Resource Allocation and Service Provisioning in Multi-Agent Cloud Robotics: A Comprehensive Survey

Mahbuba Afrin, Student Member, IEEE, Jiong Jin, Member, IEEE, Akhlaqr Rahman, Member, IEEE, Ashfaqur Rahman, Senior Member, IEEE, Jiafu Wan, Member, IEEE, and Ekram Hossain, Fellow, IEEE

Abstract—Robotic applications nowadays are widely adopted to enhance operational automation and performance of real-world Cyber-Physical Systems (CPSs) including Industry 4.0, agriculture, healthcare, and disaster management. These applications are composed of latency-sensitive, data-heavy, and compute-intensive tasks. The robots, however, are constrained in the computational power and storage capacity. The concept of multi-agent cloud robotics enables robot-to-robot cooperation and creates a complementary environment for the robots in executing large-scale applications with the capability to utilize the edge and cloud resources. However, in such a collaborative environment, the optimal resource allocation for robotic tasks is challenging to achieve. Heterogeneous energy consumption rates and application of execution costs associated with the robots and computing instances make it even more complex. In addition, the data transmission delay between local robots, edge nodes, and cloud data centres adversely affects the real-time interactions and impedes service performance guarantee. Taking all these issues into account, this paper comprehensively surveys the state-of-the-art on resource allocation and service provisioning in multi-agent cloud robotics. The paper presents the application domains of multi-agent cloud robotics through explicit comparison with the contemporary computing paradigms and identifies the specific research challenges. A complete taxonomy on resource allocation is presented for the first time, together with the discussion of resource pooling, computation offloading, and task scheduling for efficient service provisioning. Furthermore, we highlight the research gaps from the learned lessons, and present future directions deemed beneficial to further advance this emerging field.

Index Terms—Multi-robot system, Cloud computing, Edge computing, Resource allocation, Service provisioning, Computation and communication trade-off, Offloading, Task scheduling.

I. INTRODUCTION

In recent years, a significant emphasis is given on building smart Cyber-Physical Systems (CPSs) in industry, transport, healthcare and agriculture to perform complex engineering operations with limited human involvement, reduced cost and improved performance [1]. The basic component of these CPSs are the robots. A robot is an autonomous entity that can perceive the external environment, make intelligent decisions, and trigger physical actions. Although a robot is equipped with computing resources to conduct small-scale data processing, it cannot carry out large-scale computations on its own [2]. To deal with this limitation of standalone robots, the concept of multi-robot system has been emerged. In multi-robot systems, a group of robots (connected via a wired or wireless network) work collaboratively to achieve a common goal [3]. For example, in a smart factory, some robots handle the inventory while others check the safety and quality; and collectively they contribute to the production management. Nevertheless, the performance of multi-robot system is subject to the heterogeneous energy capacity of robots. In case of widespread deployment, the maintenance of persistent communication among the robots becomes difficult due to their mobility. The lack of storage within the robots causes further disruptions in exchanging and preserving the large volume of data during robot-to-robot interactions [2]. To overcome these limitations, the concept of cloud computing has been extended to multi-robot system, which is termed as cloud robotics [2], [3]. In this paradigm, cloud offers the computing resources such as virtual machines or containers and engages resources from both local robots and remote cloud data centres for scalable and extensive data processing [4].

However, the robotic systems and the cloud data centres are normally multi-hop distance apart and that is responsible for longer communication time and data transfer delay. As a result, cloud robotics often becomes less suitable for latency-sensitive operations. Edge computing can play a vital role in addressing this limitation by bringing the infrastructure, platform, and software services closer to the data sources through cloudlet [5] and fog nodes [6]–[8]. The integration of edge computing [9] with existing cloud robotics [10] paves the way of creating a new computing paradigm for executing robotic applications and their associated tasks on the physical resources at different communication hops from robot-to-cloud. Since this new paradigm simultaneously harnesses the computing capabilities at device level (robot), edge level (fog node, cloudlet) and remote level (cloud data centres), in this paper, it is termed as multi-agent cloud robotics. Figure 1 depicts the organization and benefits of a multi-agent cloud robotic system. In this system, tasks with hard real-time constraints such as robotic surgery in smart hospital [11], soft real-time tasks with less restrictive deadline such as pesticide spraying in smart farm [12], and parcel distribution in smart factory [13], [14] can be performed. The edge computing platform performs the soft real-time tasks including field image processing and path planning. On the other hand, the latency-tolerant and high storage requiring tasks such as eHealth report distribution and weather forecasting are managed by the cloud data centre.

In the context of CPSs, multi-agent cloud robotics can bring various benefits as listed below.

• By exploiting computing resources at different infrastructure
levels, it provides options to the CPSs for executing diverse robotic applications.

- It offers an abstraction to classify the tasks of robotic applications as per the characteristics of computing infrastructures.
- It provides distributed resources to process big data generated in the CPSs with reduced communication delay.
- It enhances the competency of robots in sharing knowledge using local and edge network rather than depending on the distant cloud.
- It significantly reduces the bandwidth requirements (for sending data to cloud) in CPSs dealing with multiple devices, by bringing the resources closer to the robots.

With multi-agent cloud robotics, the dependency of the CPSs to the cloud data centre as well as the load on cloud data centre (to handle multiple CPSs) is sharply decreased.

Based on the characteristics, a multi-agent cloud robotic system appears similar to the contemporary computing paradigms, namely, Mobile Cloud Computing (MCC) and Multi-access Edge Computing (MEC), where the edge computation is extensively harnessed. However, there exist explicit differences among the functionalities of these paradigms. They are discussed as follows.

A. Multi-Agent Cloud Robotics and Related Computing Paradigms

MCC facilitates resource and energy-constrained smart phones by enabling them to offload compute-intensive mobile applications to cloud data centres for execution. Through an intermediate layer of cloudlets between the smart phones and cloud data centres, MCC manages a 3-tier execution platform for the offloaded applications. MCC supports both local and global distribution of computing and application services through private resources. These services are accessed using LAN or WiFi technology. On the other hand, MEC manages a virtual server on the cellular base stations to assist faster networking, location tracking and content delivery services for the smart phone and Internet of Things (IoT) users. It usually offers a multi-tier computing platform for the applications, where the services are reachable via one-hop proximal distant access points using LTE, 3GPP or 5G network technology. Generally, the MEC infrastructures are provided by telecom operators. Unlike MCC and MEC, a multi-agent cloud robotic system follows an n-tier resource orientation, where the services are offered by harnessing local (robots), proximal (edge resources), or global (cloud data centres) computing agents. Here, the primary data producers are the sensor clusters, IoT devices and robots and the services are accessed by wireless access technologies including WiFi, LTE, Zigbee, Bluetooth, LoRa technology including 3GPP and 5G.

Moreover, compared to MCC and MEC, multi-agent cloud robotics explicitly supports artificial intelligence so that robots can make autonomous decision. Being an intelligent entity, robots are able to learn from the environment and alter their predefined mobility pattern dynamically. Conversely, MCC and MEC do not support autonomous mobility as the end devices generally operate through human intervention. These distinctive architectural and operational features make a multi-agent cloud robotic system more complicated than MCC and MEC.

Since the multi-agent cloud robotics is a complex synthesis of robotics, tele-operation, edge computing and core cloud technologies, the performance of such an environment depends on proper management and co-ordination among heterogeneous resources. Therefore, efficient resource allocation and service provisioning policies are required to fully exploit the benefits of multi-agent cloud robotics. Generally speaking, resource allocation refers to the assignment of requested applications to the competent resources, whereas service provisioning sets the base for resource allocation through resource pooling and computation offloading, and assists in task scheduling. The development of efficient resource allocation and service provisioning policies often becomes challenging due to the diverse requirements unique to multi-agent cloud robotics.

B. Challenges of Resource Allocation and Service Provisioning in Multi-Agent Cloud Robotics

The domain specific requirements and challenges for resource allocation and service provisioning in multi-agent cloud robotics are listed below:

- **Real-time learning and autonomous action**: Robotic applications require advanced artificial intelligence and machine learning techniques for situation-aware real-time and automated action. Although edge infrastructure in multi-agent cloud...
Multi-agent cloud robotics helps meeting real-time service requirements, it lacks support for installing numerous resource-enriched computing devices with large amount of processing cores and GPUs due to cost, structural and maintenance constraints. Consequently, they limit the scope of executing computer-intensive artificial intelligence and machine learning techniques. Therefore, the resource allocation and service provisioning in multi-agent cloud robotics urges rigorous classification and selection of compatible resources for robotic applications which is not mandatory for MCC and MEC dealing with predefined set of events and actions.

- **Complex data stream processing:** Multi-agent cloud robotics requires modular robotic applications with diverse programming models including map-reduce and distributed data flow to process complex data streams generated by robot-embedded multi-purpose sensors [23]. Due to resource and communication constraints, the local and proximal resources of multi-agent cloud robotics often fail to participate in such intense data processing that increases the burden on global resources. Therefore, resource allocation and service provisioning policies for a multi-agent cloud robotics system need to determine where to process data streams in real time. This significantly differs from MCC and MEC, which generally process simple batch or stream data generated by conventional sensors of smart phones and IoT devices.

- **Dynamic robotic cooperation:** There are some cases such as earthquake disaster management and submarine cable maintenance when multiple heterogeneous robots need to work collaboratively for attaining a common goal. Moreover, while operating in such adverse working environments with limited communication support, the real-time robot-to-robot and robot-to-edge or cloud interactions get hampered significantly [22]. To meet these requirements and constraints, the robotic cooperation often depends on distributed but coordinated algorithms such as federated learning and distributed learning, which adds further resource management overhead to multi-agent cloud robotics. The resource and service provisioning policies should address this issue deliberately that is not compulsory in MCC and MEC due to their simplified working environments.

- **Cross-infrastructure interoperability:** In multi-agent cloud robotics, heterogeneous and distributed computing resources complement the robots to carry out their responsibilities. These resources are owned by different service providers. For example, multi-agent cloud robotic systems can use the Google or Amazon cloud infrastructure and their edge infrastructure can be set by the cellular service providers. The resource sharing, privacy, and fault tolerance functions of these infrastructure are also managed with separate black box software systems. Therefore, the resource allocation and service provisioning policies in multi-agent cloud robotics require support for cross-infrastructure and policy-driven interoperability, which is not essential for MCC and MEC because of their association with homogeneous resource providers.

- **Synchronized decision making:** Since multi-agent cloud robotics involves computing infrastructure from different communication layers to execute robotic applications, an explicit synchronization of decision making entities belonging to these infrastructure is highly required. However, the attainment of such synchronization becomes challenging due to the intelligence interference among robots, edge, and cloud resources. Therefore, resource allocation and service provisioning policies should fairly distribute the resource management responsibilities among multiple decision making entities, which is not obligatory for MCC and MEC relying on a single decision maker.

- **On-demand computation and communication trade-off:** The offloading facility in multi-agent cloud robotics enables the robots having low energy or low processing power to access the edge and cloud infrastructure and execute the computer-intensive robotic applications. To conduct this activity, a seamless interaction among local robots, edge and remote resources are required which consumes additional communication time. It becomes more complicated because of the uncertain mobility of robots. Furthermore, the heterogeneity of resources in multi-agent cloud robotics with respect to processing cores, networking standards and energy consumption rates intensifies the appropriate resource selection problem for offloading. Therefore, resource allocation and service provisioning policies in multi-agent cloud robotics require on-demand computation and communication trade-off [24], [25]. On the other hand, in MCC and MEC, the computation and communication trade-off often depend on predefined rules because of the static mobility pattern of the users.

- **System-specific policy:** In multi-agent cloud robotics, the Quality of Service (QoS) requirements of robotic systems vary from one to another. For example, the service delivery deadline for a robotic application in smart hospital is more stringent than that of in smart farm. Similarly, for some robotic systems, higher accuracy is important such as robotic surgery, whereas for others the minimization of cost is vital such as smart factory maintenance. A generalized resource allocation and service provisioning policy is not feasible for multi-agent cloud robotics like MCC and MEC. Therefore, system-specific policies are required for efficient resource management, which is quite difficult due to the dynamics of robotic systems.

- **Energy-delay optimization:** Although computing resources from every layer of the hierarchical architecture participate in executing robotic applications, still the realization of such an environment is constrained by the energy limitations of the resources. Moreover, it is challenging to finish the application execution with the residual energy of the resources. Specifically, in a multi-agent cloud robotic system, robots can

### TABLE I: Comparison with related computing paradigms of multi-agent cloud robotics

| Dimension                  | Mobile Cloud Computing | Multi-Access Edge Computing | Multi-Agent Cloud Robotics |
|----------------------------|------------------------|-----------------------------|---------------------------|
| Geo-distribution           | Local/ Global          | Local/ Proximal             | Local/ Proximal/ Global   |
| Resource orientation       | 3-tier                 | 2/3-tier                    | n-tier                    |
| Infrastructure provider    | Private entities       | Telecom operators           | Private entities/ Cloud provider |
| Application type           | Lightweight            | Lightweight                  | Lightweight/ Heavyweight   |
| Data producer              | Mobile devices         | IoT devices/ Cellular gateways | Multiple sensors/ IoT devices/ Robots |
| General data type          | Batch                  | Simple stream               | Complex stream            |
| Service access             | LAN/ WiFi              | LTE/ 3GPP/ 5G               | WiFi/ LTE/ Zigbee/ Bluetooth/ LoRa/ 3GPP/ 5G |
| Autonomous learning support| No                     | No                          | Yes                       |
| Autonomous mobility        | No                     | No                          | Yes                       |

- **Complex data stream processing:** Multi-agent cloud robotics requires modular robotic applications with diverse programming models including map-reduce and distributed data flow to process complex data streams generated by robot-embedded multi-purpose sensors [23]. Due to resource and communication constraints, the local and proximal resources of multi-agent cloud robotics often fail to participate in such intense data processing that increases the burden on global resources. Therefore, resource allocation and service provisioning policies for a multi-agent cloud robotics system need to determine where to process data streams in real time. This significantly differs from MCC and MEC, which generally process simple batch or stream data generated by conventional sensors of smart phones and IoT devices.

- **Dynamic robotic cooperation:** There are some cases such as earthquake disaster management and submarine cable maintenance when multiple heterogeneous robots need to work collaboratively for attaining a common goal. Moreover, while operating in such adverse working environments with limited communication support, the real-time robot-to-robot and robot-to-edge or cloud interactions get hampered significantly [22]. To meet these requirements and constraints, the robotic cooperation often depends on distributed but coordinated algorithms such as federated learning and distributed learning, which adds further resource management overhead to multi-agent cloud robotics. The resource and service provisioning policies should address this issue deliberately that is not compulsory in MCC and MEC due to their simplified working environments.

- **Cross-infrastructure interoperability:** In multi-agent cloud robotics, heterogeneous and distributed computing resources complement the robots to carry out their responsibilities. These resources are owned by different service providers. For example, multi-agent cloud robotic systems can use the Google or Amazon cloud infrastructure and their edge infrastructure can be set by the cellular service providers. The resource sharing, privacy, and fault tolerance functions of these infrastructure are also managed with separate black box software systems. Therefore, the resource allocation and service provisioning policies in multi-agent cloud robotics require support for cross-infrastructure and policy-driven interoperability, which is not essential for MCC and MEC because of their association with homogeneous resource providers.

