

Received 20 February 2023, accepted 7 April 2023, date of publication 11 April 2023, date of current version 22 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3266331

RESEARCH ARTICLE

Learning and Game Based Spectrum Allocation Model for Internet of Medical Things (IoMT) Platform

SUNGWOOK KIM 

Department of Computer Science, Sogang University, Mapo-gu, Seoul 04107, South Korea

e-mail: swkim01@sogang.ac.kr

This work was supported in part by the Ministry of Science and ICT (MSIT), South Korea, through the Information Technology Research Center (ITRC) Support Program, supervised by the Institute for Information and Communications Technology Planning and Evaluation (IITP), under Grant IITP-2022-2018-0-01799; and in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant 2021R1F1A1045472.

ABSTRACT The Internet of Medical Things (IoMT) paradigm provides pervasive healthcare services in-home monitoring networks. Nowadays, these services play an imperative part in the life of human beings. However, excessive requirements of health services result in insufficient spectrum resources and service delays. In this study, a novel spectrum allocation scheme is proposed for the IoMT system platform. The main challenge of our scheme is to effectively share the limited spectrum resource while dynamically handling different service requests. To achieve a mutually desirable solution for multiple IoMT devices, our proposed scheme is designed as a bi-level control algorithm using the ideas of *multi-agent reinforcement learning (MARL)* and the *Balakrishnan-Gómez-Vohra (BGV)* solution. At the first level, each IoMT device selects its salient point according to the *MARL* model. At the second level, the spectrum resource is distributed through the *BGV* solution, which is implemented by considering the selected salient point of each device. Through the sequential interactions of intelligent devices, our bi-level control approach can effectively guide individual IoMT devices to choose cooperation strategies while optimizing the spectrum allocation process. Finally, numerical results show the effectiveness of our proposed scheme through the comparisons with benchmark protocols. We demonstrate the performance improvement of our method in terms of the normalized device payoff, IoMT system throughput and device fairness.


INDEX TERMS Internet of Medical Things, multi-agent reinforcement learning, Balakrishnan-Gómez-Vohra solution, bi-level spectrum allocation, cooperative game theory.

I. INTRODUCTION

Over recent years, wireless technology has been one of the fastest-growing technologies in the area of communications. Today, it is agreed that wireless communication is becoming one of the largest carriers of digital data around the globe. Currently, smart mobile devices that exploit seamless connectivity offered by mobile wireless networks are deployed everywhere and impact the most various of contexts. Regarding the potential use cases of wireless technology and smart

devices, the Internet of Things (IoT) has gained widespread attention and it is set to take a leading role in the future networks. Originally, the concept of IoT was most interesting in the business and industrial fields, but focus has shifted on filling homes and workplaces with smart devices. It constitutes an integral part of the future 6 generation (6G) network and has received much attention due to its great potential to deliver customer services in many aspects of modern conditions of living [1], [2].

In the IoT paradigm, various applications play a vital role to affect our daily life by connecting the physical environment to the cyberspace of communication systems. By keeping

The associate editor coordinating the review of this manuscript and approving it for publication was Jad Nasreddine .

these things in view, several applications are developed. Recently, health industry is changing drastically in developed countries as the life expectancy has abruptly raised. As the population of older adults has rapidly increased, chronic diseases are also increasingly pressuring these countries' healthcare systems. The Internet of Medical Things (IoMT) is an IoT branch dedicated to the healthcare industry. It combines both the reliability and safety of traditional medical devices and dynamicity, genericity and scalability capabilities of traditional IoT. By deploying various medical smart devices on numerous patients, IoMT is able to realize in-home monitoring remotely, and can satisfy a variety of healthcare requirements. In addition, pervasive health monitoring systems allow patients to move freely indoors without being restrained. The rapid evolution of IoMT technology has led the research community to envision a wide range of smart healthcare projects [2], [3]

In the IoMT system, wireless connected devices generate a large amount of signals, including multimodal data. Because of the huge, complex, and multidimensional nature of the medical data generated by connected healthcare devices, how to effectively transmit a huge amount of data becomes a difficult problem. In addition, IoMT devices constitute 40% of the IoT market at the end of 2020. It is expected to expand in the next couple of years due to the IoMT devices' potential contribution to provide ubiquitous health monitoring services. Understandably, this rapid increase in the number of medical devices and excessive healthcare data limit the development of IoMT. Especially, the scarcity of spectrum resources restricts the further implementation of IoMT applications, since real-time performance is required while satisfy the delay constraint of the time-sensitive tasks for medical information analysis. Therefore, adaptive spectrum sharing is a major challenge and key requirements for the success of IoMT systems. Additional requirement is the cooperation among different medical devices; it is brought by the interconnectivity of medical devices through the intelligent management policy [1], [2], [3], [4].

Motivated by the above discussion, we propose a new spectrum allocation scheme for the IoMT network system. Usually, spectrum allocation is one of the most important issues in network managements. To ensure the communication qualities, an effective dynamic spectrum sharing policy is essential to design novel IoMT control algorithms. Until now, many research efforts and mathematical methods have been unfolded to deal with spectrum allocation problems. Recently, reinforcement learning algorithms and game theory have captured the attention of researches because of their impressive abilities to model potential system dynamics as an intelligent control mechanism. Some scientist and researchers contribute comprehensive presentations of the relevant techniques to design control mechanisms from both the game theory and reinforcement learning concepts. The central idea of game theory is to model strategic interactions as a game between a set of players. This setting is often used as a

testbed for multi-agent reinforcement learning approaches. In this study, we develop a learning based game model, which is an important part in formulating, design, and successful operations for the IoMT spectrum allocation scenario.

Originally, game theory focused on purely strategic interactions. Since then, it has developed into a general framework for providing complex systems science with a systematic approach for deciding a series of Pareto-efficient strategies in cooperative situations or the best strategy in non-cooperative situations. From the perspective of the timing of behavior, games can be divided into static games and dynamic games. In static games, all players make decisions simultaneously, without knowledge of the strategies that are being chosen by other players. Unlike static games, dynamic games define the possible orders of the events and players iteratively play a similar stage game. Therefore, the players observe the outcome of the previous game round and make their decisions for the next game round. To maximize their profits, players can react adaptively to other players' decisions. Recently, dynamic games have attracted interest from fields as diverse as psychology, economics, biology, engineering and telecommunication fields [5].

Even though game theory is a prevalent tool of modeling and analyzing decision-making problems, game players are not smart in a decision-making process due to their lack of ability to interact with the environment. Therefore, game players cannot always be able to maximize their payoffs, which are consistent with their preferences among different alternative outcomes. Fortunately, the era of artificial intelligence (AI) is coming; it has become a near-ubiquitous technology in our communities, homes and workplaces that is helping to improve our day-to-day lives. AI is a term applied to computers, robots, or machines that exhibit aspects of human intelligence, reasoning and decision-making. One of the most popular AI technologies such as the reinforcement learning methods are particularly attractive to addressing the resource allocation issue. With the advent of reinforcement learning, dynamic game models have also gained importance within the AI community and computer science. In particular, dynamic game paradigm provides both a means to describe the problem setting for multi-agent learning algorithm and the tools to analyze the outcome of reinforcement learning [6].

