IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 21, 2018, accepted April 2, 2018, date of publication April 11, 2018, date of current version May 9, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2825282

A Survey of Decision-Theoretic Models for Cognitive Internet of Things (CIoT)

KHAQAN ZAHEER[1], MOHAMED OTHMAN[1], (Member, IEEE), MUBASHIR HUSAIN REHMANI[2], (Senior Member, IEEE), AND THINAGARAN PERUMAL[1], (Member, IEEE)

[1]Department of Communication Technology and Network, Universiti Putra Malaysia, Selangor 43400, Malaysia
[2]Telecommunications Software and Systems Group, Waterford Institute of Technology, X91 K0EK Waterford, Ireland

Corresponding authors: Khaqan Zaheer (khaqan82@gmail.com), Mohamed Othman (mothman@upm.edu.my), and Mubashir Husain Rehmani (mshrehmani@gmail.com)

This work was supported by the Research University Publication Fund, under the Malaysia Ministry of High Education through the Research Management Centre, Universiti Putra Malaysia.

ABSTRACT Communication technology and wide spread usage of Internet of Things (IoT) are rapidly becoming the core enabler for consumers to use their smart devices in their daily routine. These smart devices are gradually transforming the IoT scenario into a new paradigm called cognitive Internet of Things (CIoT). We believe that CIoT will revolutionize many service sectors including smart cities, transportation, healthcare, and environmental monitoring. The current development in CIoT is still elusive as it has to achieve effective interoperability and autonomous decision making. Most importantly, the decision theoretic models are seen as an enabler to achieve effective interoperability among heterogeneous CIoT objects. In this paper, we provide a survey of the existing decision theoretic models and their usage for CIoT, an architectural CIoT framework is also proposed to discuss the open issues and solution for potential challenges emerging in the area of CIoT research.

INDEX TERMS Cognitive Internet of Things (CIoT), decision theoretic models, game theory, multi-agent learning, Markov decision process, multi-arm bandit problem, optimal stopping problem.

I. INTRODUCTION

The Internet of Things (IoT) defines an emerging paradigm of cyber physical systems, in which billions of interconnected smart objects collect, analyze and exchange vast information from all over the world. Today, IoT not only functions as the substantial choice for computing and communication paradigm but also provides rich set of services in smart cities, homes, transportation and environmental monitoring [1]. Recently, the deployment of cognitive computing in IoT has gained eminent interest, where intelligence is infused into smart objects to learn a lot from the physical world. Such paradigm is called as cognitive Internet of Things (CIoT) [2]. According to the latest surveys, approximately 500 billion devices will be connected to the Internet by 2020 and we need a concrete framework of IoT which can easily fulfill the future requirements [3]. CIoT is extension of IoT paradigm which is equipped with cognitive abilities to enhance performance and achieve intelligence. Recently, IoT European Research Cluster (IERC) has published a detailed paper on IoT current and future road-map for development until 2015 and beyond 2020 [4].

The rapid development in the field of IoT in the past few years enables the smart devices to provide seamless connectivity among the objects by using the intelligent services and applications. However, the applications of IoT are still not intelligent enough to perform decision making and are dependent on human beings for cognition processing. Still, the framework of IoT is not equipped with the brain to do decision on its own. CIoT is termed as IoT paradigm unified with cognitive abilities for managing inter-operation via decision making among heterogeneous smart objects [5]. Therefore, Wu *et al.* introduced the concept of Cognitive Internet of Things [6] in which the general objects work like agents and interact with the physical environment with minimal human interaction. In this paper our main focus is to equip the existing IoT framework with human cognition which is capable of performing the intelligent decision making independently. In short, we have embedded human intelligence in the framework by adding the intelligent decision making layer in the system design for intelligent decision making. The addition of of this intelligent layer in the existing framework have several advantages including increasing the resource efficiency, saving human time and efforts, intelligence decision making, enhance service provision, self-organizing, optimization etc. to name a few.

TABLE 1. IERC current projects on internet of things [8].

Features of all current projects by IERC (Semantic Interoperability) [8]	IoT-A [8]	Probe-IT [8]	Open IoT [8]	GAMBAS [8]	IoT-est [8]	IoT-I [8]	ebbits [8]	Smart Agrifood [8]	IoT-6 [8]	iCore [8]	BUTTLER [8]	Iot@Work [8]
Integration	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Annotation	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Management	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓
Discovery	✗	✓	✓	✗	✓	✗	✗	✗	✓	✗	✓	✓
Analysis and Reasoning	✗	✗	✓	✗	✗	✗	✗	✓	✗	✗	✗	✗
Visualisation	✗	✓	✓	✓	✓	✗	✗	✓	✗	✓	✓	✓

The literature clearly showed that many projects had been done in IoT, however, the research in the field of cognitive IoT is still under development phase and requires a lot of work from the research community for its practical implementation. The most vital aspect of handling the heterogeneous objects can be managed by utlizing their semantics and ontology for the virtualization of the these objects in current architecture of CIoT [6]. A key challenge for CIoT is to overcome the heterogeneity of dissimilar objects in terms of their features and the network technologies for their interconnection [7]. In [7], Foteinos *et al.* presented the three layered IoT architecture for enabling the autonomous application and the reuse of objects across various domains. Moreover, IoT European Research Cluster (IERC) have initiated several projects on IoT that include research on current framework and semantic interoperability. IERC research cluster has provided a comprehensive insight on several projects of IoT like IoT-A, IoT-I, OPENIoT, i-Core, PROBE-IT, BUTLER. IoT@Work, IoT.EST, GAMBAS, COIN, IoT6, SmartAgrifood, CONNECT and ComVantage [8] which are summarized in Table 1. The SENSEI project [9] integrates the physical world with digital world by dividing it into three different abstractions which include resources, entities, and resource users, for addressing a large number of widely distributed wireless sensors and actuators. The key take-away is to propose the new models to enhance ontology, domain knowledge and decision making techniques based on application requirements. In [4] Vermesan *et al.* provides an indepth detail about the conceptual framework of IoT, technological trends, IoT applications and technology enablers (intelligence, communication, integration, semantic technologies etc. to name a few). Basically, the authors highlight the current research agenda, timelines and priorities in IoT that include identification technology, IoT architecture, communication technology, network technology, software, services, hardware, discovery and search engine technologies, power and storage technologies, security and privacy, and standardization in detail. The authors also provide the guidelines for future technological development in IoT and also highlight the issues including IoT standardization, Ontology based semantic standards, spectrum energy communication protocols standards, standards for communication within and outside cloud, International quality/integrity standards for data creation, data traceability and decision making in IoT. Another recent work by Vermesan and Friess [10] elaborates the current IoT advancements and also suggests the future roadmap for IoT in detail. The paper highlights the IoT strategic research and innovation agenda that includes development of smart-X applications, IoT related future technologies (clod computing and semantic technologies), networks and communication, processes, data management, security and privacy, device level energy issues, IoT related standardization and IoT protocol convergence for future development. Finally, the author highlights the current projects on IoT including Open IoT, iCORE, Compose, SmartSantander, Fitman, OSMOSE and CALIPSO [10] project in details with their results and future opportunities for IoT. Most recently in [11], Khan *et al.* have provided the comprehensive overview of cognitive radio based IoT system. In this paper, they have presented a brief overview of cognitive radio (CR) based IoT system and suggested that system is a viable solution for effective and efficient utilization of spectrum resources even in the presence of primary users (PUs). Moreover, CR technology also equipped the objects in IoT with intelligence in order to learn, think and make decisions of both physical and social worlds. They have also highlighted the potential applications of CR based IoT systems. This paper has also presented a comprehensive survey about different architectures and framework along with the functionality of each layer in the framework of CIoT systems. They have also discussed about spectrum related functionalities and opportunities for CIoT systems. The authors explain about traditional three layer architecture and functionalities related to each layer but are unable to explain about intelligent decision making in their work. Moreover, they briefly explain about VO and CVO creation but do not provide any details about objects reuse in the framework. Finally, the intelligent decision making which is the heart of CIoT system is not discussed in this work. They have also highlighted that literature on IoT has presented many frameworks for IoT, but these

lack motivations and necessity of standardization. Although, the efficient spectrum utilization using CR based concepts is explored but the implementation of decision making models (game theory, markovian decision process etc. to name a few.) is not highlighted in their work. Therefore, the need of the hour is to utilize the game theoretic decision making model for cognitive decision making about object reuse and efficient spectrum resources utilization in CIoT.

The previous literature regarding usage of game theoretic approach for an IoT-based employee performance evaluation in industry [12]. They have evaluated the performance of employees by mining data collected by the sensory nodes using the MapReduce model. This information is then used to draw automated decisions for employees using game theory. In [13], Bui *et al.* have proposed the game theoretic based real time decision making approach for IoT based traffic light control system. They have proposed the connected intersection system consisting of IoT devices and then used the game theory algorithms including Cournot Model and Stackelberg Model for intelligent decision making. In [14], Kim has proposed a new quality-of-service management scheme based on the IoT system power control algorithm which utilizes the concept of game theoretic for power allocation. Basically, they have utilized the R-learning algorithm and docitive paradigm in which the system agents can teach other agents how to adjust their power levels while reducing computation complexity and speeding up the learning process. In [15], Bhatia and Sood presented the solution of decision making in IoT assisted activity of defense personals. They have implemented the game theoretic based automated models to aid monitoring personals for efficient monitoring of social activities and analyzing it over suspicious scale. In [16], Semasinghe *et al.* presented the game theoretic mechanism for resource management in wireless IoT systems. They have utilized the game theory models including evolutionary game, bargain game mean field game and mean field auction game for the management of IoT-related resources in large-scale systems.

The previous studies also show that Markov decision process is used for intelligent decision making. In [17], Kim have presented the QoS control scheme based on Markov game model which can effectively allocate IoT resources while maximizing system performance in IoT systems. Moreover, this distributed proposed system provides step-by-step feedback process and enables adaptability and responsiveness to current IoT system conditions. In [18], Yau and Buduru proposed an intelligent planning technique for mobile IoT applications based on Markov decision process which can enhance efficiency of IoT device action planning. Moreover, it also enables mobile networks with elastic resources from various mobile clouds which are effective in supporting IoT applications. In [19], Alam *et al.* proposed an integrated reinforcement learning approach based on genetic algorithm for device and application aware SLA maintenance and management in IoT environment. This approach enables the automatic management of IoT devices with the help of genetic algorithm reinforcement learning. In [20], Xiong *et al.* addressed the resource allocation problem in the proposed SDN-based IoT network and used the semi-Markov decision process (SMDP) to maximize the expected average rewards of the network. Moreover, they have proposed the optimal solution by using the SMDP problem in a relative value iteration algorithm. In [21], Alsheikh *et al.* have presented a comprehensive survey on Markov decision process for wireless sensor network. The paper also highlighted the designs of the Markov decision process in wireless sensor networks including data exchange and topology formation, resource and power optimization, area coverage and event tracking solutions, and security and intrusion detection methods.

Early attempts on CIoT highlight the cognitive processes consisting of a three-layered ring inclusive of virtual object (VO) layer, composite virtual object (CVO) layer and service layers of IoT for service provisioning [6]. The work mentioned in [6] proposed context gathering and identifying intelligent devices as real world objects (RWOs) and subsequently sending the appropriate information or object to the Internet via gateway from any location at any time. These RWOs are represented as VO. These smart virtual objects are capable of hiding the functional and implementation details from their recipients. The proposed CIoT framework generally combines many VOs to establish CVO, which provides services to end-users as well as higher hierarchy applications [22]. CVO is a smart object which contains all the semantically portable information about the objects including their virtual object creation, functions and parameters, user-centric services, identifiers etc. to name a few [23]. The combination of CVOs opens up a new opportunity in managing dynamic objects in CIoT. The most challenging goal here is to efficiently utilize the resources by reusing the existing CVOs and combining them into intelligent representations which will minimize their time of creation and management for intelligent decision making. Numerous questions remain elusive as how to make cognitive vision of IoT as a reality. For instance, how much cognitive ability can we push to the IoT without risking the service provision of IoT applications? What developments are needed in order to ensure robust and accurate decision making of cognitive context with IoT?

Cognitive Radio Network (CRN) is a promising paradigm that optimizes radio spectrum utilization and throughput [2], has ability to perceptively perform decision based on historical information, which shares similar potential of deployment for CIoT. Similarly, a few attempts have been made in the past to present the deeper viewpoint on those mentioned decision-theoretic models for CRN as above [2], [24]. The previous literature showed that game models are best for capturing the interactions among several players. The work done by Y. Xu *et al.*, aims to provide a detailed work on decision making in cognitive radio network by the selection of appropriate channel among several channels for opportunistic spectrum access in multi-agent decision theoretic model. They proposed a complete framework for analysing,

learning and evaluating the process of object selection in cognitive radio network. In [24], Xu *et al.* discussed the game theoretic perspective of self organization and optimization for cognitive CRN. Moreover, the intelligent decision making for small cells in CRn is also explored in detail. In [23], Xu *et al.* highlighted the opportunistic spectrum access in CRN for achieving global optimization using the local interaction games. They proposed a localized selfless game in which each player tried to maximize his utility function and collection of utilities of its neighbours was proposed to achieve global optimization via local information exchange for cognitive decision making in CRN. Another work proposed in [25] also proposed the game theory based network selection in CRN. Finally, the most recent literature [26] has proposed the usage of CRN techniques for IoT based system to manage the shortage of spectrum for IoT devices. They have used different game model including inspection game, bargain game, hierarchical game models etc. to name a few according to different scenarios for intelligent decision making of channel selection in IoT.

In this paper, we provide a brief overview of current technological advancements in IoT and also suggest how to utilize the current work on object semantic, ontology, interoperability, communication, management, integration and management in our proposed novel architecture of CIoT. The paper also provides the insight knowledge of basic game models which are useful to analyze and model the user interaction with CIoT system. The most suitable game models for CIoT include repeated games, graphical game, evolutionary game, hierarchical games, coalition game and bayesian game. These game models provide guidelines for designing and modeling the non-cooperative game models for better learning and self-organizing. In the non-cooperative game model, the players (object) are autonomous and make rational decisions in order to maximize their individual utility functions. In these models, the most commonly used solution concept parameters are Nash equilibrium (NE) and correlated equilibrium [24]. Moreover, the decisions of players are distributed and autonomous which lead to self-organizing optimization. Finally, the paper also highlights the behaviour update rule which is a fundamental element for practical implementation of game model.

These proposed solutions are inherited from CRN which can be deployed on multi-agent systems using game theory specifically by implementing hybrid or multiple learning approaches for efficient decision making in CIoT. The most challenging issue of unpredictable, dynamic and incomplete information could be solved by learning derivation as well as knowledge model extraction. For example in game theory, the commonly deployed concept solutions are NE and correlated equilibrium [27]. Likewise, the decisions of the players are distributed and autonomous which can lead to self-organizing and optimization. In addition, the converged attributes can be analyzed by applying theories of Markovian process and stochastic approximation [28]–[30]. It is evident that the work presented in [4]–[10] and [12]–[30] clearly emphasizes on the importance of decision making in CIoT for optimized service provisioning.

