

Big Sensor Data: A Survey

Yin Zhang¹(✉), Wei Li², Ping Zhou², Jun Yang², and Xiaobo Shi²

¹ School of Information and Safety Engineering, Zhongnan University of Economics and Law,
Wuhan, China

yin.zhang.cn@ieee.org

² School of Computer Science and Technology, Huazhong University of Science and Technology,
Wuhan, China

{weili.epic,pingzhou.cs,junyang.cs,xiaoboshi.cs}@qq.com

Abstract. The big data model has been shown to be very useful for transforming human-generated data and producing value-added business services. On the other hand, much less research has been performed on applying big data models towards machine-generated data such as from large-scale networked sensor systems. This paper surveys the use of big data models towards networked sensor systems and its application towards smart cities and urban informatics. We discuss use-cases for big sensor data research and summarize the common strategies and solutions, identify key techniques and challenges, and discuss a practical framework for handling big sensor data. The paper concludes with a discussion on future trends and research challenges of networked-based sensor systems. The aim of this survey paper is to be useful for researchers and designers to get insights into this emerging area, and motivate the development of practical solutions for application towards smart cities.

Keywords: Big data · Sensor data · Sensor networks

1 Introduction

Big data techniques are targeted towards solving system level problems that cannot be solved by conventional methods and technologies. With the emergence of new networked sensor technologies (e.g. large scale wireless sensor networks, body sensor networks [1–4], Internet/Network/Web/Vehicle-of-Things [5]), the next generation of big data systems will need to deal with machine-generated data from these forms of networked sensor systems [6–9]. This paper is focused on big sensor data systems (a term to conceptualize the application of big data models towards networked sensors) and surveys its applications towards smart cities and urban informatics. This is becoming an important research area [10]. An estimate from IBM is that the volume of machine-generated data sources will increase to 42 % of all data by 2020, up from 11 % in 2005¹. Much less research has been conducted for big sensor data systems compared with conventional big data systems [11].

¹ <http://www.ibm.com/big-data/au/en/big-data-and-analytics/operationsmanagement.html>.

Recently, there has been much research for smart cities. Many smart cities have been established in the world such as in Santander [12], Barcelona [13], and Singapore [14]. In Australia, the state of Tasmania is experimenting with the world’s first economy-wide intelligent sensor network technology (Sense-T)². Smart city applications would generate huge amounts of data from a wide variety of data sources ranging from environmental sensors, mobile phones, localization sensors, to data generated by people from social networks. While there is general agreement that the use of these big sensor data would lead to improved services in smart cities, many research challenges remain on how to integrate and utilize these big sensor data. In this survey paper, we take a bottom-up approach and use representative use-cases for big sensor data research, deconstruct the studies to identify key techniques and challenges, and attempt to summarize the common strategies and solutions utilized by different researchers. The use-cases selected for study have been implemented in the real-world.

Figure 1 illustrates the boom of the global data volume. While the amount of large datasets is drastically rising, it also brings about many challenging problems demanding prompt solutions.

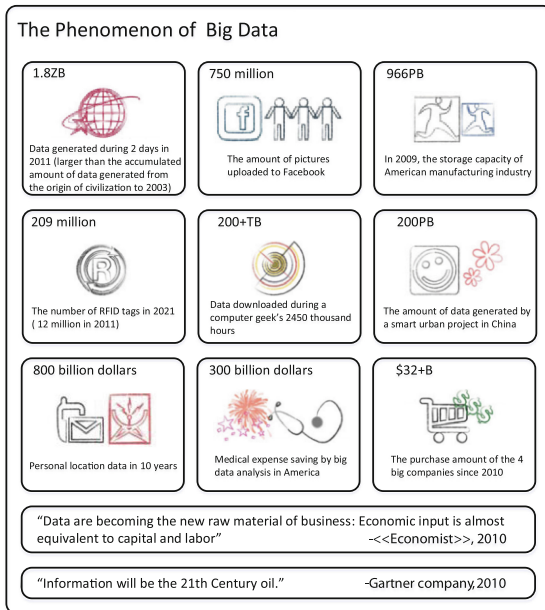


Fig. 1. The continuously increasing big data

² <http://www.sense-t.org.au/>.

