

## Review Article

# Emergent Technologies in Big Data Sensing: A Survey

Ting Zhu,<sup>1</sup> Sheng Xiao,<sup>2</sup> Qingquan Zhang,<sup>1</sup> Yu Gu,<sup>3</sup> Ping Yi,<sup>4</sup> and Yanhua Li<sup>5</sup>

<sup>1</sup>University of Maryland, Baltimore County, Baltimore, MD 21250, USA

<sup>2</sup>Hunan University, Changsha, Hunan 410082, China

<sup>3</sup>IBM Research, Austin, TX 78758, USA

<sup>4</sup>Shanghai Jiaotong University, Shanghai 200240, China

<sup>5</sup>University of Minnesota Twin Cities, Minneapolis, MN 55416, USA

Correspondence should be addressed to Ting Zhu; [zt@umbc.edu](mailto:zt@umbc.edu)

Received 2 October 2014; Revised 2 March 2015; Accepted 2 March 2015

Academic Editor: Joel Rodrigues

Copyright © 2015 Ting Zhu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

When the number of data generating sensors increases and the amount of sensing data grows to a scale that traditional methods cannot handle, big data methods are needed for sensing applications. However, big data is a fuzzy data science concept and there is no existing research architecture for it nor a generic application structure in the field of sensing. In this survey, we explore many scattered results that have been achieved by combining big data techniques with sensing and present our vision of big data in sensing. Firstly, we outline the application categories to generally summarize existing research achievements. Then we discuss the techniques proposed in these studies to demonstrate challenges and opportunities in this field. Finally, we present research trends and list some directions of big data in future sensing. Overall, mobile sensing and its related studies are hot topics, but other large-scale sensing researches are flourishing too. Although there are no “big data” techniques acting as research platforms or infrastructures to support various applications, multiple data science technologies, such as data mining, crowd sensing, and cloud computing, serve as foundations and bases of big data in the world of sensing.

## 1. Introduction

Big data, as a concept, was first proposed by META Group analyst Doug Laney in the 2001 research report [1] and his related lectures. Increasing volume (amount of data), velocity (speed of data), and variety (range of data types and sources) are used as three important characteristics to define big data. As for now, two new characters, value and veracity, are added by some organizations [2] to further illustrate the necessary properties of big data. This “5Vs” model, which is used for describing big data and its related challenges, like data capture, storage, search, sharing, transfer, analysis, and visualization, is a hot topic in current data science research field.

In the field of sensing, special issues are generated. With the exponential increasing number of data generating devices (such as computers, tablets, and sensors, especially smartphones), vast amount of data needs to be processed. Research methods for big data can be applied to various fields by utilizing sensing techniques, such as science, engineering,

medicine, health care, finance, business, and ultimately the whole society. However, currently, there is still no generic and systematic big data research model in the world of sensing.

The vision of data processing in future sensing is vague and relevant infrastructures and structures have not yet been well defined. A road map has yet to be made, even though there have been published research papers. Techniques to collect, analyze, or process sensing data are usually ameliorated from existing data sciences, and, until now, there is no clear definition to describe what is “big data.” The most intuitive understanding that comes into people’s mind is a large amount of data reflecting the space domain of data sourcing. In the 5Vs model, volume and variety are directly relevant to this understanding. In the world of sensing, large amount of data is usually gotten from a large sensing area, for example, town or city level sensing or the applications for Internet of Things.

Town or city level sensing relies not only on sensors within city infrastructures, but also on a large number of device owners willing to sense and contribute their data to

data aggregation platforms. A survey result shows that every day we create more than 2.5 quintillion bytes of data, and a prediction says that, in 2016, over 4.1 terabytes of data will be generated per day per square kilometer in urbanized land area. Furthermore, in 2016, it is estimated that 39.5 billion dollars will be spent on smart city technologies, up from 8.1 billion dollars in 2010 [3]. The pervasive use of mobile phones and other similar mobile sensing devices will account for a dominant portion of aforementioned increment. Smartphones enable everyone to collect data at any time and place. Although some sensing data may not be valuable to the sensor owner, they can be valuable to the scientific community.

Currently, building a generic sensing platform for a city scale data application faces many challenges. The first challenge is how to design a system in which users can benefit from data sharing [4, 5]. As one of the most important parts of city scale sensor, personal sensing devices are still within the “owner-is-the-user” model. Getting considerable benefits without personal information leakage is the baseline of making full use of individual sensing data, as privacy and security are general concerns. The second challenge is how to effectively collect the data scattered in the individual sensing devices. The large amount of data generated by distributed sensors typically does not have a central control or a centralized accounting device that can be notified when new data is generated.