Contributions of This Paper: This paper provides a comprehensive survey on decision theoretic models for CIoT. The major contributions of this paper are as follow:

- We provide an insight overview of decision theoretic schemes for CIoT.
- We propose a novel operational framework for large scale CIoT based on real world example.
- We provide detailed overview of game theoretic solution for cognitive decision making in IoT.
- We also propose the solution for intelligent decision making in IoT by selecting multiple decision or hybrid decision theoretic solution.
- We highlights the issues, challenges and future direction of decision theoretic solution in CIoT.

The organization of the paper is shown in Fig.1 and is described afterwards. The proposed cognitive framework with real world scenario is discussed in Section II. We propose a cognitive framework in order to facilitate the discussion on the important open challenges on the cognitive IoT and decision theoretic models. In Section III, we present the fundamental game theoretic solutions for CIoT. Moreover, the appropriate decision theoretic models for intelligent decision making are also discussed in detail in this section. The comparative analysis of decision theoretic solutions in terms of information selection, convergence speed and cost are highlighted in Section IV. In Section V, we have listed possible open issues and future directions for CIoT along with concluding remarks. Moreover, we also propose the solution of intelligent decision making in IoT by selecting multiple or hybrid decision theoretic solution for intelligent decision making in CIoT. Finally, Section VI concludes the entire paper.

II. OPERATIONAL FRAMEWORK FOR LARGE SCALE CIoT

A. MOTIVATION AND REAL WORLD SCENARIO

The current applications on a large-scale CIoT are still inefficient to perform effective and efficient communication and intelligent decision making. The previous literature clearly showed that the IoT is in development phase and research community is mainly focusing on its fundamental architecture, applications, future technologies (clod computing and semantic technologies), networks and communication, data management, security and privacy etc. to name a few [4]–[10], [12]–[21] but still the IoT framework is not smart enough to learn, think, and recognize cyber, physical and social worlds by itself. The technological advancements still do not equip the objects with brain to become smart objects that can be reused for intelligent decision making. The current applications in CIoT are still ambiguous and inefficient for effective interoperability and intelligent decision making among heterogeneous objects. To solve these issues, multi-agent decision theoretic models are proposed to cater decision layer in a federated manner. In this article, we focus specifically on the methodology to analyze, learn, design and

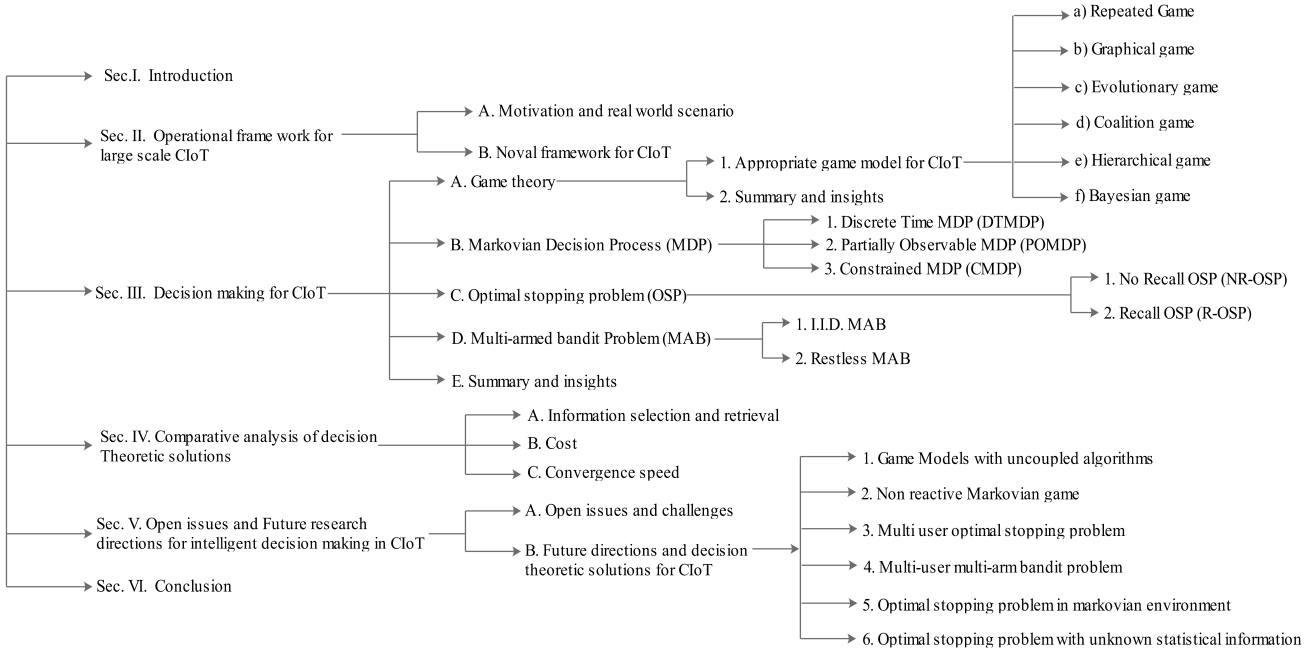

FIGURE 1. Organization of the paper.

evaluate multi-agent decision-theoretic solution selection for intelligent decision making among heterogeneous objects and channel selection in CIoT. To the best of our knowledge, there is not a single comprehensive paper that implements hybrid solution of game theory, markovian decision process, multi arm bandit and optimal stopping for decision making in IoT. Moreover, the literature also showed that research community is not focusing on this area of research for IoT. The previous studies clearly indicate that the work done in IoT and CRN for channel selection is either based on game theory or markovian decision process or optimal stopping problem or multi arm bandit problem but none of the authors has used all these techniques together for appropriate channel and object selection (reuse) for intelligent decision making in CRN. In [2], Xu *et al.* has presented a survey paper in which he proposed decision theoretic solutions for channel selection in CRN which are not suitable for IoT scenario. Therefore, in this paper we suggest the decision theoretic solution for IoT based on the concept of CRN. Moreover, we also propose the novel framework for IoT which includes a new layer of decision making in the 3 layered hierarchical architecture. The addition of this layer in our proposed architecture equipped the old architecture with brain and the decision maker will select the appropriate object and channel for intelligent decision making. This motivates us to propose a framework capable of autonomous cognitive decision making autonomously for a wide range of applications ranging from smart homes to smart cities.

Let us consider a smart city scenario in which an elderly person Bob has opted for medical assistance from the medical center. In this scenario, he is equipped with the wearable smart device capable of monitoring the patient's health like body temperature, heartbeat, blood pressure etc. to name a few and sending this information to local intelligent decision maker in his smart home as shown in Fig. 2. This local CIOT decision maker system regularly obtains the health status of the patient through the connected device. The local system is equipped with four layered architecture which is capable of decision making at local level. In this framework, the VO and their corresponding CVO are created which represent all the functionalities of RWO as discussed above. For instance, the local decision maker obtains the regular health status update of patient and creates the corresponding VO and CVO accordingly in their corresponding databases. The local decision maker at Bob's house is connected to the medical center global decision maker server which is capable of monitoring the health of patient at all the time and informing the relevant doctor about his health status by selecting the appropriate CVO from object repository. Moreover, this intelligent framework also contains a repository for objects policies which includes all the information related to these VOs and CVOs. This repository facilitates the decision maker for selecting and re-utilizing the most appropriate existing object for intelligent decision making in future. Therefore, in this smart CIoT system the decision maker not only informs the doctor about his patient's health status but also alerts the staff at hospital in case of any emergency by selecting the appropriate CVO from its repository. Moreover, the other health saving support systems like ambulance service, paramedic staff etc. to name a few are also informed about the patient's current health status by the intelligent decision maker using CVO. In short, this intelligent CIoT system is connected to all the

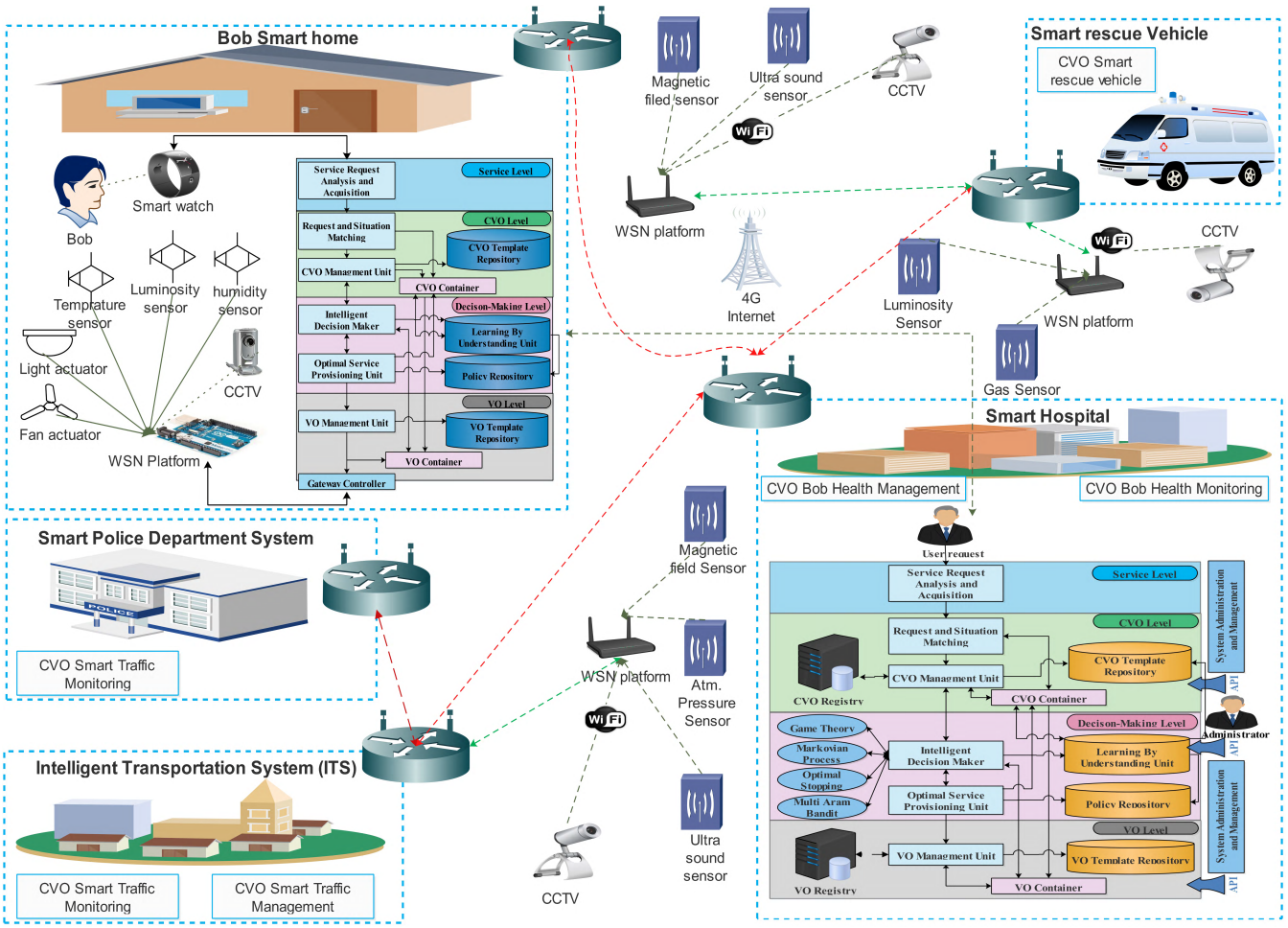

FIGURE 2. Smart city scenario using CIoT framework for intelligent decision making.

emergency service provider systems to save the life of any patient in case of emergency.

B. NOVEL FRAMEWORK FOR CIoT

The Literature review clearly shows that IoT is limited to interoperability, architecture, communication, security etc. to name a few and does not address the issues of autonomous decision making for smart city environment [8], [10]. Moreover, the current framework of IoT which consists of three layers is still inefficient for effective communication and intelligent decision making [7], [8], [10]. In this paper, a decision theoretic framework is proposed to cater the decision layer as shown in Fig.3. This novel framework is hierarchical in nature and composed of four layers with unique functionalities. The first level is known as VO level, second level is decision making level, third level is CVO level and upper most level is the service/stakeholder level. The intelligent components at each level are capable of providing self-management (configuration, healing, optimization, and protection) and learning. In other words, the entities at each level are capable of perceiving and reasoning on context for conducting associated knowledge based decision making (through associated optimization algorithms and machine learning), and autonomously adapting their behaviour and the configuration according to the derived situation. The aim of this revolutionary management framework includes:

- The implementation of intelligent learning methods that can enhance context awareness by providing the means to exploit more objects.
- The implementation of hybrid intelligent decision making algorithms that can improve the energy-efficiency by selecting the most suitable object among heterogeneous objects.
- The efficient management of heterogeneous resources of large scale system in the form of intelligent representation.
- The implementation of hybrid decision theoretic solutions for channels selection that can improve the spectrum efficiency of large CIoT system.
- The implementation of intelligent algorithms that enables the reliable service/application provision by rendering the high reliability through the ability to use heterogeneous objects in complementary manner.