2 Domains of Big Sensor Data

A definition³ for big data is given by as “Big data is high-volume, high-velocity, and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making”. An extended definition is that big data systems would involve the five Vs: (1) big volume of data (e.g. involving datasets of terabytes/petabytes), (2) variety of data types, (3) high velocity of data generation and updating, (4) veracity (uncertainty and noise) of acquired data, and (5) big value, as shown in Fig. 2. The first four V’s are concerned about data collection, preprocessing, transmission, and storage. The final V focuses on extracting value from the data using statistical and analytical methods.

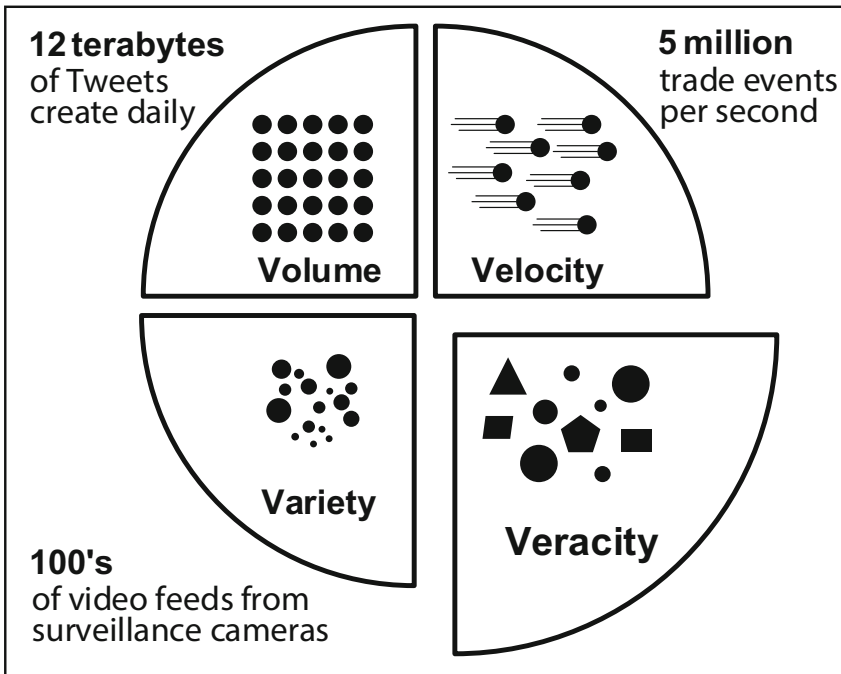


Fig. 2. The 4Vs feature of big data

3 Big Sensor Data Generation and Acquisition

If we take data as a raw material, data generation and data acquisition are an exploitation process, data storage is a storage process, and data analysis is a production process that utilizes the raw material to create new value.

³ <http://www.gartner.com/it-glossary/big-data>.

3.1 Data Generation

The data collection process consists of three modules: (1) data acquisition, (2) information extraction and cleaning, and (3) data integration, aggregation and representation. The inputs into the data collection process are the raw sensor data values $s(x,y,t)$ harvested from (multiple) sensor farms. The output of the process are the cleaned and aggregated data values $s'(x,y,t)$. The number of output data points would not be more than the number of input data points. The number of output data points may be less than the number of input data points due to data sampling and aggregation. Another characteristic is that the data collection process does not involve utilization of other data sources.

Data generation is the first step of big data. Given Internet data as an example, huge amount of data in terms of searching entries, Internet forum posts, chatting records, and microblog messages, are generated. Those data are closely related to people's daily life, and have similar features of high value and low density [15, 16]. Such Internet data may be valueless individually, but, through the exploitation of accumulated big data, useful information such as habits and hobbies of users can be identified, and it is even possible to forecast users' behaviors and emotional moods [17–20].

