Big Sensing Data Curation in Cloud Data Center for Next Generation IoT and WSN

Scalable, IoT device responsive On-Cloud data curation will match the trend.

By Chi Yang, Deepak Puthal, Saraju P. Mohanty, and Elias Kougianos

Modern sensing devices play a pivotal role in achieving data acquisition, communication and dissemination for the Internet-of-Things (IoT). Naturally, IoT applications and intelligent sensing systems supported by sensing devices, such as wireless sensor networks (WSN), are closely coupled. Modern intelligent sensing systems generate huge volumes of sensing data, well beyond the processing capabilities of common techniques and tools. Hence, to collect, manage, and process IoT big sensing data within an acceptable time duration is a new challenge for both research and industrial applications. The massive size, extreme complexity and high speed of big sensing data bring new technical requirements including data collection, data storage, data organization, data analysis and data publishing in real time when deploying real world IoT applications. To better facilitate these IoT applications, the convergent research of WSN, big data, IoT and Cloud Computing is a natural scientific development trend. In this paper, we concentrate on big sensing data curation and preparation issues with Cloud Computing under the theme of IoT. There are three especially critical issues that need to be addressed, namely scalable big sensing data cleaning, scalable big sensing data compression and Cloud based data curation response for IoT device optimization. Viewed from the IoT side, all IoT sensing devices are integrated together in an adaptive solution and upload their data onto the Cloud. The automatic responses from both the Cloud and intelligent sensors will change the status or behavior of sensing devices, hence the status of the IoT itself.

1. INTRODUCTION

With fast growing attention from both academic research and industrial communities, the IoT consists of thousands of connected devices (mostly modern sensing devices) for information monitoring, gathering, communicating, exchanging, analyzing, decision making and finally instantly responding to acquired information to intelligently control the behavior of physical devices or the factors of a real-world environment [1], [2], [3], [4]. In order to achieve the above mentioned IoT applications, several pivotal technical challenges are involved including WSN, cloud computing and big data [4], [5], [6]. It is well known that sensor devices and WSN in the IoT can generate high reliability, variety, volume, value and velocity [7] data sets. The Cloud, with its massive computing power, storage, and scalable provided software services, offers a promising platform to deal with the challenges brought by IoT big data [8], [9]. Applications of big sensing data processing in the Cloud can be encountered in different fields such as medical/health monitoring, weather forecasting, environmental monitoring, industry production, social media analysis and business analysis. **Figure 1** shows the application of integrated WSN, IoT and the Cloud. Recently, some data curation techniques have been developed for processing of sensing data on cloud data centers, such as the Sensor-Cloud platform [10]. However, those methods and platforms are far from complete, and significant work is still needed. To the best of our knowledge, when it comes to the convergence of Cloud, IoT, WSN very little published research can be found [2], [3], [12]. Hence, how to use the massive computational power and common platforms offered by the Cloud for processing big sensing data from the IoT and offering response actions to the IoT, motivates a new research direction. This paper focuses on the data curation problems of IoT big sensing data processing on the Cloud to facilitate IoT applications and provides current trends and future research scope of IoT big sensing data processing.

FIGURE 1. Application specific integration of the IoT, WSNs and Cloud.

When uploading integrated big sensing data onto a Cloud platform, a novel data cleaning technology will be developed and deployed on the Cloud, including novel sensing data error detection and recovery algorithms to capture errors, conflicts and missing data in big sensing data sets in real time. Simultaneously, recovery algorithms will also be developed and deployed on the Cloud to provide compensating solutions for the detected errors and defects. Intelligent sensing systems and WSNs can also conduct some basic error detection and recovery with the in-network computation power and storage offered by smart sensors.

