Toward Big Data in Green City

Chunsheng Zhu, Huan Zhou, Victor C. M. Leung, Kun Wang, Yan Zhang, and Laurence T. Yang

ABSTRACT

Integrating sensors and cloud computing, sensor-cloud is a very powerful system for users to obtain big data in green city. In this article, toward big data in green city, we first review the latest work concerning big data and sensor-cloud, respectively. Further, we introduce three types of sensor-cloud (i.e., PSC, ASC, and SSC) for green city. Specifically, about PSC, participatory sensing is incorporated into sensor-cloud for sensing big data. In terms of ASC, an agent is incorporated into sensor-cloud for transmitting big data. For SSC, a social network is incorporated into sensor-cloud for sharing big data. For PSC, participatory sensing is incorporated into sensor-cloud for transmitting big data. In terms of ASC, an agent is incorporated into sensor-cloud for sensing big data. In terms of SSC, a social network is incorporated into sensor-cloud for sharing big data. In this article, toward big data in green city, we first review the latest work concerning big data and sensor-cloud, respectively. Further, we introduce three types of sensor-cloud (i.e., PSC, ASC, and SSC) for green city. Specifically, about PSC, participatory sensing is incorporated into sensor-cloud for sensing big data. In terms of ASC, an agent is incorporated into sensor-cloud for transmitting big data. For SSC, a social network is incorporated into sensor-cloud for sharing big data. Finally, open research issues with respect to big data and sensor-cloud are discussed, respectively. We hope this article can serve as enlightening guidance for future research regarding big data in green city.

INTRODUCTION

Recently, with the rapid advances in green sensing, computing, and communication technologies, our city is evolving into a green city with numerous green applications (e.g., green homes, green grids, green factories). In particular, with the fast growth of urban population and fast urban development, a huge amount of data from various sources is being generated by green city. This explosive data growth in green city is making big data [1] a hot topic for the fulfillment of green city.

Incorporating sensors and cloud computing, sensor-cloud [2] is a very powerful system for users to achieve big data in green city. Particularly, as shown in Fig. 1, on one hand, various big data in our city is sensed and gathered by a variety of ubiquitous sensors (e.g., temperature sensors, humidity sensors, pressure sensors, light sensors, video sensors) and further transmitted to the cloud via a sink or sinks. On the other hand, the received big data is stored and processed by the powerful data centers in the cloud and further delivered to users on demand. Thus, with sensor-cloud, users are able to have access to the desirable big data anytime and anywhere if there is any network connection.

In this article, toward big data in green city, recent work about big data and sensor-cloud is reviewed first. Further, three types of sensor-cloud — participatory sensor-cloud (PSC), agent-based sensor-cloud (ASC), and social-sensor-cloud (SSC) — are introduced for green city. Particularly, with respect to PSC, participatory sensing [3] is incorporated into sensor-cloud for sensing big data. In ASC, an agent [4] is incorporated into sensor-cloud for transmitting big data. In terms of SSC, a social network [5] is incorporated into sensor-cloud for sharing big data. Eventually, the open research issues concerning big data and sensor-cloud are investigated. This article aims to serve as inspiriting guidance for future research with respect to big data in green city.

The rest of this article is organized as follows. The following two sections review the latest work about big data and sensor-cloud, respectively. Then we present PSC, ASC, and SSC. Following that, we investigate the open research issues on big data and sensor-cloud, respectively. The final section concludes this article.

LATEST WORK ON BIG DATA

Focusing on the network infrastructure of big data, the unique challenges in establishing a network infrastructure for geologically distributed and rapidly generated big data are discussed in [6]. Specifically, it first investigates the challenges in terms of every segment in the network highway, such as the access networks connecting data sources, the Internet backbone bridging them to remote data centers, and the dedicated networks within a data center and among data centers. Moreover, it presents two case studies empowered by networking, and highlights the promising and interesting future directions based on those case studies.