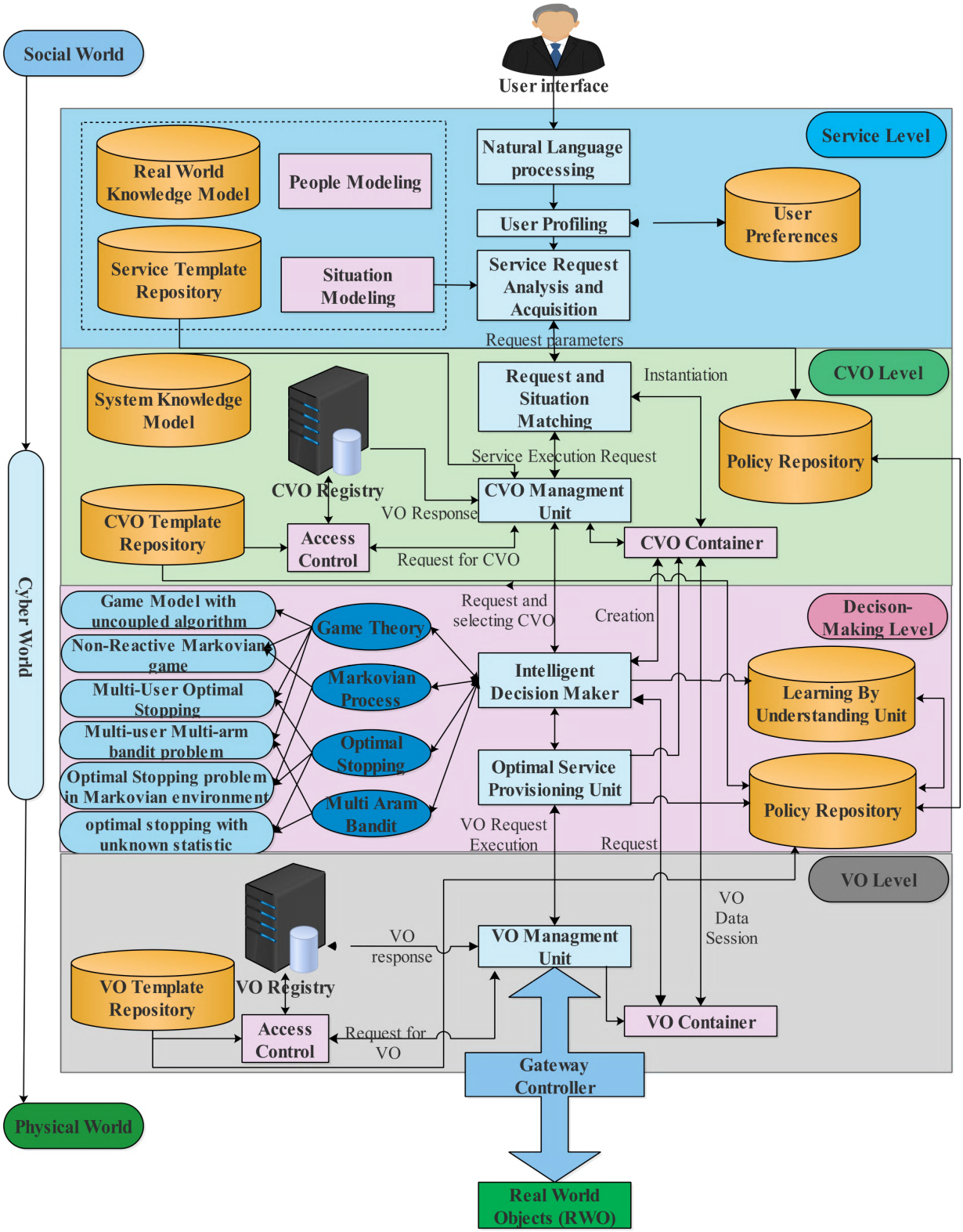

FIGURE 3. Novel CIoT framework diagram which introduce a new layer of decision making between VO layer and CVO layer in exiting architecture for intelligent decision making.

To achieve these goals, we have added an additional layer of decision making in the old architecture to support the decision making process. The addition of this layer has equipped the framework maker with brain and enables the decision maker to coordinate with other supporting databases i.e. CVO template repository, service repository, VO repository, policy repository etc. to name a few for intelligent decision making of channel selection and object reuse in CIoT. This layer

serves as a bridge between the service level and object level. The CIoT paradigm is hierarchical in nature and requires the game based algorithms for objects interlinking and management. Our proposed multi-agent framework is capable of analyzing, optimal learning and evaluating the decision-theoretic object and channel selection for intelligent decision making in CIoT. The heart of the proposed CIoT framework is intelligent decision maker which is equipped with the intelligent algorithms of game theory, markovian decision process, optimal stopping and multi-arm bandit. Our proposed framework provides an intelligent solution for decision making by combining hybrid or multiple decision-theoretic models for object and channel selection. The role of decision maker is most vital in the framework as it instructs other layers and their components for future actions related to object creation, updation, policies, etc. to name a few. For instance, the decision maker instructs the CVO management unit for the selection of existing object from CVO template repository or it can instruct the CVO container for the creation of new CVO object and updation of its policy. Moreover, the proposed framework also facilitates the decision maker for providing an adequate solution for appropriate object selection, existing object reuse and creation of smart objects by merging the existing objects into intelligent representation to minimize time for object creation, decision making, appropriate channel selection and efficient utilization of resources. This layer is also responsible for the management of policies and service provision to other layers. Moreover, this repository for objects policies includes all the information related to these VOs and CVOs. This layer also contains the database of learning by understanding where can the learning algorithms like Q learning [31], stochastic learning [32] and reinforcement learning [33] be implemented for intelligent object selection and reuse. Moreover, it also facilitates the decision maker to select the most feasible channel for communication. This database contains complete information about the semantics derivation, ontology, context of environment and knowledge discovery of objects. This component makes every object in CIoT intelligent by analysing the data and discovering some valuable patterns as knowledge. This repository also facilitates the decision maker for selecting and re-utilizing the most appropriate existing object and channel selection for intelligent decision making in future.

The first layer of our proposed architecture is VO layer which serves as a bridge between the real world objects (RWO) or IoT enabled devices. The IoT devices have sensors which collect information from the user/environment while actuators enable the devices to transfer this data to CIoT framework through gateway controller. The VO management unit contains all the functions related to VO including their creation, updation, deletion, etc. to name a few. Basically, this level provides a high-level interface to devices/objects by abstracting the complexity of the underlying CIoT infrastructure. The VO's contains all the information about discovery, exploitation and detailed description of objects/devices [8]. The information about every object is stored in the VO registry which includes the type of VO, the object that is connected to, all the functionalities and features that each VO can provide and other related information of each VO. Moreover, the VO template repository facilitates the VO registry for the creation of VO with predefined properties. This database also facilitates the policy repository for updation of new functionalities of VOs.

The CVO level is responsible for combining many VOs to establish CVO, which provides services to end-users as well as higher hierarchy applications [22]. At this level, the CVO management unit is responsible for handling all the databases/registries and performs coordinated and intelligent decision as per user requirement. All the information regarding the CVOs is found in CVO registries which includes detailed features of CVOs, VOs which are included in this CVO, all the related data related to CVO creation and its related features. The component of request and situation matching is responsible for building knowledge and experience related to all the CVOs which are created previously. This knowledge is then utilized by the intelligent decision maker to make intelligent decision robust and efficient. Basically, this component works with CVO management unit to search for the existing CVOs that can fulfil the requested service requirement. This enables the reuse of existing CVOs, increase efficiency in terms of time, and resource saving. More specifically, when the decision maker requests for a CVO then these components explore the CVO registry and try to find that particular CVO based on matching pattern. If this CVO exits then the decision maker reuses it. Otherwise it triggers the optimal composition of VO, which will dynamically create a CVO according to the requested functions and policies.

The service level enables the users to define the features of a required service/application through compatible interfaces and also provides all the functionalities for the fulfillment of requested services requirements independently. These services are composed of various features including performance, energy efficiency, etc. to name a few. Moreover, support for different parameters in terms of services like time, location, temperature etc. to name a few. must be supported by this level. This layer is composed of service request analysis and acquisition and other databases to provide these services. Initially, the service request from the user is translated by natural language processing component so that it can be understood by the remaining components of this layer. Basically, this component translate and infers the functions and policies which are requested by the user from the interface. The request analysis and acquisition component evaluates the conditions under which the services were requested and provides the corresponding parameters i.e. time, location etc. to name a few with the help of all the attached databases i.e. situation modeling, service template repository and real world knowledge model. The connected databases help in obtaining and learning information based on user preferences. This layer forwards the request of user to CVO level for the dynamic creation CVO from different VOs.

For instance, if Bob is having a heart attack, then the local decision maker at Bob's home gets the request from smart watch for assistance of medical doctor. Initially, the decision maker first searches for the appropriate existing CVO in the CVO registry. If it is found then it automatically reuses the appropriate CVO to provide medical assistance and triggers alarm at home as well as informs the global decision maker at medical center for assistance. Otherwise it combines many VOs and creates the CVO which will ultimately trigger the emergency situation condition at home as well at the hospital assistance server. The intelligent decision maker at this stage uses the game theoretic solution for selecting the appropriate object and efficient channel for transferring all the data related to Bob's current condition to the hospital assistance center. The efficient utilization of hybrid game theoretic solution not only saves time but also guarantees the secure transmission of information to both hospital and emergency service providers. The same intelligent framework and assistance system at hospital receives the information of emergency situation and makes decision accordingly. The global decision maker at hospital sends the ambulance by triggering the corresponding CVO from its respective databases. In this smart city environment, the intelligent traffic monitoring system will assist the ambulance to select the most appropriate route by using the coordination among several objects for best route to Bob's house. This coordination among the objects and selection of most suitable channel for wireless communication between wireless devices in smart home, smart hospital and smart traffic monitoring system can be achieved by using the proposed novel decision making layer that utilizes existing decision-theoretic solutions including game theory models, Markovian decision process (MDP), optimal stopping problem (OSP) and multi-arm bandit problem (MAB) for appropriate object selection and intelligent decision making in smart city environment. The intelligent framework also enables several decision makers to perform mutual coordination and management among several players in large scale CIoT by utilizing the functionalities of local interaction game and coordination ontology. In our scenario, the patient's life can be saved by using game theoretic model for global optimization using local interaction game among objects. This paper specifically focuses on decision theoretic solutions, cooperative decision making and global optimization using local interactive game model on large scale CIoT which is discussed in detail in later sections.

III. DECISION MAKING FOR CIoT

Decision making for CIoT involves continuous inference, process selection and interpretation of the meaning of acquired data accordingly. As already discussed, in CIoT applications, decisions might need to be taken e.g., informing the medical center, selecting best route for ambulance, acquiring channels for wireless transmission or even joint execution of task among multiple services. In this paper, we propose the solution for intelligent decision making which utilizes the learning capability of CRN for appropriate decision making based on historical information of smart objects in CIoT. In this paper, we analyse the suitability of four basic decision-theoretic models for CIoT applications.

A. GAME THEORY

Game theory [25] is a mathematical model for analyzing mutual interactions in multi-user decision systems. This model comprises of a fixed number of players, action commands and a utility function that maps player's actions into a real value. The models are defined as cooperative games and non-cooperative games. In cooperative game model, the players attempt towards sensible decision for maximizing utility function. Basically, the players (objects) are grouped together, and co-operate according to the agreed payoff portion. In a non-cooperative game, the solution concepts of NE and correlated equilibrium (CE) are deployed [27]. Moreover, the decisions of the players are distributed and autonomous which lead to self-organizing optimization [25].

To make mutual interactions among multiple players, we have to formulate a game for object selection. The game model is represented by $\mathcal{G} = \{N, A_n, u_n\}$, where player or object set is represented by $N = \{1, \ldots, N\}$, strategy set of objects is denoted by A_n and the utility function of n players is denoted by u_n. The pure strategy is selected by the game player when he selects a single action from his action set. Let us suppose that the strategy of object n is denoted by $a_n \in A_n$ and $a = \{a_1, \ldots, a_N\}$ represents the strategy profile of all the objects. Moreover, $u_n(a_n, a_{-n})$ represents the utility function while σ_n represents the mixed strategy of object. The notation $\sigma_n\ (a_n)$ represents the probability that object selects strategy a_n. Therefore, we can express the utility function in mixed strategy profile $\sigma = (\sigma_n, \sigma_{-n})$ as

$$u_n(\sigma_n, \sigma_{-n}) = \sum_{a \in A} \left(\prod_{n \in N} \sigma_n(a_n) \right) u_n(a) \tag{1}$$

The NE [2], [25], is the best solution for non-cooperative games in which each user maximizes its individual utility function by deviating unilaterally. In this scenario, it cannot deviate and maximize its utility function as it is supposed to know the equilibrium strategy of other players. Similarly, CE [25] also provides a perfect solution for better coordination and flexible utility function design by performing correlation among objects based on observation. CE is better than NE because it has convex sets which have the property to address fairness between the players, as compared to isolated points of NE. The non-cooperated game models are proposed in [2], [24], and [25] for NE and CE solutions. The following aspects must be taken into consideration:

- Implementation of utility function must be designed cautiously so as to avoid the users on wasting the resources [29]. This is important as game theory addresses the interaction among multiple decision-makers with no guarantee on the performance.
- Information updating and learning, specifically by achieving constant update of information rules for the

users. New learning procedures are required for a stable outcome. In this regard, our framework includes a component of learning by understanding of better practical solution about the unknown environment. The consideration here is to implement the reinforced learning algorithms or stochastic learning automata [23], [24] to practically converge to NE and CE of game for desirable solution in rapidly changing CIoT environment.

1) APPROPRIATE GAME MODEL FOR CIoT

This section provides insight knowledge of basic game models which are useful to analyse and model the user interaction with CIoT system. The paper also highlights the behaviour update rule which is most important rule for the implementation of game model. Moreover, all of them utilize the parallel sensing strategies for game based solutions. The current models suitable for CIoT include:

a: Repeated game

In CIoT, there are millions of objects which are distributed across the network and we need the repeated game which is played in finite or infinite horizon. In this intelligent game, the game players update their strategy according to their previous action-payoff. These games are best for modeling and analyzing the objects in distributed environment because the decision makers observe the environment by constantly accessing the spectrum. In [34], Xu *et al.* presented the solution for distributed channel selection in opportunistic spectrum access (OSA) system as repeated game. In [35], the authors also presented channel selection algorithm which is based on reinforcement learning for multi-user and multichannel distributed system. They performed the simulation of reinforcement learning based algorithms in static environment and showed that these algorithms converge to N.E of game. The reinforcement learning technique is implemented in time-varying spectrum environment for distributed channel selection in OSA system [36]. In [23], Xu *et al.* proposed an intelligent learning algorithm which is known as stochastic learning automata that converge to Nash Equilibrium of game.

b: Graphical game

The most efficient and effective game model suitable for large scale distributed radio network is graphical game or local interaction game [37] or spatial game [38]. In this game, the actions of a game player are only affected by its neighboring players instead of all players of the game. Basically, the transmission by the player has an affect on its neighboring players which are in its transmission range and do not affect the distant players. Therefore, it results into spatial-reuse in the cognitive radio system and can be successfully applied in CIoT. In [32], Smith proposed a regret minimization algorithm for free-use OSA system to converge to NE. Xu *et al.* [39], proposed a share-use OSA system which utilizes the intelligent learning algorithm that converges to evolutionary stable strategy (ESS). In [23], Xu *et al.* designed a share-use OSA system for minimizing the collision level and maximizing the throughput of the system. The formulated graphical games in [23] and [21] are called potential games. Marden *et al.* [40] showed that the behavior update is best response which is an average optimal solution. On other hand, in [23] Xu *et al.* formulated the spatial adaptive play [41] which is asymptotically optimal with local information exchange. Moreover, literature showed that the graphical games can be formulated as spatial congestion game [42] for opportunistic spectrum access systems. The payoff of the players in congestion game is a function of the number of players which are in contact with it and utilize the same network resources [43]. In previous literature, the authors basically tries to investigate the conditions under which a pure NE strategy can be achieved in spatial congestion games.

c: Evolutionary game

This theory was initially presented by the biologists to calculate the population dynamic [39]. The evolutionary game was then formulated based on this theory to form a evolutionary stable strategy (ESS) [39]. In [39], Xu *et al.* present the ESS to calculate robustness and define the utility function over statistics. In this paper, the ESS is categorized by robustness against invaders with the property that the population remains unchanged until ESS is reached. Moreover, this technique is also efficient in perturbations by a small number of players. This exiting game is applied successfully to wireless system and can be applied to CIoT for multiple access of objects, cooperative spectrum sensing and network selection [44]. In [45], Monderer and Shapley propose the channel selection algorithm in shared OSA system as an evolutionary game. The authors have showed that the complete network information having Bernoulli distribution in each slot can converge to an ESS when replicator dynamics are applied to it. Basically, this work provides an adequate solution for interactions among the players and dynamics of channel selection in OSA system. The robust game is presented in [24], in which the authors explained in detail about the dynamic and random deployment of cognitive small cells (CSCs). In these cognitive small cells, the utility function is defined by the expected capacity in all possible active sets of cells. The author proved it by simulation in dynamic environment that the algorithms of distributed leaning automata are successfully applied to these potential game that will converge to NE.