2. THE STATE-OF-THE-ART

In current Internet and IoT usage, we have entered the big data era of petabytes. Traditional database management tools or data processing technologies become powerless due to limited resources of computation, storage and communication. Big sensing data is generated with the features of high Volume, Value, Variety, Velocity, and Veracity from a wide range of real world applications. Since the 1980s, data generation doubles its size every forty months [7], [8]. In only one year (2012), the everyday data generation rate was equal to 2.5 quintillion (2.5×10^{18}) bytes. Currently, dataset sizes are measured in exabytes (10^{18} bytes). In 2015, there were around 10,000 exabytes of digital data being generated. Following that digital data explosion, the size of big data is expected to surpass 40,000 exabytes by the year 2020 [7], [8]. In many application fields, including meteorology forecast, connectomics, physics simulation, genomics, biological science, gene analysis and environmental research [3], [6], [7], [8], [12], big data processing can even dictate the performance of whole systems. To offer solutions for the new challenges brought by big data, more and more research attention has been attracted by big data in terms of both fundamental research and technical applications. Data curation, as a significant step towards big sensing data processing technology, is commonly deployed on a Cloud platform for achieving scalability, massive resource access, real time data analytics and behavior control for the IoT. The data flow between source sensors (in WSNs and the IoT) and the Cloud with classification is shown in **Figure 2** which provides an overview of the IoT, WSNs, and the Cloud integration with data processing at different levels.

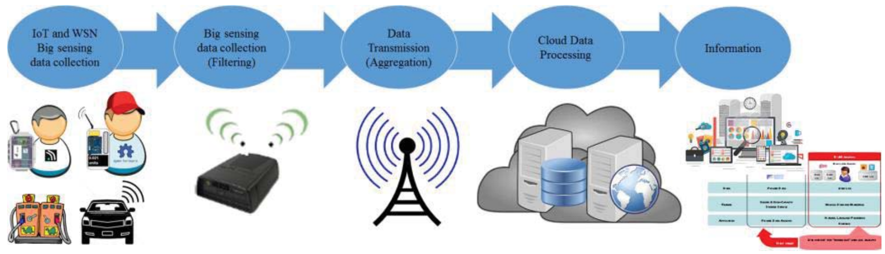

FIGURE 2. Incremental data flow between sensors and the Cloud with classifications.

Big Sensing Data

Modern sensing systems significantly change everyday life by giving us with the capability to monitor, understand and interact with the physical environment around us [1], [5], [14]. In real world applications, sensing systems are becoming much smaller, and smarter with more connectivity and more mobile capability [15]. The price for achieving the above functional improvement is higher data rates, more data storage and more powerful data analysis requirements. As a result, big data sets with high speed, high dimensions and high volume are introduced by countless sensing systems deployed in our environment [1], [7]. As important sources of big data sets, sensing systems generate sensing data with extremely large volumes which are far beyond the processing capability of common data processing software and tools. However, to collect, store, organize, analyze and publish big sensing data from modern sensing systems in real time are critical and essential targets when deploying most of real world sensing systems [16]. In addition, the privacy and security issues of big sensing are widely important [11], [17].

Cloud Computing

The National Institute of Standards and Technology (NIST) defined the Cloud as a framework for enabling convenient, on-demand network access to a shared pool of configurable computing resources such as servers, storage, networks, applications, service, etc. These computational resources should be provided and accessed in a time-efficient manner. At the same time, when acquiring these computational resources, the management effort or interactions with service providers should be as little as possible. Based on the NIST definition, there are four deployment models, three service distribution models and five important features for Cloud Computing. The important five key characteristics of cloud computing include resource pooling, broad network access, measured services, on-demand self-service and rapid elasticity. The core technologies are web service technologies, virtualization and distributed programming models such as MapReduce [18], [19], [20] and, the much newer Spark [21].

Intelligent Sensors and WSN

Sensors, processors and wireless communication devices are becoming much smaller, cheaper, smarter and reliable day by day. Reliable and inexpensive sensor systems based on better computing and communication methods are of the size of a credit card. The architecture of a node is integrated with sensing, processing and communicating. Recently, development and research trends of WSNs are mainly influenced and propelled by the advances in computing and communication [1], [2], [6], [22], [23]. In general, a wireless sensor network consists of intelligent communicating sensors deployed across a landscape for sampling and interpreting real world phenomena, then it forwards the results back to a base station or a user gateway [13], [24]. In addition, for current IoT applications, some control feedback is also expected from either the inner WSN, or a third-part intelligent system such as the Cloud to change the behavior of sensing devices in the WSN. A WSN is a typical distributed system that generates and processes correlated temporal and spatial data from multiple data sources. Under the theme of IoT, all these different sensing devices form different complex networks and generate huge amount of continuous heterogeneous data sets which require to be integrated, cleaned, classified, compressed, stored, exploited and analyzed [7], [25], [26]. These data always need third party processing such as cloud and processing results used as feedback to change sensing device behavior for optimizing IoT component performance [5].