Moreover, generated through longitudinal and/or distributed data sources, datasets are more large-scale, highly diverse, and complex. Such data sources include sensors, videos, clickstreams, and/or all other available data sources. At present, main sources of big data are the operation and trading information in enterprises, logistic and sensing information in the IoT [21, 22], human interaction information and position information in the Internet world, and data generated in scientific research, etc. The information far surpasses the capacities of IT architectures and infrastructures of existing enterprises, while its real time requirement also greatly stresses the existing computing capacity [23].

3.2 Big Data Acquisition

As the second phase of the big data system, big data acquisition includes data collection, data transmission, and data pre-processing. During big data acquisition, once we collect the raw data, we shall utilize an efficient transmission mechanism to send it to a proper storage management system to support different analytical applications. The collected datasets may sometimes include much redundant or useless data, which unnecessarily increases storage space and affects the subsequent data analysis. For example, high redundancy is very common among datasets collected by sensors for environment monitoring. Data compression technology can be applied to reduce the redundancy. Therefore, data pre-processing operations are indispensable to ensure efficient data storage and exploitation:

1. **Data Collection:** Data collection is to utilize special data collection techniques to acquire raw data from a specific data generation environment. In addition to the aforementioned three data acquisition methods of main data sources, there are many other data collect methods or systems. For example, in scientific experiments, many special tools can be used to collect experimental data, such as magnetic spectrometers and radio telescopes. We may classify data collection methods from different perspectives. From the perspective of data sources, data collection methods can be

classified into two categories: collection methods recording through data sources and collection methods recording through other auxiliary tools.

Participatory sensing is considered as an emerging application scenario for efficient data crowdsourcing from smart device equipped ordinary citizens. Liu et al. in [24–26] presented a novel resource negotiation scheme bridging between dynamic sensing tasks and heterogeneous sensors. Liu et al. in [27–29] proposed a novel framework and subsequent participant selection and incentive mechanism for participatory crowdsourcing including the smart device users, central platform and multiple task publishers. In [30], existing incentive mechanisms are extensively surveyed and future research directions are clearly given. Liu et al. [31] extensively analyzed the relationship between energy consumption and smart device user behaviors, and then proposed a novel approach to select the optimal amount of participant while considering possible user rejections. Song et al. [32] introduced an energy consumption index to quantify the average degree of how participants feel disturbed by the energy cost, and proposed a suboptimal approach for participant selection under the multi-task sensing environment. Liu et al. [33] presented a quite novel family-based healthcare monitoring system for long-term chronic disease caring. Event detection systems and energy efficient approaches are given in [34, 35] including both centralized optimal approach and fully distributed suboptimal solutions by participatory sensing. Furthermore, Zhang et al. [36] focused on privacy leakage issues of participatory sensing and presented a participant coordination based architecture and flow to successfully protect user privacy. Finally, Yurur et al. in [37] presented a few posture detection schemes by using the sensor equipped smart devices. Finally, Liu et al. [38] presented a novel concept of quality of service (QoS) index to integrate the multi-dimensional QoS requirements to ensure the degree of QoS satisfactions

2. **Data Transportation:** Upon the completion of raw data collection, data will be transferred to a data storage infrastructure for processing and analysis. As we know, big data is mainly stored in a data center. The data layout should be adjusted to improve computing efficiency or facilitate hardware maintenance [39]. In other words, internal data transmission may occur in the data center. Therefore, data transmission consists of two phases: Inter-DCN transmissions and Intra-DCN transmissions [40–42].
3. **Data Pre-processing:** Because of the wide variety of data sources, the collected datasets vary with respect to noise, redundancy, and consistency, etc., and it is undoubtedly a waste to store meaningless data. In addition, some analytical methods have serious requirements on data quality. Therefore, in order to enable effective data analysis, we shall pre-process data under many circumstances to integrate the data from different sources, which can not only reduce storage expense, but also improve analysis accuracy.

4 Big Data Storage and Analysis

4.1 Data Storage

The explosive growth of data has more strict requirements on storage and management. In this section, we focus on the storage of big data. Big data storage refers to the storage and management of large-scale datasets while achieving reliability and availability of data accessing. We will review important issues including massive storage systems, distributed storage systems, and big data storage mechanisms. On one hand, the storage infrastructure needs to provide information storage service with reliable storage space; on the other hand, it must provide a powerful access interface for query and analysis of a large amount of data.