- **Synchronized decision making:** Since multi-agent cloud robotics involves computing infrastructure from different communication layers to execute robotic applications, an explicit synchronization of decision making entities belonging to these infrastructure is highly required. However, the attainment of such synchronization becomes challenging due to the intelligence interference among robots, edge, and cloud resources. Therefore, resource allocation and service provisioning policies should fairly distribute the resource management responsibilities among multiple decision making entities, which is not obligatory for MCC and MEC relying on a single decision maker.

- **On-demand computation and communication trade-off:** The offloading facility in multi-agent cloud robotics enables the robots having low energy or low processing power to access the edge and cloud infrastructure and execute the computer-intensive robotic applications. To conduct this activity, a seamless interaction among local robots, edge and remote resources are required which consumes additional communication time. It becomes more complicated because of the uncertain mobility of robots. Furthermore, the heterogeneity of resources in multi-agent cloud robotics with respect to processing cores, networking standards and energy consumption rates intensifies the appropriate resource selection problem for offloading. Therefore, resource allocation and service provisioning policies in multi-agent cloud robotics require on-demand computation and communication trade-off [24], [25]. On the other hand, in MCC and MEC, the computation and communication trade-off often depend on predefined rules because of the static mobility pattern of the users.

- **System-specific policy:** In multi-agent cloud robotics, the Quality of Service (QoS) requirements of robotic systems vary from one to another. For example, the service delivery deadline for a robotic application in smart hospital is more stringent than that of in smart farm. Similarly, for some robotic systems, higher accuracy is important such as robotic surgery, whereas for others the minimization of cost is vital such as smart factory maintenance. A generalized resource allocation and service provisioning policy is not feasible for multi-agent cloud robotics like MCC and MEC. Therefore, system-specific policies are required for efficient resource management, which is quite difficult due to the dynamics of robotic systems.

- **Energy-delay optimization:** Although computing resources from every layer of the hierarchical architecture participate in executing robotic applications, still the realization of such an environment is constrained by the energy limitations of the resources. Moreover, it is challenging to finish the application execution with the residual energy of the resources. Specifically, in a multi-agent cloud robotic system, robots can
perform simultaneously as a service provider and a user. As a result, the energy limitations of robots significantly hamper the execution of applications compared to MCC and MEC. Therefore, selection of appropriate resources and time-efficient resource allocation are required to deliver services within the energy constraints of the resources.

- **Comprehensive business model:** Cloud computing has a widely accepted business model for subscription-oriented services. However, such business model is not feasible for edge infrastructure as it mainly deals with event-driven and localized demand. On the other hand, robotic systems often require a set of reserved resources for processing their complex stream of data. Due to such variations of service requirements, it is complicated to develop a comprehensive business model for multi-agent cloud robotics that significantly disrupts the cost and budget-satisfying resource allocation, which is generally easier to develop in MCC and MEC.

- **Security and safety:** In a multi-agent cloud robotics system, robots are able to share the resources with other robots, edge resources as well as cloud data centre. Consequently, this system can easily become a target for security threats at every communication layer (local, edge, and cloud). The number of points of data exposure is comparatively higher in multi-agent cloud robotics than in MCC and MEC. Even in a local network, if any robot becomes vulnerable to privacy and security threats, it can easily manipulate others’ data in the system because of robot itself being an intelligent entity. This type of risk is relatively lower in MCC and MEC. In addition, due to the random mobility of the robots, they are required to communicate over heterogeneous networks that gives rise to another security issue. Therefore, for efficient resource allocation and service provisioning, the selection of reliable resources is crucial. Apart from the security vulnerabilities, uncertain events such as link failures, edge node shut down, power cut and interference can obstruct the robots to access the infrastructure services. In such cases, proactive and reactive fault tolerance techniques are helpful to make the system robust; however, it is difficult to develop these techniques due to the dynamics of multi-agent cloud robotic systems.

Resource allocation and service provisioning have been well studied in different computing paradigms including MCC and MEC [13]–[19]. However, the adopted algorithms, techniques and recommendations of these paradigms cannot directly be applied to multi-agent cloud robotics due to the above-mentioned issues.

### C. Motivation of this Article and Our Contribution

A notable number of research works have been conducted to ensure efficient resource allocation and service provisioning in multi-agent cloud robotics overcoming the existing challenges. Depending on network condition, application requirements and resource availability, different robot-edge-cloud interaction techniques such as peer to peer, proxy and clone-based communication have been developed [2], [3]. Although these works have significant impact in enhancing the resource allocation and service provisioning on multi-agent cloud robotics, there are a very few efforts in the literature for categorizing them in a systemic manner. To address this issue, we have conducted an extensive survey on existing resource allocation and service provisioning policies for multi-agent cloud robotics. To the best of our knowledge, this is the first literature review on resource allocation and service provisioning problem in multi-agent cloud robotics. This work can be a notable inclusion to the existing literature, helping readers understand the state-of-the-art of multi-agent cloud robotics with its promising use cases for Industry 4.0 applications, healthcare, smart agriculture, disaster management and other robotic applications. The importance of optimal resource allocation, resource pooling, computation offloading, and task scheduling are well studied in this survey for efficient service provisioning to the multi-agent robotic systems. A summary of lessons learned from the existing works in literature are also offered to identify the research gaps to address the challenges of resource allocation and service provisioning. Moreover, a holistic framework for efficient resource allocation and service provisioning is also provided and we believe that it could add a value to the existing literature and open directions for future research. The major contributions of this paper are:

- A taxonomy of resource allocation considering the resource type, performance metrics, application structure, service model and allocation mechanism is presented in this survey.
- Existing approaches on resource pooling, computation offloading and task scheduling in this field are also well explored for efficient service provisioning.
- The lessons learned from the literature review are summarized and the gaps in efficient resource allocation and service provisioning for robotic systems in multi-agent cloud robotics are identified.
- Moreover, future research directions with a holistic framework premised on them will assist researchers to improve the state of multi-agent cloud robotics.

### D. Article Organization

The rest of the article is organized as follows. The list of abbreviations used in this survey is provided in Table I. Before going into an in-depth discussion on resource allocation and service provisioning, in Section II, we summarize the recent practices of multi-agent cloud robotics both in academia and industry along with current and future applications of this paradigm. In this section, the contributions of our survey are also compared with related surveys. In Section III a comprehensive review with the help of a taxonomy is presented particularly focusing on the research of resource allocation in multi-agent cloud robotics. Different techniques in

### TABLE II: A list of abbreviations used in this survey

| Acronym | Description |
|---------|-------------|
| BtF     | Bag-of-Tasks |
| CAN     | Controller Area Network |
| CPSs    | Cyber-Physical Systems |
| CRASF   | Cross-infrastructure Resource Allocator and Service Provisioner |
| DAG     | Directed Acyclic Graph |
| DL      | Deep Learning |
| DSN     | Data Stream Network |
| IaaS    | Infrastructure-as-a-Service (IaaS) |
| ILP     | Integer Linear Programming |
| IoT     | Internet of Things |
| MCC     | Mobile Cloud Computing |
| MFC     | Multi-access Edge Computing |
| ML      | Machine Learning |
| PaaS    | Platform-as-a-Service |
| QoS     | Quality of Experience |
| QoS     | Quality of Service |
| RaaS    | Robot-as-a-Service |
| ROS     | Robot Operating System |
| SaaS    | Software-as-a-Service |
| SLA     | Service Level Agreement |
| SLAM    | Simultaneous Localization and Mapping |
| SOA     | Service-oriented Architecture |
| UAVs    | Unmanned Aerial Vehicles |
| sIP     | Micro Internet Protocol |
| VMs     | Virtual Machines |
resource pooling, computation offloading, and task scheduling are studied in Section IV. From the lessons learned in this survey paper, the research gaps for efficient resource allocation and service provisioning are investigated in Section V. A prospective holistic framework for efficient resource allocation and service provisioning along with directions for future research are provided in Section VI. Finally, concluding remarks are drawn in Section VII.

II. Overview of Multi-Agent Cloud Robotics Advancements

In this section, the current and future applications and recent practices of multi-agent cloud robotics in industry are summarized. We also conduct a comparative study of our work with existing surveys.

A. Current and Future Applications of Multi-Agent Cloud Robotics

Multi-agent cloud robotics has significantly improved the performance of various CPSs including smart factory, remote healthcare, smart farm, and disaster management by offering real-time execution platform for different robotic applications including simultaneous localization and mapping (SLAM), robotic vision, path planning, navigation, grasping and surgical assistance [3].

1) Industry 4.0: Industry 4.0 refers to the fourth industrial revolution that assists in digitization of industrial manufacturing with the help of IoT, big data analytics, and cloud computing. In this environment, a group of robots and human work collaboratively throughout the entire industrial value chains [26–27]. As industrial robots are the key drivers for such applications, multi-agent cloud robotics inherently plays a significant role in developing these applications [28]. Smart factory is one of the prominent applications of Industry 4.0 [13], [29], [32].

Many works have concentrated on multi-agent cloud robotics for smart factory. The feasibility of multi-agent cloud robotics in gathering and processing sensory data for the navigation of autonomous vehicles in industrial environment is discussed in [32]–[35]. This concept is also exploited for ubiquitous product management [36], customer maintenance [28], [29] and material handling [40] in smart factory.

In smart factory, cloud and edge based industrial robots have to deal with heterogeneous tasks and dynamic environment like taking variety of orders from consumers, dealing with different situations and making delivery of each order [28], [29]. The dynamic scheduling of the tasks to the components in smart factory according to workload is one of the key research challenges [41]. For the task execution and resource sharing among cloud and edge-aided industrial robots, computational load scheduling is also equally important [42]. To meet these aspects, development of efficient resource allocation strategies in multi-agent cloud robotics is a fundamental demand for Industry 4.0 applications.

2) Health Care: The capabilities of multi-agent cloud robotics have also been leveraged to enable remote health monitoring [11] and elderly care support [43]–[45], and assist people with disabilities in mobility [46], [47] and communication [48], [49].

To monitor the health status of a patient in real time, a healthcare system based on robotics and cloud computing is designed in [11] which can also be controlled remotely by the doctors or carer. The authors of [48] present a robotic telepresence platform by means of cloud technology, supporting mobility impaired peoples. Their platform assists the robot with autonomous navigation capabilities and teleoperation among robot and users. The developed solution in [43] bestows health-care management services to senior citizens and improves their living standards. It supports user with indoor localization service, care reminding service, environmental monitoring service in form of robotic service. Location based and personalized assistive services to seniors are also delivered by [44]. For chronic disease management, a domiciliary reminder service for personalised medical support using hybrid cloud robotics framework is afforded in [45]. Another service is provisioned in [49] to support human-robot interaction and environmental sensing.

With the help of cloud and edge computing, service robots perform different services to overcome the limitations of healthcare [50]–[53]; however, proper allocation of resources is required to confer guaranteed QoS for time-critical task execution in multi-agent cloud robotic systems.

3) Agriculture: Smart farming is a viable solution to develop the traditional agricultural system, where real-time data gathering, processing and analysis, as well as automation technologies on the farming procedures are applied [12], [54]. Recently, the concept of multi-agent cloud robotics has been extended for smart farming with the help of smart, distributed and collaborated sensors, IoT, GPS, cloud computing, aerial and ground robots [55], [56] and offer autonomous field monitoring, pesticides spraying, weeding as well harvesting [57]–[60].

A service model using an agricultural expert cloud for ubiquitous agricultural environments is developed in [61], which yields necessary information and services for cultivating any crops on any ecosystem. Based on cloud computing, another service-oriented smart farming architecture is introduced in [62] that bestows local farm services, real-time services and cloud services including weather and map services to multiple farming areas. To collect data from deployed sensors in a smart farm including cameras, drones and soil sensors considering the bandwidth constraints, an IoT platform is presented in [63]. Another integrated system to support collection, analysis, and prediction of agricultural environment for strawberry infection prediction is outlined in [64].

Unmanned Aerial Vehicles (UAVs) play a significant role in smart farming system by providing imagery analysis and agricultural surveillance as well as on-demand communication [65]. UAVs are applied in [66] to perform vegetation detection, plant-tailored feature extraction and classification to estimate the distribution of crops and weeds in the field. Through a coalition of aerial and ground robots, how the services of multi-agent cloud robotics can be extended to smart farming is also demonstrated in [67]. However, agricultural production systems are prone to unpredictable environment (e.g., rain, temperature, humidity etc.) and unwanted events (e.g., animal diseases, pests) [68]. Still the smart farms need to integrate robots and sensor data for delivering agro-services, which are not constrained by communication and unpredictable environments. The use of smart technologies including artificial intelligence and multi-agent cloud robotics are in their infancy stage in agriculture [69], opening a way for further investigation.

4) Disaster Management: The services of multi-agent cloud robotics have been adopted widely for conducting unmanned search and rescue operations in hostile environments [70]–[74], including alpine [75] and fire-driven emergency scenarios [10], [76].

To enhance the collaboration of human rescuer and ground-aerial robots, in [73] cloud computing is augmented with robotics to provide fast, reliable and available resources for searching and rescuing operation. For the interaction between users and UAVs, another multi-agent cloud robotics architecture is presented in [70] for emergency management and monitoring service. Moreover, for
search and rescue operations in large-scale disaster, cloud resources are offered as infrastructure to robots in [74]. Multi-agent cloud robotics has also been enhanced for fire-driven emergency management service in [10] and [76] by incorporating edge resources.

In general, while providing disaster management services, a group of robots, edge and cloud resources work collaboratively towards completing a central mission. Thus, optimal allocation of tasks to the robots and offloading the tasks to the edge or cloud is required to achieve the objective successfully considering the environmental uncertainties.

Apart from the above mentioned applications, with the advancement of multi-agent cloud robotics, robots offer services in smart home system for making daily life more easier [101]–[105]. Robotic services are also offered in security inspection, grocery shopping delivery, smart transportation, entertainment as well as education and many more applications [2].

B. Recent Industrial Practices of Multi-Agent Cloud Robotics

Apart from the academia, several technology and business organizations are also focusing on the development of various solutions to integrate multi-robot systems, edge and cloud computing. For example, Google, Amazon, and Microsoft have already started extending cloud technologies to the robots. Google cloud robotics platform aims at harnessing artificial intelligence, cloud, and robotics to offer utility services to the customers. Amazon is patronizing the AWS RoboMaker that provides machine learning, monitoring and analytics support to the robots. Microsoft has developed Robot Operating System (ROS) to program the robots with higher capacity [106]. Additionally, the Honda RaaS platform aims at providing a wide range of robot and cloud-based services to support communication, robotic cooperation, and data sharing. Another company named CloudMinds works on production of robots with embedded deep learning facilities for real-time data collection and sharing using 5G. There exist other robotics companies including Fetch Robotics, InVia Robotics, Kuka, Plus One Robotics, FANUC Corporation, ABB Robotics that are currently working on warehouse automation [107], [108].

Despite of such initiatives to standardize the concept of multi-agent cloud robotics, the efficient resource allocation and service provisioning for the robotic applications is still a major concern. Therefore, a good number of research works are currently being conducted to develop various resource allocation and service provisioning policies for multi-agent cloud robotics. This survey focuses on summarizing these works in a systematic manner.