The Internet of Things (IoT)

In brief, the Internet of Things (IoT) can be characterized as a wireless network of internet-connected smart sensing devices ready to gather and transmit data with the support of embedded sensors or sensor networks. The IoT consists of intelligent physical devices connected through techniques including electronics, software, sensors, actuators, and network connectivity for data acquisition, communication and dissemination [5]. The three components of the IOT are "Things", the Internet, and connectivity. But the value (data) generated, transmitted and interpreted in the IoT plays a critical role in smoothing the breach between the digital and the physical world for setting up automated systems. In typical IoT applications, smart objects can be sensed, evaluated and controlled in a remote manner through current popular network infrastructures. This working model of the IoT means great opportunities and trends for integrating our physical real world with our computing systems through networks and smart sensors. It will inevitably make significant improvement for peoples' lives in terms of efficiency, accuracy, economic profit and greatly reduced human intervention [1], [2], [15]. Specifically, sensors and actuators are the backbone physical equipment for realizing IoT applications. According to the literature, it is estimated that the IoT will contain more than 50 billion objects by 2020 [12]. Based on those smart sensed, connected and controlled objects, heterogeneous big data sets with huge volumes are expected. To index, store and analyze those big IoT sensing data becomes more and more important.

3. A SPECIFIC EXAMPLE FRAMEWORK

With the advance of modern smart sensing technologies, big sensing data and the IoT are emerging standards applied to datasets which may not be digested with efficiency and effectiveness using common data processing software, tools and techniques. Because IoT big sensing data sets are often from various structured or unstructured physical devices with the characteristics of high volume, fast data rate and unreliable value, data curation has to be done to guarantee the data quality and minimized size. It is also important for making decisions to change the status of the IoT devices. Without successful data curation, it is impossible to calculate optimized strategies for changing device behavior for the IoT. The Cloud, with its enormous power in computation and huge storage, enables clients to deploy big data curation without requiring heavy assets. IoT big sensing data sets can be widely encountered in industrial and scientific activities. For example, high volume of big data from body sensors is generated by body monitoring equipment. These data sets are uploaded to the Cloud by U.S. hospitals. How to use those data for disease analysis with acceptable accuracy and efficiency poses an interesting topic. In order to process big sensing data efficiently on the Cloud and generate better IoT services, several critical issues should be discussed including reduction of the big data size, data quality, fast query of big sensing data and data curation result for interactions with smart sensing devices. To cope with the above challenges and issues, a framework is proposed to offer a data curation solution with the convergent study of big data, WSN, Cloud Computing and the IoT.

As shown in **Figure 3**, smart sensing devices in the IoT generate big sensing data which is filtered and aggregated as soon as sensing devices perform sampling and data communication to form big sensing data streams. At this stage, heterogeneous big sensing data sets can be integrated with some lightweight method before they are forwarded to the Cloud for more complicated data processing.

When big sensing data or data streams are uploaded to cloud data centers, the first important stage of cleaning for our proposed on-Cloud data curation framework starts. Specifically, to clean the heterogeneous data or data streams from multiple data sources, there are three main tasks to be finalized including: (1) error detection, (2) error recovery, and (3) data consistency and redundancy check. In terms of error detection, WSN big data set from sensing devices or WSNs are commonly subject to corruption and error because of the low reliability of wireless communications, signal processing inaccuracy and hardware defects in the nodes. For achieving high quality WSN application, the first step is to guarantee that the data received is accurate, logically connected and clean. However, sensor data error and cleaning is still a challenging issue and new methods are required to solve the problem. To the best of our knowledge, there are few research works on in-network WSN error detection recovery techniques which are limited by the computation power, storage, wireless network energy consumption and time latency. With the support of Cloud computing, scalable fast error detection and scalable fast error recovery will be possible for real time big sensing data cleaning. In addition, in the process of error detection and recovery, the Cloud environment is helpful in analyzing the error sources and generate feedbacks to the IoT to change sensing device behavior.