Traditionally, as auxiliary equipment of server, data storage device is used to store, manage, look up, and analyze data with structured Relational DataBase Management Systems (RDBMSs). With the sharp growth of data, data storage device is becoming increasingly more important, and many Internet companies pursue big capacity of storage to be competitive. Therefore, there is a compelling need for research on data storage.

4.2 Data Analysis

The analysis of big data mainly involves analytical methods for traditional data and big data, analytical architecture for big data, and software used for mining and analysis of big data. Data analysis is the final and the most important phase in the value chain of big data, with the purpose of extracting useful values, providing suggestions or decisions [43]. Different levels of potential values can be generated through the analysis of datasets in different fields. However, data analysis is a broad area, which frequently changes and is extremely complex. In this section, we introduce the methods, architectures and tools for big data analysis.

Traditional data analysis means to use proper statistical methods to analyze massive data, to concentrate, extract, and refine useful data hidden in a batch of chaotic datasets, and to identify the inherent law of the subject matter, so as to maximize the value of data. Data analysis plays a huge guidance role in making development plans for a country, understanding customer demands for commerce, and predicting market trend for enterprises. Big data analysis can be deemed as the analysis technique for a special kind of data. Therefore, many traditional data analysis methods may still be utilized for big data analysis.

In the dawn of the big data era, people are concerned how to rapidly extract key information from massive data so as to bring values for enterprises and individuals. Although the parallel computing systems or tools, such as MapReduce or Dryad, are useful for big data analysis, they are low levels tools that are hard to learn and use. Therefore, some high-level parallel programming tools or languages are being developed based on these systems. Such high-level languages include Sawzall, Pig, and Hive used for MapReduce, as well as Scope and Dryad LINQ used for Dryad.

5 Big Data Applications

The works in [44–47] used technical sensors embedded in the environment. The work in [44] for Barcelona aimed to expose the daily routines and patterns of people using the city bicycling program. The system collected data on when a bike is picked up or parked. The work in [45] embeds sensor into the city infrastructure. The data collected would be useful for studying the impact of air pollution on respiratory disease, and generating data to inform cycleway development. The work in [46] for Amsterdam used 2400 vehicle detector stations and 60 number plate recognition cameras to decrease the vehicle loss hours. The works in [48–52] used humans as sensors and collect data as people go about their daily routines. The work in [48] used Call Detail Records (CDRs) from a cellular network to characterize the human mobility. The work in [49] collects fine-grained environmental information in the city using data mined from crowdsourced bicycles. The work in [50] used the geolocation from photos on the Flickr social networking website to uncover the movements of tourists in Rome. The work in [51] optimize the 5G small cell networks accounting for the big data from user mobility in urban region. The works in [53, 54] are examples where both technical and human sensors are used. The work in [53] used a combination of GPS data and radio channel measurements from a cellphone network to give the instantaneous position of each mobile element. The work in [54] is for flood risk management in Brazil and used a combination of in-situ water sensors and human participatory sensing to give the water level height.

6 Conclusion, Open Issues, and Outlook

6.1 Data Collection

There are three important challenges for data gathering or collection in big sensor data systems: (1) Other than the 5V's found in conventional big data, big sensor data systems also need to consider an E (energy efficiency) to be fulfilled [55]. This requirement of energy efficiency should be applied at all stages in the big data pipeline whenever there is a non-rechargeable power source to be negotiated [56]. (2) Big sensor data systems will require the collection of data from hundreds of thousands of sensors which could be embedded in “Things” (e.g. garbage cans, street lights, etc.) which could be located and scattered anywhere in the environment. It is prohibitively expensive to implement a sensor relay infrastructure everywhere in the city for data gathering [57]. (3) The third challenge for data collection is to find ways to utilize available sensing infrastructures for new tasks or applications.