C. Related Surveys

Several surveys in the literature highlight the importance of resource allocation. However, they do not review and address the challenges of resource allocation problem in multi-agent cloud robotics. For example, only a set of approaches to assign robots for task execution in multi-agent system are reviewed in [96], [97]. In comparison, the resource allocation problem is well studied in cloud computing [77]–[80]. The cloud resource provisioning and scheduling algorithms are summarized in [82]–[85]. Similarly, resource provisioning and auction-based radio resource allocation mechanisms for wireless and mobile systems are discussed in [86] and [87], respectively. Computation offloading is another widely explored technique in the field of wireless and mobile systems [88], [89]. Existing offloading and service provisioning techniques in mobile cloud computing are comprehensively surveyed in [15], [17], [90]–[92]. Conversely, the computation offloading techniques in mobile edge computing and multi-access edge computing are briefly described in [93]–[95] and [16], [109], respectively.

On the other hand, in literature, there exist a notable number of surveys on cloud robotics. However, most of them conduct surveys on architectural perspectives [2], [3], [98]. Additionally, the complexities and limitations of cloud robotics including the necessity of efficient resource and task allocation policies are partially discussed in [99]. The robotic cooperation techniques for critical CPSs such as disaster management and smart manufacturing are reviewed in [100] and [42], respectively. Nevertheless, there does not exist any survey in the literature that summarizes the resource allocation and service provisioning techniques including resource pooling, computation offloading and task scheduling jointly for multi-agent cloud robotics. A comparative study of our survey with respect to the existing works is illustrated in Table III. In the following sections, different aspects of resource allocation and service provisioning in multi-agent cloud robotics are reviewed in detail.

III. RESOURCE ALLOCATION IN MULTI-AGENT CLOUD ROBOTIC SYSTEMS

The robotic applications of real-world systems possess an inherent demand of faster processing. The optimal allocation of resources can play a vital role in meeting this requirement of the applications to a great extent. Usually, robotic applications encapsulate single or multiple robotic tasks. Basically, resource allocation defines the assignment of a resource to a task based on its availability and the QuoS requirements of the task. However, these requirements of

| Works | Resource Allocation | Resource Discovery | Computation Offloading | Task Scheduling | Research Domain |
|-------|---------------------|-------------------|-----------------------|----------------|----------------|
| [7]   | ✓                   | ✓                 | ✓                     |                | Cloud computing |
| [9]   | ✓                   | ✓                 | ✓                     |                | Cloud computing |
| [9]   | ✓                   | ✓                 | ✓                     |                | Cloud computing |
| [8]   | ✓                   | ✓                 | ✓                     |                | Wireless and mobile systems |
| [8]   | ✓                   | ✓                 | ✓                     |                | Wireless access network |
| [8]   | ✓                   | ✓                 | ✓                     |                | Mobile cloud computing |
| [15], [17], [90]–[92] | ✓   | ✓                 | ✓                     |                | Multi-access edge computing |
| [45] | ✓                   | ✓                 | ✓                     |                | Multi-agent system |
| [5]   | ✓                   | ✓                 | ✓                     |                | Cloud robots |
| [8]   | ✓                   | ✓                 | ✓                     |                | Cloud robots |
| [100] | ✓                   | ✓                 | ✓                     |                | Cloud robots |

This survey ✓ denotes broad discussion on the respective topic. ✓ denotes partial discussion on the respective topic.
robotic tasks vary from one system to another. Furthermore, the resource availability in robot, edge and cloud instances changes with the course of time unless it is reserved. For example, the QoS requirements of tasks for a robotic surgery application is quite stringent than the tasks of a robotic parcel delivery application. Similarly, the robotic application that monitors environmental context occupies resources for a longer period compare to robot enabled autonomous irrigation application where the resource acquisition occurs only for a specific time. During such cases, the detailed exploitation of resource type, performance parameter, application structure, service model, allocation mechanism and evaluation method can guide towards the efficient resource allocation. In this work, these aspects of resource allocation are narrated thoroughly with the help of a taxonomy shown in Figure 2.

A. Resource Type

As the robotic applications consist of latency sensitive, compute and data-intensive tasks, a wide variety of resources are required to execute these applications. Based on the necessities of the tasks, allocation of three types of resources, namely, computational resource, network resource and storage are discussed in the literature.

1) Computational resource: The resources that are used for data processing and executing the manifold tasks of robotic applications are known as computational resource. Cloud instances, local robots and edge nodes are the basic computational resources.

• Cloud instances: Cloud instances are allocated to robotic applications for large-scale computation. Cloud service providers virtualize the computing servers and offer a variety of computing instances including virtual machines (VMs) and containers to the multi-agent robotic systems. The authors of [110]–[112] focus on allocating cloud instances for executing robotic applications. Moreover, cloud instances are dynamically configurable according to the resource requirements of the applications. However, the integration of robotic systems with the cloud instances depends on the category of cloud infrastructure namely public, private, edge and hybrid cloud.

  – Public cloud: In this of cloud infrastructure, the instances are accessed on-demand over the Internet [28]. It bestows the greatest level of efficiency in shared resources where the cloud users must pay as per the usage. In the literature, [113]–[127] deploy public cloud infrastructure for the robotic applications. However, the resources in public cloud often become vulnerable to security issues. Due to geographical distance, the data transfer also gets delayed while executing robotic tasks using public cloud.

  – Private cloud: In this type of cloud infrastructure, instances are accessed through a private network. It supports versatile and convenient end-to-end interaction, where the management of the resources are conducted within the local data centres. [128]–[133] adopt private cloud for their systems. Private cloud provides more security than public cloud. However, the maintenance of private cloud requires dedicated administrator, which is not always cost efficient for the multi-robot systems.

  – Hybrid cloud: It is a combination of public and private cloud infrastructure with an additional orchestration and automation support [28]. In this type of cloud infrastructure, the tasks of real-time or mission-critical applications are run on the private cloud instances while the public cloud instances deal with the delay tolerant tasks [134]. Hybrid cloud can also encapsulate edge resources as mentioned in [74], [135]–[139]. Hybrid cloud can also encapsulate edge resources as mentioned in [22], [116], [134]. Nevertheless, the goal of hybrid cloud is to ensure a scalable environment for networked robots where the instances can be extended either from public or private cloud infrastructure based on the task requirements. However, such management of instances in hybrid cloud becomes a challenging issue when there exist an explicit data-dependency among the robotic tasks.

As there exist pros and cons of each type of cloud infrastructure, the resource allocation policies for multi-agent cloud robotics require to observe them deliberately while dealing with the robotic applications.

• Local robots: The on-board CPU within the robots also supports the task execution of robotic applications. In multi-agent cloud robotic systems, local resources can be shared among the connected robots to complete the task execution in a collaborative manner. For example, the local robots are used for robotic task execution in [130]–[134]. However, the on-board computing components often become exhausted because of the size, shape, power supply, motion mode and working environments of the robots [2]. Resource re-configuration is also infeasible once the robots are built for a particular system. Despite of such constraints, it is still conducive to use the local robotic resources for the hard real-time robotic applications since they offer comparatively lesser data transfer delay.

• Edge infrastructure: The inter-communication delay in sending the computation request to cloud and receiving the response at the local robots resist the whole system in meeting the QoS requirements of robotic applications. Local robots have limited energy and computational capacity, which further hampers the execution of large-scale real-time applications. To address these limitations, the concept of edge computing is introduced between the robots and the cloud in multi-agent cloud robotics. Edge infrastructures such as cloudlet or edge cloud and fog nodes facilitate task execution in proximity of the robotic systems and consequently enhances the QoS of the applications.

  – Cloudlet or edge cloud: Basically Cloudlet or edge cloud is an extension of cloud that can provide virtualized resources and assist multi-robot systems in executing the latency sensitive and compute-intensive tasks [76]. Integrating both edge and cloud computing with the robotics, a platform is introduced in [22]. Moreover, the local robots and edge cloud are allocated for reducing communication delay in [10], [76], [143].

  – Fog node: Any personal computer, mobile device, smart edge device, car, sensors, traditional networking devices including set-top boxes, gateway routers, smart switches, proxy servers equipped with computational resources are generally used as potential fog nodes and they are deployed in a distributed manner across the edge [146], [147]. Rather than forwarding the tasks of latency sensitive robotic applications to a centralized cloudlet or private cloud, they can be processed more efficiently at the edge network using fog nodes [148]. The idea of utilizing fog nodes in multi-agent cloud robotics is addressed in [149]–[153].

To summarize, although cloud resources offer higher computational power, they include additional delay for data transfer which is not suitable for latency sensitive tasks. Although robots are feasible
2) Network resource: The physical and logical resources within a network that help connectivity, communication and data transmission are referred as network resource. In literature, the allocation of network resources are explored in [154]–[157]. A mesh orchestration of networking nodes makes the sharing of network resources possible in multi-agent cloud robotic systems. To maintain guaranteed data flow over the shared network, bandwidth, network slices and power transmitters are allocated as the network resources.

- **Bandwidth**: The amount of data that can be transmitted in a fixed amount of time defines the bandwidth of a network. Optimal bandwidth allocation is very important in multi-agent cloud robotics for retrieving and sharing multi-sensor data, communicating over the network, and getting the robotic services from cloud and edge instances with QoS guarantee. In literature, [154]–[156] focus on bandwidth allocation in multi-agent cloud robotics.

- **Networking node**: The connection points for data transmission within a network such as routers, switches, multiple antennas, wireless access points and base stations are referred as networking node [158]. Even in [157], the robots are considered as the networking nodes. However, optimal placement of the nodes in the network is very crucial for sending and receiving data in a timely manner.

- **Wireless power**: By harnessing the mobility of robots, various hazardous tasks can be done in remote environments. However, the main limitation of mobile robots is the short battery lifetime. Therefore, while conducting remote jobs, mobile robots are required to return to a base station for charging, manually battery replacement or tethering [159]. However, these operations are time-consuming. As a sustainable solution, wireless power transfer has emerged to address the energy limitations of robots and provide a backup of on-board battery [160]. In literature, there exist several techniques for wireless power transmission including radiative power transfer (microwave, laser) and non-radiative power transfer technologies are (inductive, magnetic resonance, capacitive and acoustic) [161], [162]. However, in multi-agent cloud robotics, the core challenge of transmitting wireless power is to mitigate the mutual interferences.

- **Network slice**: A network slice is an end-to-end logical network that runs on a shared physical infrastructure to provide negotiated network services. It virtualizes the network infrastructure including terminal, access network, core network and make them deployable across multiple network providers [163], [164]. In multi-agent cloud robotic systems, the utilization purpose of network infrastructure varies from one robotic system to another. For example, a multi-robot system may require high reliable data transfer, whereas the other one may require low latency with high data rate. Network slicing is an essential technology in multi-agent cloud robotics to handle such heterogeneous network service requirements of robotic systems. How the network slices will be composed and how much resource will be allocated for a particular network slice and how the network slices will be distributed across multiple robotic systems are the key research questions to solve in this domain. However, the efficient mechanisms for wireless power allocation and network slicing are still in their early phase of research and need to be explored more.

3) Storage: This type of resources are used for storing data, information sharing and collective learning while executing any robotic application. In multi-agent cloud robotics, the cloud-based data servers, foglets, built-in memory of local robots offer the storage facilities. However, most existing works in multi-agent cloud robotics pay attention to computational and network resource allocation, while a few including [165] focus on allocating storage resource. Additionally, in multi-agent cloud robotics, it is very

---

**Fig. 2**: Taxonomy of resource allocation in multi-agent cloud robotics.
crucial to select the data and their storage location so that ease of data access can be ensured during real-time interactions.

- **Cloud storage**: Cloud infrastructure inherently serves an extensive storage space to the robots. Virtualization techniques and NoSQL databases play the vital roles in storing the information in cloud storages. These information are explicitly maintained, operated, and managed by the cloud service providers. Additionally, cloud storages support collective learning by providing application programming interfaces for analysing historical behaviours according to task requirements [2]. However, in multi-agent cloud robotic system while executing robotic applications, the exchange of information should be happening in real-time. In this case, an uninterrupted Internet connectivity is a must. Moreover, there are some sensitive applications where the data privacy is regarded as one of the QoS requirements. To deal with such applications, the selection of appropriate cloud infrastructures is very important.

- **Robotic storage**: The built-in memory of the local robots can collectively support the additional storage resources for multi-agent cloud robotics. A group of connected robots usually share information among each other while executing the robotic applications in a collaborative manner. Robotic storage allows the robots to store necessary information locally so that the robots residing outside the communication range, can still participate in learning and data collection. However, even after the augmentation of robots, this type of storage offers a very limited capacity and does not support easy re-configuration. Therefore, they are considered well-suited for static environment. For dynamic environment, the cloud storage outperforms the local robotic storage.

- **Edge storage**: Although cloud computing helps robots in sharing information, it adds high latency during the execution of hard real-time applications. In the form of fog servers and foglets, edge computing offers storage resources to multi-agent cloud robotic systems. They can easily store necessary information and update data according to the application requirements using the adaptive and augmented memory [150]. Additionally, in [166] and [167], the idea of using local robot as a fog storage server is introduced. Nevertheless, the multi-agent cloud robotic system needs to be concerned about the security and privacy preservation of data while utilizing edge storage. Therefore, recently much emphasize is given on developing efficient algorithms for secured edge and fog storage management in multi-agent cloud robotics.

**B. Performance Metrics**

While allocating resources for robotic applications in multi-agent cloud robotics, different metrics including time, power, cost and QoS are exploited to monitor their performance and meet the service requirements. Various aspects of these parameters are discussed below.

1) **Time**: It is one of the important metrics to measure the efficiency of resource allocation policies. In multi-agent cloud robotics, the minimization of computation time and communication time and the timely execution of tasks within deadline is desired for the improved performance.

- **Computation time**: The time required for executing a task on computational resources is defined as computation time. It mostly depends on the processing speed of assigned resource and size of the task. The authors of [111], [141]–[143], [168] design their resource allocation schemes to minimize the application completion time.

- **Communication time**: It basically refers to the networking delay while exchanging data among different entities of multi-agent cloud robotic systems. Reduced communication time reflects the efficiency of network resources in assisting the application execution on the computational resources. [140], [142], [155], [157] address the minimization of communication time.

- **Deadline**: The maximum tolerable delay of a system in receiving application service delivery is specified as the deadline. Service delivery deadline plays a vital role in characterizing latency sensitive (real-time) and latency tolerant (non real-time) applications. It also functions as a decision parameter while allocating resources to meet the application’s QoS requirement. Among the existing works, [10], [76] emphasise on the satisfaction of application deadline while allocating resources for the tasks of robotic application.

Additionally, there are some other time-based metrics such as resource discovery time, service access time, data sensing frequency of sensors which are also required to be considered as performance indicator while allocating resources in multi-agent cloud robotics.

2) **Power**: The minimization of overall power usage is another important indicator of improving performance in multi-agent cloud robotics. Power usage is mainly focused on the energy consumption and heat generation aspects of the computing resources.

- **Energy consumption**: It is calculated in terms of computation time and per unit time energy requirements of the resources to execute a task. As robots are usually energy-constrained, low battery power can degrade the performances of multi-agent cloud robotic systems. At the same time, the energy consumption of cloud resources causes huge capital expenditure and operational cost. Hence, considering the energy parameter, appropriate resources are required to be allocated so that with less energy, desired performance of the system can be ensured. In the literature, very few works including [10], [76], [111], [112] aims at minimizing the energy consumption of resources while assigning the applications.

