Cloud Robotics: Architecture, Challenges and Applications

Guoqiang Hu, Member, IEEE, Wee Peng Tay, Member, IEEE, and Yonggang Wen, Member, IEEE

Abstract—We extend the computation and information sharing capabilities of networked robotics by proposing a cloud robotic architecture. The cloud robotics architecture leverages the combination of a virtual ad-hoc cloud formed by machine-to-machine (M2M) communications among participating robots, and an infrastructure cloud enabled by machine-to-cloud (M2C) communications. Cloud robotics utilizes elastic computing models, in which resources are dynamically allocated from a shared resource pool in the cloud, to support task offloading and information sharing in robotic applications. We propose communication protocols, and several elastic computing models to handle different applications. We discuss the technical challenges in computation, communications and security, and illustrate the potential benefits of cloud robotics in several applications.

Index Terms—robotics, elastic computing, cloud computing, and machine-to-machine communication

I. INTRODUCTION

Robotic systems have brought significant economic and social impacts to human lives over the past few decades [1]. For example, industrial robots (especially robot manipulators) have been widely deployed in factories to do tedious, repetitive, or dangerous tasks, such as assembly, painting, packaging, and welding. These preprogrammed robots have been very successful in industrial applications due to their high endurance, speed, and precision in structured factory environments. To enlarge the functional range of these robots or to deploy them in unstructured environments, robotic technologies are integrated with network technologies to foster the emergence of networked robotics.

A network robotic system refers to a group of robotic devices that are connected via a wired and/or wireless communication network [2]. Networked robotics applications can be classified as either teleoperated robots or multi-robot systems. In a telescopulated robot, a human operator controls or manipulates a robot at a distance by sending commands and receiving measurements via the communication network. Application examples include remote control of a planetary rover and remote medical surgery. In a multi-robot system, a team of networked robots complete a task cooperatively in a distributed fashion by exchanging sensing data and information via the communication network. Examples include cooperative robot manipulators, a team of networked robots performing search and rescue, and a group of micro satellites working cooperatively in a desired formation.

However, networked robotics, similar to standalone robots, faces inherent physical constraints as all computations are accomplished in the robotic network, and information access is restricted to the collective storage of the network. With the rapid advancement of wireless communications and recent innovations in cloud computing technologies, some of these constraints can be overcome through the concept of cloud robotics, leading to more intelligent, efficient and yet cheaper robotic networks. In this paper, we describe a cloud robotics architecture, some of the technical challenges, and its potential applications. Some preliminary results on optimal operation of cloud robotics are also presented.

The rest of the paper is organized as follows. In Section II, we outline various challenges and constraints in networked robotics. In Section III, we describe the architecture of our proposed cloud robotics, and elaborate on two key enabling sub-systems. Section IV addresses technical challenges in designing and operating a cloud robotics architecture. In Section V, we highlight a few important applications in robotics that will benefit from our proposed cloud robotics. We conclude and summarize this paper in Section VI.

II. CHALLENGES IN NETWORKED ROBOTICS

Networked robotics, especially the multi-robot systems as illustrated in Figure 1, distributes the workload of sensing, actuating, communication, and computation among the group of...
participating robots. It has achieved great success in industrial applications, intelligent transportation systems, and security applications. However, the advancement of networked robotics is restricted by resource, information, and communication constraints inherent in the existing framework. We discuss these constraints in detail in this section.

A. Resource Constraints

Although a robot can share its computation workload with other units in the network, the overall effectiveness of the robotic network is limited by each robot’s resources, including onboard computers or embedded computing units, memories, and storage space. Physically, these onboard computing devices are restricted by the robots’ size, shape, power supply, motion mode, and working environment. Once the robots are designed, built and deployed, it is technically challenging to change or upgrade their resource configurations.

B. Information and Learning Constraints

The amount of information a robot has access to is constrained by its processing power, storage space, and the number and type of sensors it carries. Networked robotics allows the sharing of information amongst robots connected by a communication network so that a global task can be solved or computed cooperatively using the whole network. However, networked robotics is constrained by the information observed or computed by robots in the network, and by the examples or scenarios that the network encounters, and hence limiting its ability to learn. A robotic team learning to navigate may perform very well in a static environment, where all obstacles can be mapped out with an increasing accuracy over time. On the other hand, the learning process has to be repeated once the environment changes or the robotic team is placed in a new unfamiliar environment. The map databases maintained by the robotic team is also limited by the collective amount of storage space (including memory and disk) the team has.