6.2 Data Inference

From the paper discussions, we see that a strong emphasis for big sensor data systems is towards using a variety of heterogeneous data sources or historical data to infer missing data or predict future trends in the spatial temporal sensing field. The challenge

is to find suitable models and techniques to integrate the various sources of data to solve the big data problem. They used an application from systems biology and showed the diverse data modalities which can be extracted from a single instance of DNA (e.g. the high dimensional expression data, the sparse protein interaction data, the sequence data, the annotation data, and the text mining data). Their approach used the kernel trick where data which has diverse data structures is all transformed into kernel matrices with the same size for combination.

6.3 Value Generation

We find an emphasis towards data-driven research especially using machine learning and complex network techniques. Whereas mathematical models and simulation techniques have been useful for studying the characteristics and behaviors of smaller scale systems, the move to study large-scale systems necessitate the development of new data-driven modeling techniques. Conventional mathematical and simulation models face difficulty in acquiring the correct parameters or in dealing with unpredictable and unknown factors. Amongst, the machine learning approaches, the emergence of deep learning and crossdomain techniques show potential to discover hidden insights and trends in big sensor data systems [58].

6.4 Real-World Applications

Many applications for smart cities (e.g. for earthquake/disaster early warning system, air pollution monitoring) require (near) real-time performance to serve its function [59]. Table 1 presents various typical big sensor data applications. This will drive the

Table 1. Typical Big Sensor data applications

Project	City	Sensors	Data	Analysis	Value
U-Air [60]	Beijing	GPS, environmental sensors, cell phone	Meteorological data, traffic data, human mobility data, Point-of-Interest data, road network data	Neural Network, Conditional Random Field (CRF).	Achieved the inferring of PM10 for entire Beijing in five minutes (near real-time performance) with an accuracy of 82 %.
Sensing city [61]	Christchurch	Environmental Sensors	Air pollution levels, noise levels, water use, traffic flow		Incorporate sensors into building infrastructure
Amsterdam smart city [62]	Amsterdam	2400 vehicle detector stations, 60 number plate recognition cameras	Traffic flow		Traffic flow management to decrease vehicle loss hours by as much as 10 %

“Velocity” characteristic for big sensor data systems. Currently, most (if not all) research on big sensor data systems do not consider this aspect (are performed offline), and research is conducted using historical or past data. In the future, we anticipate the research and development of big sensor data systems where real-time analytics will be performed on large volumes of recently acquired data from multiple sensor farms, and using a number of diverse and historical sources.