- **Heat generation**: It refers to the thermal effect on computing resources while executing different tasks of robotic applications. As the heat emission increases with the number of resources, it increases the probability of hardware system failures and carbon footprint. Therefore, it is essential for service provider to improve the cooling system and generate less heat while handling the application tasks. In [110], the reduction of both energy consumption and heat generation are emphasized to meet the power usage in multi-agent cloud robotic systems.

Apart from these power-related metrics, the detail analysis of residual battery lifetime of end devices and the exploitation of energy characteristics of communication medium are also required to design a power-efficient multi-agent cloud robotic system. However, such initiatives are subjected to extensive research.

3) **Cost**: This performance parameter preserves the interest of both service providers and end users in multi-agent cloud robotics. Since service providers are responsible for facilitating computing services to numerous robotic systems, they always aim at making a comprehensive profit through extreme resource utilization [169]. Similarly, the users of robotic systems demand maximum Quality of Experience (QoE) without surpassing the budget. In such circumstance, cost-aware resource allocation mechanism preserves...
the economic benefits of both service provider and users of robotic systems. Resource deployment cost and user expenses are the main cost-related factors in multi-agent cloud robotics.

- **Resource deployment cost**: The associated cost for infrastructure placement in multi-agent cloud robotics defines the resource deployment cost. It includes the expenses for placing sensors, local robots as well as creating the VMs and containers in cloud. Networking cost is also encapsulated in resource deployment cost which is directly related to the charges of data transferring, bandwidth, and networking nodes. In the literature, most of the works emphasizing the cost-related issues of multi-agent cloud robotics, aim at minimizing the resource deployment cost. For example, [110], [143], [154], [155] reduce the resource deployment cost while allocating resources for the robotic applications.

- **User expenses**: The amount paid by the robotic systems and its users to occupy the resources for executing tasks of applications is known as user expenses. In multi-agent cloud robotics, service providers determine the per unit time usage cost of resources to execute a task. On the other hand, robotic systems urge to access the desired computing services within the budget constraint. Thus, the minimization of user expenses becomes an important element while allocating resources in multi-agent cloud robotics. There are several works in the literature including [10], [141] that investigate user expenses during resource allocation.

- **Profit**: The net business gain from the revenue and service execution cost is considered as the service provider’s profit. The profit of service providers while assigning resources to the robotic tasks is discussed in [112]. Moreover, the service provider’s profit in multi-agent cloud robotics depends on the service charges and the business costs associated with a particular robotic service. These aspects are brought into attention in [165]. Furthermore, the pricing model for any computing service is determined by the operators and it depends on the resource availability. While allocating resources in multi-agent cloud robotics, the selection of pricing model also becomes an important concern as it helps to maximize profit within the budget constraints.

4) **Quality-of-Service**: The most significant performance parameter in multi-agent cloud robotics is QoS. It refers to the distribution of resources according to the requirements of the applications. QoS is driven according to the Service Level Agreement (SLA) between the service provider and robotic systems. The availability and fault tolerance of resources also play vital roles in defining the QoS.

- **Service Level Agreement (SLA)**: An agreement between the cloud service provider and robotic system as per the usage of resources is denoted as SLA. In multi-agent cloud robotics, service provider is responsible to allocate resources according to the requirements of an application and avoid the SLA violations. The performance of a system in terms of SLA depends on ratio of the number of successfully executed tasks and the number of total assigned tasks [158]. In [111], [155], [156]. QoS is enhanced by minimizing the SLA violations.

- **Availability**: The accessibility, usability, scalability of the resources while executing the applications defines the availability of resources in multi-agent cloud robotics. The service providers must ensure resource availability prior to allocating them to any robotic application. However, in [144], SLA and resource availability are simultaneously counted as the QoS parameters.

- **Fault tolerance**: This property enables a computing paradigm to continue the execution of assigned tasks even after the failure of any computing and networking entities [170]. In multi-agent cloud robotics, uncertain anomalies such as link failures, power shortage, limited bandwidth and interference can hamper the simultaneous execution of the robotic application. Hence, it is very important to ensure fault tolerant during resource allocation. Additionally, security and reliability, maintainability can also drive the fault tolerance and collectively influence resource allocation.

C. Application Structure

In multi-agent cloud robotics, a robotic application consists of single or multiple tasks. Based on the dependencies and requirements of different tasks, the robotic applications follow various application structures for execution. Directed Acyclic Graph (DAG) workflow and Bag-of-Tasks (BoT) are the mostly used models for robotic applications [172].

1) **DAG workflow**: A workflow contains a sequential series of tasks having data dependencies among them. In most cases, it is described as a DAG, where the nodes are tasks, and the edges denote the task dependencies. The execution order of tasks in a workflow can be either serialized or parallel. This type of structure depicts how the tasks are interrelated that eventually helps in allocating the resources optimally with better efficiency [10], [76]. [111], [140]–[144] adapt DAG workflow while allocating resources for the tasks of robotic applications.

2) **Bag-of-Tasks**: Generally, Bag-of-Tasks (BoT) is applied to those applications where no data dependencies exist among the tasks. BoT applications are parallel applications comprised of independent but similar tasks [172]. Identical tasks from different applications as well as from the same application can form BoT application model, which assists concurrent utilization of resources in distributed environments. In literature, [110], [112], [154]–[157]. [165] use BoT for modelling and allocating resources for multiple applications.

Nevertheless, while allocating the resources for the tasks of robotic applications, the multi-tenancy facility of resources is required to be taken into account. It helps in determining whether multiple resources will execute the same task or vice-versa. Additionally, the resources need to be allocated according to the delay sensitivity of applications even it requires to go beyond their architectural differences.

D. Service Model

There are five types of service models, namely, Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), Software-as-a-Service (SaaS), Robot-as-a-Service (RaaS) and Networking-as-a-Service (NaaS) that a multi-agent cloud robotic system can widely support.

1) **IaaS**: It refers to the pool of virtualized resources used for executing the tasks of robotic applications. Generally, virtual machine, containers, servers, storage are delivered as IaaS. The frameworks developed in [130], [138], [139] yield IaaS provided by the cloud service provider to the robots. Moreover, there are some application specific researches for disaster management [74], vision acquisition [131], where the integration of customized IaaS with robotic systems are considered explicitly. Additionally, in [115], the confederation of cloud infrastructure is considered for supporting robotic systems. While providing IaaS for the robotic
systems, the infrastructures are required to support diversity so that these resources can be beneficial to heterogeneous robots. In addition, the resource distribution among multiple robotic systems should be vibrant enough to fulfill the requirements. Furthermore, the infrastructure should be extensible in terms of adding new functionalities such as self-healing and fault tolerance.

2) PaaS: It refers to the operating systems, application runtime environments, programming interfaces, databases, web servers that help in developing robotic applications. In literature, [119], [120], [145] discuss PaaS for multi-agent cloud robotics. The PaaS highlighted in [119], [120] offer cloud aided map optimization and autonomous robot patrol for SLAM application. The feasibility of edge computing-enabled PaaS for robotic applications is investigated in [22], [145], [173]. A platform is introduced in [174] to offer robot inference and learning as-a-Service for grasp planning and objection recognition.

3) SaaS: It refers to the on-demand robotic applications over the Internet such as object recognition, path planning, map building, speech translation, knowledge base etc. In the literature, [122]–[124], [126], [128], [132], [135] discuss SaaS for multi-agent cloud robotics. In [128], robotic software packages are migrated to cloud so that multiple robots can simultaneously access them. The parallelization of robotic algorithms using cloud resources is discussed in [135]. Robotic path planning is delivered as a software in [123]. Additionally, the motion planning and control of industrial robots using cloud resources are discussed in [118], [124], [126], [132]. Cloud based computation are used in [175] for 3D robot grasping. SaaS is also offered in [176] for robot vision tasks including detection, segmentation, and object classification.

4) RaaS: It refers to robots that can be dynamically combined as-a-Service to execute specific applications [177]. The robots are distributed in different locations and can be accessed as service through multi-agent cloud robotic system for executing various robotic applications. RaaS is supplied in terms of computing, sensing, motion planning, navigation, and perception service. In literature, [117], [125], [127], [129], [137] particularly aims at serving RaaS. The idea of using a robot as an all-in-one service-oriented architecture (SOA) unit to simultaneously perform the responsibilities of service providers, service brokers, and service clients is revealed in [117], [137] enables robots and sensors to expose their functions to a virtualized resource pool and forms a cloud of robots. Robots computation facilities are offered as a cloud service to the end users in [129]. Moreover, the on-board sensors of the robots offer sensing-as-a-Service by perceiving information from the environment and support the processing and actuation, based on the sensed data [178]. Sensing capability of the robots is provided as a service in [179] for door-opening control problem in the real environment. Visual sensing is also offered in [180], [181]
by supporting the collaborative execution of image processing tasks on the robots and cloud. Furthermore, a prototype is developed for robotic grasping application in [125] to offer robotic and automation as-a-Service. Additionally, an access control policy for mobile robots is proposed in [127].

5) NaaS: To support the end devices having limited networking capacity and communication constraints, mobile robots such as drones or unmanned aerial vehicles (UAVs) provide Networking-as-a-Service (NaaS). Residing in proximity of the end devices, these robots act as either a repeater, or wireless range extender, or base station, or an access point. In literature, UAVs perform as wireless base stations in [182] that provide coverage for ground users. In [183], UAVs are deployed as flying base stations to support the coverage and throughput of wireless communication. UAV serves as a data ferry between the source base station and multiple destination receivers in [184]. UAVs facilitate data transmission of IoT devices in [185], [186]. UAV can also play the role of aerial cloudlet to collect and process the computation tasks offloaded by ground users [171]. Moreover, drones and UAVs are expected to be an important component of 5G or beyond 5G cellular architectures to make wireless broadcast or point-to-multipoint transmissions easier [187]. However, while offering NaaS in multi-agent cloud robotics, several issues such as the energy consumption minimization of service robots, flying time minimization with optimal trajectory planning of UAVs need to be focused for efficient resource allocation and service provisioning.

In literature, there exist other research works that deal with multiple service models simultaneously to facilitate various robotic applications. For example, [118], [121], [134], [138] offer IaaS, PaaS and SaaS collectively to multi-agent cloud robotics. Additionally, in [113], the architecture of Rapyuta platform is discussed that can be used by the developers to model and design robotic applications, and the applications running on this platform can be directly accessed as a software service. In most of the cases, IaaS, PaaS, and SaaS are jointly exploited for web applications. However, there exist significant differences between general web applications and robotic applications. Web applications are typically stateless, single processes that use a request-response model to talk to the client. On the other hand, robotic applications are state-driven, multi-processed, and require a bidirectional communication with the client [113]. Therefore, while provisioning services in multi-agent cloud robotics, the selection of appropriate service models as per the requirements of a particular robotic system is very important.

E. Mechanism

The successful execution of robotic applications in multi-agent cloud robotics largely depends on the efficient selection of resource allocation mechanisms. The design of resource allocation mechanism varies in harmony with the system environment and application requirements. In the literature of multi-agent cloud robotics, the formal optimization, approximation method, evolutionary computing, game theory, heuristic and deep learning techniques are adopted widely for resource allocation.

1) Formal optimization: It helps to solve any optimization problem with single objective or multi-objective under certain constraints. In multi-agent cloud robotics, depending on the key performance metrics of a particular system, the optimization method is selected. For example, in a smart factory, the main objective of resource allocation is to reduce the resource deployment cost. On the other hand, in a smart farm, the resource allocation objective is to complete the trajectory of the UAVs within their battery lifetime.

Based on the characteristics of optimization function such as convex (minimization) or concave (maximization) and the relation among the constraints, the choice of optimization method (linear programming, integer programming, non-linear programming, combinatorial, stochastic) is made in multi-agent cloud robotic system. For example, [142] linearizes a mixed-integer non-linear problem into an integer linear programming (ILP) model and reduces the computation and communication time during robotic application execution. Similarly, [142] allocates resources with minimal latency. However, for complicated optimization problem, it is not always feasible to find the solution satisfying all the constraints, especially when the resource allocation is required to conduct in real-time.

2) Approximation method: Approximation methods help to find a solution of any complex optimization problem within a reasonable time frame [189]. In literature, heuristic, greedy, fuzzy, dynamic programming are used for solving resource allocation problem in multi-agent cloud robotics. For example, a heuristic approach that allocates computational tasks among the cloud and the local robots and minimizes execution time and cost is applied in [141]. The approach is evaluated by simulations and the results show that it helps to select computing resources efficiently and reduces the time and cost by 77% for using the local resources only. While allocating the computational resources, [111] also uses a heuristic algorithm and finds a near-optimal value for time and energy consumption. This algorithm is evaluated by simulations and the results depict that it saves 99.3% time, reduces 23.8% energy consumption, and satisfy QoS. Greedy algorithm is another kind of heuristic that takes the best immediate or local solution of an optimization problem and subsequently improves it to find a global optimum. In the literature, [110] follows greedy algorithm as heuristic for computational resource allocation. The simulation results show that it provides 22% better near-optimal solution. Furthermore, [165] uses max-heap as a heuristic for storage allocation in multi-agent cloud robotics and minimizes the cost of servers. Similarly, [156] offers fuzzy logic-based heuristic for bandwidth allocation in multi-agent cloud robotics. Although heuristic approaches assist to find the optimal or near optimal solution in faster way, they are unable to provide the best optimal solution. Hence, it is only suitable in approximating the exact solution.

3) Evolutionary computing: It is one of the popular methods for solving NP-hard optimization problem which is also applicable to multi-agent cloud robotics. For example, in an autonomous oil factory maintenance application, to support the on-demand mobility and the path planning of the robots as well as the selection of access points for sending data to cloud simultaneously become complicated. This joint optimization problem can be easily solved with the help of evolutionary computing. In literature, a genetic algorithm (GA) based strategy is formulated for product transportation in cloud-based manufacturing that optimizes the computational load of robots, overall cost and processing time concurrently [143]. The simulation results depict that this strategy reduces computation time and cost up to 10%. Another GA based resource deployment method is used in [144] that organizes multiple resources for exchanging messages and executes robotic applications meeting their QoS. This method is evaluated by both simulations and experiments on real testbed. The results show that the solution is able to satisfy the QoS requirements of 98% tasks. As GA based approaches are based on random initial population, they often fail to reach the most optimal solution satisfying all objectives and constraints. Therefore, non-dominated sorting genetic algorithm-II (NSGA-II) [190] is adopted in [10], [76] for optimizing multiple resource
allocation objectives simultaneously. The energy consumption is optimized by \[76\] while executing the tasks of robotic applications within its deadline. According to simulation results, this solution reduces the task execution time by 15% and the energy consumption of resources by 10%. In \[10\], this solution is further improved by considering application deadline, energy consumption of resources and the user expenses concurrently. Simulations results demonstrate up to 18% improvement in optimizing time, energy, and cost. Both solutions initially select better population for evolution so that the stopping criteria is met in quicker time. Basically, evolutionary computing methods always follow a population-based trial and error approach to find a solution, which sometimes becomes time consuming and infeasible for real-time applications.