Although there has been a surging interest in studying game theory and multi-agent reinforcement learning, no prior work has particularly focused on the sophisticated combination of these two control paradigms. In this study, our major goal is to develop a novel hybrid spectrum allocation scheme for the IoMT network system. In the proposed scheme, each IoMT device works simultaneously as a dynamic game player and learning agent. Based on the complicated interactions between game model and learning algorithm, we can get a good performance balance among various IoMT devices for different medical services.

The remainder of the paper is organized as follows. In Section II, the technical background for game theory

and multi-agent reinforcement learning are described. In Section III, we review the related work. Section IV illustrates the IoMT system platform, and formulates the spectrum allocation problem. And then, our proposed scheme is designed as a bi-level control algorithm to find an effective solution. In Section V, numerical results show the performance improvement of our proposed scheme, and the comparisons against the state-of-the-art benchmark protocols. Finally, conclusions and future expectations are drawn in Section VI.

II. TECHNICAL CONCEPTS AND MAIN CONTRIBUTIONS

As a subfield of game theory, cooperative games concern with the game players forming alliances and working together to seek to achieve their common goals. At the 1950, the classical model proposed by J. Nash has been one of the most successful paradigms of cooperative game theory. It has been the foundation stone of an extensive theoretical literature, and the solution idea that Nash defined and characterized has been widely applied. More broadly, Nash's approach has been deeply influential in demonstrating the power of the cooperative game model in the search for well-behaved solutions, and beyond that, it has been an important source of inspiration in the development of different cooperative game solutions. Until the present day, a number of game theory researchers have enriched the Nash's original idea through the introduction of a salient point [7], [8].

Recently, Balakrishnan, Gómez, and Vohra developed a new game solution concept, which is called as the *BGV* solution in this paper, by introducing a new salient point into bargaining problems. The *BGV* solution describes a cooperative situation where the salience of the reference point mutes or tempers the negotiators' aspirations. Therefore, the context of the cooperative bargain will affect the manner in which the salience point influences the negotiated outcome. To put it more concretely, the *BGV* solution employs two points; disagreement point and aspiration point. The disagreement point works as an anchor point, and the aspiration point is referenced as a tempered utopia or modified utopia point. The *BGV* solution is the intersection between a ray connecting the disagreement point with the aspiration point, and the bargaining set's boundary. This is equivalent to saying that the *BGV* solution chooses the maximum individually rational utility profile at which each negotiator's utility gain from his disagreement point has the same proportion to the utility difference between his aspiration point and his disagreement point [8], [9].

A multi-agent system is defined as a group of smart agents that sense and interact with an environment and act for pre-defined goals. It brings a new paradigm to design various control applications. Due to the dynamics and complexity of environments, reinforcement learning techniques have been developed for multi-agent systems in order to improve the performance of each agent and the whole system. In recent years, *multi-agent reinforcement learning (MARL)* has been proposed as an extension of reinforcement learning in a multi-agent domain. Current researches claim that *MARL*

can be treated as a fusion of policy search techniques on feasible sets, dynamic programming and game theory. More recently, the combination of game theory and *MARL* has become a research hotspot to explore the coordination and competition among multiple agents. The design of *MARL* models often involves a game-theoretic approach, so-called *learning games*, and demonstrate the existence of solutions by game theory. Therefore, we should adopt an effective learning technique to obtain a fair-efficient game solution in the dynamically changing multi-agent environment [10].

In this study, we aim to optimize the IoMT system performance by adopting the *BGV* solution and *MARL* model. The objective of formulated optimization problem is to maximize the normalized device payoff and network throughput while balancing the fairness of IoMT devices. The considered control scheme is divided into two levels, i.e., learning-level mechanism and bargaining-level mechanism. In the learning-level mechanism, each IoMT device decides its salient point through the *MARL* model. In the bargaining-level mechanism, the limited spectrum resource is shared among IoMT devices based on the fair-efficient manner; the *BGV* solution is applied to solve the spectrum allocation problem. For the efficient operation of IoMT system infrastructure, we guide selfish IoMT devices toward a socially optimal outcome. To satisfy this goal, the methodology adopted in our scheme is a coordinative learning game. The significant major contributions of the paper are summarized as follows:

- We construct a novel spectrum allocation scheme for health monitoring applications. Based on the fundamental concepts of the *BGV* solution and *MARL* model, we develop a bi-level control algorithm in the IoMT network platform.
- For the learning-level mechanism, each individual IoMT device intelligently learns the current system condition, and selects its salient point for the health monitoring applications. This process is operated in a parallel and distributed manner.
- For the bargaining-level mechanism, a cooperative game is considered to share the spectrum resource for individual IoMT devices. Considering the individual rationality of devices, the idea of *BGV* solution is adopted to reach a consensus with reciprocal advantage. By using a dynamic control fashion, this allocation process is operated in a centralized manner.
- Our jointly designed scheme enables the sequential interactions among multiple IoMT devices to achieve a mutually desirable solution. Our bi-level approach sophisticatedly combines two control mechanisms, which act iteratively and cooperatively to strike an appropriate system performance.
- Performance evaluations demonstrate the effectiveness of our proposed scheme in comparison with the existing state-of-the-art spectrum allocation protocols. Simulation analysis and numerical results show the efficiency of our hybrid approach in terms of the normalized device payoff, IoMT system throughput and device fairness.

III. RELATED WORK

Many previous researches have investigated health-aware application services. Since the concept of IoMT was first introduced, one of the most important issues is to effectively allocate the limited spectrum resource in the IoMT platform. Therefore, a few research papers have been published recently to solve the spectrum allocation problem [3], [11], [12]. The paper [3] proposes the *Health Monitoring for IoMT (HMIoMT)* scheme for in-home health monitoring services. In this scheme, the IoMT system is divided into intra-Wireless Body Area Networks (WBANs) and beyond-WBANs. For intra-WBANs, the Nash bargaining solution is adopted to optimally allocate the spectrum resource in a centralized manner. In beyond-WBANs, patients compete for channel and computation resources to process the health monitoring packets. By considering individual rationality and potential selfishness, the scheduling problem is formulated effectively. To obtain the strategy profile, a potential game is proposed in a decentralized manner to get the Nash equilibrium. Finally, performance evaluations demonstrate the effectiveness of the *HMIoMT* scheme with respect to the system-wide cost and the benefit of patients [3].

In [11], the *Resource Management for IoMT (RMIoMT)* scheme is designed to investigate the minimization problem for healthcare services in the IoMT system by considering quality-of-service (QoS) requirement, power limit, and wireless fronthaul constraint. Taking account of the distinct time sensitivity of medical data, a low-complexity algorithm is introduced to accelerate solution efficiency. Specifically, matching theory is adopted for the channel assignment between APs and IoMT devices. When the match process starts, each device first chooses the favorite channel according to its preference sequence. After being refused by channels, the device will go on sending access request to the following one in its preference sequence. When the set of unmatched IoMT devices is empty, the channel matching is finished. Finally, the simulation results reveal the effectiveness of the *RMIoMT* scheme [11].

Y. Yang et al. develop the *Distributed Learning for IoMT (DLIoMT)* scheme for the effective allocation of communication resource [12]. First, authors study the unique characters of IoMT and construct a distributed resource allocation problem. Second, they form a repeat game model by specifying the interactions of APs in the IoMT system. To reach a Nash equilibrium, each AP performs the strategy selection process based on probability learning algorithm, which is called as a multi-criticality strategy learning algorithm. Given the feature of medical data, medical criticality is divided into several categories. Each AP learns the strategy selection probabilities for different categories via the multi-criticality strategy learning algorithm. Multiple IoMT devices can transmit medical data to different APs, which treat these medical data packets equally. This setting guarantees the cooperative actions among IoMT devices. Finally, performance evaluations indicate the effectiveness of the *DLIoMT* scheme in terms of various aspects [12].