References

1. Chen, M., et al.: Body area networks: a survey. *Mob. Netw. Appl.* **16**(2), 171–193 (2011)
2. Fortino, G., Galzarano, S., Gravina, R., Li, W.: A framework for collaborative computing and multi-sensor data fusion in body sensor networks. *Inf. Fusion* **22**, 50–70 (2015)
3. Fortino, G., Giannantonio, R., Gravina, R., Kuryloski, P., Jafari, R.: Enabling effective programming and flexible management of efficient body sensor network applications. *IEEE Trans. Hum. Mach. Syst.* **43**(1), 115–133 (2013)
4. Wan, J., Zou, C., Ullah, S., Lai, C.F., Zhou, M., Wang, X.: Cloud-enabled wireless body area networks for pervasive healthcare. *IEEE Netw.* **27**(5), 56–61 (2013)
5. Zheng, K., Zheng, Q., Chatzimisios, P., Xiang, W., Zhou, Y.: Heterogeneous vehicular networking: a survey on architecture, challenges, and solutions. *IEEE Commun. Surv. Tutor.* **17**(4), 2377–2396 (2015)
6. Taleb, T., Ksentini, A.: On alleviating MTC overload in EPS. *Ad Hoc Netw.* **18**, 24–39 (2014)
7. Lin, K., Xu, T., Song, J., Qian, Y., Sun, Y.: Node scheduling for all-directional intrusion detection in SDR-based 3D WSNs. *IEEE Sens. J.* (2016). doi:[10.1109/JSEN.2016.2558043](https://doi.org/10.1109/JSEN.2016.2558043)
8. Lin, K., Chen, M., Deng, J., Hassan, M.M., Fortino, G.: Enhanced fingerprinting and trajectory prediction for IoT localization in smart buildings. *IEEE Trans. Autom. Sci. Eng.* **13**(3), 1294–1307 (2016)
9. Hossain, M.S.: Cloud-supported cyber-physical localization framework for patients monitoring. *IEEE Syst.* **PP**, 1–10 (2015). doi:[10.1109/JSYST.2015.2470644](https://doi.org/10.1109/JSYST.2015.2470644)
10. Zhang, Y.: GroRec: a group-centric intelligent recommender system integrating social, mobile and big data technologies. *IEEE Trans. Serv. Comput.* (2016). doi:[10.1109/TSC.2016.2592520](https://doi.org/10.1109/TSC.2016.2592520)
11. Zheng, K., Yang, Z., Zhang, K., Chatzimisios, P., Yang, K., Xiang, W.: Big data-driven optimization for mobile networks toward 5G. *IEEE Netw.* **30**(1), 44–51 (2016)
12. Sanchez, L., Muñoz, L., Galache, J.A., Sotres, P., Santana, J.R., Gutierrez, V., et al.: SmartSantander: IoT experimentation over a smart city testbed. *Comput. Netw.* **61**, 217–238 (2014)
13. Bakici, T., Almirall, E., Wareham, J.: A smart city initiative: the case of Barcelona. *J. Knowl. Econ.* **4**(2), 135–148 (2013)
14. Kloeckl, K., Senn, O., Ratti, C.: Enabling the real-time city: LIVE Singapore! *J. Urban Technol.* **19**(2), 89–112 (2012)
15. Lin, K., Wang, W., Wang, X., Ji, W., Wan, J.: Qoe-driven spectrum assignment for 5 g wireless networks using sdr. *IEEE Wirel. Commun.* **22**(6), 48–55 (2015)
16. Lin, K., Song, J., Luo, J., Ji, W., Hossain, M.S., Ghoneim, A.: GVT: green video transmission in the mobile cloud networks. *IEEE Trans. Circ. Syst. Video Technol.* (2016). doi:[10.1109/TCSVT.2016.2539618](https://doi.org/10.1109/TCSVT.2016.2539618)