C. Communication Constraints

Common protocols for machine-to-machine (M2M) communications include proactive routing, which involves the periodic exchange of messages so that routes to every possible destination in the network is maintained [3], and ad hoc routing, which forms a dynamic route to a destination node only when there is a message that needs to be sent [4]. Proactive routing incurs high computation and memory resources in the route discovery and maintenance process. Ad hoc routing protocols suffer from high latency as a route has to be established before a message can be sent, and are not practical if the network topology is highly dynamic. These drawbacks are significant in mobile robotic networks, and may lead to severe performance degradation.

D. From Networked Robotics To Cloud Robotics

Networked robotics can be considered as an evolutionary step towards cloud robotics, i.e., cloud-enabled networked robotics, which leverages emerging cloud computing technologies to transform networked robotics. The design objective is to overcome the limitations of networked robotics with elastic resources in a centralized cloud infrastructure.

Cloud computing provides a natural venue to extend the capabilities of networked robotics. NIST [5] defines cloud computing as “a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.” Through its three service models (i.e., software, platform and infrastructure), it would enable tremendous flexibility in designing and implementing new applications for networked robotics.

Several research groups have started to explore the use of cloud technologies in robotic applications. For example, research groups at Google have developed smart phone driven robots that can learn from each other via the cloud [6]. A research group at Singapore’s ASORO laboratory has built a cloud computing infrastructure to generate 3-D models of environments, allowing robots to perform simultaneous localization and mapping (SLAM) much faster than by relying on their onboard computers [7].

III. Cloud Robotics

In this section, we first describe a system architecture for cloud robotics, and then focus on the two key enabling sub-systems: the M2M communication framework and the elastic computing architecture. Our cloud robotics differentiates from existing solutions in that it leverages two complementary clouds (i.e., a virtual ad-hoc cloud and an infrastructure cloud).

A. System Architecture

In Figure 2, we illustrate the system architecture for our proposed cloud robotics. The architecture is organized into two complementary tiers: a machine-to-machine (M2M) level and a machine-to-cloud (M2C) level.
On the M2M level, a group of robots communicate via wireless links to form a collaborative computing unit (i.e., a virtual ad-hoc cloud). The benefits of forming a collaborative computing unit are multi-fold. First, the computing capability from individual robots can be pooled together to form a virtual ad-hoc cloud infrastructure. Second, within the collaborative computing unit, information can be exchanged for collaborative decision making in various robot-related applications. Finally, it allows robots not within communication range of a cloud access point to access information stored in the cloud infrastructure or send computational requests to the cloud.

On the M2C level, the centralized cloud infrastructure provides a pool of shared computation and storage resources that can be allocated elastically for real-time demand. The elastic computing model allows the group of networked robots to offload computation-intensive tasks for remote execution, resulting in “remote-brain” robots. Moreover, the benefits of a large volume of storage provided by the centralized cloud are two-fold. First, it can unify a large volume of information about the environment, which can be organized in a format usable by robots. Second, it can provide an extensive library of skills or behaviors that are related to task requirements and situational complexities, making it feasible to learn from the history of all cloud-enabled robots.

B. M2M Communication Architecture

Robots in a network can communicate if they are within communication range of each other, and with the cloud servers if the robots are close to access points of the cloud infrastructure. A wireless M2M communication network can be formed by robots working cooperatively with each other to route and relay information. We call either a robot or an access point a node in the M2M network.

Several standards like Zigbee, Bluetooth, and WiFi Direct have been developed for short range wireless communications between robots. For long range communications, radio frequency and microwave communication technologies may be used.