d: Coalition game

The coalition game is designed to form a group or cluster of set of players to achieve better coordination and increased payoff in distributed system. This game is also useful in large scale CIoT as it coordinates among the distributed decision makers and objects. In [46], Xu *et al.* proposed that different countries can form coalition to improve their human potentials while players can also form coalition to improve their sensing performance. The problem of spectrum sensing and access in opportunistic spectrum access for the partitioned network is discussed in [33]. From the previous literature

review, the benefits of coalition game are highlighted which include: (i) coalition game provides better performance and throughput by sensing different channels and sharing their information to reduce time for sensing. (ii) the interference can be reduced by joint coalition of players for the channel access. (iii) the channel capacity can be improved by joint coalition of the players by distributing their total power over multiple channels. In short, the performance of the system can be improved by using coalition game by coordination and cooperation among the players [47].

e: Hierarchical game

All the game models discussed in previous sections are highlighting the interaction of players with equal priority or having no hierarchy due to distributed structure of the network. Moreover, the CIoT system are hierarchical in nature and require the hierarchical game based algorithms for objects interlinking and management. The most useful game model for hierarchical network is stackelberg game [47]. In this game model, the game players are represented by the leaders and several followers who compete for certain resource. In this game, the leader takes an action based on a situation and the followers take actions according to the leader or follow its own actions. Moreover, both the leaders and followers can not deviate from the stackelberg Equilibria [48]. In some cases, both leader and followers maximizes their utility function or leader has no utility and his (resource user) aim is to maximize the accumulated utility of the followers. The results of these researches clearly showed that the efficiency of NE can be improved significantly by using Stackelberg equilibria. In [49], Xiao *et al.* proposed the intervention rule for hierarchical game model. In this model, the leader selects the proposed intervention rule and then the followers choose their action according to the selected intervention rule of the leader. This hierarchical game model is called intervention game model in which leader (resource user) regulates the resources which are shared among the followers. In [24], Xu *et al.* use the hierarchical game to propose cluster-based hierarchical structure for self-organization and optimization in large-scale networks. They presented the rule for calculating the computational complexity in cluster-based hierarchical game based on Q-learning approach [50]. The results showed that the cluster based hierarchical approach is most suitable for dense network. The need of the hour is to design the application for distributed channel selection in CIoT system which is based on hierarchical games for better resource allocation and joint power control.

f: Bayesian game

Finally, the Bayesian game model is also a suitable game model for the CIoT [51]. This evolutionary game model is very useful in the situations where each player is uncertain about the other player knowledge and payoff. This exiting game model represents the fusion of game theory and probability theory to solve the problem of incomplete and uncertain information. In any game, the players have their own private information and it is hidden from other players but this information has a great impact on overall game play. The players use this private information and knowledge which is generated through probability distribution to act optimally according to the situation [52]. In [53], Huang and Krishnamurthy have used the Bayesian game to model the network selection problem. In this research, the players of the game are the users and their action set is represented by the selection of available access network. The information about the preferences of other user is partial and uncertain. The previous literature clearly showed that the Bayesian game model used by the players can reach Bayesian NE in an environment with incomplete information.

2) SUMMARY AND INSIGHTS

The repeated games are suitable for modelling and analyzing the objects in distributed environment because the decision makers observe the environment by constantly accessing the spectrum. Basically, the game players update their strategy according to their action-payoff history in previous play. In this game, the reinforcement learning based channel (object) selection algorithm for multi-user and multichannel distributed system can be applied that converges to NE of game. On the other hand, the most efficient and effective game model suitable for large scale distributed radio network is graphical game in which the actions of a game player are only affected by its neighboring players instead of all players of the game. Moreover, this game can be formulated as spatial congestion game in which the pay-off of a player is a function of the number of players which interact with it and use the same network resources to converge to NE. In this game, the exponential learning algorithm can be applied for the selection of appropriate object among several objects to maximize the throughput and minimize the collision level.

In evolutionary game, the game players use the evolutionary stable strategy (ESS) to provide an adequate solution for interactions among the players and dynamics of channel selections in OSA system. Moreover, the ESS is used to calculate robustness and define the utility function over statistics. On the other hand, the coalition game is designed to form a group or cluster for set of players to achieve better coordination and increased pay-off in distributed system. This game is useful in large scale CIoT as it coordinates among the distributed decision makers and objects. This intelligent game provides better performance and throughput by sensing different channels and sharing their information to reduce time for sensing. Moreover, the interference can be reduced by joint coalition of players for the channel access. Moreover, the channel capacity can be improved by joint coalition of the players by intelligently distributing their total power over multiple channels.

The Hierarchical game model is the most appropriate game model for CIoT because of its hierarchical nature and requires the hierarchical game based algorithms for objects interlinking and management. The most important game model for hierarchical network is Stackelberg game in which the leader

takes an action based on a situation and the followers take actions according to the leader or follow its actions. In this game model, both leader and followers maximize their utility function or the leader has no utility and his (resource user) aims to maximize the accumulated utility of the followers which results into efficiency of NE significantly. Finally, the Bayesian game model is also suitable for the CIoT as it is very useful in the situations where each player is uncertain about the other player knowledge and pay-off. The players in the game use their private information and knowledge which is generated through probability distribution to act optimally according to the situation.The information about the preferences of other user is partial and uncertain. The Bayesian game model used by the players can reach Bayesian Nash equilibrium in an environment with incomplete information.

B. MARKOVIAN DECISION PROCESS

MDP provides an excellent framework for modeling decision making in CIoT for several objects of multi-periods. Basically, it calculates the probability that the system will transfer from one state to another state based on maximum discounted reward for optimal policy calculation. The probability of outcome does not rely on the previous state of the players. Markov decision process describes an environment for reinforcement learning in which the environment is fully observable i.e. the current state completely characterizes the process. The problem with sequential decision making of multi-periods in CIoT can be well-solved by Markov decision process (MDP) models. The most attractive feature of this technique is the formulation of spectrum sensing and channel selection as an MDP problem. This section highlights the brief overview of three basic types of models namely: the discrete time MDP (DTMDP), partially observable MDP (POMDP) and constrained MDP [2], [24] that enables intelligent decision making in CIoT.

1) DISCRETE TIME MARKOVIAN DECISION PROCESS (DTMP)

In DTMP model, the process can be observed periodically and classified into one of the possible states, an action from the possible action is taken and as a result, the process will return a state in which the user has to switch. The basic elements of Discrete time MDP model are defined as:

- The time interval $k = 0, 1, 2, \ldots$
- Set of states $s \in S$.
- Set of actions $a \in A$.
- Reward function $R : S \times A \rightarrow \mathbb{R}$. Moreover, the received reward is represented as $R(s, a)$ when player perform action a at state s.

In this model, when a player performs an action a at state s, the system will transit from state s to s' in the next period using the stochastic transition model probability $Pr(s'|s, a)$. The game player calculates the maximum discounted reward for state s by mapping the states to actions for optimal policy $\pi(s)$. It is given by

$$V^*(s) = \max_{\pi} E\left[\sum_{k=0}^{\infty} \gamma^k R(s_k, a_k) | s_0 = s, a_k = \pi(s_k)\right] \quad (2)$$

where $E[.]$ represents the expected operation and optimal discount factor is represented by $\gamma \in [0, 1]$. Moreover, the maximum discounted future reward for state and action pair (s, a) is calculated by the Q-value function which is given by

$$Q^*(s, a) = \max_{\pi} E\left[\sum_{k=0}^{\infty} \gamma^k R(s_k, a_k) | s_0 = s, a_0 = a, a_k > 0 = \pi(s_k)\right] \quad (3)$$

This method provides the optimal policy which is stationary and deterministic. Here, the stationary represents the optimal action which is denoted by $\pi^*(s)$ for every state s, while deterministic represents the single action per state which is $\pi^*(s)$:

$$\pi^*(s) = \arg_{a \in A} \max Q^*(s, a) \quad (4)$$

where the value iteration or known as Q-learning technique is used to calculate the optimal Q-value $Q^*(s, a)$ for each state-action pair (s, a) [21], [54]. Basically, the system states are observed by each of the player in decision period for DTMDP model.

For instance in our practical example of real world scenario, the intelligent decision maker periodically observes the health status or state s of Bob and takes action a when he is having a heart attack. In short, the decision maker will facilitate the game players to change their states s from normal to new state s' of heart attack on receiving the recent reading from the sensors. Therefore, the utilization of DTMP algorithms by the decision maker will enable Bob to take action a to transit from normal state s to heart attack state s' in the next period using the stochastic transition model probability $Pr(s'|s, a)$. The biggest benefit of implementing the DTMDP models in decision making layer is that the system state is completely observed by the players in each decision period. The decision maker also facilitates the game player to calculate the maximum discounted reward for state s by mapping the states to action for optimal policy $\pi(s)$ as discussed previously. Moreover, the implementation of Q-learning technique is used to calculate the optimal Q-value $Q^*(s, a)$ for each state-action pair (s, a) [21], [54]. The decision maker facilitates each player for taking an appropriate action according to the health status of each player. The addition of decision making layer into the framework and utilization of DTMP algorithm facilitates Bob for taking action when his health status changes from normal to heart attack by selecting the related CVO and channel for communication by using Q learning technique to alert the medical assistance system for sending help to the user. Moreover, the utilization of DTMP algorithm also facilitates the decision maker to

inform the doctor about his patient's health status and also alert the hospital staff in case of any emergency by selecting the appropriate CVO from its repository and appropriate channel for communication. Finally, the other health saving support systems like ambulance service, paramedic staff etc. to name a few are also informed about the patient's current health status by the intelligent decision maker using CVO and appropriate channel selection. In short, this intelligent CIoT system is connected to all the emergency service provider systems to save the life of any patient in emergency.

The previous literature has shown that the correlation between different objects can be calculated using availability vector which is capable of identifying the state information of each object. Moreover, this model is also used to predict the channel's availability as well as user activity. Chen *et al.* [55] formulates MDP model using availability vector to calculate the activities of PU. The results showed that the usage collision probability and channel availability vector facilitate the SUs to sense multiple channels sequentially in a slot and also reduce the sensing overhead.

2) PARTIALLY OBSERVABLE MDP (POMDP)

In POMDP, the agents partially observe the state of the system and then the dynamics of the system are determined by markovian decision process. Basically, the agents rely on probability distribution over set of possible states, based on a set of observation and probabilities of the Markovian process. In short, the agents partially observe the state and are not sure about the current state or current state emit observation. Moreover, the system states are partially perceived by each player during decision period for POMDP model. Such observation and unknown states create problems for the MDP. In this technique, the main goal for an agent is to choose such actions at each time step which will result into the maximization of its expected future discounted reward:

$$E\left[\sum_{t=0}^{\infty}\gamma^{t}r_{t}\right] \tag{5}$$

As discussed in the previous section, the agents are only concerned about the largest expected reward and perform the action based on this reward when $\gamma = 0$. On the other hand, it tries to maximize the expected sum of future rewards when $\gamma = 1$.

One of the solution for this problem is the interaction of the user with environment and collecting the observation. Then, the agent has to update its belief in true state by updating the probability distribution of its current state. This can be achieved by creating the vector which includes the conditional probability of continuous updating of each state of the system after each decision [56]. Hence, the optimal policy can be calculated by the user based on the maintained belief vector [56], [57]. In this model, every agent has to wait at some state for decision update by the other agents and then choose individual actions simultaneously according to decision. Moreover, the agents keep the record of each state and decision through information exchange mechanism. An ideal solution for distributed decision approach for multiple cooperative agents is provided by the Decentralized Partially Observable MDP (DEC-POMDP) [58]. This intelligent model allows each agent to observe only a small part of the system state or each agent can only access local information. Moreover, the joint optimal policy based on approximation of agents is intractable because the decision made by each agent is independent and they do not know the actions and states of other agents [59]. Finally, This model enables each agent to observe its local state instead of all the local and global states, particularly suiting the CIoT requirement in application areas like smart cities, which consist of various heterogeneous objects and dynamic services.

In [60], Zhao *et al.* have presented the framework and algorithm which is capable of perfect and imperfect spectrum sensing. The simulation results proved that the throughput is improved significantly using this algorithm. In [61], Chen *et al.* presented an intelligent strategy as a constrained POMDP for jointly sensing and accessing the OSA system. Basically, they present the separation principle which confirms that the optimal strategy of sensing and access leads to optimal solution. In [31], Unnikrishnan and Veeravalli have used the greedy algorithm for channel selection and derived an optimal policy for calculating the channel availability statistics. This algorithm will enable the user to calculate the real time statistics to ensure that the collision constraint is minimized. In [62], Hoang *et al.* has calculated the spectrum sensing and access control problem. The results showed that the total reward of a slot can be calculated by sensing duration and probability that the channel is idle. This reward is then used to calculate the false alarm, mis-detection and energy consumption in that particular slot. Basically, the author calculates the duration of sensing in each time slot to maximize the net reward.

3) CONSTRAINED MDP (CMDP)

This intelligent model is capable of solving the constraints imposed by the user in sequential decision problems by analyzing them and modelling them in an appropriate way to remove the constraints. Practically, there are always constraints imposed by the users in practical scenarios e.g. the interference imposed by users etc. to name a few. The sequential decision making problems are well managed by the CMDP model. The CMDP model is quite close to the MDP except the difference is that there is additional computational cost in calculating the policies. The cost for calculating the optimal policy p is represented by:

$$\begin{aligned} &\min\ C(p) \\ &s.t.\ \ D(p) \leq V \end{aligned} \tag{6}$$

where cost functions are represented by vector $D(u)$ having dimension Nc, of constant values. The Constrained MDP is equal to the linear program by using the discounted cost and

is represented as:

$$\begin{aligned} &\min\ \langle \rho, C \rangle \\ &s.t.\ \ \langle \rho, D_n \rangle \leq V_n,\ \ n = 1, \ldots, Nc \\ &\quad \rho \in Q \end{aligned} \tag{7}$$

where the cost vectors are represented by C and D_n with dimension $|K|$. On the other hand, the ρ vector is defined as:

$$\begin{cases} \sum\limits_{y \in S} \sum\limits_{a \in A(y)} \rho(y,a)(\delta_s(y) - \alpha \mathcal{P}_{yas}) = (1-\alpha)\beta(s) & \forall s \in S \\ \rho(y,a) \geq 0, & \forall y, a \end{cases} \tag{8}$$

where s is the state at which the action a is performed for calculating the probability of selection of each element $\rho(s,a) \in Q$. Hence, the summation of $\rho(s,a)$ converges to 1 for all $\rho \in Q$. We can calculate the stationary optimal policy p by:

$$p(a|s) = \frac{\rho(s,a)}{\sum\limits_{a \in A(s)} \rho(s,a)} \tag{9}$$

The previous literature showed that the CDMP can be used to maximize the throughput along with minimizing the collision probability requirement [63]. Moreover, the new attractive heuristic algorithms like memory-less access and greedy access can be utilized for optimal solution of decision making in CIoT. The memory-less access algorithm [63] is optimal in the scenarios where the collision constraints are tight. Moreover, the maximum throughput region can also be obtained using this algorithms [64]. The result of this study showed that in case of tight collision constraints, the outer and inner bounds match. These studies modeled Pu traffic but were unable to cater the QoS of the SUs. In [65], Niyato *et al.* have considered not only PU traffic but also catered SU to calculate the maximum probability of collision, maximum packet loss and packet delay for vehicular ad-hoc nodes. Hence, this technique can enable the opportunistic spectrum access, object selection for channel reservation and clustering control from hierarchical Markovian model.