17. Chen, M., Zhang, Y., Li, Y., Mao, S., Leung, V.C.: EMC: emotion-aware mobile cloud computing in 5G. *IEEE Netw.* **29**(2), 32–38 (2015)
18. Hossain, M.S., Muhammad, G., Alhamid, M.F., Song, B., Al-Mutib, K.: Audio-visual emotion recognition using big data towards 5 g. *Mob. Netw. Appl.* (2015). doi:[10.1007/s11036-016-0685-9](https://doi.org/10.1007/s11036-016-0685-9)
19. Zhou, L.: Specific versus diverse computing in media cloud. *IEEE Trans. Circ. Syst. Video Technol.* **25**(12), 1888–1899 (2015)
20. Zhou, L., Rodrigues, J.J., Oliveira, L.M.: QoE-driven power scheduling in smart grid: architecture, strategy, and methodology. *IEEE Commun. Mag.* **50**(5), 136–141 (2012)
21. Chen, M., Wan, J., González, S., Liao, X., Leung, V.C.: A survey of recent developments in home M2 M networks. *IEEE Commun. Surv. Tutor.* **16**(1), 98–114 (2014)
22. Hossain, M.S., Muhammad, G.: Cloud-assisted industrial internet of things (iiot)-enabled framework for health monitoring. *Comput. Netw.* **101**, 192–202 (2016)
23. Taleb, T., Ksentini, A., Kobbane, A.: Lightweight mobile core networks for machine type communications. *IEEE Access* **2**, 1128–1137 (2014)
24. Liu, C.H., Leung, K.K., Bisdikian, C., Branch, J.W.: A new approach to architecture of sensor networks for mission-oriented applications. In: *SPIE Defense, Security, and Sensing*, International Society for Optics and Photonics, p. 73490L (2009)
25. Liu, C.H., He, T., Lee, K.W., Leung, K.K., Swami, A.: Dynamic control of data ferries under partial observations. In: *IEEE WCNC*, pp. 1–6 (2010)
26. Liu, C.H., Fan, J., Branch, J.W., Leung, K.K.: Toward qoi and energy-efficiency in internet-of-things sensory environments. *IEEE Trans. Emerg. Top. Comput.* **2**(4), 473–487 (2014)
27. Liu, C.H., Fan, J., Hui, P., Wu, J., Leung, K.K.: Toward QoI and energy efficiency in participatory crowdsourcing. *IEEE Trans. Veh. Technol.* **64**(10), 4684–4700 (2015)
28. Liu, C.H., Hui, P., Branch, J.W., Bisdikian, C., Yang, B.: Efficient network management for context-aware participatory sensing. In: *Proceedings of the 2011 8th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (Secon)*, pp. 116–124 (2011)
29. Liu, C.H., Yang, B., Liu, T.: Efficient naming, addressing and profile services in internet-of-things sensory environments. *Ad Hoc Netw.* **18**, 85–101 (2014)
30. Gao, H., Liu, C.H., Wang, W., Zhao, J., Song, Z., Su, X., et al.: A survey of incentive mechanisms for participatory sensing. *IEEE Commun. Surv. Tutor.* **17**(2), 918–943 (2015)
31. Liu, C.H., Zhang, B., Su, X., Ma, J., Wang, W., Leung, K.K.: Energy-aware participant selection for smartphone-enabled mobile crowd sensing. *IEEE Syst. J.* (2015). doi:[10.1109/JSYST.2015.2430362](https://doi.org/10.1109/JSYST.2015.2430362)
32. Song, Z., Zhang, B., Liu, C.H., Vasilakos, A.V., Ma, J., Wang, W.: QoI-aware energy-efficient participant selection. In: *Proceedings of the 2014 Eleventh Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, pp. 248–256 (2014)
33. Liu, C.H., Wen, J., Yu, Q., Yang, B., Wang, W.: HealthKiosk: a family-based connected healthcare system for long-term monitoring. In: *Proceedings of 2011 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 241–246 (2011)
34. Liu, C.H., Zhao, J., Zhang, H., Guo, S., Leung, K.K., Crowcroft, J.: Energy-efficient event detection by participatory sensing under budget constraints. *IEEE Syst. J.* (2016). doi:[10.1109/JSYST.2016.2533538](https://doi.org/10.1109/JSYST.2016.2533538)
35. Zhang, B., Song, Z., Liu, C.H., Ma, J., Wang, W.: An event-driven qoi-aware participatory sensing framework with energy and budget constraints. *ACM Trans. Intell. Syst. Technol.* **6**(3), 42 (2015)
36. Zhang, B., Liu, C.H., Lu, J., Song, Z., Ren, Z., Ma, J., Wang, W.: Privacy-preserving QoI-aware participant coordination for mobile crowdsourcing. *Comput. Netw.* **101**, 29–41 (2016)