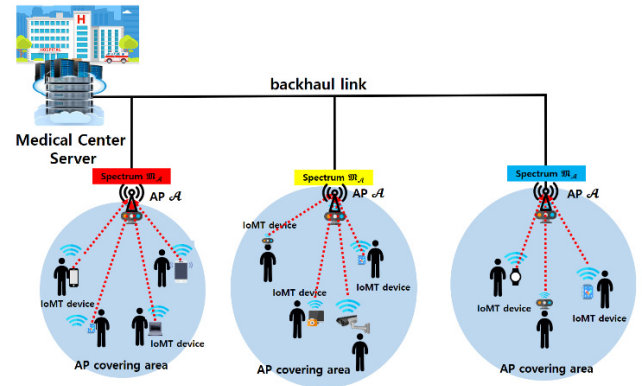


FIGURE 1. The infrastructure of IoMT system.

The earlier schemes in [3], [11], and [12] have been studied the spectrum allocation problem for the IoMT network platform. Even though some researchers tackled the IoMT resource allocation problem, they did not consider the combination of cooperative bargaining game and *MARL* algorithm for medical-aware application services. Compared to the above existing *HMIoMT*, *RMIoMT* and *DLIoMT* schemes, this article considers the ideas of *BGV* solution and distributed *MARL* model for controlling the activities of system agents, and develop a twofold spectrum allocation algorithm to ensure the performance of whole IoMT network system. To the best of our knowledge, our proposed scheme is the first in the literature to investigate the hybrid learning game mechanism, and guides intelligent IoMT devices toward a socially optimal outcome.

IV. HEALTHCARE-AWARE SPECTRUM ALLOCATION FOR THE IoMT PLATFORM

In this section, we consider a bi-level control mechanism for healthcare-aware application services in the IoMT platform. First, we formulate a learning game model for the spectrum allocation problem. And then, the fundamental ideas of the *BGV* solution and *MARL* model are introduced. Finally, we describe our proposed algorithm in detail.

A. IoMT SYSTEM INFRASTRUCTURE AND A LEARNING GAME MODEL

Based on the wearable technology, IoMT devices are always used for the individual health monitoring in daily life. In this subsection, we introduce the architecture of IoMT platform and formulate a bi-level learning game model for health-related application services. The architecture of heterogeneous IoMT network infrastructure is shown as Fig. 1. It consists of n access points (APs), i.e., $\mathbb{A} = \{A_1, \dots, A_n\}$ and m IoMT devices, i.e., $\mathbb{D} = \{D_1, \dots, D_m\}$. With low transmission power and cost, APs are envisioned to provide resilient communication services for multiple IoMT devices, which can gather, generate, fuse, analyze, and send medical data to their corresponding APs. Each $A_{1 \leq i \leq n} \in \mathbb{A}$ has its covering area, and \mathbb{D}_{A_i} is the set of A_i 's corresponding

IoMT devices where $\mathbb{D}_{\mathcal{A}_i} \subset \mathbb{D}$; they are randomly distributed in a given geographical area, and require a lot of spectrum resources for sending the medical data sensed from the patients every day. The \mathcal{A}_i has its wireless spectrum capacity ($\mathfrak{M}_{\mathcal{A}_i}$), and dynamically allocate the $\mathfrak{M}_{\mathcal{A}_i}$ for its corresponding IoMT devices. The IoMT device in the $\mathbb{D}_{\mathcal{A}_i}$, i.e., $\mathcal{D}_j \in \mathbb{D}_{\mathcal{A}_i}$, generates its health-aware data ($\mathcal{W}_{\mathcal{D}_j}$). To transmit the $\mathcal{W}_{\mathcal{D}_j}$, it leads to the conflict situation to share the $\mathfrak{M}_{\mathcal{A}_i}$ among the devices in the $\mathbb{D}_{\mathcal{A}_i}$.

In the proposed scheme, the first-level *MARL* learning model \mathbb{M} and the second-level cooperative game model \mathbb{G} are formulated. Through the \mathbb{M} and \mathbb{G} , multiple IoMT devices are sequentially interacted with each other to reach a consensus. Formally, we define the tuple entities in our dual-phase control scheme, such as

$$\{\mathbb{M}, \mathbb{G}\} = \left\{ \mathbb{A}, \mathbb{D}, \left\{ \mathbb{M}_{\mathcal{D}_k} \mid \mathcal{D}_k \in \mathbb{D}, a_{1 \leq k \leq l}^{\mathcal{D}_j} \in \mathcal{L}_{\mathcal{D}_j}, \mathcal{R}_a^{\mathcal{D}_j} \right\}, \left\{ \mathbb{G}_{\mathcal{A}_i} \mid \mathcal{A}_i \in \mathbb{A}, \mathcal{D}_j \in \mathbb{D}_{\mathcal{A}_i}, \mathfrak{M}_{\mathcal{A}_i}, \mathcal{S}_{\mathcal{D}_j}, \mathcal{U}_{\mathcal{D}_j}(\cdot) \right\}, T \right\}.$$

- \mathbb{A} and \mathbb{D} represent the set of APs and the set of IoMT devices, respectively.
- At the first-level, the $\mathbb{M}_{\mathcal{D}_k}$ is developed as a *MARL* model for the $\mathcal{D}_k \in \mathbb{D}$. It is operated in a distributed manner to learn the best action of \mathcal{D}_k .
- In the $\mathbb{M}_{\mathcal{D}_k}$, $a_k^{\mathcal{D}_j}$ is the \mathcal{D}_j 's k^{th} action, and $\mathcal{L}_{\mathcal{D}_j}$ is the \mathcal{D}_j 's action set, which consists of total l actions. $\mathcal{R}_a^{\mathcal{D}_j}$ is the \mathcal{D}_j 's reward function with the joint action a .
- At the second-level, each individual AP executes the $\mathbb{G}_{\mathcal{A}_i}$ in a centralized manner with its corresponding devices $\mathcal{D}_j \in \mathbb{D}_{\mathcal{A}_i}$.
- In the $\mathbb{G}_{\mathcal{A}_i}$, the \mathcal{A}_i 's spectrum resource ($\mathfrak{M}_{\mathcal{A}_i}$) is shared by devices in $\mathbb{D}_{\mathcal{A}_i}$. The $\mathcal{D}_j \in \mathbb{D}_{\mathcal{A}_i}$ is a game player, and $\mathcal{S}_{\mathcal{D}_j}$ and $\mathcal{U}_{\mathcal{D}_j}(\cdot)$ are his strategy and utility function, respectively.
- The $\mathbb{M}_{\mathcal{D}_k}$ and $\mathbb{G}_{\mathcal{A}_i}$ are reciprocally interdependent, and work together. It is noteworthy that we formulate the $\mathcal{A}_i - \mathbb{D}_{\mathcal{A}_i}$ association in an iterative manner.
- Discrete time model $T \in \{t_1, \dots, t_c, t_{c+1}, \dots\}$ is represented by a sequence of time steps. The length of t_c matches the event time-scale of $\mathbb{M}_{\mathcal{D}_k}$ and $\mathbb{G}_{\mathcal{A}_i}$.