Discussing the framework of big data, a novel framework to deliver big data over content-centric mobile social networks with a satisfactory quality of experience is proposed in [7]. The characteristics and challenges of mobile big data are presented first. Then a content-centric network architecture to deliver mobile big data in mobile social networks is proposed. With that, how to choose the relay node for forwarding data packets is introduced by defining priorities of data packets. Lastly, simulation results are demonstrated on the performance of the proposed framework.

Investigating the cost minimization of big data, the authors in [8] try to investigate the cost minimization problem by a joint optimization of three factors (i.e., task assignment, data placement, and data movement) for big data services in geo-distributed data centers. Specifically, to describe the task completion time considering both data transmission and computation, a 2D Markov chain is proposed, and the average task completion time
is derived in closed form. Then the problem is modeled as mixed-integer nonlinear programming, and an efficient solution is proposed to linearize it. Finally, extensive simulation results validate the efficiency of the proposed proposal.

Studying the role of big data, the authors in [9] first show the state-of-the-art communication technologies and smart-based applications in the context of smart cities. Then visions of big data analytics for supporting smart cities are presented by discussing the ways that big data fundamentally alters urban populations at different levels. After that, a future business model of big data for smart cities is introduced. Eventually, the business and technological challenges of the proposed business model are observed.

Considering the application of big data, a real-time big data analytical architecture for remote sensing satellite application is proposed in [10]. Particularly, taking into account of both real time and offline data processing, the introduced architecture consists of three main units: a remote sensing big data acquisition unit; a data processing unit; and a data analysis decision unit. Further, a detailed analysis regarding the proposed analytical architecture is performed using Hadoop, on the remotely sensed big data for land and sea area using an Earth observatory system.

**LATEST WORK ABOUT SENSOR-CLOUD**

Concerning the pricing of sensor-cloud, five sensor-cloud pricing models are proposed in [11]. Specifically, the following elements are considered by each sensor-cloud pricing model to charge the user: 1) user’s lease period; 2) sensor-cloud’s working time; 3) the sensor-cloud resources used by the user; 4) the volume of data achieved by the user; 5) the sensor-cloud path that transmits data from the sensor network to the user. With that, the characteristics and case studies of the proposed sensor-cloud pricing models are investigated, followed by a review regarding a user behavior study.

Regarding the security of sensor-cloud, a risk assessment framework utilizing attack graphs for sensor networks in a sensor-cloud is proposed in [12]. Particularly, the impacts of attacks on sensor networks are reviewed by the proposed framework first. Then reasonable timeframes that predict the degradation of the sensor network’s security performance (e.g., confidentiality, integrity, and availability) are estimated by the proposed framework. With that, the assessment results of the proposed framework are validated in various simulated attack scenarios.

About the quality of service of sensor-cloud, [13] focuses on scheduling a particular cloud data center that congregates data from various virtual sensors and further delivers the data to users. Specifically, the scheduling of the cloud data center is conducted considering several network constraints (e.g., data migration cost, data delivery cost, and service delay). Then an optimal decision rule for selecting a particular cloud data center is determined to satisfy the quality of service. Experimental results about the proposed scheduling mechanism in real-time sensor-cloud scenarios are shown.

With respect to the modeling of sensor-cloud, [14] pays attention to theoretical characterization and analysis of sensor-cloud. Particularly, a mathematical formulation of sensor-cloud is presented. Then a detailed analysis is performed, taking into account the performance metrics (e.g., energy consumption, fault tolerance, and lifetime of a sensor). Further, the evaluation of the cost effectiveness of sensor-cloud is conducted. Analytical results with the proposed formulation are also demonstrated, in terms of a sensor’s average lifetime, a sensor’s average energy consumption, and a user’s average expenditure.

For the application of sensor-cloud, [15] discusses target tracking with sensor-cloud. Specifically, how to schedule sensors to track targets in the presence of overlapping coverage is studied. A dynamic mapping algorithm is proposed, based on the Theory of Social Choice for achieving an unbiased mapping of sensors to targets. The uniformity of the dynamic mapping algorithm is demonstrated, while every target is covered by the sensors.