37. Yurur, O., Liu, C.H., Moreno, W.: Unsupervised posture detection by smartphone accelerometer. *Electron. Lett.* **49**(8), 562–564 (2013)
38. Liu, C.H., Leung, K.K., Gkelias, A.: A generic admission-control methodology for packet networks. *IEEE Trans. Wirel. Commun.* **13**(2), 604–617 (2014)
39. Chen, M., Zhang, Y., Li, Y., Hassan, M.M., Alamri, A.: AIWAC: affective interaction through wearable computing and cloud technology. *IEEE Wirel. Commun.* **22**(1), 20–27 (2015)
40. Peng, L.: On the future integrated datacenter networks: designs, operations, and solutions. *Opt. Switch. Netw.* **19**, 58–65 (2016)
41. Peng, L., Sun, Y., Chen, M., Park, K.: Datacenter-oriented elastic optical networks: architecture, operation, and solutions. *TIIS* **8**(11), 3955–3966 (2014)
42. Peng, L., Youn, C.H., Tang, W., Qiao, C.: A novel approach to optical switching for intradatacenter networking. *J. Lightwave Technol.* **30**(2), 252–266 (2012)
43. Zhang, Y., Chen, M., Mao, S., Hu, L., Leung, V.C.: Cap: community activity prediction based on big data analysis. *IEEE Netw.* **28**(4), 52–57 (2014)
44. Froehlich, J., Neumann, J., Oliver, N.: Measuring the pulse of the city through shared bicycle programs. In: *Proceedings of UrbanSense08*, pp. 16–20 (2008)
45. Zhang, Y.: iDoctor: personalized and professionalized medical recommendations based on hybrid matrix factorization. *Future Gener. Comput. Syst.* (2016). doi:[10.1016/j.future.2015.12.001](https://doi.org/10.1016/j.future.2015.12.001)
46. Zhang, Y.: Health-CPS: healthcare cyber-physical system assisted by cloud and big data. *IEEE Syst. J.* (2015). doi:[10.1109/JSYST.2015.2460747](https://doi.org/10.1109/JSYST.2015.2460747)
47. Yuan, W., Deng, P., Taleb, T., Wan, J., Bi, C.: An unlicensed taxi identification model based on big data analysis. *IEEE Trans. Intell. Transp. Syst.* **17**(6), 1703–1713 (2016)
48. Becker, R., Cáceres, R., Hanson, K., Isaacman, S., Loh, J.M., Martonosi, M., et al.: Human mobility characterization from cellular network data. *Commun. ACM* **56**(1), 74–82 (2013)
49. Chen, M., Ma, Y., Song, J., Lai, C.F., Hu, B.: Smart clothing: connecting human with clouds and big data for sustainable health monitoring. *Mob. Netw. Appl.* (2016). doi:[10.1007/s11036-016-0745-1](https://doi.org/10.1007/s11036-016-0745-1)
50. Chen, M., Hao, Y., Qiu, M., Song, J., Wu, D., Humar, I.: Mobility-aware caching and computation offloading in 5 g ultra-dense cellular networks. *Sensors* **16**(7), 974 (2016)
51. Ge, X., Ye, J., Yang, Y., Li, Q.: User mobility evaluation for 5G small cell networks based on individual mobility model. *IEEE J. Sel. Area Commun.* **34**(3), 528–541 (2016)
52. Carreño, P., Gutierrez, F.J., Ochoa, S.F., Fortino, G.: Supporting personal security using participatory sensing. *Concurr. Comput. Pract. Exp.* **27**(10), 2531–2546 (2015)
53. Calabrese, F., Colonna, M., Lovisolo, P., Parata, D., Ratti, C.: Real-time urban monitoring using cell phones: a case study in Rome. *IEEE Trans. Intell. Transp. Syst.* **12**(1), 141–151 (2011)
54. Chen, M.: NDNC-BAN: supporting rich media healthcare services via named data networking in cloud-assisted wireless body area networks. *Inf. Sci.* **284**, 142–156 (2014)
55. Chen, M., Mao, S., Liu, Y.: Big data: a survey. *Mob. Netw. Appl.* **19**(2), 171–209 (2014)
56. Ge, X., Yang, B., Ye, J., Mao, G., Wang, C.X., Han, T.: Spatial spectrum and energy efficiency of random cellular networks. *IEEE Trans. Commun.* **63**(3), 1019–1030 (2015)
57. Chen, M., Hao, Y., Li, Y., Lai, C.F., Wu, D.: On the computation offloading at ad hoc cloudlet: architecture and service modes. *IEEE Commun. Mag.* **53**(6), 18–24 (2015)
58. Ji, W., Li, Z., Chen, Y.: Joint source-channel coding and optimization for layered video broadcasting to heterogeneous devices. *IEEE Trans. Multimed.* **14**(2), 443–455 (2012)
59. Foschini, L., Taleb, T., Corradi, A., Bottazzi, D.: M2M-based metropolitan platform for IMS-enabled road traffic management in IoT. *IEEE Commun. Mag.* **49**(11), 50–57 (2011)

60. Zheng, Y., Liu, F., Hsieh, H.P.: U-Air: when urban air quality inference meets big data. In: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1436–1444 (2013)
61. Sensing City Website (2016). <http://sensingcity.org>
62. Amsterdam Smart City Website (2016). <http://amsterdamsmartcity.com>