Game-based hierarchical multi-armed bandit learning algorithm for joint channel and power allocation in underwater acoustic communication networks

Song Han\textsuperscript{a}, Xinbin Li\textsuperscript{a,b}, Lei Yan\textsuperscript{a}, Zhixin Liu\textsuperscript{a}, Xinping Guan\textsuperscript{a,b}

\textsuperscript{a} Institute of Electrical Engineering, Yanshan University, Qinhuangdao 066004, China
\textsuperscript{b} School of Electronic Information and Electrical Engineering, Shanghai Jiaotong University, Shanghai 200030, China

\section{A R T I C L E   I N F O}

Article history:
Received 21 April 2017
Revised 28 October 2017
Accepted 3 February 2018
Available online xxx

Communicated by Prof. W. K. Wong

Keywords:
Underwater acoustic communication networks
Learning algorithm
Game theory
MAB
Power allocation

\section{A B S T R A C T}

This study considers a joint channel and power allocation for multiple users in underwater acoustic communication networks as a formulated multiplayer MAB game. This study also proposes hierarchical learning algorithms, which do not need any prior environmental information and direct information exchange among users, to improve the learning ability. In upper sub-learning, each user generates a strategy through the traditional UCB1 strategy. In lower sub-learning, the concept of virtual learning information, which can be obtained as the reward of the last actual played strategy, is introduced to enrich the learning information. Users can enhance their learning ability by learning the outdated virtual learning information in lower sub-learning. As a result, the learning time it takes to achieve the NE is effectively decreased, and the cost of the algorithm is reduced. A distributed optimal NE selection mechanism is proposed to avoid falling into an inadequate local extreme value. Simulation results show high convergence speed and achieved utility of the proposed algorithm.

\section{1. Introduction}

Underwater communication technology has recently been the focus of considerable attention from the marine industry and academia \cite{1-3}. Underwater communication networks can facilitate a wide range of civilian and military applications, including underwater environment monitoring, mine reconnaissance, disaster prediction, coastline protection, and target tracking \cite{4,5}. Acoustic communication has been proven to be the only practical method for underwater application because of the high attenuation of radio waves caused by the conductivity of water \cite{6}. This finding motivates more study on underwater acoustic communication networks (UACNs).

UACNs face greater connatural challenges than radio communication \cite{7}. (1) Limited bandwidth: current acoustic communication system is up to 40 km kbps for the range-rate product; (2) limited energy: most existing acoustic modems are battery powered and the batteries are very hard to be changed or recharged in the underwater environment. Thus, in the view of overhead, the low-cost totally distributed resource allocation algorithm is more significant in UACNs. Furthermore, in contrast to radio communication networks, the fundamental question of underwater acoustic resource allocation has been an elusive one, mainly because of the lack of well-established channel models \cite{8}. In details, the speed of sound underwater is approximately 1500 m/s, which is $2 \times 10^5$ times slower than the speed of radio and it leads to more complex channel characteristics. In contrast to radio communication networks, the prior channel information is harder to obtain in UACNs.

Although the initial applications involved only single transmitter–receiver pair communication, underwater communications today are associated with complex systems that require multiple devices, which use sounds, to exist in the same area \cite{9}. The existence of multiple users leads to more limited resources and more complex communication environments. Therefore, a high resource utilization rate and a low overhead resource allocation approach for multiuser UACNs are urgently needed.

The resource allocation problem has been widely investigated in radio communication networks. The studies reported in \cite{10,11} considered the joint channel and power allocation problem as a resource optimization problem. However, the schemes based on traditional optimization theory cannot be applied in a distributed manner. The centralized scheme needs a central pro-
cessing unit to manage the networks. However, building a central processing unit in an underwater environment is difficult. Thus, the resource allocation schemes for UACNs must be distributed. Many distributed schemes based on game theory have been investigated for the resource allocation problem. In [12], the authors modeled the joint power allocation and user scheduling problem in multicell networks as a potential game and presented an efficient decentralized iterative algorithm to obtain the global optimal solution (NE). Ref. [13] formulated a novel game-theoretic framework to study the optimal power control for interference management in ultra-dense small cell networks. By properly designing the utility function with dynamic pricing and low-complexity distributed power control algorithm, leading to desirable performance outcomes. The game-based algorithms in [12,13] need only the local information exchange, thereby decreasing the complexity of the algorithms and the cost of management. In [14–16], the authors presented efficient game-based algorithms from different views, such as outage possibility, pricing power, and interference control. Nevertheless, the proposed algorithms in [14–16] considered only one channel model. Thus, these algorithms cannot be applied in multichannel orthogonal frequency division multiple access (OFDMA) systems, which can achieve a high width utilization rate [17]. Although joint channel and power allocation is more complex than single resource allocation, it is more effective for wireless networks because of the OFDMA system. Iterative water (IW) algorithm is a classical joint allocation algorithm [18]. Many studies [19,20] based on the IW algorithm have been well investigated. However, the algorithms for radio communication networks assumed that the designated user must have perfect knowledge of some channel gains [12,13] or channel selection strategies of all other users [19,20]. In UACNs with limited bandwidth and low rate, the cost of the aforementioned assumption is high. Furthermore, all of the IW-based algorithms depend on the assumption that all of the users can sense the interference information in a timely manner. However, in the face of a complicated underwater acoustic communication environment, the timely sense cannot be satisfied. Thus, the traditional IW-based algorithms cannot be easily applied in UACNs.