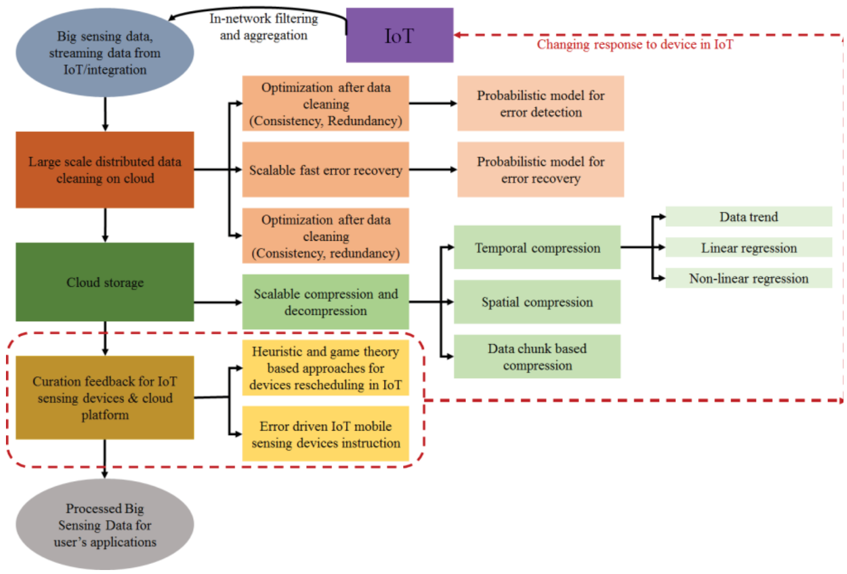

FIGURE 3. IoT generated big data curation steps in a cloud data center.

In addition to data cleaning on the Cloud, data compression should be performed for reducing big sensing data size, and reducing the future data processing time. Under the theme of compression, spatiotemporal features or other sensing data correlations can be exploited. As shown in **FIGURE 3**, several data compression techniques have been or will be developed. For example, by discovering spatial correlations in big data [20], multiple cluster structures can be obtained from a graph data set, then all edges in a cluster can share similar time series of data of a graph. Based on that partition, the workload within a cluster can be greatly reduced by the similarities of time series based inference. Temporal data compression can be carried according to the order of data items or time series based prediction using temporal correlation. In addition, the data prediction models can be improved and modified according to application requirement. By compression on the Cloud, sensing data size can be significantly reduced compared to only in-network lightweight data suppression by sensing devices themselves [27]. Meanwhile the proposed data suppression technologies in this framework should be able to guarantee acceptable data accuracy [27], [28].

As shown in figure 3, after the sensing data cleaning and compression stages, the third component of feedback based IoT sensing device and Cloud optimized control follows. Specifically, in the process of big sensing data cleaning and compression, some optimization issues have been already detected. For example, how to assign the workload of data cleaning or compression between sensing devices and the Cloud to exploit the full potential of the Cloud and smart sensors is an interesting optimization problem. In other words, the trade-off between using the Cloud or using smart sensor devices for computing will be. Furthermore, because both data cleaning and compression buffer and filter big sensing data, some data analytical functions can be performed here. From **Figure 3**, at the stage of curation feedback for IoT devices, the filtering process will be combined with more data analysis techniques for generating feedbacks. Following the proposed cleaning and compression in the framework, heuristic or game theory based algorithms could be adopted for designing adaptive mechanisms in sensing devices of the IoT. Furthermore, data errors are an indication of device failure or network defect. To understand these errors as a data curation result can be useful in changing IoT devices to other status in both hardware and software levels. For example, for certain mobile sensing devices network systems, our big sensing data curation not only finds and corrects the errors for the data set from them, it also sends feedback to mobile sensing devices in the network to move to right places and to maintain healthy network topology.

4. THOUGHTS FOR FUTURE DIRECTIONS

In the process of building up the proposed On-Cloud IoT big sensing data curation framework, different research scopes and aspects should be discussed, as shown in **FIGURE 4**. Specifically, the following research objectives should be achieved.

Sensing Data Error Classification
To detect WSN sensing data errors, a categorization is performed first to formally define error types. With that classification, the network features for the cluster-head network WSN topology are analyzed and used for error detection. Specifically, in big sensing data cleaning, we use the scale-free topology of the network for error detection. Based on that topology constraint, error detection and recovery strategies can be designed within limited spatial-temporal data blocks rather than traversing the complete big sensing data set.