A network of robots is often formed dynamically and in an ad-hoc manner. There is no central controller to coordinate the communication flow in the network. Robotic units may leave and join the network, or may become unavailable because of unpredictable failures or obstructions in the environment. Furthermore, the network is highly dynamic if robots are mobile. All these considerations make the design of effective routing protocols difficult and impractical in some scenarios. Gossip algorithms [8] are randomized methods designed to transmit a message from a source to a destination without any explicit route discovery mechanism. If two nodes are within communication range, we say that they are neighbors. When a robot wants to send a message to a destination node (either another robotic unit or a cloud access point), it randomly chooses one of its neighbors and transmits the message together with a header that contains the identifier of the destination and itself, and a time value indicating the validity period of the message. In another variant of the protocol, the message is simply broadcast to all neighbors, but depending on the application, this may incur high communication load in the network. We will focus on the protocol that chooses a random neighbor in this paper. At every time step, each node randomly chooses a neighbor to retransmit messages that are not intended for itself and are still valid. After a sufficient number of time steps, all the messages will be relayed to their destinations with high probability.

We propose the use of gossip protocols for M2M communications in cloud robotic architectures. Gossip protocols do not require route discoveries and maintenance, and are thus suited for highly dynamic mobile robotic networks. These protocols are also very simple to implement, and require minimal additional computation and memory resources. However, the trade-off is that gossiping may result in a high message latency if the network conductance is low. In cloud robotics, however, this problem is significantly mitigated as the cloud serves as a central super node for the M2M communication network. As we will see in Section IV-B, the time required for a message to be disseminated in a network is greatly reduced by the existence of a super node. Alternatively, a hybrid gossip algorithm can be used in which routes to frequently accessed nodes like the group leader in the proxy-based cloud computing model (see Section III-C), can be maintained.

C. Elastic Cloud Computing Architecture

Our proposed cloud robotics is built on the combination of a virtual ad-hoc cloud formed by a group of networked robots and a centralized cloud. This unique combination offers us a great flexibility in designing computing models, tailored for
specific applications. We focus on the following three elastic computing models (cf. Figure 3):

- **Peer-Based Model**: each robot and each virtual machine (VM) in the cloud is considered as a computing unit. These robots and VMs form a fully distributed computing mesh. A task can be divided into smaller modules for execution over a subset of the nodes in the computing mesh.

- **Proxy-Based Model**: in the group of networked robots, one unit functions as a group leader, communicating with a proxy VM in the cloud infrastructure, to bridge the interaction between the robotic network and the cloud. The set of computing units is organized into a two-tier hierarchy.

- **Clone-Based Model**: each robot corresponds to a system-level clone in the cloud. A task can be executed in the robot or in its clone. The set of robotic clones also form a peer-to-peer network with better connectivity than the physical ad-hoc M2M network. Moreover, this model allows for sporadic outage in the physical M2M network.

Each of these elastic computing models exhibits different properties in robustness to network connections, interoperability and mobility management (see Table I). Robustness refers to the network connectivity between the set of networked robots and the centralized cloud infrastructure. The clone-based model has the maximum number of linkages from robots to the cloud, and is thus the most robust; the proxy-based model is the least robust in terms of network connectivity; and the peer-based model falls between these two extreme cases. Interoperability refers to the additional complexity required in operating a cloud robotics infrastructure with an existing robotic network. The proxy-based model is the most interoperable model, because of its hierarchical structure; while the clone-based model is the least interoperable. Mobility refers to the capability of supporting mobile robots. The peer-based model supports the most flexibility for mobility, because VMs can be instantiated anywhere in the cloud infrastructure; while the clone-based model is the least flexible, because complicated VM migration mechanisms are required to support robot mobility.

<table>
<thead>
<tr>
<th>Model</th>
<th>Robustness</th>
<th>Interoperability</th>
<th>Mobility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peer-based</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
</tr>
<tr>
<td>Proxy-based</td>
<td>Low</td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td>Clone-based</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
</tr>
</tbody>
</table>

TABLE I

**Comparisons of different computing models.**

IV. TECHNICAL CHALLENGES

In this section, we discuss some specific challenges and research considerations in the cloud computation model and the M2M communication network.

A. Computation Challenges

One of the key benefits of cloud robotics is the capability of offloading computationally intensive tasks to the cloud for execution. However, the decision to offload a specific task requires a unified framework that can handle a list of complex issues. First, the offloading strategy should consider various factors, including the amount of data exchanged, and the delay deadline to complete the task. Second, the decision should also consider whether it is more advantageous to execute the task within the group of networked robots, given the presence of cloud resources. Finally, given a pool of cloud resources spread across different data centers, it is a challenge to allocate virtual machines optimally to execute the offloaded task and to manage live VM migrations.