Only a few studies focused on the resource allocation algorithms in UACNs. The focal point of the most relevant studies [8,9,21] is channel estimation to obtain channel information. Then, the estimated channel information is used to allocate the resource. In [21], a medium access control (MAC) protocol was presented to obtain channel information, and Nash equilibrium (NE) was employed to adjust transmission power. The proposed protocol improved the throughput of the networks. Nevertheless, only one channel was considered in the model. In [9], the authors presented a cross-protocol to solve the joint resource allocation problem in acoustic networks with the condition that only one user was allowed to access a single channel. Thus, the methods presented in [9,21] are inapplicable to multichannel free-access UACNs. In [8], an effective channel estimation method was used to determine whether the channels are good or bad, and the transmission power was uniformly allocated to the good channels. The method effectively improved the robustness of the UACNs. The previously presented analysis reveals that most of the previous studies focused on obtaining channel information, rather than utilizing the resource allocation algorithm, mainly because the underwater acoustic communication environment is complex and well-established statistical channel models are lacking [8].

When the a priori information is poor, the resource allocation problem is often solved by using learning algorithms. The multi-armed bandit (MAB) is a machine-learning decision-making theory for the model including unknown variables [22,23]. For instance, the UCB1 algorithm can solve the MAB problem by analyzing the history data, rather than obtaining the unknown variables. The MAB has been reported to solve the resource allocation problem in cognitive radio networks [24,25]. The works of [24,25] employed MAB to route and select spectrum access. For power allocation, a low-cost UCB-based cognitive water-filling algorithm, which considered one user and multichannel networks, was presented in [26] to solve the MAB problem. In [27], a MAB-based algorithm was presented for multichannel and multiple-user networks with the condition that only one channel can be accessed by one user. The aforementioned MAB-based algorithms effectively overcame the poor a priori channel information and improved the performance of wireless networks. However, the learning time of the existing MAB-based algorithms for radio networks is relatively long. Generally, the learning time it takes to achieve the expected strategy is more than 10,000. The UACNs cannot handle a high learning cost. The stringent requirements are the main reason that a few studies focused on underwater acoustic resource allocation algorithm [28]. Consequently, this study presents a MAB algorithm with high learning ability for UACNs.

This study addresses the issue of joint channel and power allocation with multiple channels and users in UACNs. This study presents a resource allocation algorithm that is distributed and has low information exchange, low cost, and low dependence on channel information. First, a multiplayer MAB game is formulated to model the problem. In this model, the objective of the users is to maximize their own utility. Given that the UCB1 algorithm is a classical approach employed to solve the MAB problem, the users optimize their strategy by using an improved game-based UCB1 algorithm. Second, two hierarchical learning algorithms are presented to enhance the learning ability of users. Specifically, in the proposed hierarchical learning algorithms, the concept of “virtual learning information”, which can be obtained as the actual played history data, is introduced. In contrast to the one-tier traditional UCB1 algorithm, the user learns the outdated virtual learning information in the added lower sub-learning. Thus, the learning ability can be improved, and the learning time can be reduced. Then, a distributed optimal NE selection mechanism is presented to avoid falling into an inadequate local extreme value. Finally, the analysis and comparison of the proposed algorithms are shown. The numerical results show the high learning ability and superior performance of the proposed hierarchical learning algorithms.

The main contributions of this study are as follows:

1) We formulate a multiplayer MAB game to study the joint channel and power allocation problem in OFDMA UACNs. And we present learning algorithms to solve the formulated game. Specifically, the proposed learning algorithms can be performed in totally distributed manner, and the users do not need any prior information and direct information exchange, thereby decreasing the cost of resource allocation.

2) We present a concept of “virtual learning information” to enrich the learning information, thereby improving the learning ability. Specifically, the “virtual learning information” can be obtained by utilizing only the local information.

3) Different from the traditional single-tier learning algorithm, we present the lower learning where the players learn the “virtual learning information” in the proposed hierarchical learning algorithms. The proposed two-tier learning manner effectively improves the learning ability.

4) We present an optimal NE selection mechanism to avoid falling into an inadequate local extreme value and achieve the optimal NE. Specifically, the proposed mechanism does not bring any new information exchange and can be performed distributedly.

The rest of the study is organized as follows. Section 2 describes the model for UACNs and Multiplayer MAB game. In Section 3, the joint resource allocation problem formulation is presented. The details of the proposed algorithms and the analysis are shown in...