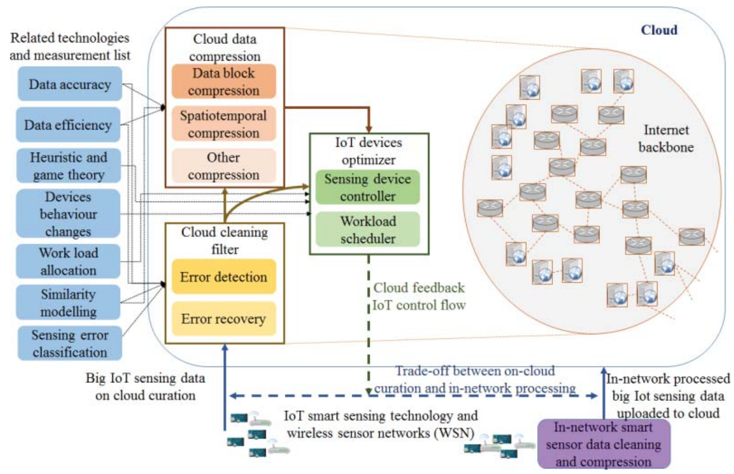

FIGURE 4. On-Cloud IoT big sensing data curation framework with associated properties.

Distributed Error Detection
For fast detection of data errors in big sensing data, and to make use of the cluster location feature of data, novel data error detection techniques are designed by exploiting the storage and computing potential of Cloud. The error detection and localization can be significantly accelerated because the detecting algorithm makes full use of complex network system topology and isolates the searching and comparing inside a sub-structure with high confidence. These detection and localization tasks are distributed to the Cloud with the MapReduce tool. The trade-offs between error detection efficiency and detection accuracy will also be considered and analyzed.

Distributed Error Recovery
A novel approach can be developed based on the prediction of recovery replacement data by making multiple data source approximations. The approximation process will use coverage information carried by data units to limit the algorithm in a small cluster of sensing data instead of the whole data spectrum. Specifically, in each sensing data cluster, a Euclidean distance based approximation is proposed to calculate a time series prediction curve. With the calculated time series, a detected error can be recovered with a predicted data value approximately. The proposed error recovery approach should achieve high accuracy in data approximation to replace the original data error. At the same time, with MapReduce based implementation for scalability, the experimental results also show significant efficiency on time saving.

Scalable and Distributed Compression
To offer novel compression methods on the Cloud for processing big sensor data sets, two important factors should be taken into consideration. Firstly, because of the sensor data strong topology and graph features, the compression should exploit them. Secondly, traditional compression is not scalable in terms of the Cloud environment. How to make it more scalable becomes critical. In our research, we combine our graph based compression and MapReduce to achieve the objective of distributed compression. Different data compression models will be developed and adopted. The compression can happen at data unit level, time series level, or even compound data blocks level.

Time Efficiency
Many big data processing applications have time limitation or high time efficiency requirement. In this paper, we highlighted that there is need of lightweight algorithms including compression and cleaning. So, under this theme, optimization, approximation and fast processing techniques are discussed.

Data Similarity

The clustering algorithm is developed for partitioning node sets and can compute the similarity between two time series. If two time series are similar to each other, they can be used for operations including mutual replacement and recovery. etc. So, to define appropriate similarity models is critical. Traditional similarity definitions and models should be calibrated before deployment. In our framework, distance calculation can evaluate similarity, hence to carry out further operations among time series, temporal and spatial prediction models should be designed. Currently, we have proposed some temporal prediction models based on data trend method and regression based method. However, there is still some disadvantages for our provided prediction models when applying them for cluster topologies. For instance, under the situation where two time series have similarity with sinusoidal functions, the regression based prediction models may totally miss effects. To offer a solution for the problem caused by the prediction models based on standard regression, we plan to offer a novel prediction model based on improved data regression with historical time series record. Suppose that there exist two time series, denoted as $T_1\{t_{11}, t_{12}, \ldots, t_{1m}\}$, and $T2\{t_{21}, t_{22}, \ldots, t_{2m}\}$ being involved in the similarity computation. m is the data sampling time stamp. The solution is designed for predicting the average dissimilarity of data trend for T_1 and T_2 in the next m rounds. A dissimilarity vector $V(d_1, d_2, \ldots, d_m)$ is calculated, where $d_i=t_{1i}-t_{2i}$. With the trends vector V, we redesign a regression model with a novel weight assignment to calculate the average dissimilarity from T_1 and T_2 [29].

