship (ICE2T), Kuala Lumpur, Malaysia, 18–20 September 2017; pp. 1–5. [CrossRef]
58. Kherbache, M.; Sobirov, O.; Maimour, M.; Rondeau, E.; Benyahia, A. Reinforcement Learning TDMA-Based MAC Scheduling in the Industrial Internet of Things: A Survey. *IFAC-PapersOnLine* **2022**, *55*, 83–88. [CrossRef]
59. Kulin, M.; Kazaz, T.; De Poorter, E.; Moerman, I. A Survey on Machine Learning-Based Performance Improvement of Wireless Networks: PHY, MAC and Network Layer. *Electronics* **2021**, *10*, 318. [CrossRef]
60. Sun, Y.; Peng, M.; Zhou, Y.; Huang, Y.; Mao, S. Application of Machine Learning in Wireless Networks: Key Techniques and Open Issues. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 3072–3108. [CrossRef]
61. Cao, X.; Yang, B.; Huang, C.; Yuen, C.; Di Renzo, M.; Han, Z.; Niyato, D.; Poor, H.V.; Hanzo, L. AI-Assisted MAC for Reconfigurable Intelligent-Surface-Aided Wireless Networks: Challenges and Opportunities. *IEEE Commun. Mag.* **2021**, *59*, 21–27. [CrossRef]
62. Sharma, H.; Haque, A.; Blaabjerg, F. Machine Learning in Wireless Sensor Networks for Smart Cities: A Survey. *Electronics* **2021**, *10*, 1012. [CrossRef]
63. Bithas, P.S.; Michailidis, E.T.; Nomikos, N.; Vouyioukas, D.; Kanatas, A.G. A Survey on Machine-Learning Techniques for UAV-Based Communications. *Sensors* **2019**, *19*, 5170. [CrossRef] [PubMed]
64. Winston, P.H. *Artificial Intelligence*; Addison-Wesley Longman Publishing Co., Inc.: Upper Saddle River, NJ, USA, 1992.
65. Xiong, Z.; Zhang, Y.; Niyato, D.; Deng, R.; Wang, P.; Wang, L.C. Deep Reinforcement Learning for Mobile 5G and Beyond: Fundamentals, Applications, and Challenges. *IEEE Veh. Technol. Mag.* **2019**, *14*, 44–52. [CrossRef]
66. Watkins, C.; Dayan, P. Technical Note: Q-Learning. *Mach. Learn.* **1992**, *8*, 279–292. [CrossRef]
67. Beletsioti, G.A.; Papadimitriou, G.I.; Nicopolitidis, P.; Varvarigos, E.; Mavridopoulos, S. A Learning-Automata-Based Congestion-Aware Scheme for Energy-Efficient Elastic Optical Networks. *IEEE Access* **2020**, *8*, 101978–101992. . ACCESS.2020.2996279. [CrossRef]
68. Narendra, K.S.; Thathachar, M.A.L. Learning Automata—A Survey. *IEEE Trans. Syst. Man Cybern.* **1974**, *SMC-4*, 323–334. [CrossRef]