B. FUNDAMENTAL IDEAS OF BGV SOLUTION AND MARL MODEL

To characterize the fundamental idea of *BGV* solution, the following notations will be used. Let $n > 1$ be a fixed natural number and define $\mathbb{N} = \{1, \dots, n\}$ and \mathbb{R}, \mathbb{R}^n denote the sets of all real numbers and the n -fold Cartesian product of \mathbb{R} , respectively. For any $x \in \mathbb{R}^n$, $i \in \mathbb{N}$ and $l \in \mathbb{R}$, let (l, x_{-i}) represent the vector $y \in \mathbb{R}^n$ such that $y_i = l$ and $y_j = x_j$ for any $j \neq i$. Vector inequalities are treated as follows. $x \geq y$ mean that $x_i \geq y_i$ for all $i \in \mathbb{N}$, $x > y$ indicates that $x \geq y$ and $x \neq y$ and $x \gg y$ means $x_i > y_i$. We denote $x \cdot y$ as the scalar product of the vectors, that is $x \cdot y = \sum_{i=1}^n (x_i \times y_i)$. An n -person bargaining problem with a reference point is a

triple (\mathcal{S}, d, r) where \mathcal{S} denotes the set of feasible outcomes, d is the disagreement point, and r is the reference point. We assume that $d, r \in \mathcal{S} \subset \mathbb{R}^n$, and $\exists x \in \mathcal{S}$ with $x > d$ and $r \geq d$. The assumption $d \in \mathcal{S}$ means that players are able to agree to disagree, the assumption $r \in \mathcal{S}$ means that the reference point is feasible, and the assumption $r \geq d$ means that the reference point is individually rational [8], [9].

By assuming that there exists $x \in \mathcal{S}$ with $x > d$, we rule out degenerate problems where no agreement can make all agents better-off than the disagreement outcome. Let $\Upsilon(\mathcal{S}, x)$ denote the aspiration vector such that for every $i \in \mathbb{N}$ and every $x \in \mathcal{S}$: $\Upsilon_i(\mathcal{S}, x) \equiv \max \{l \in \mathbb{R} \mid (l, x_{-i}) \in \mathcal{S}\}$. Accordingly, $\Upsilon(\mathcal{S}, d)$ is the ideal point and $\Upsilon(\mathcal{S}, r)$ is the tempered aspirations point. Let Σ^n be the class of all bargaining problems with a reference point. A solution concept for such problems is a function $\mathcal{F}: \Sigma^n \rightarrow \mathbb{R}^n$ that associates each $(\mathcal{S}, d, r) \in \Sigma^n$ with a unique point of \mathcal{S} . Finally, the mathematical definition of *BGV* solution, i.e., *BGV* (\mathcal{S}, d, r) , is given by [8] and [9]:

$$\begin{aligned} BGV(\mathcal{S}, d, r) &= (\lambda^* \cdot \Upsilon(\mathcal{S}, r)) + ((1 - \lambda^*) \cdot d) \\ \text{s.t., } &\begin{cases} \exists (\mathcal{S}, d, r) \in \Sigma^n \\ \lambda^* = \max \{ \lambda \in [0, 1] \mid ((\lambda \cdot \Upsilon(\mathcal{S}, r)) \\ + ((1 - \lambda) \cdot d)) \in \mathcal{S} \} \end{cases} \quad (1) \end{aligned}$$

Geometrically, the *BGV* solution is the maximum point of the bargaining set on the line segment connecting two points, which are $\Upsilon(\mathcal{S}, r)$ and d . The *BGV* solution is characterized by a collection of desirable axioms like as, *weakly Pareto optimal (WPO)*, *respect of symmetry (RS)*, *scale invariance (SI)*, *equal tempered aspirations conditional monotonicity (ETACM)*, *independence of trivial reference points (LSCCP)*, and *continuity (C)*. To explain these axioms, let Π^n be the class of one-to-one mappings $\pi: \mathbb{N} \rightarrow \mathbb{N}$. Given $x \in \mathbb{R}^n$, and $\pi \in \Pi^n$, let $\pi(x)$ denote the vector $(x_{\pi(i)})_{i \in \mathbb{N}}$. Given $\mathcal{S} \subset \mathbb{R}^n$, let $\pi(\mathcal{S}) \equiv \{y \in \mathbb{R}^n \mid \exists x \in \mathcal{S}, y = \pi(x)\}$. And, we define a solution function is any function $f: \Sigma^n \rightarrow \mathbb{R}^n$ satisfying $f(\mathcal{S}, d, r) \in \mathcal{S}$ for every $(\mathcal{S}, d, r) \in \Sigma^n$, and the $f(\mathcal{S}, d, r)$ is a solution point of the bargaining process [8], [9].

- **WPO**: For every (\mathcal{S}, d, r) , its *weakly Pareto optimal* set is defined as **WPO** $(\mathcal{S}) = \{y \in \mathcal{S} \mid x \gg y \text{ implies } x \notin \mathcal{S}\}$.
- **RS**: For each $(\mathcal{S}, d, r) \in \Sigma^n$, if for each $\pi \in \Pi^n$, $\mathcal{S} = \pi(\mathcal{S})$, then $f(\mathcal{S}, d, r)$ has equal coordinates.
- **SI**: Let Λ^n denote the class of profiles of affine transformations that act independently player by player. For each $(\mathcal{S}, d, r) \in \Sigma^n$, and each $\lambda \in \Lambda^n$, then $f(\lambda(\mathcal{S}), \lambda(d), \lambda(r)) = \lambda(f(\mathcal{S}, d, r))$.
- **ETACM**: For each pair (\mathcal{S}, d, r) and (\mathcal{S}', d', r') in the domain of d -comprehensive problems, if $\mathcal{S} \subset \mathcal{S}'$, $(d, r) = (d', r')$ and $\Upsilon(\mathcal{S}, r) = \Upsilon(\mathcal{S}', r')$ then $f(\mathcal{S}, d, r) \leq f(\mathcal{S}', d', r')$.
- **LSCCP**: For each (\mathcal{S}, d, r) in the domain of d -comprehensive problem, such that $\Upsilon(\mathcal{S}, r) = \Upsilon(\mathcal{S}, d)$, then $f(\mathcal{S}, d, r) = f(\mathcal{S}, d, d)$.

- **C** : It says that the solution should be a continuous function of the problem; there are no jumps in game players' preferences.

Multi-agent learning deals with the interaction between multiple agents, which are acting in a common dynamical environment. One of the most popular solutions for *MARL* is Q-learning. Main motivation of Q-learning based *MARL* is to achieve an agreement to receive maximum reward; this cooperative learning problem has attracted much interest in the last decade. The job of *MARL* algorithm is to estimate Q value for every action in every state. However, some cases, the environment does not have to be represented by states, only the action space; it becomes simpler while estimating an expected value of a single reward for each action available to the learning agent. In order to make the learning system fully distributed and effectively model the behavior of intelligent agents, the single state *MARL* has been introduced where the Q values of actions are effectively the estimation of the usefulness of the actions in the next step of the multi-agent learning process [13], [14].