Data Accuracy

Based on previous work of sensor data processing on the Cloud [23], [25], [26], fast detection of data errors and recovery in big sensing are quite challenging. For example, it is still an open and challenging topic to quickly locate and find sensor errors in a WSN by using cloud computational power. In our work, we initiate the process of error detection by using error models. If we compare with error detection in WSN in-network, our approach utilizes the massive data processing ability of Cloud to accelerate speed of error detection [28]. Furthermore, in more effective manner the topology features of complex networks are also analyzed in combination with the Cloud. Our proposed solution mainly focuses to achieve significance in time performance gains in detection of error with high accuracy with more consideration for optimized control of sensing devices in the IoT [30].

Scalability

As analyzed above, we can improve the efficiency in compression, cleaning and evaluation of big sensor data by deploying our data processing lifecycle on cloud data centers. However, to implement those techniques and make them scalable on the Cloud, our potential new design should offer better scalability. As a matter of fact, by the implementation with MapReduce and Hadoop, we can guarantee the scalability of the techniques in the given lifecycle.

Heuristic and Game Theory based Methods

WSN applications are quite customer oriented applications [31]. In other words, the understanding of sensor data is different from user to user. So, it is possible to get knowledge from domain experts during the data processing. In other words, heuristic methods can be developed for guiding the sensing device behavior in the IoT. Furthermore, when changing the sensing device behavior using our proposed data curation feedbacks, game theory can be involved for selecting among multiple candidate devices or systems in the IoT.

Optimized Workload Allocation between Cloud and IoT devices

Based on different processing capability and process models, data curation tasks can be assigned to the Cloud or smart sensing devices dynamically. Specifically, the trade-off between Cloud and device curation should be studied with the consideration of sensing data stream changes. Then, an adaptive approach will be developed to dynamically evaluate and schedule different tasks of sensing data curation. An optimized scheduling result will be generated to benefit both data curation targets and future IoT services.

Device Behavior Changes based on Data Curation

Sensing device behavior changes can happen in both the physical behavior of IoT devices such as location, and software level behavior such as data rate, communication topology and other adaptive algorithms. For example, in a sensing data curation process, if our Cloud-based error detection locates a sensing data source (device) reporting continuous errors, it can send feedback to the IoT to actuate some mobile sensing device to replace that data source to sustain a healthy IoT sensing data service.

5. CONCLUSIONS

This article presented a general roadmap for IoT big sensing data curation on the Cloud with background studies. This roadmap was designed according to current research trends, limitations and future directions. It combines different technologies from several fields including WSN, IoT, big data and the Cloud. The interaction and correlation between those four technologies was analysed. Especially the correlation between big data and WSN in IoT, the correlation between big data and Cloud, and the correlation between Cloud and WSN in IoT were investigated with a logical connected integrity. In addition, the interactions between Cloud and IoT were discussed. To the best of our knowledge, it was first proposed in this roadmap how to use Cloud big data curation results for sensing device manipulation, including mobile sensors and actuating equipment in the IoT. The design and construction of this roadmap would benefit current real world IoT big data processing applications such as medical/health monitoring, weather forecasting, environmental monitoring, industry production, social media analysis and business analysis where the users could access seamless IoT data services and realize remote physical device control through their mobile devices such as cell phones in a pervasive environment. In other words, all the complicated computation and communication of big data curation among sensing devices, IoT and Cloud platform, could be hidden from end users.

ABOUT THE AUTHORS

Chi Yang (chiyangit@gmail.com) received his Master of Science by Research degree (in computer science) from Swinburne University of Technology, Melbourne, Australia, and received his Ph.D in computer science at the University of Technology, Sydney (UTS), Australia. He is also a doctorate student in the University of Western Australia. Currently, Dr. Chi Yang is working for the Stratus project (https://stratus.org.nz/), as a full-time Postdoc. Research Fellow in the Unitec Institution of Technology, Auckland, New Zealand. Dr. Chi Yang has published research papers in different high-quality research conferences, journals and transactions. He is also PC member of different international conferences, such as TrustCom, ACISP and BDSE. His major research interests include WSN, IoT, Big Data Processing, Could Computing, parallel & distributed computing, privacy & security and XML data streams.