C. OPTIMAL STOPPING PROBLEM

The optimal stopping model deals with the problem of sequentially selecting the actions which are based on a sequence of observed random variables to maximize the expected reward and to minimize the expected cost. Basically, this process observes each variable based on its reward and performs a stopping action to minimize the cost. Hence, we can say that the observed variables are used to calculate the current reward. It provides an adaptive and efficient spectrum discovery mechanism by constantly observing the fixed channels with their overhead consideration. The OSP model has sequence of:

- The random variables, $X_1, X_2, \ldots, X_N$, whose joint distribution is known and their realizations are denoted by $x_1, x_2, \ldots, x_N$.
- The mapping function in which the observed random variables are mapped to real-valued rewards, i.e.,

$$y_1(x_1), y_2(x_1, x_2), \ldots, y_N(x_1, \ldots, x_N) \tag{10}$$

In real world scenario, when the player observes the nth variable, x_n and stops, then he will receive a known reward $y_n(x_1, \ldots, x_n)$. Moreover, the future reward i.e., $y_m(x_1, \ldots, x_n, x_m)$, is random and unknown, if we choose to proceed to observe and stop in the m^{th} variable where $m > n$. The backward induction model is very useful in solving the OSP models with finite horizon [66]. Our goal is to stop at stage N. Firstly, the optimal rule at stage $N-1$ is calculated by the user. Then he can proceed to stage $N-2$ to calculate the optimal rule at this stage along with information about the rule of previous stage i.e. $N-1$. Basically, the backward induction model is used for finding the optimal rule by backward iterations to calculate the optimal rule for the future stages [67]. The stage value can be calculated as:

$$V_n = \begin{cases} y_N(x_1, \ldots, x_N), & \text{if } n = N \\ \max \left\{ y_n(x_1, \ldots, x_n), \right. & \\ \left. E[V_{n+1} | \{x_i\}_{i=1}^n] \right\}, & \text{if } n < N \end{cases} \tag{11}$$

$E[V_{n+1} | \{x_i\}_{i=1}^n]$, represents the expected reward for the future stages using the optimal rules. The expected reward can be maximized if we stop at stage n if $y_n(x_1, \ldots, x_n) \geq E[V_{n+1} | \{x_i\}_{i=1}^n]$, or to continue otherwise. This technique can be implemented by using two models: No Recall-OSP and Recall-OSP (R-OSP).

1) NO RECALL OSP (NR-OSP)

In NR-OSP models, the decision maker performs decision on the basis of current variable by simplifying it as $y_n(x_n)$. Basically, the decision making is done on the basis of currently observed variable while the recalling of previously observed variable is not allowed. The backward induction model can be easily created for NR-OSP models. In [68], Sabharwal *et al.* have explained in detail about channel having good quality and also throw light on the issue of time overhead for exploring multiple channels. They have considered the Rayleigh block-fading identically and independently. They have implemented the NR-OSP model to explore the channels to achieve the expected throughput. In [69], Jiang *et al.* have utilized the NR-OSP model to explore the sensing order and channel quality. The results showed when the probabilities of channels availability are in descending order then the optimal solution for channel sensing can not be guaranteed. Therefore, they have introduced the dynamic programming method for finding the optimal sensing order but this will lead to higher computational complexity. In [70], Cheng and Zhuang have used NR-OSP to sense each channel according to its descending achievable rate. To summarize all these OSP model, we conclude that the user released the accessed channels for a pre-defined duration and then followed the same sensing process in next iteration. In [71], Li *et al.* have presented two

dimensional NR-OSP model for sequential sensing and channel access in time frequency domain. Basically, they have used NR-OSP model and finite state Markovian channels for calculating when and which channel has to be accessed and released by the user. They have formulated the problem according to three actions i.e. firstly utilizes the current available channel and then continue to sense the available channels and finally when the new channel is sensed then this current channel is freed accordingly.

2) RECALL OSP (R-OSP)

In this model, the usage of the previously observed variable is allowed and the decision is supported by the previously observed state. Basically, the R-OSP observes the previous variable to recall $x_k, k \leq n$ by decision maker. The backward induction model is not feasible for this technique because of the large computational complexity which increases exponentially as the number of decision horizons increases. Therefore, we have to modify the NR-OSP with k-stage look ahead to overcome the computation complexity. This k-stage look-ahead rule enables users to sense the successive k stages and then stop. The most intelligent and simple solution for this problem is known as 1-SLA rule which makes the truncated version of the problem [67]. Moreover,this rule provides an optimal solution for monotone R-OSP models [66].

In [72], Jia *et al.* have presented the R-OSP model for finding the trade-off between idle channels and multiple channels sensing overhead. They have applied k-SLA rule, where $k = 1, 2$ are used. The results have shown that 1-SLA rule is excellent because it is very close to optimal solution. In [73], Chang and Liu have used the R-OSP model to calculate the optimal strategy using 2-SLA rule to achieve optimal solution within finite steps. Basically, they have taken three action i.e. they used an observed channel and sensed the unobserved channels and then used those unobserved channels using the 2-SLA rule to find the optimal solution. In [74], Kim and Giannakis have presented R-OSP model in which the SUs adopted the parallel sensing technique. The results have shown that the sensing capability increases when the sampling duration is increased which leads to more spectrum opportunities for SUs. Moreover, the collision constraints imposed by the PUs can also be minimized by using the dynamic programming to find the optimal rule for the choosing best time to stop sensing and the best set of channels to access. In [67], Xu *et al.* have used R-OSP model for finding the energy-efficient channels. Basically, they have utilized the 1-SLA rule for the maximization of throughput. Finally, in [75] and [76], the authors have proposed the R-OSP model using 1-SLA rule which enables the collision avoidance among SUs. Moreover, they have also presented the solution for trade-off between channel exploration and exploitation in OSA systems.

D. MULTI-ARMED BANDIT PROBLEM

Multi-armed bandit (MAB) problem provides best learning technique for choosing one or more objects among several objects whose statistical information is unknown. Basically, it explores the statistics about resources during the decision process and maximizes the current reward on the basis of current estimated statistics. The classical MAB problem has players who are playing with K arms in equally divided time slots. The real-valued reward is obtained by choosing any one of the arm to play by the player at each time slot. The main purpose of MAB is to model a learning policy π based on the history information of decision to maximize the cumulative reward. The performance is evaluated by using regret matrix. This matrix calculates the reward loss of the policy and is written as:

$$R_\pi(T) = T\mu^* - \sum_{t=1}^{T} r_{t\pi}(t) \tag{12}$$

where the maximum expected reward is represented by $\mu^* = \max\{\mu_k\}$, $\pi(t)$ represents the selected arm at time t, and $r_{t\pi}(t)$ shows the reward in time slots n. The main purpose is to make $R_\pi(T)$ as smaller so that the time regret will to 0 i.e $\frac{R_\pi(T)}{T} \to 0$, as $T \to \infty$. Consequently, this will result into maximum time-averaged reward.

The solution for expected regret in term of linear arms is asymptotically logarithmic with time as $O(K \log T)$ is proposed by Lai and Robbins [77]. Their work is then further extended for multiple arm solution in [78]. Gittins and Gittins proposed index policies based on sample mean and briefly described the upper confidence bound case (UCB1) algorithm in [79]. In this algorithm, the arm with the highest index $\hat{\mu_k}(T)+\sqrt{\frac{2\log T}{m_k}}$ is chosen at each decision epoch T, where the measured expected reward of arm k in each epoch T is $\hat{\mu_k}(T)$ and m_k represent arm which is played k number of times. The first part of the equation represents exploitation while the second part is for exploration because of the less chances of that arm to be played. MAB is extensively studied in previous literature and it is categorized into rested and restless MAB. In former type, the state of arm evolves when it is played otherwise it is frozen. While on the other hand, the later type involves the arm state evolution and it is independent of the actions. For restless MAB the optimal policy is calculated by using highest Gittin's [80] index policy of playing the arm each time. While on the other hand, the restless MAB is calculated by using highest Whittle's index [81] policy of playing the arm each time. This technique can be implemented by using two models: i.i.d. MAB and restless MAB. These two models are explained briefly in next section.

1) I.I.D. MAB

This model is used to model each arm as an i.i.d. process with an unknown distribution and unknown mean. In this model, the players are unaware about means or distribution of rewards from different arms. In this scenario, the players do not have any dedicated control channels for communication. Therefore, the case in which two players select the same arm, neither of them gets any reward. In [82], Lai *et al.* have modeled each channel as an arm and applied a UCB1

algorithm on it. The results have shown that the probability of selecting a channel is directly proportional to the measured expected reward. This work is extended in [83], Lai *et al.* have calculated the user-channel matching problem in which various transmission rates for multiple users are designed as MAB problem. In this paper, the authors performed a user and channel matching based on an arm and applied UCB1 algorithm which is then scaled with all the matching profiles. Basically, the authors have modified the UCB1 algorithm based on the correlation between different arms. The simulation results are order-optimal because the regret increases logarithmically in time while increases polynomially in the number of channels. In [84], Anandkumar *et al.* have utilized the ϵ-greedy algorithm for the selection of collision free channel selection profile. Moreover, they have used adaptive random UCB1 algorithms which enables the SUs to randomly choose a channel only if the collision occurs in the previous slot otherwise it proceeds with UCB1 algorithm. The results showed that this policy is capable of handling the collision among multiple SUs which are selecting the same channel and has a logarithmic order. In [85], Liu and Zhao have designed N-parallel algorithm which enables the players to access multiple channels simultaneously. In [35], Montanari and Saberi have proposed a decentralized optimal learning policy which is order-optimal because it has a logarithmic order. In [86], Gai and Krishnamachari have proposed selective learning policy based on two algorithms for the case of prioritized users and equal access policy for all users. Basically, they have provided a solution for determining the access policy in distributed system which is based on selective learning policy of k-th largest expected reward or UCB1.

2) RESTLESS MAB

The models in which the arm state evolution is continuous and it is independent of the actions of players is called restless MAB. In Restless MAB, a player selects K arms out of N arms to play at each time. Basically, the state of the arm decides the reward for the player when the it is played and transmitted according to Markov rules regardless of their active or passive states. These Markovian dynamics are known to the players in advance. The main purpose of this model is to design an optimal arm selection policy which will maximize the long term reward. The performance of this model is measured by the regret for arm selection policy. Basically, the regret is measured in terms of reward loss for the scenario in which the player knows which K arms are the most rewarding arms and always plays these K best arms.

The previous literature [87] showed that the PU activities can be modeled as a Markovian process and the channel quality exhibits the Markovian properties. However, these statistics must be known in advance in restless MAB models. In [88], Tekin and Liu presented the restless MAB solution for single user in which they utilized the UCB1 algorithms. Basically, they have modified the previous UCB1 algorithm for formulating the regenerative cycle algorithms and mean index policy to get the expected regret in logarithmic order. In [89], Wang and Chen have designed an intelligent myopic policy which provides an optimal solution to the problem. The literature has shown that one of the major limitation of previous restless MAB models is that they are capable of handling the single user scenarios and the policies which are used previously are not applicable to multi-user scenarios.

E. SUMMARY AND INSIGHTS

Game theory provides a mathematical solution for analyzing the mutual interaction among users in multi-user decision systems. Basically, game models provide an efficient framework for the selection of the most appropriate object among several objects. In cooperative game model, the players (objects) are grouped together, and co-operate according to the agreed pay-off portion for decision making and maximizing utility function. In a non-cooperative game, the concepts solution of NE and CE are deployed. Moreover, game theory provides utility functions for appropriate object selection, policy for multiple access, self-organizing optimization, rate adaptation etc. to name a few among several objects to achieve the optimal solution. MDP on the other hand provides the framework for modelling decision making in CIoT for several objects of multi-periods. Basically, it calculates the probability of the system to transfer from one state to another state based on maximum discounted reward for optimal policy calculation. The probability of outcome is not dependent on the previous state of the player or system. POMDP is an extension of MDP in which the player or object does not know its current state and uses the probability distribution about its belief state. Basically, the system takes an action based on its belief state, it observes what happens next and updates its belief state according to the expected reward.

Optimal stopping problem (OSP) refers to the sequential problem of deciding actions which maximizes the projected reward, based on observed pattern of random variables. Such process basically observes each variable based on its reward and performs a stopping action to minimize the cost. It delivers adaptive and capable spectrum discovery mechanism by constantly observing the fixed channels with their overhead consideration. One the other hand, selection among several objects whose statistical information is known is provided by the multi-armed bandit (MAB) problem. This efficient learning technique explores the statistics of objects during the decision process and maximizes the reward on the basis of current estimation statistics about objects. Basically, this technique utilizes more time on estimating the statistics of each object during decision making for better rewards for the objects based on statistical information and decision making.

IV. COMPARATIVE ANALYSIS OF DECISION THEORETIC SOLUTIONS

This section presents the comparative analysis of all the four decision theoretic methods for the practical implementation in our proposed CIoT framework. It also provides a detailed insight review of the above discussed decision theoretic knowledge assisted learning solutions in terms of

their information retrieval, performance and convergence speed.

A. INFORMATION SELECTION AND RETRIEVAL

In CIoT, the information exchange is the most challenging part because of the dynamic, uncertain and always changing condition of the IoT ecosystem. Moreover, the problem of heterogeneous, non-linear, high dimensional and parallel processing of objects data adds more complexities to it. From practical aspect, the spectrum access systems must consider the environmental conditions which include the information about the spectrum, channel quality and traffic demand of user. The essential criteria is to apply the cognitive radio standards related to spectrum allocation, nodes selection, transmit power control and communication protocols for the effective usage of radio resources by the objects [2]. The information exchange in CIoT can be static or dynamic, local or global, realistic or unrealistic, and known or unknown. Firstly, the game theory provides an excellent solution for capturing the interaction and coordination among multiple users (objects) and their behaviors. The quantity, precision and accuracy jointly characterize the quality of information [44]. Quantity means how useful is the information which is obtained from a specific task. Precision is defined as the proportion of relevance of information among all the information. Finally, the accuracy is referred to as information relevance to the decision maker. The need of the hour is to practically design the coupled game learning solution which requires information about the users and uncoupled algorithms which are capable of decision making based on the local information of the users (objects). The dynamic and correlated states are well addressed by the theory of Markovian process [87]. Basically, it observes the state of each object, shares it with other objects and makes optimal joint policy to reach a global stable state. The scenarios in which the statistical information is unknown, a priori is well managed by the multi-arm bandit problem [91]. It provides considerably best learning techniques for choosing one or more objects among several objects whose statistical information is unknown. The problem of uncertain realization of unexplored channels is well addressed by the theory of optimal stopping problem [81]. In fact, this process observes each variable (object) based on its reward and performs a stopping action to minimize the cost.