We advocate an optimization framework to determine the optimal task execution strategies. Specifically, our objective is to minimize the amount of energy consumed by the robot, under the constraint that the task should be completed within a specified deadline. The fundamental trade-off lies between the energy consumed for executing the task by the on-board CPU within the robot and the energy consumed transmitting the amount of data to the cloud for remote execution.

In our initial investigation, we have considered the two alternative choices of standalone execution by the robot and cloud execution. We assume the following energy consumption models. For the standalone execution, Dynamic Voltage Scaling (DVS) [9] is assumed to minimize the total energy usage for the computing task; for the cloud execution, we assume a polynomial energy consumption model in which the amount of energy consumed to transmit s bits across a wireless channel with fading coefficient $g$ is proportional to $s^n/g$, where $n$ depends on the coding scheme. It can be shown that, for a given task profile of $L$ bits of data and a delay deadline of $T$, the minimum energy consumed for the standalone execution...
is \( \mathcal{E}_c^* \sim L(T/L)^2 \), and the minimum transmission energy for cloud execution is \( \mathcal{E}_c^* \sim L(T/L)^{(n-1)}. \)

Using the above results, one can determine the optimal operational region for either task execution model by simply comparing the energy usage for a specific task. For example, the optimal task execution region is illustrated in Figure 4 for \( n = 2.5 \). The boundary between the two optimal operational regions is a line (i.e., \( L/T = \text{const} \)), where \( L/T \) can be considered as an effective data transmission rate. In this case, when the effective transmission rate is larger than a threshold, the cloud execution is optimal; otherwise, the standalone execution is optimal. In-depth theoretical and numerical analysis for this simple task offloading scheme can be found in our technical paper [10], and is omitted here due to limited space.

### B. Communication Challenges

As discussed in the previous section, the choice of standalone or cloud execution depends on the delay sensitivity of the task. The communication delay introduced in sending the computation request to the cloud has to be factored into the decision. In this section, we provide upper bounds for the time required for a message to be delivered with high probability. Packet delivery failures and communication outage are inherent in any wireless communication systems. A communication network based on gossip protocols can thus be regarded as a system with higher failure rates, but significantly lower overheads. Furthermore, the additional increment in the failure rate depends on the network topology, and the use of the cloud as a super node in the network effectively controls this rate.

In the gossip protocol, each node chooses a neighbor randomly to transmit a message. Suppose that node \( i \) chooses node \( j \) with probability \( P_{ij} \), where a zero probability implies that the two nodes are not within communication range, and are therefore not neighbors. It can be shown [8] that the communication delay of disseminating a message from a single node in the M2M network to all \( N \) nodes in the network is \( O(\log N/\Phi) \), where \( \Phi \) is the conductance of the network, given by

\[
\Phi = \min_{S: |S| \leq N/2} \frac{\sum_{i \in S, j \in S^c} P_{ij}}{|S|}.
\]

In general, the worst case communication delay is \( O(N \log N) \). However, in our cloud robotics architecture, M2M links are expected to be short range, so that we can partition the M2M network into teams of robots with some maximum size \( M \), with each team having at least one link to the cloud super node. For example, in the clone-based model where all robots have communication links to the cloud, the size \( M = 1 \). We see that the conductance of such a network is at least \( 1/M \), so that the delay is bounded by \( O(M \log N) \). Typically \( M \) is either constant or grows slowly with the size \( N \) of the network. Furthermore, task offloading is typically to immediate neighbors or to the cloud. Therefore, in the worst case, the time required for M2C communications is \( O(M \log M) \).