The single state *MARL* model \mathbb{M} is a tuple $\langle N, \{A_i\}_{i \in N}, \{R_i\}_{i \in N} \rangle$, where N is a collection of agents, A_i is the set of actions available to the agent i , and R_i is its payoff. Multiple agents simultaneously choose their actions from their own action sets and, receive their payoffs on the basis of the actions performed by all the agents. Let $\mathbf{a}^t = [a_1^t, a_2^t, \dots, a_n^t]$ be the joint action executed at iteration t , where \mathbf{a}_{-i}^t is the joint action of all the agents except the agent i . At each learning stage, an agent's experience is characterized not only by its own action and payoff but also from all the actions actually executed by other agents in the multi-agent environment. Therefore, agents can learn to coordinate their actions through the environmental feedback. The desired outcome of this process is to let the multiple agents collectively learn their best actions that maximize the total system profit. Usually, the Q-value is the expected cumulative reward for taking a particular action. For each pair \langle joint-action, action \rangle , the agent i 's Q-value, i.e., $Q_i(\mathbf{a}^t, a_i^t)$, is represented as follows [13], [14].

$$\begin{aligned}
 & Q_i(\mathbf{a}^t, a_i^t) \\
 &= (1 - \alpha) \cdot Q_i(\mathbf{a}^{t-1}, a_i^t) + \alpha \cdot \max_{a_i^{t+1}} \{ \mathcal{R}_t + \gamma \cdot Q_i(\mathbf{a}^t, a_i^t) \}
 \end{aligned} \tag{2}$$

where $\alpha \in [0, 1]$ is the learning rate, and $\gamma \in [0, 1]$ is the discount factor. To jointly reach a consensus, multiple intelligent agents should overcome the defect of selfishness through the collaboration.

C. THE PROPOSED SPECTRUM ALLOCATION SCHEME FOR THE IoMT PARADIGM

To develop our spectrum allocation scheme, we construct a bi-level control algorithm for each IoMT device; the first level is implemented as a learning model \mathbb{M} , and the second

level is designed as a cooperative game model \mathbb{G} . During time steps, \mathbb{M} and \mathbb{G} are operated sequentially, and they interactive with each other to reach a consensus. From the traditional Q-learning function, it is clear that a state-action pair is important. However, the state-action pair in Q learning cannot effectively model the real-world IoMT network system. In order to make the learning system fully distributed and effectively model the network's physical behavior, we adopt the single-state Q-learning mechanism. In this mechanism, the formulation of state-action pairs is less of an issue. It is originally proposed to solve stateless games in computer science, and effectively carried out in a dynamically changing environment. Based on each action's Q value, an agent can select his action, and the Q value of the selected action is updated by receiving a reward. Comparing with the conventional Q learning, the information of the successor state is irrelevant. This approach dramatically reduces the complexity of the learning model while enhancing the applicability of *MARL* Q-learning in a distributed manner [16].

For the first control process, we design a single-state multi-agent Q-learning model \mathbb{M} . Since the agents in a single-state environment is stateless, we need a simple reformulation of the equation (2). In this formulation, each agent maintains each action's Q value, which is updated after each learning step according to the reward received for the action. In our stateless setting, we assume that each IoMT device has his action set \mathcal{L}_D , which consists of device's available salient points. For example, the Q-value of \mathcal{D}_j 's k^{th} action, i.e., $Q_k^{\mathcal{D}_j}(\mathbf{a}_{-\mathcal{D}_j}, a_k^{\mathcal{D}_j})$, provides an estimate of the value of performing the joint action $\mathbf{a} = (\mathbf{a}_{-\mathcal{D}_j}, a_k^{\mathcal{D}_j})$. The \mathcal{D}_j updates its estimate $Q_k^{\mathcal{D}_j}(\cdot)$ value based on the experience sample $\langle \mathbf{a}, \mathcal{R}_a^{\mathcal{D}_j} \rangle$ where $\mathcal{R}_a^{\mathcal{D}_j}$ is the \mathcal{D}_j 's reward of \mathbf{a} . Simply, $\mathcal{R}_a^{\mathcal{D}_j}$ is same as the \mathcal{D}_j 's utility function $\mathcal{U}_{\mathcal{D}_j}(\cdot)$ in the second-level cooperative game \mathbb{G} ; it is defined in the equation (4). Finally, our single-state *MARL* function is defined as follows [15], [17]

$$\begin{aligned}
 & Q_{1 \leq k \leq l}^{\mathcal{D}_j}(\mathbf{a}_{-\mathcal{D}_j}, a_k^{\mathcal{D}_j}) \\
 &= (1 - \alpha) \cdot Q_k^{\mathcal{D}_j}(\mathbf{a}_{-\mathcal{D}_j}, a_k^{\mathcal{D}_j}) + \alpha \cdot \mathcal{R}_a \\
 &= Q_k^{\mathcal{D}_j}(\mathbf{a}_{-\mathcal{D}_j}, a_k^{\mathcal{D}_j}) + \alpha \cdot (\mathcal{R}_a - Q_k^{\mathcal{D}_j}(\mathbf{a}_{-\mathcal{D}_j}, a_k^{\mathcal{D}_j})) \tag{3}
 \end{aligned}$$

A major challenge for the selection of actions is to strike a balance between exploring and exploiting. In our scheme, we have chosen the Boltzmann strategy; each agent chooses an action to perform in the next iteration with a probability that is based on its current estimate of the usefulness of that action [15], [17].

For the second control process, we develop a cooperative game model \mathbb{G} . In the $\mathbb{G}_{\mathcal{A}_i}$, individual devices in the $\mathbb{D}_{\mathcal{A}_i}$ share the $\mathfrak{M}_{\mathcal{A}_i}$ in a centralized manner. In this game, the spectrum allocation process is operated based on the concept of BGV solution. In the viewpoint of $\mathcal{D}_j \in \mathbb{D}_{\mathcal{A}_i}$, the utility

functions, i.e., $\mathcal{U}_{\mathcal{D}_j}(\cdot)$, is defined as follows:

$$\begin{aligned} & \mathcal{U}_{\mathcal{D}_j}(\Gamma_{\mathcal{A}_{\mathcal{D}_j}}, \mathcal{S}_{\mathcal{D}_j}, \mathcal{J}_{\mathcal{D}_j}) \\ &= \exp\left(\mu \times \log\left(\frac{\min(\mathcal{W}_{\mathcal{D}_j}, \mathcal{S}_{\mathcal{D}_j})}{\mathcal{W}_{\mathcal{D}_j}}\right)\right) \times \frac{1}{\mathcal{J}_{\mathcal{D}_j}} \\ & \text{s.t., } \sum_{\mathcal{D}_j \in \mathbb{D}_{\mathcal{A}_i}} \mathcal{S}_{\mathcal{D}_j} \leq \mathfrak{M}_{\mathcal{A}_i} \end{aligned} \quad (4)$$

where μ is an adjustment parameter for the $\mathcal{U}_{\mathcal{D}_j}(\cdot)$, and $\mathcal{J}_{\mathcal{D}_j}$ is the service sensitivity of \mathcal{D}_j 's application. $\mathcal{W}_{\mathcal{D}_j}$ and $\mathcal{S}_{\mathcal{D}_j}$ are the requested spectrum amount and allocated spectrum amount, respectively. According to (1) and (4), the spectrum resource $\mathfrak{M}_{\mathcal{A}_i}$ is shared based on the *BGV* solution idea. For the \mathcal{D}_j , the allocated spectrum amount, i.e., $BGV_{\mathcal{D}_j}(\cdot)$, is given by:

$$\begin{aligned} & BGV_{\mathcal{D}_j}(\mathcal{D}_j \in \mathbb{D}_{\mathcal{A}_i} \mid \mathbb{U}, d_{\mathbb{D}_{\mathcal{A}_i}}, \mathbf{r}_a) \\ &= (\lambda^* \cdot \Upsilon(\mathbb{U}, \mathbf{r}_a)) + ((1 - \lambda^*) \cdot d_{\mathbb{D}_{\mathcal{A}_i}}) \\ & \text{s.t., } \lambda^* \\ &= \begin{cases} \mathbb{U} = \{\mathcal{D}_k \in \mathbb{D}_{\mathcal{A}_i} \mid \dots, \mathcal{U}_{\mathcal{D}_k}, \dots\} \\ \Upsilon(\mathbb{U}, \mathbf{r}_a) = \Upsilon_{\mathcal{D}_j \in \mathbb{D}_{\mathcal{A}_i}} \\ \quad = \max\left\{r_{\mathcal{D}_j}, \max\left\{t \in \mathbb{R} \mid (t, \mathbf{r}_{-\mathcal{D}_j}) \in \mathbb{U}\right\}\right\} \\ \max\left\{\lambda \in [0, 1] \mid ((\lambda \cdot \Upsilon(\mathbb{U}, \mathbf{r}_a)) + ((1 - \lambda) \cdot d_{\mathbb{D}_{\mathcal{A}_i}})) \in \mathbb{U}\right\} \end{cases} \end{aligned} \quad (5)$$

where \mathbf{r}_a is a reference point to temper the IoMT devices' aspirations. In our proposed scheme, the \mathbf{r}_a is defined as devices' actions where $\mathbf{r}_a = \{\mathcal{D}_k \in \mathbb{D}_{\mathcal{A}_i} \mid \dots, r_{\mathcal{D}_j}, \dots\}$ and $r_{\mathcal{D}_j}$ is the \mathcal{D}_j 's selected action, i.e., $a_{1 \leq k \leq l}^{\mathcal{D}_j}$. The $\mathbf{r}_{-\mathcal{D}_j}$ is the all devices' selected actions except the \mathcal{D}_j .

D. MAIN STEPS OF OUR BI-LEVEL SPECTRUM ALLOCATION ALGORITHM

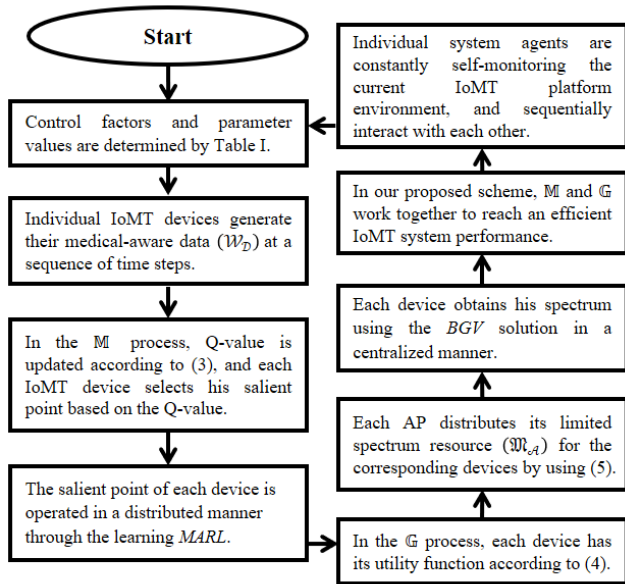
The coordination of multi-agents has become increasingly popular in the reinforcement learning and game theory. Because of its generality and robustness, the combination of learning algorithm and game theory has attracted recent attention. However, how to integrate the learning model and game solution has not yet been tackled in depth. In this study, we focus our attention on a learning game to solve the spectrum allocation problem in the IoMT system platform. Our main interest is in the application of *MARL* to the sequential decision problem, which is being controlled by multiple intelligent IoMT devices. For the computation simplicity, we focus on the single state *MARL* model [15]. It is a little different approach compared to the straightforward application of traditional *MARL*. Multiple IoMT devices repeatedly play a cooperative game in which they independently select their individual actions based on the *MARL*. The chosen actions

at any point constitute a joint coordinated action. Based on these actions, IoMT devices play a cooperative repeated game based on the *BGV* solution. Our proposed bi-level control paradigm sophisticatedly combines the *MARL* model and the *BGV* solution; they mutually dependent and act cooperatively to obtain a fair-efficient system performance. In addition, we can provide the ability to practically respond to current IoMT system conditions while keeping the computation complexity under the control. It is a suitable approach for real-world network operations. The primary steps of our proposed scheme are described as follows, and they are described by the following flowchart:

- Step 1:** Based on the experimental settings in the Section V and Table 1, control factors and parameter values are determined to carry out the numerical experiments.
- Step 2:** At a sequence of time steps, individual IoMT devices in the \mathbb{D} generate their medical-aware data ($\mathcal{W}_{\mathcal{D}}$), which are transmitted to their corresponding APs.
- Step 3:** At the first control process, each IoMT device selects his salient point based on the Q-value. According to (3), the Q-value is updated based on the experience sample of joint action and reward.
- Step 4:** To decide the best salient point of each device, the learning *MARL* model \mathbb{M} is operated in a distributed manner.
- Step 5:** At the second control process, each AP distributes its spectrum resource ($\mathfrak{M}_{\mathcal{A}}$) for the corresponding devices. Each device has its utility function according to (4); it is also used as the device's reward.
- Step 6:** At the first control process, each IoMT device selects his salient point based on the distributed *MARL* process. According to this decision, the second control process distributes the spectrum resource based on the idea of *BGV* solution.
- Step 7:** Each device obtains his spectrum by using (5). For the spectrum sharing problem in the IoMT platform, the cooperative game \mathbb{G} is executed in a centralized manner.
- Step 8:** In our proposed scheme, the \mathbb{M} and \mathbb{G} are sophisticatedly combined based on the reward and utility functions. Therefore, the \mathbb{M} and \mathbb{G} work together to reach an efficient system performance.
- Step 9:** During a sequence of time steps, individual system agents are constantly self-monitoring the current IoMT platform environment, and sequentially interact with each other in the both distributed and centralized fashions. For the next iteration, it proceeds to Step 2.

V. PERFORMANCE EVALUATION

In this section, we present the simulation results and discuss the performance of our proposed bi-level spectrum allocation algorithm. By using the MATLAB software, we model our system, and simulations are run. To outline the benefits of



FLOWCHART 1. Flowchart of the proposed algorithm.

our approach, we show a detailed comparative analysis with other competing protocols of *HMIoMT* [3], *RMIoMT* [11] and *DLIoMT* [12]. Simulation parameters and their values are summarized in Table 1, and the simulation environment and system scenario are given as follows:

- Simulated the AP assisted IoMT system platform consists of ten APs and hundred IoMT devices, i.e., $|\mathbb{A}| = 10$, and $|\mathbb{D}| = 100$.
- Ten APs are deployed in the network coverage area, and individual IoMT devices are randomly distributed over there.
- Each IoMT device $\mathcal{D}_{1 \leq j \leq 100}$ generates different health-aware application data ($\mathcal{W}_{\mathcal{D}_j}$) where the arrival process of $\mathcal{W}_{\mathcal{D}_j}$ is the rate of Poisson process (ρ). The offered range is varied from 0 to 3.0.
- Individual IoMT devices can directly contact with their corresponding APs; they communicate with the APs through wireless spectrum links.
- The total spectrum resource of each AP ($\mathcal{M}_{\mathcal{A}}$) is 100 Gbps.
- The $\mathcal{L}_{\mathcal{D}}$ is the set of each device's available salient points where it consist of 6%, 8%, 10%, 12% and 14% of $\mathcal{W}_{\mathcal{D}}$.
- The disagreement points for bargaining process, i.e., $d_{\mathbb{D}_A}$, are zeros.
- We assume the absence of physical obstacles in the AP's coverage area.
- The spectrum allocation process is specified in terms of basic spectrum units (BSUs) where one BSU is 64 Mbps in this study. Therefore, each AS has 1600 BSUs.
- The AP assisted IoMT system performance measures obtained on the basis of 100 simulation runs are plotted as functions of the Poisson process (ρ).