4) Game theory: This approach assists in finding an optimal outcome from a set of choices by analysing the costs and benefits of each independent participants within a system, especially in a competitive scenario \[191\]. In multi-agent cloud robotics, sometimes it may happen that multiple applications compete to use the same set of available resources. For example, UAVs are required in a smart farm for crop health monitoring, pesticide control as well as livestock management applications which arises a competition among the applications to occupy the resources. Therefore, to solve this kind of issue, some works in the literature apply game theoretic approaches. For example, a bandwidth allocation strategy using Stackelberg game is described in \[155\] and the performance of the strategy is evaluated through simulations. In addition, \[155\] allocates bandwidth in multi-agent cloud robotics with a task-oriented pricing mechanism. Experimental results accumulated from testbed and simulations show that the time and cost increase with number of network nodes. For bandwidth allocation, auction mechanism is also investigated in \[157\]. Both simulation and real testbed results depict that the mechanism is efficient to increase bandwidth usage by 35.9%. K-means clustering, and auction-based game models are used in \[140\] to assign robots and allocate resources for search and rescue operations. The evaluation study is conducted by means of simulations as well as testbed, and the results illustrate that the speed and scalability of the system increases. Generally, by identifying either Nash equilibrium or the best response, or the pareto efficiency, game theory approaches solve complex problems. However, these approaches assume that the agents or players make rational decisions at all the times, which is not obvious in many cases. In multi-agent cloud robotic systems, it is not always possible to interact with the players consistently as the communication is a constraint in such environment. Considering the system environments and application types, it is also challenging to define and solve a game theoretic model of resource allocation problem in real-time.

5) Deep learning: The formal optimization and approximation method including heuristics, meta-heuristics such as evolutionary approaches are computationally expensive to be run on local robots or edge resources. In addition, game theoretic approaches are not always suitable due to the heterogeneity of resources and the communication overhead. To overcome these issues, machine learning (ML) based approaches are also explored in the literature. When the optimization problem becomes complex and the exploitation of contextual information is important to make resource allocation decisions, ML approach with deep learning (DL) can be a viable solution. However, for training and testing, a large amount of data is required in DL approach which is not always available in a dynamic environment with limited storage. Therefore, as a combination of ML and DL, deep reinforcement learning is becoming a promising technique for multi-agent cloud robotics to allocate resources for the robotic tasks with discrete and continuous state and action space. For example, in a smart factory, sometimes robots may require computation support, sometimes may require data offloading support and in some cases may require storage support. Depending on the dynamic environment, heterogeneous types of services are requested in the system. To deal with this uncertain demand, learning based resource allocation mechanism will be an attainable solution. Reinforcement learning based mechanism for computational resource allocation is used in \[112\] that optimizes the energy consumption of resources and profit of service providers. The simulation results illustrate that RL scheme performs better than GA under the condition of limited cloud computing resources. Nevertheless, this approach becomes infeasible when the reward function or the objective function is ambiguously defined without assessing the system environment and other constraints.

In recent years multi-agent cloud robotics has become one of the major fields of interest for both academia and industries. Table \[IV\] highlights a brief summary of the reviewed papers in respect of resource allocation. Although many important aspects are identified in the existing literature, there exist some other issues that are required to be addressed for further improvement in this domain. Since multi-agent cloud robotic systems are comprised of heterogeneous devices with different capabilities, the allocation of resources needs to be efficient enough to satisfy the tasks requirements. For mission critical application such as fire-fighting or robotic surgery, low latency communication and reliable connectivity is required. Thus, reliable, and delay-aware resource allocation technique is necessary while selecting allocation mechanism. Since the resources are energy constrained, energy-efficient resource allocation policies are essential to meet the QoS in dynamic network environments. Additionally, the performance of most of the resource allocation policies are evaluated through simulations. Only a few works perform experiments on real-world testbeds. Moreover, it is evident that real-world robotic applications possess an inherent need for faster processing. Optimal allocation of resources and efficient service provisioning for the robotic systems will boost the successful execution of these applications. In the following section, different aspects of service provisioning such as appropriate resource pooling, proper task selection for offloading and scheduling the offloaded tasks are discussed in detail.

IV. SERVICE ProvisionING IN MULTI-AGENT Cloud Robotic Systems

To efficiently allocate resources for the tasks of robotic applications, a multi-robotic system and the resource manager of edge and cloud environments work collaboratively as shown in Fig. \[3\] This process initiates from the multi-robotic system by requesting the resource manager to discover resources for the robotic applications. While making such requests, the multi-robotic system also forwards the application profile such as their resource requirements and dependencies to the resource manager. Based on the application profile, the resource manager provisions resources for the multi-robotic system. Resource pooling includes the activation or reservation of the resources and the creation of resource pool. However, after provisioning, the resource specifications and configurations are forwarded to the multi-robotic system. Based on the resource specification, the multi-robotic system decides which task of the robotic application is required to offload. A resource manager can receive a huge number of offloaded tasks from different multi-robotic systems. After receiving the offloaded tasks, the resource
manager defines the scheduling order of the tasks based on their QoS requirement such as the deadline constraints. In parallel, the resource manager allocates resources to the tasks according to their availability. Depending on such interactions between multi-robot system and resource manager, the pooling of resources and the offloading and scheduling of the tasks are regarded as the pre-processing operations for the resource allocation. In this work, these pre-processing operations for resource allocation are collectively termed as service provisioning.

A. Resource Pooling

Resource pooling defines the arrangement of the resources for task processing and storing data to provide services to the users. For realizing resource pooling, a service provider at first needs to locate and retrieve the existing resources across multiple infrastructure domains in multi-agent cloud robotics, which is commonly known as resource discovery. After discovering, resource scaling through aggregation, amalgamation and shifting is required to make the pool of resources compatible for hosting robotic applications. However, for creating any resource pool, two major research questions are required to be addressed. They are discussed below.

- **How secured the resources are?:** While selecting resources from local, edge and cloud infrastructure to create resource pool, the security of the resources needs to be ensured. The security issues can be classified with multiple aspects. Firstly, it is essential to protect the physical resources from malicious attacks. While sharing data with other robots, edge resources and cloud data centre, unauthorized access to data needs to be prevented by strong integrity and confidentiality protection [192]. It is necessary to ensure that the robots only interact with legitimate servers and reliable devices. Reversely, with the help of authentication scheme edge or cloud data centre will only establish communication link with authorized robots and reject illegal communication requests. Privacy preservation is another fundamental requirement for data transmission among robots, robot-to-edge/cloud, and edge-to-cloud. Robust encryption mechanism is a viable solution to maintain data privacy. In multi-agent cloud robotics, for service provisioning, wireless access services from heterogeneous networks from multiple service providers may be required. Consequently, communication and transmission process become riskier for their inherent security vulnerabilities. Therefore, while creating the resource pool, trust-worthy resource selection is mandatory for efficient service provisioning. In addition, efficient security measures are required to prevent vulnerabilities both at resource level and communication level. Different mechanisms such as trust establishment, reputation-based trust, trust measurement techniques [2], [100], [193]–[195] have been well studied in literature to provide trust-worthy resources. Recently, blockchain technology has become popular to address the security and privacy concerns in multi-agent cloud robotics [196]–[200].

- **How efficiently resources can interact?:** While creating the resource pool in multi-agent cloud robotics, the selection of communication networks and protocols play an important role for efficient data exchange among resources and to maintain the QoS of the applications. Existing networking protocols have been widely used for wired or wireless communications in multi-agent cloud robotic systems [2]. Nevertheless, the efficiency of such communication depends on the working environment and application scenario. Consequently, depending on the contexts, researchers have adopted miscellaneous communication protocols to reduce data transmission delay and offer better QoS. Robots can communicate among themselves and with cloud using middleware like ROS [114], [120], [123], [128], [133], [138], [201]–[205]. However, this type of middleware cannot always support secured communication. Another key challenge for communication in a dynamic environment is to optimize route discovery and route maintenance with minimum computation time and resource requirement. Gossip protocol is recommended in [2] for robot-to-robot and robot-to-cloud communications that are particularly suited for mobile robots. Data stream network (DSN) and controller area network (CAN) are adopted for real-time communication in [22] and [132], respectively. Micro Internet protocol (uIP) is adopted in [131] to support simplified communication between robot and cloud for data transmission. TCP or IP socket based communication also benefits robots for faster communicate with cloud [125], [126]. However, for delay-sensitive applications, it is always time consuming to send data to cloud rather than communicating with edge resources. Using short-range wireless communications technologies such as Zigbee, Bluetooth or Wi-Fi direct, robots get services from the edge

---

**Fig. 3**: Interaction between multi-robot system and resource manager during resource allocation and service provisioning.
resources \[134\], \[145\]. A heartbeat protocol is implemented in \[206\] to deal with network latency of multi-agent cloud robotics. Network bandwidth usage can be substantially preserved by processing the data on edge resources \[173\], \[207\]. Packet delivery failure and communication outage are inherent drawbacks of wireless communications \[2\]. Therefore, backup mechanisms are required, and the system needs to be robust enough to recover from unwanted events.

1) Resource Discovery: Resource discovery indicates the arrangement of the resources for task processing and storing data. The amount of resource required for a robotic system is estimated according to the resource availability and the characteristics of the system in multi-agent cloud robotics. Service providers apply different resource discovery policies to meet the dynamic and static demands of robotic systems. Depending on the characteristics of the demands two basic type of resource arrangements such as on-demand and reserved are made to find the available resource pool from the cloud, edge, or local resources.

- **On-demand arrangement:** On-demand resource discovery allows the robotic systems to access available resources on-the-fly when they are needed. This arrangement is selected for the execution of robotic applications where the workload varies or changes very frequently. To deal with the dynamic workload, on-demand approach is exploited in \[110\], \[112\], \[141\], \[156\], \[157\], \[165\]. Moreover, a robotic system is billed on pay-per-use basis to utilize edge or cloud resources by the service provider. The idle or unused resources can also be provisioned for the robotic systems as spot-instances. The price of these instances changes periodically, depending on the supply and demand of the spot instances among multiple robotic systems. To get access to the spot-instances, a particular system is required to bid the current price. When multiple robotic systems target for the same service or the same resources from cloud data centre or edge server, spot-instances help a particular system to execute their tasks immediately. However, spot-instance discovery is yet to be explored for multi-agent cloud robotic systems.

- **Reserved arrangement:** Robotic systems arrange resources prior to executing any application rather than sending resource request on-demand to the service provider. The robotic systems harness the resources for a fixed contract period with a static price. In this stable resource arrangement, robotic system negotiates with the service provider for a particular service and the provider provisions resources in advance for that service. Basically, reserved provisioning is applicable for the robotic applications that have predictable and unchanging workload. In literature, considering the static workload, \[10\], \[76\], \[111\], \[140\], \[142\], \[143\], \[154\], \[155\] prefer reserved arrangements of resources.

Additionally, efficient resource discovery helps in enhancing the competency of edge or cloud based robotic services.

2) Resource Scaling: In multi-agent cloud robotics, the scattered and isolated geo-distribution of discovered resources makes the creation of resource pool challenging. In this case, resource scaling is required to enhance the suitability of the resources in provisioning the services. Different techniques for resource scaling including resource aggregation, resource amalgamation, and resource shift are applicable in multi-agent cloud robotics.

- **Resource aggregation:** It means combining the same type of resources to create resource pool with increased power \[208\]. It is a commonly used technique to prepare resource for service provisioning in multi-agent cloud robotics. For example, memory from robot, edge and cloud storage can be aggregated together to store large amount of data while dealing with big data in CPSs. Similarly, for computation intensive tasks, CPUs from multiple resources are augmented to create a more powerful processor. However, resource aggregation only supports the same type of resource augmentation, which is not always capable of meeting the resource demand.

- **Resource amalgamation:** It aims at combining individual capabilities of resources to offer greater dimension of resources \[208\]. For example, while monitoring a field, the images from the on-board cameras of different robots can be combined to create an image with higher resolution and using edge or cloud resources with higher processing capabilities 3-D effects can be added to get more clearer view. During resource amalgamation, as multiple devices participate in sharing resources, the management of their heterogeneity and cross platform operations adds further overhead to the system.

- **Resource shift:** It refers to moving resources from one device to another device \[208\], e.g., migrating VM from one cloud server to another server for executing robotic applications. Even the connectivity can also be shared by making additional access links available to any poorly connected robot or edge resource. However, while shifting resources, the communication time, and the compatibility of the destination device to host the robotics applications, also need to be taken into account.

To summarize, efficient resource pooling is a basic requirement in multi-agent cloud robotics. The security, accessibility, availability, and compatibility of the resources need to be assessed while creating the resource pool as it sets the foundation for further service provisioning operation including computation offloading and task scheduling.

B. Computation Offloading

As local robots in multi-agent cloud robotic systems have limitations in computational power, storage, and energy, it is often required to move the compute-intensive tasks to resource enriched cloud or edge instances for successful robotic application execution. The idea of computation offloading from local robots to cloud resources is derived from mobile cloud computing \[209\]. However, the service orchestration, application requirements and resource orientation differ from mobile cloud computing to multi-agent cloud robotics. More specifically, the key distinction between two paradigms is the robot’s unique ability to move on-demand, which allows them to actively access better communication links for offloading. Therefore, it is infeasible to directly apply the mobile cloud-based offloading techniques to a multi-agent cloud robotic system. Additionally, a robot determines its position in respect of other robots which facilitates in local offloading as well as edge or
### TABLE V: Comparison among computation offloading schemes in multi-agent cloud robotics

| Criteria                  | Scheme       | Pros                                                                 | Cons                                                                 | Open Issues                                                                 |
|---------------------------|--------------|----------------------------------------------------------------------|----------------------------------------------------------------------|-----------------------------------------------------------------------------|
| Time of decision making   | Re-active    | • Offers urgent action                                               | • Effective solution for shorter period                             | • Balance between pro-active and re-active decision making                  |
|                           | Pro-active   | • Provides long term solution                                         | • Stringent deadline to make decision                              | • Selection of appropriate algorithm                                        |
|                           | Full         | • Better energy optimization                                         | • Computation overhead                                              |                                                                             |
|                           | Full         | • Higher data dependency                                             | • Requires validation prior to setting                               |                                                                             |
|                           | Full         | • Higher bandwidth requirement                                        | • Trade-off between computation and communication                    |                                                                             |
|                           | Partial      | • Synchronized decision                                              | • Vulnerability on single point of failure                          |                                                                             |
|                           | Multiple     | • Security vulnerability                                              | • Burden on single resource                                         |                                                                             |

| Offloading type           | Partial      | • Supports parallel processing with load distribution                | • Time consuming for task partition and result aggregation           |                                                                             |
|                           | Full         | • Easier decision-making                                             | • Higher bandwidth requirement                                     |                                                                             |
|                           | Full         | • Less operational latency                                           | • Burden on single resource                                         |                                                                             |
|                           | Full         | • Simple decision making                                             | • Synchronization of decisions                                     |                                                                             |
|                           | Multiple     | • Prolong network lifetime                                           | • Security vulnerability                                            |                                                                             |
|                           | Multiple     | • Better energy optimization                                         | • Burden on single resource                                         |                                                                             |
|                           | Multiple     | • Synchronization of decisions                                       | • Burden on single resource                                         |                                                                             |
|                           | Multiple     | • Security vulnerability                                              | • Burden on single resource                                         |                                                                             |

| Number of decision maker  | Single       | • Less communication overhead                                        | • Lack of system overview                                           |                                                                             |
|                           | Single       | • Simple decision making                                             | • Burden on single resource                                         |                                                                             |
|                           | Multiple     | • Prolong network lifetime                                           | • Optimal selection of robots                                        |                                                                             |
|                           | Multiple     | • Better energy optimization                                         | • Synchronized decision making                                       |                                                                             |

Cloud-based offloading. In multi-agent cloud robotic system, local robot itself can actively participate as a resource provider to support compute-intensive tasks execution offloaded by other robots. Thus, it is important to investigate the role of mobility and communication while offloading a task in multi-agent cloud robotics. On the other hand, it is difficult to make optimized offloading decisions due to the delay constraint, mobility of robots, and additional data transfer and computation cost. A trade-off among these parameters ensure the improved performance of robotic systems. For efficient computation offloading in multi-agent cloud robotics, two important questions need to be addressed. They are discussed below.