B. COST

Performance is also an important aspect in CIoT learning decision theoretic models. The strategy will take long to update the player information based on history action payoff to explore all the possible selections. The performance in these models is measured according to the cost of resource utilization and action switching. The literature showed that the information exchange which involves the resource utilization causes extra overhead on network resources like power, updating, discovery, time etc. to name a few [81]. Moreover, it is also affected by the action switching in the network which involves the hardware re-synchronization and re-configuration according to the newly updated chosen action or strategy [90]. The solution for reducing the resource and network utilization is to design such uncoupled algorithms which are dependent on local information and are capable of combining the decision theoretic games model with MDP, MAB and OSP for multi-user system. The literature indicates that uptill now, only the solution for coupled algorithms is reported, which is useless for multi-agent system [2]. In this paper we proposed the optimal solution for decision making in CIoT which is a combination of multiple decision theoretic models i.e game theory, MAB, MDP and OSP for achieving better performance in multi-user system. The action switching cost can be reduced by including the action cost in the optimization objectives. This can be included explicitly into the problem formation for optimal stopping problem which considers the overhead of resource selection and discovery [91]. Moreover, other decision theoretic solutions including game model, MDP and MAB problem will automatically converge to optimal solution by the trial pay-off history of the players (objects) as presented in Table 2. This behavior leads to enormous switching cost as it considers the history of players repeatedly [90]. Therefore, only optimal stopping theory provides better performance and less cost as compared to other three decision theoretic models.

C. CONVERGENCE SPEED

A better learning technique should be designed to increase the convergence speed and minimizing the cost. The convergence speed is the most important aspect that should be considered in designing an efficient learning algorithm as it reduces the information overhead and cost by adapting the environment. These models only consider the convergence property and do not achieve the convergence speed which is important in the development of practical system. Moreover, the background study shows that they will achieve convergence only when the iteration number increases at large amount [2], [81]. Therefore, they create a large overhead due to asymptotical convergence and cost which is not suitable for practical system implementation. Finally, the optimal stopping problem provides the decision solution for just one-shot, which means that it is unable to provide any solution for convergence speed.

V. OPEN ISSUES AND FUTURE DIRECTIONS FOR INTELLIGENT DECISION MAKING IN CIoT

This section highlights the open issues and challenges related to intelligent decision making in CIoT. Moreover, we outline the future directions for the solution of these challenges on a large scale CIoT. Basically, we present a practical solution for optimal learning and provisioning algorithm which is formed by combination of multiple decision theoretic models i.e game theory, MDP, MAB and OSP for intelligent decision making in CIoT.

TABLE 2. Comparative analysis of decision theoretic models.

	Objective	Information Retrieval	Cost	Convergence Speed
Game Models	Interaction between multiple objects for appropriate object selection	The action of one user (object) has direct effects on the action of other users (objects) [2], [44].	Resource utilization for information exchange among users. Moreover, it also consider action switching cost before convergence [2], [90].	Game model only consider the asymptotical convergence for few algorithms [2], [90].
Markovian Decision Process	Formulation of spectrum sensing and channel selection as a MDP problem.	The information about the system state is dynamic and correlate [2], [87].	Consider only the switching cost for selection [2], [87].	Asymptotical convergence is studied [2], [87].
Optimal Stopping Problem	Calculate the trade-off between the cost of sequential sensing for objects and the expected reward according to their cost	There is uncertain realization of unexplored channels [2], [91].	Switching cost exist. Basically, it observes each variable based on their reward and perform a stopping action to minimize the cost [2], [91].	✗
Multi-arm Bandit Problem	Best learning technique for unknown statistical information of environment as it presents the tradeoff between exploitation and exploration	The statistical information about objects and environment is unknown a priori [2], [81].	The selection cost is well managed by this technique for unknown objects, channels and environment. [2], [81].	Asymptotical convergence is studied [2], [81].

A. OPEN ISSUES AND CHALLENGES

The future of emerging CIoT technology is quite exciting and bright although it is in developmental stage and needs a lot of research in this field to achieve practical solutions. There are still many challenges and issues which are presented in this section as follows:

- The current solution for CIoT has imperfect sensing and impact on decision result. The channel sensing and selection of CIoT is still explored separately in CIoT domain. Here, the core objective is to achieve joint execution of channel sensing and selection in federated manner.
- An important feature is to incorporate the user demands in the metric formulation for intelligent decision making. The current solution for decision making is only considering the optimization of allocated resources and is unable to highlight the user demand. It is of utmost importance to consider the user demands for specific object and form an optimization matrix for intelligent decision making.
- The intensive research on the learning algorithms showed that they can bring their strategies up-to-date on the basis of their action pay-off information. This strategy takes a long time to explore all the actions of the players and converge to a stable solution. The potential outcome is to design the knowledge based upon learning technologies which increases the convergence speed for better performance [24].
- The current game-theoretic models are unable to handle the large scale CIoT network due to lack of knowledge and self-organization. The requirement is to formulate a new game theoretic model which can handle the social behaviors and manage self-organization. It is necessary to deploy a bio-inspired system which consumes a localized selfless game while each player maximizing utilities and collecting utilities of its neighbors was proposed to achieve global optimization via local information exchange [23]. The prospective outcome here is to design a model based on social behaviors to address the most challenging issue of self-organization optimization.In large-scale CIoT, the globalized information interchange between the players is not appropriate. Infact the players have to rely on the local information which can be extracted by utilizing the game model. The need of the hour is to propose a scheme in which the game players are altruistic and share all the information with their neighbors to achieve optimization and self-organizaion. [2].
- It is desirable to achieve global optimization using local coordination and interaction games for the objects with dynamically unpredictable and incomplete information constraints resolution for cognitive decision making.
- Last but not the least, this new emerging field is under developmental phase and the need of the hour is to conduct an extensive research and apply it to a practical system. A lot of efforts are required from academia and industry to develop the application and practical system of CIoT for different scenarios ranging from smart home to smart city.

B. FUTURE DIRECTIONS AND DECISION THEORETIC SOLUTION FOR CIoT

The previously presented decision models including game theory, MDP, OSP and MAB, mostly addressed only one challenge for decision making in CIoT. This paper proposes the solution for optimal learning and provisioning by combining multiple or hybrid types of decision-theoretic solutions

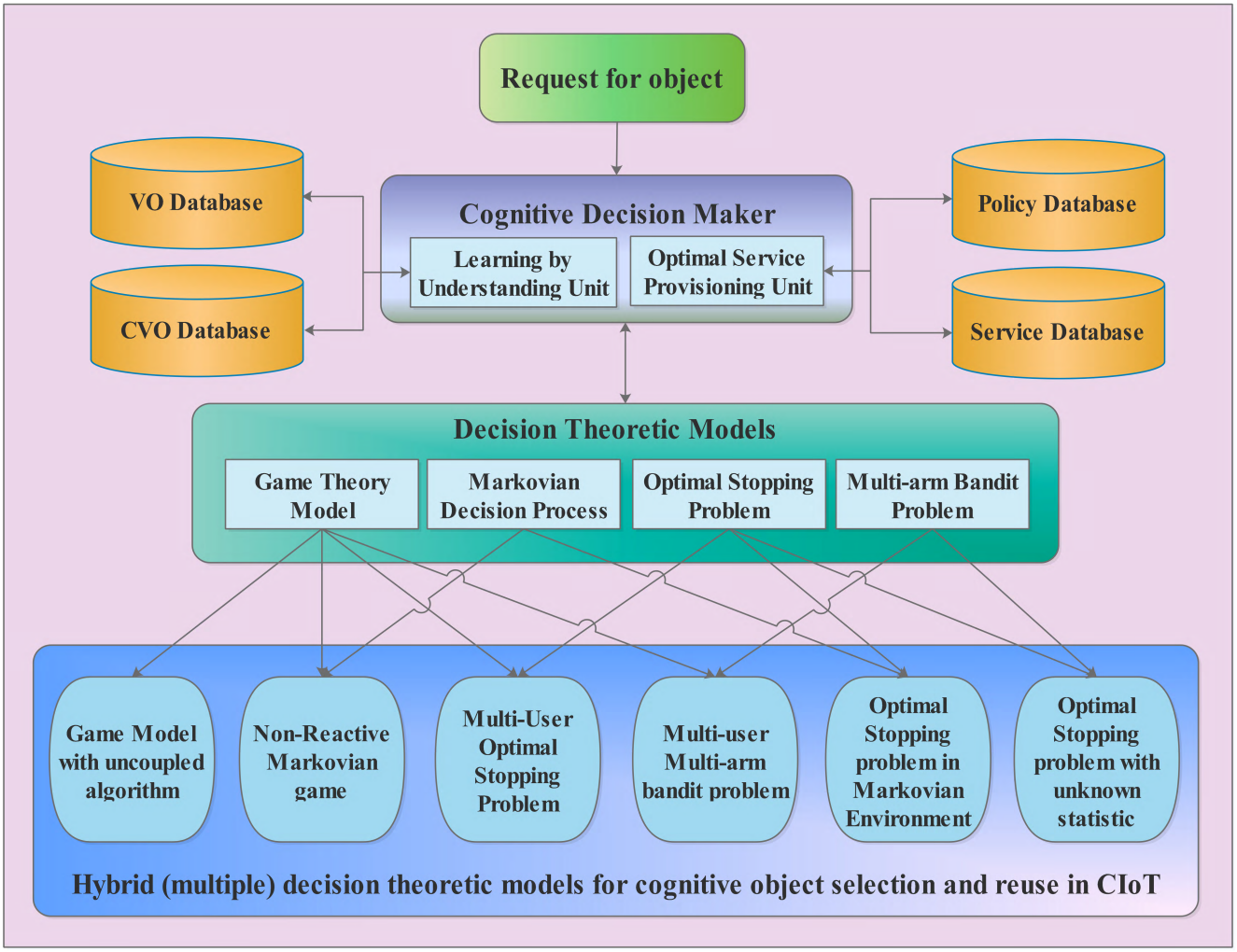

FIGURE 4. Multiple (hybrid) decision theoretic models for cognitive decision making of object selection in CIoT.

for intelligent decision making in large scale CIoT as shown in Fig. 4.

1) GAME MODELS WITH UNCOUPLED ALGORITHMS

The current game models usually monitor the environment and statistical information of other players (objects). The cognitive property of CIoT allows the objects perceiving action cycle which includes the objects' information, environment's information and heterogeneous objects sensing etc. to name a few as mentioned in above sections. In this framework, the observation on multi-user object and channel selection is done through perceptive action cycle, extract useful information from partial feedback and then fine-tune their behaviors as well as policies towards some desirable solution using stochastic or repeated games. The hierarchical nature of CIoT framework requires the new uncoupled learning algorithms like Stackelberg game [47], [48] as discussed in section IIIe. The Hierarchical game models are the most appropriate game model for CIoT because of their hierarchical nature and require the hierarchical game based algorithms for objects interlinking and management. In previous section, we have mentioned that the network resource utilization overhead and traditional game theory algorithms are coupled and designed for single user's scenario. Hence, it is necessary to design the update rule carefully that contains neighbor information which guarantees the convergence towards the desirable solution [24]. Moreover, it is also desirable to propose uncoupled algorithms which utilize the local partial action pay-off information of the neighboring players and objects for the desirable solution. One of the best solution currently available in previous literature for CIoT is the Stackelberg game model based on uncoupled algorithm which utilizes the action pay-off of neighboring objects for the desirable solution. In this game model, the leader takes an action based on a situation and the followers take actions according to the leader or follow its actions. In this game model, both leader and followers maximize their utility function or leader has no utility and he (resource user) aims to maximize the accumulated utility of the followers which results into efficiency of NE significantly. On the other hand, when this local information is not

available or is unknown priori, we need to consider the individual action pay-off information for the desirable solution. The action of one player is affected by the action of other player in uncoupled algorithms which makes it more difficult for the game models to achieve optimality and convergence. The take-away of this model is to carefully design the utility function and fully couple an algorithms like learning automata to achieve self-organization and optimization for practical implementation [2], [24]. Finally the cost of information exchange among the users as shown in Table 2 is quite high for the traditional coupled algorithms. Therefore, it is the need of the hour to design the game models with uncoupled algorithms which can provide optimality and convergence of objects and channel selection for intelligent decision making in CIoT.

2) NON REACTIVE MARKOVIAN GAME

The cognitive decision making can be achieved by combining the game theory with Markovian decision process. This amazing approach provides the distributed solution of optimal learning and provisioning among the multiple objects for appropriate object and channel selection efficiently. Basically, this approach involves multiple players, competing for resources, and relies on the current state of the player and does not require the information about the neighboring players in Markovian environment. In this model, the current state and actions of the game players jointly determines the system state in classical Markovian environment [70]. In [70], Cheng and Zhuang have presented a stochastic approximation algorithm that can adaptively estimate the NE policies and track such policies for non-stationary problems where the statistics of the channel and user parameters evolve with time. Basically, the system state changes are totally dependent on the current actions chosen by the players and hence it is termed as reactive Markovian game. In case of CIoT, the game involved in the system state is non-reactive in Markovian environment. In this scenario the spectrum utilization and occupancy state of CIoT is totally dependent on the objects or players instead of their actions. The previous literature presented some solutions for classical reactive Markovian game models which include stochastic approximation and value iteration as discussed in section III b. need to be modified concisely in case of CIoT. Therefore, such drawback can be solved by developing effective solutions via non-reactive Markovian game for cognitive decision making in IoT. Moreover, there are multi-agents and objects in CIoT which are distributed across different domains. It is desirable that the solutions must be non-reactive for multiple domains in which the information about other users is not desirable to perform intelligent decision making. The consideration here is to implement the reinforced learning algorithms or stochastic learning automata to practically converge to NE and CE of game for desirable solution in rapidly changing CIoT environment.