Deepak Puthal (Deepak.Puthal@uts.edu.au) is a Lecturer in the School of Computing and Communications at University of Technology Sydney (UTS), Australia. His research interests include cyber security, Internet of Things, distributed computing, and wireless communications. Puthal has a PhD in Computer Science from UTS, Australia. He has published in several international conferences and journals including IEEE and ACM transactions.

Saraju P. Mohanty (saraju.mohanty@unt.edu) is a Professor at the Department of Computer Science and Engineering, University of North Texas. He is an inventor of 4 US patents. He is an author of 220 peer-reviewed research articles and 3 books. He is currently the Editor-in-Chief (EiC) of the IEEE Consumer Electronics Magazine. He currently serves on the editorial board of 6 peer-reviewed international journals including IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems and ACM Journal on Emerging Technologies in Computing Systems. Prof. Mohanty has been the Chair of Technical Committee on Very Large Scale Integration (TCVLSI), IEEE Computer Society (IEEE-CS) to oversee a dozen of IEEE conferences. He serves on the steering, organizing, and program committees of several international conferences. More about him can be available from: http://www.smohanty.org.

Elias Kougianos (eliask@unt.edu) is Professor in Engineering Technology at the University of North Texas. He obtained his Ph.D. in electrical engineering from Louisiana State University in 1997. He is author or co-author of over 120 peer-reviewed journal and conference publications. He is a Senior Member of IEEE.

REFERENCES

[1] Y. Sun, H. Song, A. J. Jara and R. Bie, "Internet of Things and Big Data Analytics for Smart and Connected Communities", *IEEE Access*, vol(4), pp. 766-773, 2016.

[2] D. Tracey, "A Holistic Architecture for the Internet of Things, Sensing Services and Big Data", in *Proceeding of the 13rd IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing,* pp. 546-553, 2013.

[3] T. Zhu, S. Xiao, Q. Zhang, Y. Gu, P. Yi and Y. Li, "Emergent Technologies in Big Data Sensing: A Survey", *International Journal of Distributed Sensor Networks*, 2015(8), 2015.

[4] S. P. Mohanty, U. Choppali, and E. Kougianos, "Everything You wanted to Know about Smart Cities", *IEEE Consumer Electronics Magazine*, Volume 6, Issue 3, July 2016, pp. 60-70.

[5] IEC White Paper for "Internet of Things: Wireless Sensor Networks" http://www.iec.ch/whitepaper/pdf/iecWP-internetofthings-LR-en.pdf, accessed on March, 01, 2017.

[6] The Continuum: Big Data, Cloud & Internet of Things, https://www.ibm.com/blogs/internet-of-things/big-data-cloud-iot/, accessed on March, 01, 2017.

[7] Editor Summary, "Big data: science in the petabyte era: Community cleverness Required", *Nature* 455 (7209):1, 2008.

[8] S. Tsuchiya, Y. Sakamoto, Y. Tsuchimoto, V. Lee, "Big Data Processing in Cloud Environments," *FUJITSU Science and Technology Journal*, 48(2), pp. 159-168, 2012.

[9] M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, M. Zaharia, "A view of cloud computing," *Communications of the ACM,* 53(4) (2010), 50-58.

[10] M. Yuriyama and T. Kushida, "Sensor Cloud Infrastructure," in *Proceedings of the 13th International Conference on Netowrk-Based Information Systems*, pp.1-8, 2010.

[11] D. Puthal, S. Nepal, R. Ranjan, and J. Chen, "Threats to Networking Cloud and Edge Datacenters in the Internet of Things," *IEEE Cloud Computing*, vol. 3, no. 3, pp. 64-71, 2016.