**THREE TYPES OF SENSOR-CLOUD**

**PSC**

As an emerging sensing and computing paradigm, participatory sensing [3] tasks everyday mobile devices (e.g., cellular phones) to form interactive and participatory sensor networks, which enable public and professional users to gather and analyze as well as share local knowledge. In other words, participatory sensing combines the ubiquity of mobile devices with the sensing capabilities of sensor networks, targeting the seamless collection of data from a large number of user-carried devices.

Induced by the participatory sensing concept, as shown in Fig. 2a about PSC, users play the role of sensors by sensing and gathering data with mobile devices and further transmitting the data to the cloud. After the cloud stores and processes the received data, the processed data is delivered to users on demand. Moreover, clouds share the data with each other by communications between them.

In such a way, when offering big data to users, the following benefits can be achieved with PSC:

- The number of sensors needed to be deployed to sense and gather data can be reduced greatly, since a lot of data is sensed and gathered with user-carried devices.
• The amount of data that needs to be gathered can be decreased substantially due to a certain number of data being shared through cloud-to-cloud communications.

ASC

In the field of artificial intelligence, agent-based system technology [4] is a very powerful paradigm to conceptualize, design, and implement software systems. Specifically, characterized by acting autonomously on behalf of their users, agents are computer programs that address a growing number of complex problems across distributed and open environments.

Motivated by the agent concept, as illustrated in Fig. 2b concerning ASC, there are agents near the sensors. The agents are in charge of transmitting the data to the cloud, while the sensors sense, gather, and transmit the data to the agents. After the cloud stores and processes the received data, the processed data is delivered to users on demand. In addition, the clouds share the data with each other via communications between them.

In such a manner, when providing big data to users, the following advantages can be obtained with ASC:

• The energy consumption of sensors can be reduced substantially, because the multihop communications from the sensors to the sink/sinks are eliminated with the agents.
• The speed at which the cloud receives the sensory data can be faster, since the data sensed by the sensors is directly transmitted to the cloud via the nearby agents rather than indirectly transmitted to the cloud via a faraway sink or sinks.
• The volume of data that needs to be collected can be decreased substantially, because a certain number of data is shared via cloud-to-cloud communications.

SSC

A social network [5] is a network in which social members (e.g., individuals, organizations) interact with each other. Particularly, in a social network, the resources and services can be shared among social members with established relationships.

Based on the social network concept, as demonstrated in Fig. 2c regarding SSC, the users form social networks and the sensors form sensor networks. The data sensed and gathered by the sensors in the sensor networks is transmitted to the cloud via the sink or sinks in the sensor networks. After the cloud stores and processes the received data, the processed data is delivered to the social networks on demand. Then the users in the social networks share the received data. Moreover, the clouds shares the data with each other through communications between them.

In such a way, when offering big data to users, the following benefits can be obtained with SSC:

• The energy consumption of cloud can be decreased greatly, since the desirable data is delivered from the cloud to the social networks once rather than from the cloud to various users multiple times.
• The amount of data that needs to be gathered can be reduced substantially, due to a certain number of data being shared through cloud-to-cloud communications.

CASE STUDY

As presented in Fig. 3, in green city, various entities (shopping malls, markets, retail stores, restaurants, playgrounds, parks, traffic routes, factories, plants, universities, schools, companies, etc.) exist. In some places (e.g., shopping malls, markets, retail stores, restaurants, playgrounds, parks, traffic routes) where there is high diversity and mobility of users, PSCs can be applied to sense the big data. For some sites (e.g., factories, plants) where there is high stability in terms of the type of users and the expected performance of sensors, ASCs can be deployed to transmit the big data. Regarding some places (e.g., universities, schools, companies) in which there is a high similarity of users leading to an easily shared network, SSCs can be formed to share the big data.
Open Research Issues on Big Data

Integration of Big Data: Big data is in various data formats from various sources. For big data, one of the most important issues is integrating such a large amount of data with different formats from multiple sources. Therefore, integration of big data is one of the key challenges to be addressed, especially if some data are incomplete, incorrect, or in the wrong format.