1) **How offloading decision should be made?** The offloading performance largely depends on how the offloading decision is made. This decision can be made either reactive or proactive, based on the time when the decision is made.

   - **Re-active scheme:** In this case, the offloading decision is taken after a situation has arisen where offloading is necessary. In such a situation, a quick or immediate decision making is required based on system demand.
   - **Pro-active scheme:** This is a predictive scheme which anticipates computation requirements and takes a rational decision before any event has occurred. Using predictive analytics, the system context, tasks, and resource requirements are envisaged before actual offloading request is generated.

   With re-active decision making, the deadline to make the decision can be very stringent, which can obstruct to make an effective decision. On the other hand, despite the consumption of additional time for decision making, with the help of predictive analytics, pro-active offloading gives a long-term solution to the system at the cost of additional computation overhead. Therefore, a balance between pro-active and re-active decision making is necessary. Selection of appropriate algorithm to implement this decision making is required for better performance.

2) **How actual offloading should be conducted?** The performance of offloading techniques is also affected by the number of offloaded tasks. It is the responsibility of robots to conduct either full or partial offloading based on the assessment of network profile (bandwidth, access point), device profile (battery status, local storage), and system objectives (minimize cost, distance, time, energy).

   - **Partial offloading:** In this offloading approach, rather than forwarding the whole application, a part of the application is transferred from local robot to remote instances for execution. The remaining parts are either managed by local robots or sent to other computing infrastructure like edge.
   - **Full offloading:** This is a conventional approach for offloading in multi-agent cloud robotics, where the application is transferred completely to the remote cloud for execution and after execution the results are sent back to local robots. It increases the burden on communication link by sending a large volume of data along with the application to the remote destination.

Partial offloading supports parallel processing and consequently allows load distribution among the resources. The waiting time to process the data on a single resource is also reduced. However, data dependency among the resources becomes higher as tasks are partitioned among multiple resources. It consumes additional time for task partition and result aggregation from different resources. On the contrary, in full offloading, since the entire computation is offloaded, the burden of task partitioning diminishes and the decision making becomes easier. Nevertheless, it incurs higher bandwidth to send the full task and consumes higher energy of single resource. Therefore, a trade-off exists between computation and communication in offloading decision making. In addition, energy-delay optimization is also required for making efficient offloading decision.

Another crucial criterion of offloading is to distribute the offloaded tasks among the robots. It basically depends on the available local resources (robots) and their orientation. Regarding this feature, the offloading decision-making factors are classified as follows:

   - **Single robot:** For a single robot, the offloading decision depends on the location and the appropriate offloading approach that ensures optimal transfer of information between robot and the computing instances.
   - **Multiple robot:** In this approach, multiple robots work collaboratively to complete the offloading operation. It requires to make a balance between robot-to-robot coordination and robot-to-cloud or edge communication.

Using a single robot, the offloading decision becomes simple and requires less communication overhead. However, for a single robot, it is not always possible to get the complete overview of the system and make efficient decision to maintain the QoS. In addition, the failure of a single robot can affect the full system. On the other side, offloading decision with multiple robots offers prolonged network lifetime and better energy optimization. Yet, the synchronization of the decisions from multiple robots becomes challenging and the system becomes vulnerable to security issues due to higher exposure of data. Therefore, optimal selection of robots is essential in making offloading decision.

To sum up, the main challenge in computation offloading is to trade-off between the local and the remote computation and communication cost in multi-agent cloud robotics. It is also challenging to efficiently offload a task and receive the response, since the robots can move on-demand to multiple locations during the application execution period. Since the offloading performance depends on...
the network resources, it is required for the robots to make a balance between mobility and bandwidth utilization. Finally, the identification of an appropriate approach to perform the offloading based on the arbitrary system context requires further research. The pros and cons of different offloading schemes are summarized in Table VI.

In the context of multi-agent cloud robotics, computation offloading is considered as a trending topic. Therefore, over the last few years, plenteous research has been conducted in this area. Initial studies such as [113], [135], [219], [220] give particular attention to the infrastructure and framework for offloading in multi-agent cloud robotics. Platforms to support offloading the compute-intensive tasks from the robotic network to cloud and edge resources are also furnished in [114], [136]. Subsequently, other studies including [143], [221] exploit the optimal approach for efficient task offloading between local robot and remote computing instances. Additionally, task partitioning has been studied and established in [222], [223], where the applications are decomposed in tasks and distributed among the robots and cloud resources. Furthermore, based on each decision-making criterion, the offloading performance can vary significantly. To deal with such cases, different offloading approaches according to the resource architecture, application profile and performance parameters are also accentuated in the literature. Table VI presents the summary of some notable works in computation offloading for multi-agent cloud robotics that refer the aforementioned attributes in detail. As seen in Table VI, most of the existing approaches perform full application offloading to cloud, whereas very few including [114], [210], [215], [218] apply partial offloading. Another variation is seen in the number of engaged robots for offloading operations. For example, in [177], the authors consider single robot methods.

In addition, most of the offloading algorithms follow heuristic approach ([111], [210]) and evolutionary method ([4], [213], [216]), whereas others adapt greedy algorithms ([40], [123]), game theory ([214], [217]) and dynamic programming ([212], [224]) to make offloading decisions. There also exist differences among the existing works in selecting the offloading objectives. Minimization of energy ([4], [141], [210], [218]), latency ([111], [214]), distance or movement cost ([123], [216]) and bandwidth or communication cost ([114], [212], [217]) are considered as the potential offloading objectives.

With the help of offloading, the management of resources become distributed among the local robots, edge and cloud. Based on the features of offloaded tasks, resources are usually allocated from the available resource pool. If the tasks to be offloaded are selected appropriately by the robots, task scheduling and service provisioning can be conducted efficiently at the edge or cloud. Nevertheless, if the resource of each cloud or edge server is limited, then offloading becomes inefficient. It turns into worst when all robots select the same server as the destination. Hence, a detailed study is required so that the offloading techniques in multi-agent cloud robotics can deal with such cases deliberately and improve the application QoS.

C. Task Scheduling

Multi-agent cloud robotics enables the local robots to outsource their computing tasks and storage requirements to the edge or cloud and ensures enhanced capacities and higher performance for the robotic systems. One of the important research issues is to determine how resource providers can efficiently handle the overwhelming requests from different robotic systems when the amount of resource is limited. With the help of resource discovery and resource scaling, the resource pool is identified for the robotic systems while scheduling determines in which order the robotic tasks will arrive and leave the resources. A task scheduling approach becomes efficient when it avoids longer scheduling delay and waiting time for the requested services. It is also liable for arranging incoming service requests in a certain manner so that the available resources are properly utilized. To ensure the highest utilization of available resources, the tasks should be dispatched to the resources reasonably. Usually, task scheduling is conducted after a fixed period or dynamically according to task arrival rate, availability of resources and workload on the resources.

1) Static scheduling: In this approach, tasks are scheduled after a specific time interval, where communication cost and computation cost of each task are estimated prior to its assignment on the edge or cloud. Then, these costs are compared with the cost of each task on different resources. In the literature, [10], [76], [111], [140], [142], [143], [154] follow this task scheduling strategy.

| Work | Offloading Type | Decision Maker | Methodology | Objective | Evaluation Method | Results | Use Case |
|------|----------------|----------------|-------------|-----------|------------------|---------|----------|
| [210]| Partial        | Single         | Multi-level decision using heuristic | Minimize time and energy | Testbed | Saves 0.052s and 19.63f than local execution. | Object recognition and motion control. |
| [213]| Full           | Multiple       | Greedy approach | Minimize path cost | Testbed | Find the shortest path faster than local robot. | Path planning. |
| [114]| Partial        | Multiple       | Cloud based task execution | Minimize network latency | Prototype | Only 1% false positive result. | Object recognition and grasping. |
| [211]| Partial        | Multiple       | Full path planning using heuristic | Minimize time | Simulation | Offers 90% better result. | Path planning in smart home and healthcare. |
| [212]| Full           | Single         | Dynamic parallel algorithm | Minimize communication cost | Prototype | Performs better than only robotic execution. | Vision based navigation assistant. |
| [213]| Partial/Full   | Single         | Genetic algorithm | Minimize energy and network lifetime | Theoretical analysis and simulation | Lifetime of the network is prolonged | Generalized robotic applications. |
| [40] | Full           | Multiple       | Context-aware greedy approach | Minimize energy and cost | Simulation. | Context awareness performs better in material handling | Smart factory. |
| [214]| Full           | Multiple       | Market-based management strategy | Minimize time | Testbed, Simulation | Significant improvement in bandwidth usage and load balancing. | Generalized real-time robotic tasks. |
| [215]| Partial        | Multiple       | QoS-aware game theory | Minimize latency and energy | Simulation | Stable performance gain with increasing number of robots. | Generalized robotic applications. |
| [4]  | Partial/Full   | Single         | Genetic algorithm | Minimize energy, time and distance | Simulation | Communication and mobility-aware offloading provide better performance. | Smart factory. |
| [7]  | Full           | Multiple       | Genetic algorithm | Minimize overall system energy | Simulation | Reduce energy consumption. | Smart factory. |
| [217]| Full           | Multiple       | Deep RL based off-loading | Minimize communication cost | Simulation | Supports offloading for mobile robots. | Generalized robotic applications. |
| [218]| Partial        | Multiple       | Hierarchical iterative approach | Minimize energy, computation time | Theoretical analysis and simulation | Effectively reduce the energy consumption. | SLAM. |
2) Dynamic scheduling: According to this approach, task scheduling is conducted randomly without any prior time or cost estimation. In literature, [110], [112], [141], [155]–[157], [165] schedule the tasks dynamically.

The efficient resource utilization of a robotic system depends more on the scheduling and load balancing methodologies than the random allocation of resources. For instance, to share resources among cloud-aided industrial robots in a smart factory, dynamic scheduling of the tasks and computational load balancing are equally important [31], [42]. In literature, very few works are concerned about the scheduling policies in multi-agent cloud robotics. Moreover, due to lack of holistic monitoring of the resources, scheduling techniques are tedious to implement in a multi-agent cloud robotic system. In future, appropriate selection of task scheduling strategies for successful service provisioning to the robotic systems should be explored extensively. In the following sections, the gaps of existing literature are explicitly identified from the lessons learned and multidisciplinary future research directions are discussed for further improvement of this domain.

V. LESSONS LEARNED

In this paper, multi-agent cloud robotics is reviewed from resource allocation and service provisioning perspectives and the lessons learned by this review are summarized in this section.

1) It is clear that multi-agent cloud robotics enhance the performance of associate CPSs by enabling the execution of robotic applications in different infrastructure levels. These robotic applications are composed of various tasks. On the other hand, the computing resources within a multi-agent cloud robotic system are not equally competent to execute all types of tasks due to their heterogeneity. For example, the latency-sensitive tasks are preferable to execute at the robots or edge computing infrastructure whereas the compute and storage-intensive tasks require forwarding to cloud infrastructure. In such cases, the partitioning of robotic applications based on the characterististics of the tasks can be a viable solution, although it incurs additional management overhead in aggregating the outcome of the tasks at the receiver end.

2) A significant amount of energy is required to run the robots and edge resources. Moreover, the arrangement of grid-based energy is costly when the robotic system works in remote places. To deal with such constraints different policies have already been developed to optimize the energy consumption in multi-agent cloud robotics. However, such reactive approaches are not feasible when the computing environment and corresponding resources depend on the renewable energy. Since the availability of renewable energy is subjected to the context of physical environment, the proactive resource allocation and service provisioning based on such uncertain parameter also becomes very complicated.

3) Mobility is one of the definitive feature of multi-agent cloud robotics. However, the mobility pattern of robots varies from one robot to another. For example, an UAV changes its location more frequently than a ground robot. In such cases, if the mobility-driven service provisioning requests are initiated from the robots, it will add a significant communication and data management overhead to the system which will be quite tedious to deal with at the run-time. A centralized approach to move the computation as per the location of robots can be an alternative solution. However, it requires a complete understanding of the network. Moreover, the monitoring of each robot in an individual manner can be challenging task when the mobility is uneven. Therefore, dynamic switching between these two solutions is highly preferable to ensure efficient mobility driven resource allocation and service provisioning.

4) As noted, a multi-agent cloud robotic system is a combination of robotic systems, edge and cloud computing. Most likely, these infrastructures have their own resource manager that controls the internal functions. The co-existence of multiple resource manager incurs further synchronization problem. Through deliberate negotiation, this problem can be solved to a great extent. However, it limits the scope of making independent decisions for autonomic resource management. Therefore, a fair and adjustable distribution of resource management responsibilities in multi-agent cloud robotics is highly recommended.

5) The computing infrastructures and execution platforms in multi-agent cloud robotics are highly exposed at different communication layers including edge-network layer, core-network layer and cloud communication layer. Consequently, it broadens the possibility of various security threats including data leakage and tempering. Since multi-agent cloud robotics deal with real-time use cases, heavy weight security features often slow down the interactions among different entities within the computing environment. Therefore, it is preferable to develop lightweight security features. However, it has a limitation in guaranteeing scalable solutions which is subjected to the extensive research.

6) Only a handful of initiatives have been taken to realize the economic potential of multi agent cloud robotics. Since a universally accredited business model is yet to be developed, it is very difficult to make a comprehensive profit in multi agent cloud robotics satisfying the interest of all participating service providers. On different note, the rigid intention of profit enhancement sometimes urges to relax the important QoS parameters that degrades the trust between robotic system users and service providers significantly. Therefore, in most of cases, QoS requirements are given higher priority while provisioning services and allocating resources. It not only reduces the SLA violations but also resists the relinquish rate. Consequently, it improves the economic benefits of the service providers in multi-agent cloud robotics.

7) There exist different frameworks to solve the resource allocation and service provisioning problem in multi-agent cloud robotics. Most of the frameworks conduct limited evaluation and hardly ensure low latency data-flow between robot and cloud. Furthermore, these frameworks are application-specific and their in-built software systems are not always adaptable to decentralized resource architecture and they often fail to deal with the real-world environmental constraints during robot to cloud interaction. Additionally, most of the existing works focus on computation offloading to the cloud as a part of the service provisioning. The edge resources are not set as the destination for offloading the computation. Besides, the importance of resource provisioning plan and determination of task scheduling period are barely investigated in these works.

The aforementioned lessons learned from the literature review helps in identifying the research gaps in multi-agent cloud robotics. In the following section a holistic framework for resource allocation and service provisioning along with some future research directions are provided that can address these research gaps to a great extent.