<table>
<thead>
<tr>
<th>Model</th>
<th>M2M</th>
<th>M2C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peer-based (expander)</td>
<td>( O(\log N) )</td>
<td>( O(\log N) )</td>
</tr>
<tr>
<td>Proxy-based</td>
<td>( O(M \log N) )</td>
<td>( O(M \log M) )</td>
</tr>
<tr>
<td>Clone-based</td>
<td>( O(\log N) )</td>
<td>( O(1) )</td>
</tr>
</tbody>
</table>

**TABLE II**

Comparisons of worst-case communication delays for different computing models.

depends on the particular network topology. We assume that for any subset of nodes in the network, there exists at least a fraction \( \alpha \) with communication links to the cloud. Such a network belongs to the class of expander graphs.

The above analysis applies for a static network. We have shown that it generalizes to similar results for mobile networks with switching topologies [11].

### C. Optimization Framework

We have considered the simple scenario of standalone versus cloud execution of a task. In general, the task offloading decision should made among three execution strategies, including:

- standalone execution by the individual robot,
- collaborative execution by the group of networked robots, and
- cloud execution.

In some cases, a hybrid model including partial execution with all these strategies is possible. We aim to develop an optimization framework involving all execution modes with communication and execution costs included, to find the optimal execution strategy. In particular, the optimal strategy should take into consideration the time-varying nature of the wireless M2M communication network and the latency introduced by the gossiping protocol. In addition, only the minimal required set of information should be communicated, and this depends on the particular application. It is of interest to investigate what information should be stored locally versus on the cloud.

### D. Security Challenges

Trust and security issues are major considerations in cloud robotics. Specifically, our solution faces two major security challenges due to its cloud implementation.

We need the VM environment to be trust-worthy. A malicious VM can subtly sabotage an important task without the robot being aware of the damage. In military applications, the robotic unit has to identify a trust-worthy VM infrastructure to connect and to avoid malicious infrastructures (e.g., battlefield communication vehicles from an enemy). In general, three approaches can be adopted to cope with this problem, including:

1. Trust establishment: the user performs some pre-use actions to check a VM’s host environment.
2. Trust measurement: some root-of-trust components that do not belong to the cloud platform provider (e.g., from hardware vendors or virtualization software providers)
monitor the VM, and securely report trust measurements to a user or third party.

3) Reputation-based trust: the user verifies the VM infrastructure by the service provider’s identity and then relies on legal, business or other external considerations to infer trust.

A robot needs trust to launch task delegation on a public cloud, especially when the computation and network traffic incur monetary costs. The computing environments in the cloud should be verifiable by a user or a trusted party, e.g., to ensure there is no hidden or malicious code running besides the delegated tasks. Moreover, confidential data may be stored in the public cloud storage, while logically private to clone devices. Therefore, strong integrity and confidentiality protection are needed to secure application data.

V. APPLICATIONS

Future robotic applications will largely benefit from cloud robotics, which provides the following advantages over traditional networked robots.

- Ability to offload computation-intensive tasks to the cloud. The robots only have to keep necessary sensors, actuators, and basic processing power to enable real-time actions (e.g., real-time control). The battery life is extended, and the robotic platform becomes lighter and less expensive with easier to maintain hardware. The maintenance of software onboard the robots also becomes simpler, with less need for regular updates. As the cloud hardware can be upgraded independently from the robotic network, the operational life and usefulness of the robotic network can be easily extended.

- Access to vast amounts of data. The robots can acquire information and knowledge to execute tasks through databases in the cloud. They do not have to deal with the creation and maintenance of such data.

- Access to shared knowledge and new skills. The cloud provides a medium for the robots to share information and learn new skills and knowledge from each other. The cloud can host a database or library of skills or behaviors that map to different task requirements and environmental complexities. The RoboEarth project [12] is trying to turn this into a reality.

Due to these advantages, cloud robotics has a wide range of potential applications in data-intensive or computation-intensive tasks in the areas of intelligent transportation, environment monitoring, health care, smart home, entertainment, education, and defense. In this section, we discuss the opportunities and challenges that cloud robotics brings to traditional robotic applications. Specifically, we focus on three robotic applications: SLAM, grasping, and navigation.

A. SLAM

SLAM [13] refers to a technique for a robot or an autonomous vehicle to build a map of the environment without a priori knowledge, and to simultaneously localize itself in the unknown environment. SLAM, especially vision-based SLAM and cooperative SLAM, are both data intensive and computation intensive. The steps such as map fusion and filtering for state estimation can be processed in a parallel fashion. Thus, these tasks can be offloaded to the cloud. For example, a grid based FastSLAM is implemented in a cloud computing framework as reported in [7]. As demonstrated in [7], the cloud can substantially improve the implementation speed of SLAM.