Processing Platforms of Big Data: Substantial big data applications are based on big data analysis, which usually requires huge processing capability. As a result, the hardware and software platforms for big data should offer communication and computing capabilities, with high and stable performance. In addition, the software platforms should be compatible with different kinds of hardware platforms.

Management of Big Data: Big data that have been created at ever increasing rates are characterized by volume, variety, and velocity. Thus, there is great potential in terms of obtaining valuable insights from big data. For proper and efficient utilization of big data, it is critical to design appropriate and effective big data management techniques based on various applications.

Open Standards of Big Data: Open standards are a significant element for ensuring that different systems can interoperate with each other. For big data in multiple systems, open-standards-based technologies should be explored. With open-standards-based technologies, various big data systems can exchange and utilize information cohesively.

Security and Privacy of Big Data: One essential issue for big data is security and privacy. Different data need different levels of security policies to protect them against malicious attacks. Furthermore, various data require different levels of privacy policies for preventing unauthorized access. In addition, for different big data transmissions over various types of networks, different levels of security and privacy policies are demanded.

Quality of Service of Big Data: In terms of big data, the quality of service of big data is always a fundamental concern for users. Quality of service is affected by various factors in the specific application scenarios, while satisfactory quality of service is different for various users. Therefore, it might be promising to investigate techniques that can predict and control the quality of service of big data for various application scenarios and different users.

Open Research Issues about Sensor-Cloud

Framework of Sensor-Cloud: Sensor-cloud is a system with a number of interesting applications (e.g., vital signs monitoring, water quality monitoring, fault diagnostics). For various applications of sensor-cloud, the frameworks of sensor-cloud are generally different. Thus, it is worthwhile to explore the framework of sensor-cloud, considering the specific application requirement (throughput, response time, etc.).

Energy Efficiency of Sensor-Cloud: Energy efficiency is always a crucial factor for sensor-cloud, since most sensors are generally equipped with non-rechargeable batteries, and the implementation of cloud inherently involves tremendous energy consumption. Therefore, mechanisms (e.g., duty scheduling for sensor networks, energy-efficient medium access control for sensor networks, energy-efficient data transmission from sensor networks to cloud, energy-efficient job scheduling in cloud) to save the energy consumption of sensor-cloud are in urgent need.

Resource Management of Sensor-Cloud: There are various resources (e.g., sensing resources, storage resources, computing resources) in sensor-cloud. Maintaining the resources, utilizing the resources, and eventually offering desirable data to users is a challenging task. In this regard, resource management strategies (e.g., resource allocation, resource optimization, resource coordination) for sensor-cloud are very important.

Localization of Sensor-Cloud: Localization is a necessary technique for sensor-cloud, since the data sensed and gathered by the sensors will be useless if the sensors have no idea of their geographical locations. Particularly, for a sensor-cloud in which there are a large number of sensors, a localization process that finds the position of sensors is needed because it will be very expensive to use GPS for many sensors. As a result, localization mechanisms for sensor-cloud should be investigated.

Mobility of Sensor-Cloud: Mobility is a fundamental attribute that should be considered for sensors. Mobile sensor networks have attracted a lot of attention from both the academic and industrial communities, and they have demonstrated several benefits (e.g., better energy efficiency, improved coverage, enhanced target tracking) over static sensor networks. Similarly, taking into account the mobility of sensors in sensor-cloud, the performance of mobile sensor-cloud is worthy of study, in contrast to static sensor-cloud.

Testbed of Sensor-Cloud: Regardless of the size of sensor-cloud, there is considerable cost with respect to the deployment, operation, and maintenance of sensor-cloud. To better use the installed capacity of sensor-cloud to reduce the underutilization of sensor-cloud to the barest minimum, it is very valuable to explore a testbed of sensor-cloud, on which the performance of sensor-cloud can be tested.