VI. A HOLISTIC FRAMEWORK AND FUTURE RESEARCH DIRECTIONS

A holistic framework encapsulates different hardware and software components that simplify the end-to-end interaction among
the associated entities and facilitate the integration of various resource and service management policies in a scalable manner. Fig. 4 depicts such a framework for resource allocation and service provisioning in a multi-agent cloud robotic system. As already mentioned, in this environment, the computing platform for robotic applications is extended to multiple infrastructure levels. This framework exploits the edge infrastructure for latency-sensitive tasks as the edge computing nodes are in the proximity of robotic systems and their dynamic augmentation for on-demand resource allocation is quite easier through ad-hoc networking. Moreover, the framework prefers to schedule compute-intensive tasks to cloud infrastructure as the datacentres host powerful computing servers and the resources can be reserved for a certain period. However, the main entity of the proposed holistic framework is a Cross-infrastructure Resource Allocator and Service Provisioner (CRASP). CRASP is also composed of several components which are deployed in distributed manner across different infrastructure levels. CRASP is provided with a robust and reliable communication link that interconnects its components logically. The brief discussion of these components in different infrastructure level is discussed below.

A. Robotic System Level

In this infrastructure level, the Offloading Decision Maker of CRASP resides that analyses the possibility and determines the benefits of task offloading in multi-agent cloud robotics. To perform these operations, the Application Profiler assist the Offloading Decision Maker by providing meta-data regarding the application architecture and programming model and the Task Classifier checks whether that task is latency-sensitive or compute-intensive. However, prior to start task offloading, the Dispatcher and Negotiator connects the edge and cloud infrastructure and perceives the state of the computing platform. Later, based on the offloading decisions, the Dispatcher and Negotiator forwards the tasks and data to the corresponding infrastructure. Moreover, while the data are processing remotely, the Dispatcher and Negotiator also guides the Flow Controller to tune the input data transmission rate as per the context of the processing destinations.

B. Edge Infrastructure Level

In this level, a specialized Scheduler component of CRASP is placed that operates the execution of latency-sensitive tasks on edge computing nodes. Usually, these nodes interact with each other by forming ad-hoc clusters. The Resource Discoverer helps to identify a suitable node from such cluster to execute a task. The Network Monitor support this operation by updating the network status periodically. Based on the outcome of Resource Discoverer, the Resource Allocator assign the task to the selected edge node. However, in most of the cases, latency-sensitive tasks originate from even-driven physical actions. Therefore, CRASP consider the allocation of resources at the edge infrastructure as an on-demand operation. Additionally, during uncertain scenarios such as node failure, power outage and resource shortage, the scheduled tasks to the edge infrastructure are required to be forwarded to the cloud. In such contexts, the Resource Discover directly communicates with the cloud-based resource allocator of CRASP and solve the issue.

C. Cloud Infrastructure Level

In this level, CRASP focuses on scheduling compute-intensive tasks. Since the compute-intensive tasks are expected to have
a longer period of execution time, the Resource Allocator of CRASP at this level prefers to reserve the resources rather than dynamically provisioning them. However, to perform this operation, the Resource Allocator interacts with the Resource Pool that contains the references of all computing resources within the cloud infrastructure. After the allocation, the Resource Allocator grasp the status of task execution time-to-time with the help of Resource Monitor.

However, there exist extensive research scopes to improve this framework which are discussed in the following subsections.

1) Event-driven resource allocation and service provisioning: The execution of an application for a robotic system can trigger the execution of another application. For example, a robotic livestock monitoring application in a smart farm can trigger a robotic emergency management application. In this case, both the applications should be executed simultaneously. However, such arrangement often gets disrupted due to the resource constraints of multi-agent cloud robotics. Therefore, efficient resource allocation and service provisioning policies are required to deal with the event-driven requirements of robotic systems.

2) Energy-efficient resource consolidation and scaling: Energy is one of the major concerns for any computing paradigms, especially when it is accumulated from renewable sources. Since the availability of renewable energy is subjected to the environmental contexts, the computing infrastructures are required to be adaptable to their sudden changes. In multi-agent cloud robotics, it can be attained by consolidating the resources when the supply of renewable energy is poor and scaling up the resources when the opposite happens. However, it is not such straightforward. Dynamic consolidation and scaling of resources in multi-agent cloud robotics alter the network topology and incurs additional resource management overhead. Therefore, the resource allocation and service provisioning policies require to observe these issues deliberately which demand extensive research.

3) Balance between pro-active and re-active offloading decision: In multi-agent cloud robotics, mobility of robots is one of the key factors that need to be considered for making the computation offloading decision. For static mobility pattern of the robots, proactive offloading decision provides feasible solution as it helps anticipating system behaviour more accurately. However, in most of the cases, it is not obvious that robots will maintain a static pattern. To deal with such cases, re-active offloading decision making is required, despite of their effectiveness for shorter period. To deal with the dynamics of multi-agent cloud robotics, a balance between pro-active and re-active decision making is required so that with the help of predictive analytics the system can react in uncertain events with higher accuracy for a longer period.

4) Allocation of network slices for 5G enabled multi-agent cloud robotics: 5G cellular communication has already created a significant buzz in both industry and academia. Unlike traditional cellular communications, the physical network in a 5G system is virtualized in multiple slices. These slices are used to transmit data traffic for different applications. The amount of network bandwidth to be allocated to each network slice depends on the priority of the applications. However, in multi-agent cloud robotics, the level of necessity for an application can change very frequently. For example, in a smart factory when an anomaly happens, the robotic application investigating the location of the fault runs in high priority. Soon after identifying the fault location, the emergency management application gets the higher priority of execution. Therefore, the resource allocation and service provisioning policies for multi-agent cloud robotics should be intelligent to make and tune prioritization of applications in run-time. Since such prioritization depends multiple parameters including the QoS requirements of the applications and user expectations, the policies require detailed exploitation of these parameters to ensure the air distribution of bandwidth on network slices. In future, this concept can be extended for 5G and beyond environment to support mission critical applications using Tactile Internet.

5) Mobility-as-a-Service: In a geographically large-scale robotic system, the intensity of network connectivity is not uniform at all the locations. For example, in a smart farm, the robots working far from the access points receive poor network signal strength compared to others. At the same time, the inconsiderate deployment of access points will increase the mutual interference. Therefore, it is required to orchestrate the access points on ad-hoc basis. To resolve this issue, the mobility of robots can also be used as a service. The robots especially the UAVs can act as portable access points or signal booster for the robots receiving poor networking signal. However, the existing works in the literature consider mobility as a constraint of the multi-agent cloud robotics and to enable the mobility-as-a-service in this domain, significant research effort is required.

6) Real-time resource augmentation: As noted, multi-agent cloud robotics incorporate resources from different service providers. Since there exists a black box interface between the resource management policies of these service providers, a significant number of administrative operations is required to perform during resource provisioning in multi-agent cloud robotics. These operations are often time consuming that resist the real-time interactions among the corresponding entities. Interoperable resource allocation and service provisioning policies can solve this issue to a great extent. However, such policies need to observe the individual interests of each providers which is subjected to detail research.

7) Pricing model for resource consumption: Since there exists no accredited business model for the consuming the resources in multi-agent cloud robotic systems, it is very difficult to boost the revenue of service providers and facilitate the incentives for the users. Additionally, it obstructs the scope of providing compensation for SLA violations. Therefore, a detailed pricing model for multi-agent cloud robotic systems is required. However, such pricing model is difficult to developed as the computing components within a multi-agent cloud robotic system are highly heterogeneous and their operational cost in per unit time very significantly. Extensive research towards this direction can be a significant contribution to the existing literature.

8) Lightweight security measures during data management: There are some robotic use cases where sensitive data are exchanged. For example, in robotic medical assistance, the electronic health report contains a significant amount of private information. Sometimes this information requires to be accessed by different professional and organization bodies including insurance, pharmacist, and doctors. In such cases, an easy access to data is necessary for making real-time decisions. At the same time, the data access should be made secure so that unauthorized modification can be prevented. However, existing security measures such as blockchain and 128-bit asymmetric key cryptography are highly computation-intensive that slows down the real-time interactions to a great extent while identifying the authorized data access and modification. Therefore, lightweight security measures are needed for multi-agent robotic systems that not only secure data management but also support real-time interactions.
We have presented a survey and the research outlook on resource allocation and service provisioning in multi-agent cloud robotics. The recent development and research on multi-agent cloud robotics both in academia and industry have been reviewed. As a prerequisite of efficient resource allocation and service provisioning, the concepts of resource pooling, computation offloading, and task scheduling have been discussed separately along with their challenges. In addition, the lessons learned from the survey have been summarized and a holistic framework for resource allocation and service provisioning in multi-agent cloud robotics has been presented along with several potential research directions. We believe that this comprehensive survey will serve as a useful reference and provide guidelines for further investigation and advancement of multi-agent cloud robotics.

ACKNOWLEDGMENT

This work is partially supported by Australian Research Council Discovery Project Grant DP190102828. The authors would also like to thank Data61, CSIRO, Australia for funding and supporting this work. In addition, the work of E. Hossain was supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC).

REFERENCES

[1] R. Rajkumar, I. Lee, L. Sha, and J. Stankovic, “Cyber-physical systems: the next computing revolution,” in Design Automation Conference. IEEE, 2010, pp. 731–736.

[2] G. Hu, W. P. Tay, and Y. Wen, “Cloud robotics: architecture, challenges and applications,” IEEE Network, vol. 26, no. 3, pp. 21–28, 2012.

[3] J. Wan, S. Tang, H. Yan, D. Li, S. Wang, and A. V. Vasilakos, “Cloud robotics: Current status and open issues,” IEEE Access, vol. 4, pp. 2797–2807, 2016.

[4] A. Rahman, J. Jin, A. Rahman, A. Cricenti, M. Afrin, and Y-N. Dong, “Energy-efficient optimal task offloading in cloud networked multi-robot systems,” Computer Networks, vol. 160, pp. 11 – 32, 2019.

[5] M. R. Mahmud, M. Afrin, M. A. Razzaza, M. M. Hassan, A. Aleleiwi, and M. Alrubai, “Maximizing quality of experience through context-aware mobile application scheduling in cloudlet infrastructure,” Software: Practice and Experience, vol. 46, no. 11, pp. 1525–1545, 2016.

[6] C. Mouradian, D. Nabousli, S. Yangu, R. H. Githo, M. J. Morrow, and P. A. Polakos, “A comprehensive survey on fog computing: State-of-the-art and research challenges,” IEEE Communications Surveys & Tutorials, vol. 20, no. 1, pp. 416–464, 2017.

[7] M. Shiraz, A. Gani, R. H. Khokhar, and R. Buyya, “A review on distributed application processing frameworks in smart mobile devices for mobile cloud computing,” IEEE Communications Surveys & Tutorials, vol. 15, no. 3, pp. 1294–1313, 2012.

[8] A. Rahman, J. Jin, A. Rahman, A. Cricenti, M. Afrin, and Y-N. Dong, “Robot and cloud-assisted industrial warehouses,” Computer Networks, vol. 130, pp. 158–168, 2017.

[9] T. Taleb, K. Samandis, B. Mada, H. Flinck, S. Dutta, and D. Sabella, “On multi-access edge computing: A survey of the emerging 5g network edge cloud architecture and orchestration,” IEEE Communications Surveys & Tutorials, vol. 19, no. 3, pp. 1657–1681, 2017.

[10] M. Othman, S. A. Madani, S. U. Khan et al., “A survey of mobile cloud computing application models,” IEEE Communications Surveys & Tutorials, vol. 16, no. 1, pp. 393–413, 2013.

[11] H. T. Dinh, C. Lee, D. Niyato, and P. Wang, “A survey of mobile cloud computing: architecture, applications, and approaches,” Wireless Communications and Mobile Computing, vol. 13, no. 18, pp. 1587–1611, 2013.

[12] T. Taleb, K. Samadis, B. Mada, H. Flinck, S. Dutta, and D. Sabella, “On multi-access edge computing: A survey of the emerging 5g network edge cloud architecture and orchestration,” IEEE Communications Surveys & Tutorials, vol. 19, no. 3, pp. 1657–1681, 2017.

[13] M. Othman, S. A. Madani, S. U. Khan et al., “A survey of mobile cloud computing application models,” IEEE Communications Surveys & Tutorials, vol. 16, no. 1, pp. 393–413, 2013.

[14] H. T. Dinh, C. Lee, D. Niyato, and P. Wang, “A survey of mobile cloud computing: architecture, applications, and approaches,” Wireless Communications and Mobile Computing, vol. 13, no. 18, pp. 1587–1611, 2013.

[15] M. Shiraz, A. Gani, R. H. Khokhar, and R. Buyya, “A review on distributed application processing frameworks in smart mobile devices for mobile cloud computing,” IEEE Communications Surveys & Tutorials, vol. 15, no. 3, pp. 1294–1313, 2012.

[16] T. Taleb, K. Samandis, B. Mada, H. Flinck, S. Dutta, and D. Sabella, “On multi-access edge computing: A survey of the emerging 5g network edge cloud architecture and orchestration,” IEEE Communications Surveys & Tutorials, vol. 19, no. 3, pp. 1657–1681, 2017.
C. Hong and D. Shi, “A cloud-based control system architecture for multi-UAV,” in *Proceedings of the 3rd International Conference on Robotics, Control and Automation*. ACM, 2018, pp. 25–30.

Y. Li, H. Wang, B. Ding, P. Shi, and X. Liu, “Toward QoS-aware cloud robotic framework and its implementation,” in *2016 Int'l IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/CloudCom/IoP/SmartWorld)*. IEEE, 2016, pp. 33–40.

R. Arunugam, V. R. Enti, L. Bingbing, W. Xiaojun, K. Baskaran, F. F. Kong, A. S. Kumar, K. D. Meng, and G. W. Kit, “Davinci: A cloud computing framework for service robots,” in *2010 IEEE international conference on robotics and automation*. IEEE, 2010, pp. 3084–3089.

B. Hu, H. Wang, P. Zhang, B. Ding, and H. Che, “Cloudoid: A cloud framework for transparent and QoS-aware robotic computation outsourcing,” in *2017 IEEE 10th International Conference on Cloud Computing (CLOUD)*, 2017, pp. 114–121.

J. Furukawa, K. Kamei, C. Sharma, T. Miyashita, and N. Hagita, “UNR-PF: An open-source platform for cloud networked robotic services,” in *2012 IEEE/SICE International Symposium on System Integration (SII)*. IEEE, 2012, pp. 945–950.

S. A. Miratabzadeh, N. Gallardo, N. Gamez, K. Haradi, A. R. Puthussery, R. Arumugam, V. R. Enti, and S. A. Miratabzadeh, “Fourth-person sensing for a service robot,” in *Towards Autonomous Robotic Systems (Autonomy and Artificial Robotics for Efficient, Fluent and Robust Human-Robot Interaction)*, pp. 4566.

F. Chen, L. Wang, W. Chung, C. Rhee, Y. Shim, H. Lee, and S. Park, “Door-opening control of a service robot using the multifingered robot hand,” *IEEE Transactions on Industrial Informatics*, vol. 17, no. 6, pp. 3374–3381, 2020.