TABLE 1. System parameters used in the simulation experiments.

Factor	Value	Description
n	10	total number of APs
m	100	total number of IoMT devices
$\mathcal{M}_{\mathcal{A}}$	100 Gbps	total wireless spectrum resource of each AP
BSU	64 Mbps	the minimum amount of spectrum resource allocation
α	0.2	a learning rate for the <i>MARL</i>
μ	1.2	an adjustment parameter for the $\mathcal{U}_{\mathcal{D}}(\cdot)$

Set	Values	Description
$\mathcal{L}_{\mathcal{D}}$	{6%, 8%, 10%, 12%, 14% of $\mathcal{W}_{\mathcal{D}}$ }	the set of each device's available salient points
$d_{\mathbb{D}}$	{..., 0, ...}	disagreement point values for devices

Application type	$\mathcal{J}_{\mathcal{D}}$	$\mathcal{W}_{\mathcal{D}}$	Service time
$\mathcal{W}_{\mathcal{D}} \in \text{I}$	0.9	64 Mbps	45 t
$\mathcal{W}_{\mathcal{D}} \in \text{II}$	1.2	320 Mbps	50 t
$\mathcal{W}_{\mathcal{D}} \in \text{III}$	0.8	192 Mbps	25 t
$\mathcal{W}_{\mathcal{D}} \in \text{IV}$	1	256 Mbps	40 t
$\mathcal{W}_{\mathcal{D}} \in \text{V}$	1.1	128 Mbps	15 t
$\mathcal{W}_{\mathcal{D}} \in \text{VI}$	1.3	384 Mbps	30 t

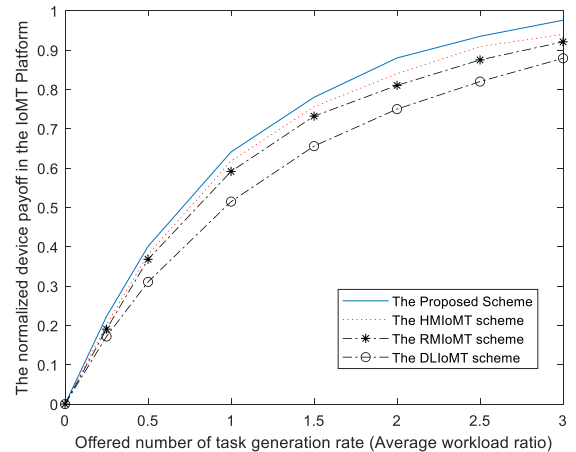


FIGURE 2. The normalized device payoff in the IoMT platform.

To evaluate the proposed solution, we compare its performance in terms of normalized device payoff, IoMT system throughput and device fairness. Table 1 shows the control parameters and system factors used in the simulation.

The normalized device payoff in the IoMT Platform is illustrated in Fig. 2. The results reflect the trend of device payoff when implementing the different spectrum allocation protocols. In the viewpoint of end users, the device payoff is a key factor to evaluate the service quality. For low average health-aware workload rates, the device payoff is virtually the same as for all protocols. However, as the workload rate increases, the device payoff of our proposed scheme is better than the *HMIoMT*, *RMIoMT* and *DLIoMT* schemes.

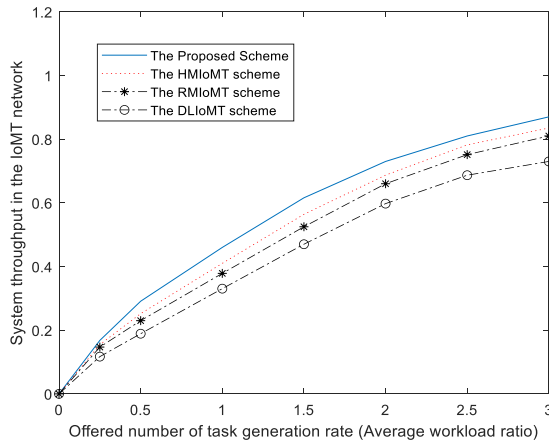


FIGURE 3. System throughput in the IoMT network platform.

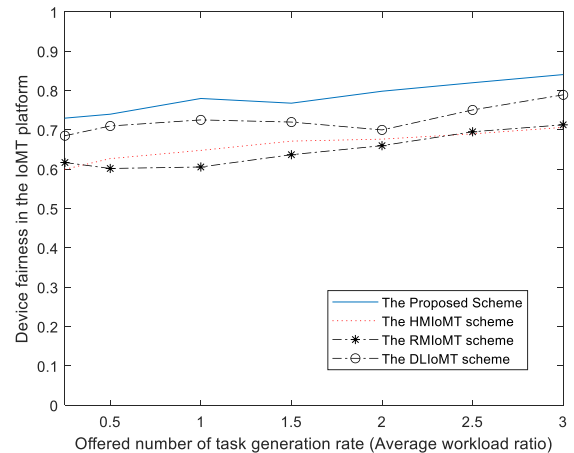


FIGURE 4. Device fairness in the IoMT platform.

The reason may be the fact that IoMT devices in our approach dynamically adjust their salient points to share the limited AP spectrum resource. Based on the *MARL* mechanism, multiple IoMT devices get the current learning information, and work together in a step-by-step interactive manner to achieve a mutually desirable solution. Under the dynamic changing IoMT system environments, our single state *MARL* method gains a significant advantage for the spectrum allocation problem.

Fig. 3 demonstrates the effectiveness of our proposed scheme with respect to the IoMT system throughput. In the viewpoint of system operators, system throughput is a main performance criterion to evaluate the system efficiency. With the rise of workload rate in the IoMT device, the system throughput increases. It is intuitive correct. As the workload rate increases, the system throughput of our scheme is much higher than that of other schemes. In our proposed scheme, IoMT devices select their actions in a distributed manner, but actual spectrum allocation process is operated in a centralized manner. By a sophisticated combination of the *MARL* model and the *BGV* solution, our bi-level control approach is quite adaptable to maximize the system throughput while adapting dynamic network changes.

Fig. 4 reveals the comparison of device fairness in the IoMT platform. To verify the device fairness for different schemes, we compare its normalized fairness index. Simulation results show the excellency of our proposed scheme for the fairness issue. To allocate the available spectrum, the major goal of *BGV* solution is to get a fair-efficient solution. Through the *BGV* solution, we investigate the fairness issue while generating maximum system efficiency. Therefore, the \mathcal{M}_A is efficiently allocated while ensuring the fairness among IoMT devices. Due to this reason, our proposed scheme can maintain an excellent device fairness.