69. Li, Y. Deep Reinforcement Learning: An Overview. *arXiv* **2017**, arXiv:1701.07274.
70. Khan, Z.A.; Samad, A. A Study of Machine Learning in Wireless Sensor Network. *Int. J. Comput. Netw. Appl.* **2017**, *4*, 105–112. [CrossRef]
71. De Rango, F.; Cordeschi, N.; Ritacco, F. Applying Q-learning approach to CSMA Scheme to dynamically tune the contention probability. In Proceedings of the 2021 IEEE 18th Annual Consumer Communications Networking Conference (CCNC), Las Vegas, NV, USA, 9–12 January 2021; pp. 1–4. [CrossRef]
72. Lee, D.J.; Deng, Y.; Choi, Y.J. Back-off Improvement By Using Q-learning in IEEE 802.11p Vehicular Network. In Proceedings of the 2020 International Conference on Information and Communication Technology Convergence (ICTC), Jeju Island, Republic of Korea, 21–23 October 2020; pp. 1819–1821. [CrossRef]
73. Chen, Y.W.; Kao, K.C. Study of Contention Window Adjustment for CSMA/CA by Using Machine Learning. In Proceedings of the 2021 22nd Asia-Pacific Network Operations and Management Symposium (APNOMS), Tainan, Taiwan, 8–10 September 2021; pp. 206–209. [CrossRef]
74. Kim, T.W.; Hwang, G.H. Performance Enhancement of CSMA/CA MAC Protocol Based on Reinforcement Learning. *J. Inf. Commun. Converg. Eng.* **2021**, *19*, 1–7.
75. Ali, R.; Shahin, N.; Zikria, Y.B.; Kim, B.S.; Kim, S.W. Deep reinforcement learning paradigm for performance optimization of channel observation–based MAC protocols in dense WLANs. *IEEE Access* **2018**, *7*, 3500–3511. [CrossRef]
76. Lee, S.; Chung, S.H. Unslotted CSMA/CA mechanism with reinforcement learning of Wi-SUN MAC layer. In Proceedings of the 2022 Thirteenth International Conference on Ubiquitous and Future Networks (ICUFN), Barcelona, Spain, 5–8 July 2022; pp. 202–204. [CrossRef]
77. Barbosa, P.F.C.; Silva, B.A.d.; Zanchettin, C.; de Moraes, R.M. Energy Consumption Optimization for CSMA/CA Protocol Employing Machine Learning. In Proceedings of the 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring), Antwerp, Belgium, 25–28 May 2020; pp. 1–5. [CrossRef]
78. Zhao, Y.; Hu, J.; Yang, K.; Cui, S. Deep Reinforcement Learning Aided Intelligent Access Control in Energy Harvesting Based WLAN. *IEEE Trans. Veh. Technol.* **2020**, *69*, 14078–14082. [CrossRef]
79. Bayat-Yeganeh, H.; Shah-Mansouri, V.; Kebriaei, H. A multi-state Q-learning based CSMA MAC protocol for wireless networks. *Wirel. Netw.* **2018**, *24*, 1251–1264. [CrossRef]
80. Aboubakar, M.; Roux, P.; Kellil, M.; Bouabdallah, A. An Efficient and Adaptive Configuration of IEEE 802.15.4 MAC for Communication Delay Optimisation. In Proceedings of the 2020 11th International Conference on Network of the Future (NoF), Bordeaux, France, 12–14 October 2020; pp. 1–7. [CrossRef]
81. Cho, S. Rate adaptation with Q-learning in CSMA/CA wireless networks. *J. Inf. Process. Syst.* **2020**, *16*, 1048–1063.
82. Cho, S. Reinforcement Learning for Rate Adaptation in CSMA/CA Wireless Networks. In *Advances in Computer Science and Ubiquitous Computing*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 175–181.

83. Zhang, M.; de Alfaro, L.; Garcia-Luna-Aceves, J. Using reinforcement learning in slotted aloha for ad-hoc networks. In Proceedings of the 23rd International ACM Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems, Alicante, Spain, 16–20 November 2020; pp. 245–252.
84. Zhang, L.; Yin, H.; Zhou, Z.; Roy, S.; Sun, Y. Enhancing WiFi Multiple Access Performance with Federated Deep Reinforcement Learning. In Proceedings of the 2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall), Virtual, 18 November–16 December 2020; pp. 1–6. [CrossRef]
85. Mennes, R.; Claeys, M.; De Figueiredo, F.A.P.; Jabandžić, I.; Moerman, I.; Latré, S. Deep Learning-Based Spectrum Prediction Collision Avoidance for Hybrid Wireless Environments. *IEEE Access* **2019**, *7*, 45818–45830. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.