A. Kehoe, A. Matsukawa, S. Candido, J. Kuffner, and K. Goldberg, “Cloud-based robot grasping with the google object recognition engine,” in *2013 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2013, pp. 4263–4270.

A. Bacci, A. Gallo, and G. Venti, “Cloud, fog, and dew robotics: architectures for next generation applications,” in *2019 7th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering (MobileCloud)*. IEEE, 2019, pp. 13–21.

I. Stoica, and K. Goldberg, “Fog robotics algorithms for distributed motion planning using lambdaserverless computing,” in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 4232–4238.

L. Wang and M. Q.-H. Meng, “A game theoretical bandwidth allocation mechanism for cloud robotics,” in *Proceedings of the 10th World Congress on Intelligent Control and Automation*. IEEE, 2012, pp. 3828–3833.

L. Wang, M. Liu, and M. Q.-H. Meng, “A pricing mechanism for task oriented resource allocation in cloud robotics.” Springer International Publishing, 2016, pp. 3–31.

L. Wang, M. Liu, and M. Q.-H. Meng, “A hierarchical auction-based mechanism for real-time resource allocation in cloud robotic systems,” *IEEE Transactions on Cybernetics*, vol. 47, no. 2, pp. 473–484, 2016.

S. H. H. Malini, M. S. A. Latiff, Y. Coulibaly et al., “Recent advancements in resource allocation techniques for cloud computing environment: a systematic review,” *Cluster Computing*, vol. 20, no. 3, pp. 2489–2533, 2017.

W. C. Cheah, A. S. Watson, and B. Lennox, “Limitations of wireless power transfer technologies for mobile robots,” *Wireless Power Transfer*, vol. 6, no. 2, pp. 175–189, 2019.

N. Shinohara, “Power without wires,” *IEEE Microwave magazine*, vol. 12, no. 7, pp. S64–S73, 2011.

L. J. Chen, W. I. S. Tong, B. Meyer, A. Abdolkhani, and A. P. Hu, “A contactless charging platform for swarm robots,” in *IECON 2011-37th Annual Conference of the IEEE Industrial Electronics Society*. IEEE, 2011, pp. 4088–4093.
A. T. Khan, X. Cao, S. Li, and Z. Milosevic, “Blockchain technology
V. Lopes and L. A. Alexandre, “An overview of blockchain integration with
R. Yang, F. R. Yu, P. Si, Z. Yang, and Y. Zhang, “Integrated blockchain and
A. K. Bozcuo˘glu, G. Kazhoyan, Y. Furuta, S. Stelter, M. Beetz, K. Okada, and
ROSbridge: ROS
N. DeMarinis, S. Tellex, V. P. Kemerlis, G. Konidaris, and R. Fonseca,
I. Afanasyev, A. Kolotov, R. Rezin, K. Danilov, A. Kashevnik, and V. Jotsov,
S. Fu, Y. Tang, N. Zhang, L. Zhao, S. Wu, and X. Jian, “Joint unmanned aerial
G. N. DeSouza and A. C. Kak, “Vision for mobile robot navigation: A survey,”
B. Li, Z. Mi, Y. Guo, Y. Yang, and M. S. Obaidat, “Partial computing of-
M. Inaba, “The exchange of knowledge using cloud robotics,”
A. Y. Ng, “ROS: an open-source robot operating system,” in
Cloud-based realtime robotic visual SLAM,” in
International Conference on Robotics and Automation (ICRA). IEEE, 2019, pp. 1982–1988.
S. L. K. C. Gudi, S. Ojha, B. Johnston, J. Clark, and M.-A. Williams, “Fog robotics for efficient, fluent and robust human-robot interaction,” in 2018 IEEE 17th International Conference on Network Computing and Applications (NCA). IEEE, 2018, pp. 1–5.
M. Katz, F. H. Fitzek, D. E. Lucani, and P. Seeling, “Mobile clouds as the building blocks of shareconomy: Sharing resources locally and widely,” IEEE Vehicular Technology Magazine, vol. 9, no. 3, pp. 63–71, 2014.
E. Cuervo, A. Balasubramanian, D. Cho, A. Wolman, S. Sarouf, R. Chandra, and P. Bahl, “Marvel: Making smartphones last longer with code offload,” in Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services. ACM, 2010, pp. 49–62.
D. Xu and X. Peng, “Multi-level offloading decision on efficient object tracking for humanoid robot,” Advanced Robotics, vol. 30, no. 15, pp. 979–991, 2016.
D. Berenson, P. Abbeel, and K. Goldberg, “A robot path planning framework that learns from experience,” in 2012 IEEE International Conference on Robotics and Automation, 2012, pp. 3671–3678.
J. Salmerón-Garcı, P. Inigo-Blasco, F. Du, D. Cagı̈s-Muñez et al., “A tradeoff analysis of a cloud-based robot navigation assistant using stereo image processing,” IEEE Transactions on Automation Science and Engineering, vol. 12, no. 2, pp. 444–445, 2015.
Y. Gao, Z. Mi, Y. Yang, and M. S. Obaidat, “An energy sensitive computation offloading strategy in cloud robotic network based on GA,” IEEE Systems Journal, no. 99, pp. 1–11, 2018.
L. Wang, M. Liu, and M. Q.-H. Meng, “Real-time multisensor data retrieval for cloud robotic systems,” IEEE Transactions on Automation Science and Engineering, vol. 12, no. 5, pp. 507–518, 2015.
Z. Hong, H. Huang, S. Guo, W. Chen, and Z. Zheng, “QoS-aware cooperative computation offloading for robot swarms in cloud robotics,” IEEE Transactions on Vehicular Technology, vol. 68, no. 4, pp. 4027–4041, 2019.
A. Rahman, J. Jin, A. Rahman, T. Cricenti, and A. Kulkarni, “Communication-aware cloud robotic task offloading with on-demand mobility for smart factory systems,” IEEE Transactions on Industrial Informatics, vol. 15, no. 5, pp. 2501–2511, 2019.
S. Chinchali, A. Sharma, J. Harrison, A. Elhafsi, D. Kang, E. Peragam, E. Cidon, S. Katti, and M. Pavone, “Network offloading policies for cloud robotics: a learning-based approach,” arXiv preprint arXiv:1902.05703, 2019.
B. Li, Z. Mi, Y. Guo, Y. Yang, and M. S. Obaidat, “Partial computing off-
I. Ounmanakinde and R. Vikash, “Development of a survivable cloud multi-
M. Katz, F. H. Fitzek, D. E. Lucani, and P. Seeling, “Mobile clouds as the building blocks of shareconomy: Sharing resources locally and widely,” IEEE Vehicular Technology Magazine, vol. 9, no. 3, pp. 63–71, 2014.
E. Cuervo, A. Balasubramanian, D. Cho, A. Wolman, S. Sarouf, R. Chandra, and P. Bahl, “Marvel: Making smartphones last longer with code offload,” in Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services. ACM, 2010, pp. 49–62.
D. Xu and X. Peng, “Multi-level offloading decision on efficient object tracking for humanoid robot,” Advanced Robotics, vol. 30, no. 15, pp. 979–991, 2016.
D. Berenson, P. Abbeel, and K. Goldberg, “A robot path planning framework that learns from experience,” in 2012 IEEE International Conference on Robotics and Automation, 2012, pp. 3671–3678.
J. Salmerón-Garcı, P. Inigo-Blasco, F. Du, D. Cagı̈s-Muñez et al., “A tradeoff analysis of a cloud-based robot navigation assistant using stereo image processing,” IEEE Transactions on Automation Science and Engineering, vol. 12, no. 2, pp. 444–445, 2015.
Y. Gao, Z. Mi, Y. Yang, and M. S. Obaidat, “An energy sensitive computation offloading strategy in cloud robotic network based on GA,” IEEE Systems Journal, no. 99, pp. 1–11, 2018.
L. Wang, M. Liu, and M. Q.-H. Meng, “Real-time multisensor data retrieval for cloud robotic systems,” IEEE Transactions on Automation Science and Engineering, vol. 12, no. 5, pp. 507–518, 2015.
Z. Hong, H. Huang, S. Guo, W. Chen, and Z. Zheng, “QoS-aware cooperative computation offloading for robot swarms in cloud robotics,” IEEE Transactions on Vehicular Technology, vol. 68, no. 4, pp. 4027–4041, 2019.
A. Rahman, J. Jin, A. Rahman, T. Cricenti, and A. Kulkarni, “Communication-aware cloud robotic task offloading with on-demand mobility for smart factory systems,” IEEE Transactions on Industrial Informatics, vol. 15, no. 5, pp. 2501–2511, 2019.
S. Chinchali, A. Sharma, J. Harrison, A. Elhafsi, D. Kang, E. Peragam, E. Cidon, S. Katti, and M. Pavone, “Network offloading policies for cloud robotics: a learning-based approach,” arXiv preprint arXiv:1902.05703, 2019.
B. Li, Z. Mi, Y. Guo, Y. Yang, and M. S. Obaidat, “Partial computing off-
I. Ounmanakinde and R. Vikash, “Development of a survivable cloud multi-
M. Katz, F. H. Fitzek, D. E. Lucani, and P. Seeling, “Mobile clouds as the building blocks of shareconomy: Sharing resources locally and widely,” IEEE Vehicular Technology Magazine, vol. 9, no. 3, pp. 63–71, 2014.
E. Cuervo, A. Balasubramanian, D. Cho, A. Wolman, S. Sarouf, R. Chandra, and P. Bahl, “Marvel: Making smartphones last longer with code offload,” in Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services. ACM, 2010, pp. 49–62.
D. Xu and X. Peng, “Multi-level offloading decision on efficient object tracking for humanoid robot,” Advanced Robotics, vol. 30, no. 15, pp. 979–991, 2016.
D. Berenson, P. Abbeel, and K. Goldberg, “A robot path planning framework that learns from experience,” in 2012 IEEE International Conference on Robotics and Automation, 2012, pp. 3671–3678.
Jiong Jin (IEEE M’11) received the B.E. degree with First Class Honours in Computer Engineering from Nanyang Technological University, Singapore, in 2006, and the Ph.D. degree in Electrical and Electronic Engineering from the University of Melbourne, Australia, in 2011. From 2011 to 2013, he was a Research Fellow in the Department of Electrical and Electronic Engineering at the University of Melbourne. He is currently a Senior Lecturer in the School of Software and Electrical Engineering, Faculty of Science, Engineering and Technology, Swinburne University of Technology, Melbourne, Australia. His research interests include network design and optimization, edge computing and distributed systems, robotics and automation, and cyber-physical systems and Internet of Things as well as their applications in smart manufacturing, smart transportation, and smart cities.

Akhlaqur Rahman received his PhD degree in Electrical and Electronic Engineering from Swinburne University of Technology, Melbourne, Australia in 2019. He received his B.Sc in Electrical and Electronic Engineering from American International University, Bangladesh in 2012. Before joining Swinburne, Akhlaqur was a Lecturer and Coordinator for the department of Electrical and Electronic Engineering (from 2013 to 2014) at Uttara University in Bangladesh. Akhlaqur is currently working as Lecturer at Engineering Institute of Technology (EIT), Australia. Akhlaqur’s research focus is on cloud networked robotics, industrial automation, network optimization, smart manufacturing, and Internet of Things (IoT). He is a Member of IEEE.

Ashfaqur Rahman is a Principal Research Scientist and Team Leader at Data61 division of CSIRO, Australia. He is a data scientist and working as a researcher for nearly 15 years. His key research areas are machine learning and Data mining, More specifically Ensemble learning and fusion, Evolutionary Algorithm based network optimization, Distributed Machine Learning for Big Data, Feature selection/weighting methods, Image segmentation and classification. Dr. Rahman received his Ph.D. degree in Information Technology from Monash University, Australia in 2008. He has published over 100 peer-reviewed journal articles, book chapters and conference papers. He supervises several PhD students in collaboration with Swinburne University of Technology and Federation University. He served as reviewer of prestigious journals and conferences. He is the associate editor of Elsevier journal 'Information Processing in Agriculture'. He is involved in organization of key events including IEEE DICTA [2018, 2013], MSLDA [2018, 2017, 2016, 2014]. He is a Senior Member of IEEE.

Ekram Hossain is a Professor and the Associate Head (Graduate Studies) in the Department of Electrical and Computer Engineering at University of Manitoba, Canada. He is a Member (Class of 2016) of the College of the Royal Society of Canada. Also, he is a Fellow of the Canadian Academy of Engineering. Dr. Hossain’s current research interests include design, analysis, and optimization beyond 5G cellular wireless networks. He was elevated to an IEEE Fellow “for contributions to spectrum management and resource allocation in cognitive and cellular radio networks”. He received the 2017 IEEE ComSoc TGCNC (Technical Committee on Green Communications & Computing) Distinguished Technical Achievement Recognition Award “for outstanding technical leadership and achievement in green wireless communications and networking”. Dr. Hossain has won several research awards including the 2017 IEEE Communications Society Best Survey Paper Award and the 2011 IEEE Communications Society Fred Ellersick Prize Paper Award. He was listed as a Clarivate Analytics Highly Cited Researcher in Computer Science in 2017, 2018, 2019, and 2020. Currently he serves as the Editor-in-Chief of IEEE Press and an Editor for the IEEE Transactions on Mobile Computing. Previously, he served as the Editor-in-Chief for the IEEE Communications Surveys and Tutorials (2012-2016). He is a Distinguished Lecturer of the IEEE Communications Society and the IEEE Vehicular Technology Society. He serves as the Director of Magazines for the IEEE Communications Society (2020-2021). Also, he is an elected member of the Board of Governors of the IEEE Communications Society for the term 2018-2020.

Jiafu Wan (M’11) has been a Professor with the School of Mechanical and Automotive Engineering, South China University of Technology (SCUT), since 2015. His research interests include cyber-physical systems, Industry 4.0, smart factory, Industrial Big Data, Industrial Robot, and the Internet of Vehicles. He has directed 20 research projects, including the National Key Research and Development Project, the National Natural Science Foundation of China, and the Joint Fund of the National Natural Science Foundation of China and Guangdong Province, the High-level Talent Project of Guangdong Province, and the Natural Science Foundation of Guangdong Province. So far, he has published more than 140 scientific papers, including over 100 SCI-indexed papers, over 40 IEEE Transactions/Journal papers, 19 ESI Highly Cited Papers, and four ESI Hot Papers. His research results have been published in some famous IEEE Journals and Magazines, such as the IEEE Transactions on Industrial Informatics, the IEEE/ASME Transactions on Mechatronics, the IEEE Communications Surveys and Tutorials, the IEEE Communications Magazine, the IEEE Transactions on Intelligent Transportation Systems, the IEEE Network, the IEEE Wireless Communications, the IEEE Systems Journal, the IEEE Sensors Journal, and the IEEE Internet of Things Journal. According to Google Scholar, his published work has been cited more than 7400 times (H-index = 43). His other SCI citations, the sum of times cited without self-citations reached 2000 times (H-index = 28), according to the Web of Science Core Collection. He is a Member of IEEE and Senior Member of the CMES and CCF. He is the General Chair for the IndustrialIoT 2016 and CloudComp 2016. He is an Associate Editor of the IEEE/ASME Transactions on Mechatronics, the Journal of Intelligent Manufacturing, the IEEE Access, the Elsevier Computer Networks, Mobile Networks and Applications, the Computers and Electrical Engineering, and the Microprocessors and Microsystems.