Conclusion

Targeted to serve as enlightening guidance for future research regarding big data in green city, this article has investigated sensor-cloud, which is a very powerful system for users in green city.
to achieve big data. Particularly, toward big data in green city, this article has reviewed the recent work about big data and sensor-cloud. Moreover, this article has presented three types of sensor-cloud (i.e., PSC, ASC, and SSC) for green city. In terms of PSC, participatory sensing is incorporated for sensing big data. Regarding ASC, the agent is incorporated for transmitting big data. Concerning SSC, a social network is incorporated for sharing big data. Finally, this article has outlined open research issues regarding big data and sensor-cloud.

**ACKNOWLEDGMENT**

This work was partially supported by funding from the Natural Sciences and Engineering Research Council of Canada, the ICICS/TELUS People & Planet Friendly Home Initiative at the University of British Columbia, TELUS, and other industry partners. This work was partially supported by the National Natural Science Foundation of China (NSFC) under Grant 61602272 and Grant 61572262, and the NSF of Jiangsu Province under Grant BK20141427. This work was partially supported by the project IoTSec — Security in IoT for Smart Grids, with number 248113/O70 part of the IKTPLUS program funded by the Norwegian Research Council. This research is partially supported by the projects 240079/F20 funded by the Research Council of Norway. H. Zhou is the corresponding author.

**REFERENCES**


**BIographies**

CHUNSHENG ZHU [S’12, M’16] (chunsheng.tom.zhu@gmail.com) is a professor in the College of Computer and Information Technology, China Three Gorges University. He is also a visiting scholar at the University of British Columbia. He received his Ph.D. degree from the Department of Control Science and Engineering, Zhejiang University, China. His research interests include mobile social networks, opportunistic mobile networks, and wireless sensor networks.

VICTOR C. M. LEUNG [S’75, M’89, SM’97, F’03] is a professor in the Department of Electrical and Computer Engineering and holder of the TELUS Mobility Research Chair, University of British Columbia. His research is in the areas of wireless networks and mobile systems. He is a fellow of the Royal Society of Canada, a Fellow of the Canadian Academy of Engineering, and a Fellow of the Engineering Institute of Canada.

HUIAN ZHOU is a professor in the College of Computer and Information Technology, China Three Gorges University. He is also a visiting scholar at the University of British Columbia. He received his Ph.D. degree from the Department of Control Science and Engineering, Zhejiang University, China. His research interests include mobile social networks, opportunistic mobile networks, and wireless sensor networks.

KUN WANG [M’13] is an associate professor in the School of the Internet of Things, Nanjing University of Posts and Telecommunications, China. He received his Ph.D. degree from the School of Computing, Nanjing University of Posts and Telecommunications in 2009. In 2016, he was a research fellow with the School of Computer Science and Engineering, University of Aizu, Fukushima, Japan. His current research interests include information security, ubiquitous computing, and wireless communications technologies.

YAN ZHANG [SM’10] is a professor in the Department of Informatics, University of Oslo, Norway. He is also a chief research scientist at Simula Research Laboratory, Norway. He is an Associate Technical Editor of IEEE Communications Magazine, an Editor of IEEE Transactions on Green Communications and Networking, and an Editor of IEEE Communications Surveys & Tutorials. His current research interests include next-generation wireless networks leading to 5G, green, and secure cyber-physical systems.

LAURENCE T. YANG [M’97, SM’15] is a professor in the Department of Computer Science, St. Francis Xavier University, Canada. His research interests include parallel and distributed computing, embedded and ubiquitous/pervasive computing, and big data. He has published more than 220 papers in various refereed journals (around 40 percent in top IEEE/ACM transactions and journals). His research has been supported by the National Sciences and Engineering Research Council and the Canada Foundation for Innovation.