Simulation analysis in Fig. 2 to Fig. 4 can confirm the superiority of our proposed scheme than the existing *HMloMT*, *RMloMT* and *DLloMT* schemes. By employing the bi-level control paradigm, the learning *MARL* method and cooperative

game model are mutually dependent and act cooperatively to capture dynamic interactions among IoMT devices and APs. Therefore, we can achieve a desirable solution between conflicting requirements.

VI. SUMMARY AND CONCLUSION

In this paper, we have investigated the IoMT-assisted health monitoring system with limited spectrum resource, and propose a new spectrum allocation scheme based on a bi-level control approach. In our method, we include a *MARL* process to select each device's salient point, and employ a cooperative game process to share the spectrum resource. For the *MARL* process, individual IoMT devices intelligently learn their best salient points under the dynamic IoMT network environment. Learning process is operated in parallel and distributed manner. For the cooperative game process, the *BGV* solution is implemented to share the spectrum resource among multiple IoMT devices. Sharing process is operated in a centralized fashion. These two control processes are sophisticatedly combined, and work together and act iteratively to strike an appropriate system performance. Taking accounting of contradictory service requirements, the major challenge of our bi-level approach is to reach a consensus with reciprocal advantages. In addition, we adopt the single state *MARL* model to significantly reduce the computation complexity compared to a classical *MARL* method. Finally, performance evaluations demonstrate the effectiveness of our proposed scheme for the IoMT system platform, and we can get positive benefits for health-aware application services than other existing *HMloMT*, *RMloMT* and *DLloMT* methods.

As a future work, we plan to incorporate the security issue in the IoMT system when IoMT devices send data packets to APs. Usually, the information related to IoMT devices is strictly private and confidential. The lack of security awareness among IoMT devices and the risk of several intermediary attacks for accessing health information severely endanger the use of IoMT system. Therefore, we should guarantee IoMT devices' security and privacy. Furthermore, blockchain

and smart contract technologies can be explored to improve the health monitoring services. In addition, we will propose a new distributed learning algorithm for the *MARL* process. By using new learning methods, the reliability and robustness of each agent can be improved.

AVAILABILITY OF DATA AND MATERIAL

Please contact the corresponding author at swkim01@sogang.ac.kr.

COMPETING INTERESTS

The author, Sungwook Kim, declares that there are no competing interests regarding the publication of this article.

REFERENCES

- [1] A. Ghubaish, T. Salman, M. Zolanvari, D. Unal, A. Al-Ali, and R. Jain, "Recent advances in the Internet-of-Medical-Things (IoMT) systems security," *IEEE Internet Things J.*, vol. 8, no. 11, pp. 8707–8718, Jun. 2021.
- [2] A. Gatooullat, Y. Badr, B. Massot, and E. Sejdić, "Internet of Medical things: A review of recent contributions dealing with cyber-physical systems in medicine," *IEEE Internet Things J.*, vol. 5, no. 5, pp. 3810–3822, Oct. 2018.
- [3] Z. Ning, P. Dong, X. Wang, X. Hu, L. Guo, B. Hu, Y. Guo, T. Qiu, and R. Y. K. Kwok, "Mobile edge computing enabled 5G health monitoring for Internet of Medical things: A decentralized game theoretic approach," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 12, pp. 463–478, Feb. 2021.
- [4] T. Zhang, M. Liu, T. Yuan, and N. Al-Nabhan, "Emotion-aware and intelligent Internet of Medical things toward emotion recognition during COVID-19 pandemic," *IEEE Internet of Things Journal*, vol. 8, no. 21, pp. 16002–16013, Nov. 2021.
- [5] S. Kim, *Game Theory Applications in Network Design*. Hershey, PA, USA: IGI Global, 2014.
- [6] D. Wang, W. Zhang, B. Song, X. Du, and M. Guizani, "Market-based model in CR-IoT: A Q-probabilistic multi-agent reinforcement learning approach," *IEEE Trans. Cognit. Commun. Netw.*, vol. 6, no. 1, pp. 179–188, Mar. 2020.
- [7] J. Wang, Y. Hong, J. Wang, J. Xu, Y. Tang, Q.-L. Han, and J. Kurths, "Cooperative and competitive multi-agent systems: From optimization to games," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 5, pp. 763–783, May 2022.
- [8] W. Thomson, "On the axiomatic theory of bargaining: A survey of recent results," *Rev. Econ. Design*, vol. 26, pp. 491–542, Dec. 2022.
- [9] E. Karagözoğlu, K. Keskin, and E. Özcan-Tok, "Between anchors and aspirations: A new family of bargaining solutions," *Rev. Econ. Des.*, vol. 23, nos. 1–2, pp. 53–73, 2019.
- [10] Y. Hu, Y. Gao, and B. An, "Multiagent reinforcement learning with unshared value functions," *IEEE Trans. Cybern.*, vol. 45, no. 4, pp. 647–662, Apr. 2015.
- [11] Y. Qiu, H. Zhang, and K. Long, "Computation offloading and wireless resource management for healthcare monitoring in fog-computing-based Internet of Medical things," *IEEE Internet Things J.*, vol. 8, no. 21, pp. 15875–15883, Nov. 2021.
- [12] Y. Yang, X. Wang, Z. Ning, J. P. C. J. Rodrigues, X. Jiang, and Y. Guo, "Edge learning for Internet of Medical things and its COVID-19 applications: A distributed 3C framework," *IEEE Internet Things Mag.*, vol. 4, no. 3, pp. 18–23, Sep. 2021.
- [13] E. M. D. Cote, A. Lazaric, and M. Restelli, "Learning to cooperate in multi-agent social dilemmas," in *Proc. 5th Int. Joint Conf. Auto. Agents Multiagent Syst.*, 2006, pp. 1–8.
- [14] N. Morozs, T. Clarke, and D. Grace, "Distributed heuristically accelerated Q-learning for robust cognitive spectrum management in LTE cellular systems," *IEEE Trans. Mobile Comput.*, vol. 15, no. 4, pp. 817–825, Apr. 2016.
- [15] C. Claus and C. Boutilier, "The dynamics of reinforcement learning in cooperative multiagent systems," in *Proc. AAAI/IAAI*, 1998, pp. 746–752.
- [16] T. Jiang, Q. Zhao, D. Grace, G. Alister Burr, and T. Clarke, "Single-state Q-learning for self-organized radio resource management in dual-hop 5G high capacity density networks," *Trans. Emerg. Telecommun. Technol.*, vol. 27, no. 12, pp. 1628–1640, 2016.
- [17] S. Kapetanakis and D. Kudenko, "Reinforcement learning of coordination in cooperative multi-agent systems," in *Proc. AAAI*, 2002, pp. 326–331.



SUNGWOOK KIM received the B.S. and M.S. degrees in computer science from Sogang University, Seoul, Republic of Korea, in 1993 and 1995, respectively, and the Ph.D. degree in computer science from Syracuse University, Syracuse, NY, USA, in 2003, supervised by Prof. Pramod K. Varshney. He was a Faculty Member with the Department of Computer Science, ChoongAng University, Seoul. In 2006, he returned to Sogang University, where he is currently a Professor with the Department of Computer Science and Engineering. He is also the Research Director of the Network Research Laboratory. His research interests include resource management, online algorithms, adaptive quality-of-service control, and game theory for network design.

• • •