B. Grasping

Robotic grasping has been an active research topic over a few decades. If the full 3-D model of the object is precisely known, then various methods can be applied to synthesize the grasp. If the object is unknown or not precisely known, the problem is much more challenging, and involves the access and preprocessing of vast amounts of data and can be computationally intensive. Recently, information-based or data-driven grasping methods [14] have been developed to enable robotic grasping for any hand and any object. These methods require access to large databases. By offloading this task to the cloud, grasping can be facilitated without requiring vast amounts of computing power, data, and storage space on the robotic platform. In addition, model knowledge of new objects learned by different robots can be shared in the cloud for future usage by other robots.

C. Navigation

Robotic navigation refers to a robot’s activity to determine its own position with respect to a certain reference and then to plan a path to reach a desired location. It can involve a combination of tasks such as localization, path planning, and mapping. Basically, there are two types of approaches: map-less approaches and map-based approaches [15]. Map-less approaches rely on the observations of the perception sensors for navigation. Due to the limited onboard resources, these approaches usually suffer from reliability issues. Map-based robotic navigation is relatively reliable if a precise map is available. It can either use a known map or build a map during the navigation. However, the process of building the map requires large amounts of storage space and is computationally intensive. On the other hand, the process of searching a map requires access to large amounts of data, which is challenging if the navigation area is large. Cloud robotics provides a very promising solution for future cloud-enabled navigation that avoids these two challenges. The cloud can not only provide storage space to store the large amount of map data, but also provide processing power to facilitate the building and searching of the map quickly. Through the cloud, commercially available maps (e.g., Google maps) can also be leveraged to develop reliable, agile, and long-range autonomous navigation solutions.

VI. CONCLUSIONS

We have proposed a cloud robotics architecture to address the constraints faced by current networked robots. Cloud robotics allows robots to share computation resources, information and data with each other, and to access new knowledge.
and skills not learned by themselves. This opens a new paradigm in robotics that we believe leads to exciting future developments. It allows the deployment of inexpensive robots with low computation power and memory requirements by leveraging on the communications network and the elastic computing resources offered by the cloud infrastructure. Applications that can benefit from the cloud robotics approach are myriad and includes SLAM, grasping, navigation, and many others that we have not discussed, like weather monitoring, intrusion detection, surveillance, and formation control.

REFERENCES


Guoqiang Hu is an Assistant Professor in the School of Electrical and Electronics Engineering at Nanyang Technological University in Singapore. Prior to his current position, he was a postdoc research associate at University of Florida in 2008 and an assistant professor at Kansas State University from 2008 to 2011. He received his B.Eng, M.Phil, and Ph.D degrees from University of Science and Technology of China, the Chinese University of Hong Kong, and University of Florida in 2002, 2004, and 2007, respectively. His research interest is in the analysis, control, and design of distributed intelligent systems. He is a member of IEEE.

Wee Peng Tay is an Assistant Professor in the School of Electrical and Electronics Engineering at Nanyang Technological University in Singapore. He received the BS degree in Electrical Engineering and Mathematics, and the MS degree in Electrical Engineering from Stanford University in 2002. He received the PhD degree in Electrical engineering and Computer science from the Massachusetts Institute of Technology in 2008. His main research interests are in distributed signal processing and algorithms, data fusion and decision making in ad hoc networks, machine learning and applied probability. He is a member of IEEE.

Yonggang Wen is an Assistant Professor in the School of Computer Engineering at Nanyang Technological University (NTU) in Singapore. Prior to his present position, he has held R&D positions in networking companies in the USA, including Cisco and Lucent. He received his Ph.D degree in electrical engineering and computer science from Massachusetts Institute of Technology (MIT) in 2008, his M.Phil. degree in information engineering from Chinese University of Hong Kong (CUHK) and B.Eng. degree in electronic engineering and information science from University of Science and Technology of China (USTC) in 2001 and 1999, respectively. His research interests are in cloud computing, content networking and green networks. He is a member of IEEE.