[12] Big data and the Internet of Things: Two sides of the same coin? https://www.sas.com/en_au/insights/articles/big-data/big-data-and-iot-two-sides-of-the-same-coin.html, accessed on March, 01, 2017.
[13] M. L. Rajaram, E. Kougianos, S. P. Mohanty, and U. Choppali, "Wireless Sensor Network Simulation Frameworks: A Tutorial Review", *IEEE Consumer Electronics Magazine*, Volume 6, Issue 2, April 2016, pp. 63-69.
[14] S. Mukhopadhyay, D. Panigrahi and S. Dey, "Data aware, low cost error correction for wireless sensor networks," in *Proceedings of the IEEE Wireless Communications and Networking Conference*, pp.2494-2497, 2004.
[15] S. P. Mohanty, "iVAMS: A Paradigm Shift System Simulation Framework for the IoT Era", Keynote Presentation, *17th IEEE International Conference on Thermal, Mechanical and Multi-Physics Simulation and Experiments in Microelectronics and Microsystems*, 2016.
[16] X. L. Dong and D. Srivastava, "Big data integration," in *Proceedings of the 29th IEEE International Conference on Data Engineering*, pp. 1245-1248, 2013.
[17] D. Puthal, S. Nepal, R. Ranjan, and J. Chen, "DLSeF: A Dynamic Key-Length-Based Efficient Real-Time Security Verification Model for Big Data Stream," *ACM Transactions on Embedded Computing Systems*, vol 16, no. 2, PP. 51, 2017.
[18] B. Li, E. Mazur, Y. Diao, A. McGregor, P. Shenoy, "A platform for scalable one-pass analytics using mapreduce," in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, 2011, pp. 985-996.
[19] R. Kienzler, R. Bruggmann, A. Ranganathan, N. Tatbul, "Stream as you go: The case for incremental data access and processing in the cloud," *IEEE ICDE International Workshop on Data Management in the Cloud*, 2012.
[20] K. Shim, "MapReduce Algorithms for Big Data Analysis," in *Proceedings of the VLDB Endowment*, 5(12), pp. 2016-2017, 2012.
[21] Spark, http://spark.apache.org/, accessed on February 28, 2017.
[22] S. Mukhopadhyay, D. Panigrahi, and S. Dey, "Model Based Error Correction for Wireless Sensor Networks," *IEEE Transaction on Mobile Computing*, vol. 8(4), pp. 528-543, September, 2008.
[23] S. Slijepcevic, S. Megerian, and M. Potkonjak, "Charaterization of Lacation Error in Wireless Sensor Networks: Analysis and Application," in *Proceeding of the 2nd International Conference on Information Processing in Sensor Networks*, pp. 593-608, 2003.
[24] M. L. Rajaram, E. Kougianos, S. P. Mohanty, and P. Sundaravadivel, "A Wireless Sensor Network Simulation Framework for Structural Health Monitoring in Smart Cities", in *Proceedings of the 6th IEEE International Conference on Consumer Electronics - Berlin*, 2016, pp. 78-82.
[25] D. J. Wang, X. Shi, D. A. Mcfarland, and J. Leskovec, "Measurement Error in Network Data: A re-classification," *Social Networks*, vol. 34(4), pp. 396-409, October, 2012.
[26] K. Ni, N. Ramanathan, M. N. H. Chehade, L. Balzano, S. Nair, S. Zahedi, G. Pottie, M. Hansen, M. Srivastava, and E. Kohler, "Sensor Network Data Fault Types," *ACM Transactions on Sensor Networks*, vol. 5(3), May, 2009.
[27] C. Yang and J. Chen, "A Scalable Data Chunk Similarity based Compression Approach for Efficient Big Sensing Data Processing on Cloud," *IEEE Transactions on Knowledge and Data Engineering*, 2016, DOI: 10.1109/TKDE.2016.2531684.
[28] C. Yang, C. Liu, X. Zhang, S. Nepal and J. Chen, "A Time Efficient Approach for Detecting Errors in Big Sensor Data on Cloud," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 2, pp. 329-339, 2015.
[29] C. Yang, X. Zhang, C. Liu, J. Pei, K. Ramamohanarao and J. Chen, "A Spatiotemporal Compression based Approach for Efficient Big Data Processing on Cloud," *Journal of Computer and System Sciences*, vol. 80, no. 8, pp.1563-1583, 2014.
[30] N. Laptev, K. Zeng and C. Zaniolo, "Very fast estimation for result and accuracy of big data analytics: The EARL system," in *Proceedings of the 29th IEEE International Conference on Data Engineering*, pp. 1296-1299, 2013.
[31] S. Sharma et al., "Rendezvous based Routing Protocol for Wireless Sensor Networks with Mobile Sink," *The Journal of Supercomputing*, vol. 73, no. 3, pp.1168-1188, 2017.