3) MULTI-USER OPTIMAL STOPPING PROBLEM

The previous section provides deep insight knowledge about optimal stopping problem showing clearly that it is very efficient for resource selection and discovery of objects to be single user rather than multi agent systems. The CIoT on other hand involves multi-agent cooperation and selection. Therefore, the only problem is the interaction among multi agents for better resource selection and discovery of objects. Such problem can be solved by hybrid model which merges optimal stopping problem with game theory for the desirable multi-agent systems. The literature has shown that there is little reported existing work in [91] and [92], considered the multiple agents worth numeric simulation. Still the multi-agent optimal stopping problems have not been reported. The outcome is to develop the practical solution for multi-user optimal stopping problem for intelligent decision making in CIoT. The most effective solution presented in previous literature which can be utilized with minor changes is the implementation of stochastic recall algorithm which will intelligently recall the previous states and statistical information of the player or objects optimally as discussed in section IIIc. This adaptive stochastic recall algorithm (ASRA) [66] can also capture the collision among multiple secondary users. The combination of game theory with ASRA will enable the decision maker to re-utilize the previous information about players states, their statistical information, channel utilization, objects related to their action etc. to name a few for intelligent decision making.

4) MULTI-USER MULTI-ARMED BANDIT PROBLEM

This technique is suitable for the single user CRN system but is not suitable for CIoT system as it involves multi-users (objects) having multi-arm bandit problem. This technique can be deployed in our CIoT framework by combining with evolutionary game theory models. Multi-user bandit is very useful technique as it enables cognition which is the heart of the system by learning without the statistical information about other objects and environment. Existing CIoT solutions like [93], [94], consider the multi-user scenario but these solutions are more application specific and hence cannot be applied effectively to more generic scenarios of CIoT as dicussed in Section IIId. Therefore, we suggest to combine the game theory model with multi-arm bandit problem to achieve a more general and efficient solution for learning. The literature review provides its example in [93], but these research shows the initial work that has to be improved for the practical implementation in CIoT. The authors presented the method to efficiently calculate the orthogonality among the users. Moreover, they have calculated the optimal system regret having a logarithmic order with time that converges to maximum throughput of known channel model and centralized users. Similar work presented in [94] presented the orthogonality among multiple users according to time. Both the paper's work provides excellent methods and

policies to learn about the players in stationary environment. These approaches can be applied on the CIoT environment by combining it with game theory which enables multi-armed bandit problem to find the excellent learning policy in non-stationary environment. The need of the hour is to conduct more research on this hybrid technique for the practical implementation of the CIoT in near future.

5) OPTIMAL STOPPING PROBLEM IN MARKOVIAN ENVIRONMENT

The previous section presented the optimal stopping problem in detail which shows that the random variables are distributed independently which means they are spread randomly and have no dependence on each other. Same is the case with the IoT environment in which the objects are scattered independently and the decision makers are also multiple and are distributed across the network. This technique possesses many overheads of cost and complexity for multi user scenario for maximizing the reward [2], [23], [66]. In [23], Xu *et al.* proposed an intelligent learning algorithm which is known as stochastic learning automata that converges to N.E of game. Moreover, a bio-inspired system that consumes a localized selfless game while each player maximizing utilities and collecting utilities of its neighbors was proposed to achieve global optimization via local information exchange. In [66], the authors used the recall OSP model in which the usage of the previously observed variable is allowed and the decision is supported by the previously observed state. They have used 1-SLA rule which can continuously sense the stages and information of the players. Moreover, this rule provides an optimal solution for monotone R- OSP models as discussed in Section IIIc. This can be applied to the CIoT by formulating it with the Markovian decision process to minimize the cost and maximize the reward for decision maker. The combination of this strategy and priori environmental information maximizes expected rewards which depends on a piecewise-deterministic process which gives the posterior likelihoods of the unobserved Markovian environment. This exciting new combination also creates new fundamental challenges like the sensing priorities and sequence, serves as an important requirement in CIoT domain. More specifically, any CIoT implementation should consider the sensing priorities in each epoch as adaptive and optimized based on the observation of Markovian environment. Ideally, a practical hybrid optimal stopping problem with Markovian game model can be solved by providing adequate solution on cost minimizations of the objects.

6) OPTIMAL STOPPING PROBLEM WITH UNKNOWN STATISTICAL INFORMATION

In the previous section it is clearly visible that optimal stopping problem requires object's statistical information to compute the cost and rewards every object accordingly. Such problem can be easily resolved by combining OSP with MAB problem. Likewise, the computational complexity can be reduced by decoupling in each epoch for joint decision execution task. The joint decision execution task is divided into two separate phases i.e., determining sensing priority and retrieving rule derivation [31]. In [84], Anandkumar *et al.* have utilized the greedy algorithm for the selection of collision free channel selection profile. Moreover, they have used adaptive random UCB1 algorithms which enable the SUs to randomly chooses a channel only if the collision occur in the previous slot otherwise it proceeds with UCB1 algorithm. On the other hand, 1-SLA rule which can continuously sense the stages and information of the players. These efficient learning techniques explore the statistics of objects during the decision process and maximize the reward on the basis of current estimation statistics about objects as discussed in Section III d. Basically, this technique utilizes more time on estimating the statistics of each object during decision making for better rewards for the objects based on statistical information and decision making. In this approach, the MAB is used to formulate the priority for sensing across epochs while OSP is used to retrieve the rule across each epoch. In addition, this approach will decrease the computational complexity in OSP and by just calculating the access rule of selected pattern in each epoch. The expected solution is to design the hybrid technique by combining OSP with MAB for the solution of unknown statistical information problem in CIoT.

VI. CONCLUSION

This paper has discussed the current state of the art and has opened challenges in the emerging field of CIoT. We have investigated the usage of decision theoretic models which will empower CIoT paradigm, aiming at improving the interoperability among heterogeneous smart objects. Overall, the CIoT with decision theoretic models would allow for the joint execution of heterogeneous smart objects in a federated manner. This in turn would be able to provide a fundamental platform for practitioners who are interested to gain an insight on CIoT and their related decision theoretic models as a means for harmonizing decision making among smart objects.

REFERENCES

[1] A. Taivalsaari and T. Mikkonen, "A roadmap to the programmable world: Software challenges in the IoT era," *IEEE Softw.*, vol. 34, no. 1, pp. 72–80, Jan./Feb. 2017.

[2] Y. Xu, A. Anpalagan, Q. Wu, L. Shen, Z. Gao, and J. Wang, "Decision-theoretic distributed channel selection for opportunistic spectrum access: Strategies, challenges and solutions," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 4, pp. 1689–1713, 4th Quart., 2013.

[3] I. Yaqoob *et al.*, "Internet of Things architecture: Recent advances, taxonomy, requirements, and open challenges," *IEEE Wireless Commun.*, vol. 24, no. 3, pp. 10–16, Jun. 2017.

[4] O. Vermesan *et al.*, "Internet of Things strategic research roadmap," in *Internet of Things: Global Technological and Societal Trends*, vol. 1, O. Vermesan, P. Friess, P. Guillemin, S. Gusmeroli, H. Sundmaeker, and A. Bassi, Eds. Gistrup, Denmark: River Publishers, 2011, pp. 9–52.

[5] A. Aijaz and A. H. Aghvami, "Cognitive machine-to-machine communications for Internet-of-Things: A protocol stack perspective," *IEEE Internet Things J.*, vol. 2, no. 2, pp. 103–112, Apr. 2015.

[6] Q. Wu *et al.*, "Cognitive Internet of Things: A new paradigm beyond connection," *IEEE Internet Things J.*, vol. 1, no. 2, pp. 129–143, Apr. 2014.

[7] V. Foteinos, D. Kelaidonis, G. Poulios, P. Vlacheas, V. Stavroulaki, and P. Demestichas, "Cognitive management for the Internet of Things: A framework for enabling autonomous applications," *IEEE Veh. Technol. Mag.*, vol. 8, no. 4, pp. 90–99, Dec. 2013.
[8] M. Serrano, P. Barnaghi, F. Carrez, P. Cousin, O. Vermesan, and P. Friess. (2015). *Internet of Things IoT Semantic Interoperability: Research Challenges, Best Practices, Recommendations and Next Steps*. [Online]. Available: http://www.internet-of-thingsresearch.eu/pdf/IoT_Cluster_Strategic_Research_Agenda_2009.pdf
[9] V. Tsiatsis*et al.*, "The SENSEI real world Internet architecture," in *Towards the Future Internet—Emerging Trends From European Research*. Amsterdam, The Netherlands: IOS Press, 2010, pp. 247–256.
[10] O. Vermesan and P. Friess, *Internet of Things Applications: From Research and Innovation to Market Deployment* (River Publishers Series in Communications). Aalborg, Denmark: River Publishers, 2014. [Online]. Available: https://books.google.com.my/books?id=kw2doAEACAAJ
[11] A. A. Khan, M. H. Rehmani, and A. Rachedi, "Cognitive-radio-based Internet of Things: Applications, architectures, spectrum related functionalities, and future research directions," *IEEE Wireless Commun.*, vol. 24, no. 3, pp. 17–25, Jun. 2017.
[12] N. Kaur and S. K. Sood, "A game theoretic approach for an IoT-based automated employee performance evaluation," *IEEE Syst. J.*, vol. 11, no. 3, pp. 1385–1394, Sep. 2017.
[13] K.-H. N. Bui, J. E. Jung, and D. Camacho, "Game theoretic approach on Real-time decision making for IoT-based traffic light control," *Concurrency Comput., Pract. Exper.*, vol. 29, no. 11, p. e4077, 2017. [Online]. Available: http://dx.doi.org/10.1002/cpe.4077
[14] S. Kim, "R-learning-based team game model for Internet of Things quality-of-service control scheme," *Int. J. Distrib. Sensor Netw.*, vol. 13, no. 1, pp. 1–10, 2017. [Online]. Available: https://doi.org/10.1177/1550147716687558
[15] M. Bhatia and S. K. Sood, "Game theoretic decision making in IoT-assisted activity monitoring of defence personnel," *Multimedia Tools Appl.*, vol. 76, no. 21, pp. 21911–21935, Nov. 2017. [Online]. Available: https://doi.org/10.1007/s11042-017-4611-3
[16] P. Semasinghe, S. Maghsudi, and E. Hossain, "Game theoretic mechanisms for resource management in massive wireless IoT systems," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 121–127, Feb. 2017.
[17] S. Kim, "Learning-based QoS control algorithms for next generation Internet of Things," *Mobile Inf. Syst.*, vol. 2015, 2015, Art. no. 605357.
[18] S. S. Yau and A. B. Buduru, "Intelligent planning for developing mobile IoT applications using cloud systems," in *Proc. IEEE Int. Conf. Mobile Services*, Jun. 2014, pp. 55–62.
[19] M. G. R. Alam, S. F. Abedin, A. K. Bairaggi, A. Talukder, and C. S. Hong, "An autonomic SLA management for IoT networks," in *Proc. Korea Computer Congr.*, 2016, pp. 507–509.
[20] X. Xiong, L. Hou, K. Zheng, W. Xiang, M. S. Hossain, and S. M. M. Rahman, "SMDP-based radio resource allocation scheme in software-defined Internet of Things networks," *IEEE Sensors J.*, vol. 16, no. 20, pp. 7304–7314, Oct. 2016.
[21] M. Abu Alsheikh, D. T. Hoang, D. Niyato, H.-P. Tan, and S. Lin, "Markov decision processes with applications in wireless sensor networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 3, pp. 1239–1267, 3rd Quart., 2015.
[22] P. Vlacheas *et al.*, "Enabling smart cities through a cognitive management framework for the Internet of Things," *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 102–111, Jun. 2013.
[23] Y. Xu, J. Wang, Q. Wu, A. Anpalagan, and Y.-D. Yao, "Opportunistic spectrum access in cognitive radio networks: Global optimization using local interaction games," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 2, pp. 180–194, Apr. 2012.
[24] Y. Xu, J. Wang, Q. Wu, Z. Du, L. Shen, and A. Anpalagan, "A game-theoretic perspective on self-organizing optimization for cognitive small cells," *IEEE Commun. Mag.*, vol. 53, no. 7, pp. 100–108, Jul. 2015.
[25] R. Trestian, O. Ormond, and G.-M. Muntean, "Game theory-based network selection: Solutions and challenges," *IEEE Commun. Surveys Tuts.*, vol. 14, no. 4, pp. 1212–1231, 4th Quart., 2012.
[26] S. Kim, "New game paradigm for IoT systems," in *Game Theory: Breakthroughs in Research and Practice*. Hershey, PA, USA: IGI Global, 2017, p. 120.
[27] P. S. Sastry, V. V. Phansalkar, and M. Thathachar, "Decentralized learning of Nash equilibria in multi-person stochastic games with incomplete information," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 24, no. 5, pp. 769–777, May 1994.
[28] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Hoboken, NJ, USA: Wiley, 1994.
[29] F. Fu and M. V. D. Schaar, "Learning to compete for resources in wireless stochastic games," *IEEE Trans. Veh. Technol.*, vol. 58, no. 4, pp. 1904–1919, May 2009.
[30] R. B. Myerson, *Game Theory: Analysis of Conflict*. Cambridge, MA, USA: Harvard Univ. Press, 1991.
[31] J. Unnikrishnan and V. V. Veeravalli, "Algorithms for dynamic spectrum access with learning for cognitive radio," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 750–760, Feb. 2010.
[32] J. Smith, *Evolution and the Theory of Games*. Cambridge, U.K.: Cambridge Univ. Press, 1982.
[33] W. Saad, Z. Han, T. Basar, M. Debbah, and A. Hjørungnes, "Coalition formation games for collaborative spectrum sensing," *IEEE Trans. Veh. Technol.*, vol. 60, no. 1, pp. 276–297, Jan. 2011.
[34] Y. Xu, Q. Wu, and J. Wang, "Game theoretic channel selection for opportunistic spectrum access with unknown prior information," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2011, pp. 1–5.
[35] A. Montanari and A. Saberi, "Convergence to equilibrium in local interaction games," in *Proc. 50th Annu. IEEE Symp. Found. Comput. Sci.*, Oct. 2009, pp. 303–312.
[36] S. H. A. Ahmad, C. Tekin, M. Liu, R. Southwell, and J. Huang, "Spectrum sharing as spatial congestion games," *CoRR*, pp. 1–11, Nov. 2010.
[37] H. Li and Z. Han, "Competitive spectrum access in cognitive radio networks: Graphical game and learning," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Apr. 2010, pp. 1–6.
[38] M. Azarafrooz and R. Chandramouli, "Distributed learning in secondary spectrum sharing graphical game," in *Proc. IEEE Global Telecommun. Conf.*, Dec. 2011, pp. 1–5.
[39] Y. Xu, Q. Wu, J. Wang, N. Min, and A. Anpalagan, "Distributed channel selection in CRAHNs with heterogeneous spectrum opportunities: A local congestion game approach," *IEICE Trans.*, vol. 95-B, no. 3, pp. 991–994, 2012.
[40] J. R. Marden, G. Arslan, and J. S. Shamma, "Cooperative control and potential games," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 6, pp. 1393–1407, Dec. 2009.
[41] C. Tekin, M. Liu, R. Southwell, J. Huang, and S. H. A. Ahmad, "Atomic congestion games on graphs and their applications in networking," *IEEE/ACM Trans. Netw.*, vol. 20, no. 5, pp. 1541–1552, Oct. 2012.
[42] B. Vöcking, "Congestion games: Optimization in competition," in *Proc. 2nd Algorithms Complex. Durham Workshop*, 2006, pp. 9–20.
[43] H. Tembine, E. Altman, R. El-Azouzi, and Y. Hayel, "Evolutionary games in wireless networks," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 3, pp. 634–646, Jun. 2010.
[44] X. Chen and J. Huang, "Evolutionarily stable spectrum access," *IEEE Trans. Mobile Comput.*, vol. 12, no. 7, pp. 1281–1293, Jul. 2013.
[45] D. Monderer and L. S. Shapley, "Potential games," *Games Econ. Behavior*, vol. 14, no. 1, pp. 124–143, 1996.
[46] Y. Xu, J. Wang, Q. Wu, A. Anpalagan, and Y.-D. Yao, "Opportunistic spectrum access in unknown dynamic environment: A game-theoretic stochastic learning solution," *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, pp. 1380–1391, Apr. 2012.
[47] X. Kang, R. Zhang, and M. Motani, "Price-based resource allocation for spectrum-sharing femtocell networks: A Stackelberg game approach," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 538–549, Apr. 2012.
[48] M. Razaviyayn, Y. Morin, and Z.-Q. Luo, "A Stackelberg game approach to distributed spectrum management," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar. 2010, pp. 3006–3009.
[49] Y. Xiao, J. Park, and M. V. D. Schaar, "Repeated games with intervention: Theory and applications in communications," *IEEE Trans. Commun.*, vol. 60, no. 10, pp. 3123–3132, Oct. 2012.
[50] M. Bennis and D. Niyato, "A Q-learning based approach to interference avoidance in self-organized femtocell networks," in *Proc. IEEE Global Commun. Conf. Workshops (GLOBECOM)*, Dec. 2010, pp. 706–710.
[51] K. Fahimullah and S. Hassan, "Game-theory based wireless access point selection scheme," in *Proc. IEEE Silver Jubilee Int. Multi Topic Symp. (SIMTS)*, Mar. 2010, pp. 1–6.
[52] M. Osborne and A. Rubinstein, *A Course in Game Theory*. Cambridge, MA, USA: MIT Press, 1994.
[53] J. W. Huang and V. Krishnamurthy, "Transmission control in cognitive radio as a Markovian dynamic game: Structural result on randomized threshold policies," *IEEE Trans. Commun.*, vol. 58, no. 1, pp. 301–310, Jan. 2010.

[54] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *J. Artif. Intell. Res.*, vol. 4, no. 1, pp. 237–285, Jan. 1996.

[55] D. Chen, S. Yin, Q. Zhang, M. Liu, and S. Li, "Mining spectrum usage data: A large-scale spectrum measurement study," in *Proc. 15th Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, 2009, pp. 13–24.

[56] S. Ross, J. Pineau, S. Paquet, and B. Chaib-Draa, "Online planning algorithms for POMDPs," *J. Artif. Intell. Res.*, vol. 32, pp. 663–704, 2008.

[57] C. Boutilier, "Sequential optimality and coordination in multiagent systems," in *Proc. IJCAI*, 1999, pp. 478–485.

[58] D. S. Bernstein, R. Givan, N. Immerman, and S. Zilberstein, "The complexity of decentralized control of Markov decision processes," *Math. Oper. Res.*, vol. 27, no. 4, pp. 819–840, Aug. 2000.

[59] C. Amato, G. Chowdhary, A. Geramifard, N. K. Üre, and M. J. Kochenderfer, "Decentralized control of partially observable markov decision processes," in *Proc. 52nd IEEE Conf. Decision Control*, Dec. 2013, pp. 2398–2405.

[60] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP framework," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 3, pp. 589–600, Apr. 2007.

[61] Y. Chen, Q. Zhao, and A. Swami, "Joint design and separation principle for opportunistic spectrum access in the presence of sensing errors," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 2053–2071, May 2008.

[62] A. T. Hoang, Y. C. Liang, D. T. C. Wong, Y. Zeng, and R. Zhang, "Opportunistic spectrum access for energy-constrained cognitive radios," *IEEE Trans. Wireless Commun.*, vol. 8, no. 3, pp. 1206–1211, Mar. 2009.

[63] Q. Zhao, S. Geirhofer, L. Tong, and B. M. Sadler, "Opportunistic spectrum access via periodic channel sensing," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 785–796, Feb. 2008.

[64] S. Chen and L. Tong, "Maximum throughput region of multiuser cognitive access of continuous time Markovian channels," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 10, pp. 1959–1969, Dec. 2011.

[65] D. Niyato, E. Hossain, and P. Wang, "Optimal channel access management with QoS support for cognitive vehicular networks," *IEEE Trans. Mobile Comput.*, vol. 10, no. 5, pp. 573–591, Apr. 2011.

[66] T. L. Lai and H. Robbins, "Asymptotically efficient adaptive allocation rules," *Adv. Appl. Math.*, vol. 6, no. 1, pp. 4–22, Mar. 1985.

[67] Y. Xu, J. Wang, Q. Wu, Z. Zhang, A. Anpalagan, and L. Shen, "Optimal energy-efficient channel exploration for opportunistic spectrum usage," *IEEE Wireless Commun. Lett.*, vol. 1, no. 2, pp. 77–80, Apr. 2012.

[68] A. Sabharwal, A. Khoshnevis, and E. Knightly, "Opportunistic spectral usage: Bounds and a multi-band CSMA/CA protocol," *IEEE/ACM Trans. Netw.*, vol. 15, no. 3, pp. 533–545, Jun. 2007.

[69] H. Jiang, L. Lai, R. Fan, and H. V. Poor, "Optimal selection of channel sensing order in cognitive radio," *IEEE Trans. Wireless Commun.*, vol. 8, no. 1, pp. 297–307, Jan. 2009.

[70] H. T. Cheng and W. Zhuang, "Simple channel sensing order in cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 4, pp. 676–688, Apr. 2011.

[71] B. Li *et al.*, "Optimal frequency-temporal opportunity exploitation for multichannel ad hoc networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 23, no. 12, pp. 2289–2302, Dec. 2012.

[72] J. Jia, Q. Zhang, and X. Shen, "HC-MAC: A hardware-constrained cognitive MAC for efficient spectrum management," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 106–117, Jan. 2008.

[73] N. B. Chang and M. Liu, "Optimal channel probing and transmission scheduling for opportunistic spectrum access," *IEEE/ACM Trans. Netw.*, vol. 17, no. 6, pp. 1805–1818, Dec. 2009.

[74] S.-J. Kim and G. B. Giannakis, "Sequential cooperative sensing for multichannel cognitive radios," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar. 2010, pp. 2950–2953.

[75] Y. Xu, L. Shen, A. Anpalagan, Q. Wu, J. Wang, and Y. Xu, "Energy-efficient exploration and exploitation of multichannel diversity in spectrum sharing systems," *Trans. Emerg. Telecommun. Technol.*, vol. 23, no. 8, pp. 701–706, 2012.

[76] Y. Xu, Q. Wu, J. Wang, A. Anpalagan, and Y. Xu, "Exploiting multichannel diversity in spectrum sharing systems using optimal stopping rule," *ETRI J.*, vol. 34, no. 2, pp. 272–275, 2012.

[77] V. Anantharam, P. Varaiya, and J. Walrand, "Asymptotically efficient allocation rules for the multiarmed bandit problem with multiple plays-Part II: Markovian rewards," *IEEE Trans. Autom. Control*, vol. AC-32, no. 11, pp. 977–982, Nov. 1987.

[78] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Mach. Learn.*, vol. 47, no. 2, pp. 235–256, 2002.

[79] J. C. Gittins, "Bandit processes and dynamic allocation indices," *J. Roy. Stat. Soc. B (Methodol.)*, vol. 41, no. 2, pp. 148–177, 1979.

[80] P. Whittle, "Restless bandits: Activity allocation in a changing world," *J. Appl. Probab.*, vol. 25, pp. 287–298, Jan. 1988.

[81] H. Li, "Multi-agent Q-learning for competitive spectrum access in cognitive radio systems," in *Proc. 5th IEEE Workshop Netw. Technol. Softw. Defined Radio Netw. (SDR)*, Jun. 2010, pp. 1–6.

[82] L. Lai, H. El Gamal, H. Jiang, and H. V. Poor, "Cognitive medium access: Exploration, exploitation and competition," *CoRR*, vol. 10, pp. 239–253, Oct. 2007.

[83] L. Lai, H. El Gamal, H. Jiang, and H. V. Poor, "Cognitive medium access: Exploration, exploitation, and competition," *IEEE Trans. Mobile Comput.*, vol. 10, no. 2, pp. 239–253, Feb. 2011.

[84] A. Anandkumar, N. Michael, and A. Tang, "Opportunistic spectrum access with multiple users: Learning under competition," in *Proc. 29th Conf. Inf. Commun. (INFOCOM)*, Mar. 2010, pp. 803–811.

[85] K. Liu and Q. Zhao, "Distributed learning in multi-armed bandit with multiple players," *IEEE Trans. Signal Process.*, vol. 58, no. 11, pp. 5667–5681, Nov. 2010.

[86] Y. Gai and B. Krishnamachari, "Decentralized online learning algorithms for opportunistic spectrum access," *CoRR*, pp. 1–6, Apr. 2011.

[87] S. Filippi, O. Cappe, and A. Garivier, "Optimally sensing a single channel without prior information: The tiling algorithm and regret bounds," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 1, pp. 68–76, Feb. 2011.

[88] C. Tekin and M. Liu, "Online learning in opportunistic spectrum access: A restless bandit approach," in *Proc. IEEE INFOCOM*, Apr. 2011, pp. 2462–2470.

[89] K. Wang and L. Chen, "On optimality of myopic policy for restless multi-armed bandit problem: An axiomatic approach," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 300–309, Jan. 2012.

[90] M. A. Khan, H. Tembine, and A. V. Vasilakos, "Game dynamics and cost of learning in heterogeneous 4G networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 1, pp. 198–213, Jan. 2012.

[91] Y. Xu, Z. Gao, J. Wang, and Q. Wu, "Multichannel opportunistic spectrum access in fading environment using optimal stopping rule," in *Proc. Int. Conf. Wireless Commun. Appl.*, 2011, pp. 275–286.

[92] A. Anandkumar, N. Michael, A. Tang, and A. Swami, "Distributed algorithms for learning and cognitive medium access with logarithmic regret," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 4, pp. 731–745, Apr. 2011.

[93] K. Liu and Q. Zhao, "Cooperative game in dynamic spectrum access with unknown model and imperfect sensing," *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, pp. 1596–1604, Apr. 2012.

[94] B. Li, P. Yang, J. Wang, Q. Wu, X.-Y. Li, and Y. Liu, "Observation vs statistics: Near optimal online channel access in cognitive radio networks," in *Proc. IEEE 9th Int. Conf. Mobile Adhoc Sensor Syst. (MASS)*, Oct. 2012, pp. 458–462.

KHAQAN ZAHEER received the B.Sc. degree (Hons.) in computer engineering from the Punjab University of Information and Technology, Lahore, Pakistan, and the M.Sc. degree in mobile and satellite communication engineering from Bradford University, U.K. He is currently pursuing the Ph.D. degree in network, parallel and distributed computing with the Universiti Putra Malaysia. He joined Government College University, Lahore, in 2009, and continued his career as a Lecturer with the Computer Science Department. He joined with the Computer Science Department, COMSATS Institute of Information Technology, Lahore, in 2011, as an Assistant Professor. His areas of interest are cognitive radio network, network coding, IoT, game theory, and decision theoretic models.

MOHAMED OTHMAN received the Ph.D. degree (Hons.) from the National University of Malaysia. He is currently a Professor in computer science with the Department of Communication Technology and Network, Universiti Putra Malaysia (UPM), and prior to that he was a Deputy Director of the Information Development and Communication Center, where he was an Incharge for UMPNet network campus, uSport Wireless Communication Project, and the UPM DataCenter. He is also an Associate Researcher and a Coordinator of high speed machine with the Laboratory of Computational Science and Informatics, Institute of Mathematical Science, UPM. In 2017, he received an Honorable Professor from South Kazakhstan Pedagogical University, Shymkent, Kazakhstan, and also as a Visiting Professor with South Kazakhstan State University, Shymkent, and L. N. Gumilyov Eurasian National University, Astana, Kazakhstan. He already authored over 250 National and International journals and over 200 proceeding papers. His main research interests are in the fields of computer network, parallel and distributed computing, high-speed interconnection network, network design and management (network security, wireless and traffic monitoring), consensus in IoT, and mathematical model in scientific computing. He is a member of the IEEE Computer Society, the IEEE Communication Society, Malaysian National Computer Confederation, and Malaysian Mathematical Society. He was a recipient of the Best Ph.D. Thesis in 2000 by Sime Darby Malaysia and Malaysian Mathematical Science Society. On top of that he has filed six Malaysian, one Japan, one South Korea, and three U.S. patents.

MUBASHIR HUSAIN REHMANI received the B.Eng. degree in computer systems engineering from the Mehran University of Engineering and Technology, Jamshoro, Pakistan, in 2004, the M.S. degree from the University of Paris XI, Paris, France, in 2008, and the Ph.D. degree from University Pierre and Marie Curie, Paris, in 2011. He served for five years as an Assistant Professor with the COMSATS Institute of Information Technology, Wah Cantt, Pakistan. He is currently with the Telecommunications Software and Systems Group, Waterford Institute of Technology, Waterford, Ireland. He has authored/ edited two books published by IGI Global, USA, one book published by CRC Press, USA, and one book with Wiley, U.K. He was a recipient of Best Researcher of the Year 2015 of COMSATS Wah Award in 2015. He received the certificate of appreciation as an Exemplary Editor of the IEEE Communications Surveys and Tutorials for the year 2015 from the IEEE Communications Society. He was also a recipient of the Best Paper Award from the IEEE ComSoc Technical Committee on Communications Systems Integration and Modeling and the IEEE ICC 2017. He consecutively received research productivity award from 2016 to 2017 and also ranked 1 in all Engineering disciplines from Pakistan Council for Science and Technology, Pakistan. He was also a recipient of the Best Paper Award in 2017 from Higher Education Commission, Pakistan. He served for three years (from 2015 to 2017) as an Associate Editor of the IEEE Communications Surveys and Tutorials. He currently serves as an Associate Editor of the IEEE *Communications Magazine*, Elsevier the *Journal of Network and Computer Applications*, and the *Journal of Communications and Networks*. He is also serving as a Guest Editor of Elsevier *Ad Hoc Networks*, Elsevier *Future Generation Computer Systems*, the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, and Elsevier *Pervasive and Mobile Computing*. He is currently an Area Editor of the IEEE Communications Surveys and Tutorials.

THINAGARAN PERUMAL received the Ph.D. degree in smart technology and robotics from the Universiti Putra Malaysia (UPM). He is currently a Senior Lecturer with the Department of Computer Science, Faculty of Computer Science and Information Technology, UPM. His research interests are towards interoperability aspects of smart homes, Internet of Things, wearable computing, and cyber-physical systems.