Internet of Things (IoT) is a much broader concept which was formally proposed by Kevin Ashton in 2009 [6] as a technique for uniquely identifiable objects and their virtual representations in an Internet-like structure. This concept later develops into a worldwide architecture for sensing, computing, and communication. Such large amount of computing and communication resources enables sensing, capturing, collecting, and processing real-time data from billions of distributed devices and serves a great number of applications including health care, climate monitoring, earthquake detection, volcano monitoring, power grid control, smart home, and business intelligence [7]. In the prospective future, IoT will not be restricted to uniquely identifiable objects and their virtual representations. It will include billions of devices which pour vast amount of data to our existing network. Sensor networks increasingly enable applications and services to interact with the physical world; such services may be located across the Internet from sensing networks. Internet techniques, cloudy services, and smart assets are being used to store and analyze these data to improve networks’ features, such as scalability and availability, which are required by future sensor networks that contain millions or even billions of devices.

Beside the “spacial domain,” “time domain” sensing data management is also a hot topic in data science. Real-time processing of large amount of sensing data normally requires very high computing abilities and large-scale hardware infrastructures. Even with sufficient resources, it is still challenging to reliably compile large-scale time-stamped data set. As examples in [8] demonstrated, the physical restrictions in the measurement systems, the limitations of computing abilities, the energy capacity, and the difficulties

posed by certain measurement problems, will result in data loss, data errors, and ambiguities in data inferences. Long period sensing data analysis and storage are also important research topics in “time domain,” especially in the field of environmental monitoring and object behavior analysis [9]. Remote sensing technologies are widely applied in environment related research fields. The data acquired and accumulated (usually in the form of images) requires large storage space and highly efficient analysis methods. For object behavior analysis, various techniques are applied and usually long term monitoring is required. Take [9] as an example; the accurate and continuous monitoring of lakes and inland seas is applied to analyze impact of climate changes and human activities on the terrestrial water resources since 1993.

In the rest of this survey paper, we first introduce the applications that motivate the big data sensing research in Section 2 and then summarize the existing techniques for big data sensing in Section 3 and propose the future research directions in Section 4. Finally, we conclude this paper in Section 5.

## 2. Applications

In this section, we first introduce smartphones enabled big data applications including Internet of Things, crowd sensing, environment monitoring, and health monitoring. Then, we discuss the common issue of smartphone enabled applications.

*2.1. Applications Enabled by Smartphones.* Today’s smartphones serve not only as important communication devices, but also as computing and sensing devices with rich sets of embedded sensors, such as accelerometers, digital compasses, gyroscopes, GPS, microphones, and cameras. Generally, combining growing computing abilities, these sensors are enabling new applications across a wide variety of domains, such as human health care, social networks, safety, environmental or climate monitoring, and transportation. They lead to a new research area called mobile phone sensing [3, 10–12]. As the number of smartphone users increases rapidly across the whole world, large amount of data is generated, transferred, aggregated, and analyzed. The ubiquity of mobile phones and the increasing size of the data generated by sensors and applications lead to a new research domain across computing and social science. Big data, as a data science to process high volume information, is consequently involved in this field. Researchers have begun to address big data issues by using large-scale mobile data as an input to characterize and understand real-life phenomena, including individual traits, human mobility, communication, and interaction patterns.

*2.1.1. Smartphones for Internet of Things.* Semantic-oriented vision, as one of the broader visions of Internet of Things (IoT), emphasizes on data integration and management from vast number of smart devices, such as smartphones, pads, sensor nodes, and other devices with the ability to send out information [13]. As one of the most important constituent parts of IoT, smartphones can not only provide

more information than other devices, but also act as information collecting and distributing terminals. How to integrate diverse information is a big challenge of utilizing smartphones for IoT. In [14], the authors proposed an approach to optimize data collection performance by updating routing structure of smartphones, which can also be applied to large amount of data processing in IoT.

Mobile data collected from wireless sensor networks are strongly spatial correlated; however, traditional methods are usually in static setting and the so-called optimal data collection trees are fixed and their performance suffers from link problems when mobile users change virtual sinks. The model proposed in this paper initializes an optimized tree and updates it according to users' accessing virtual sinks by locally modifying the previously constructed data collection tree. Their model is easy to implement, has low cost, and provides real-time data acquirement even when updating the tree structure. Similar techniques can be applied to vast amount of data collection and distribution structures by dynamically modifying the mobile access routing structure to achieve optimal performance [15, 16]. Similar to [14], the authors proposed a model for data collection by using smartphones in [17]. Instead of optimizing data accessing routing, this paper focuses on construction of data center and relative database. By connecting smartphones and data center to the Internet, users can monitor sensor information remotely and in real-time.

*2.1.2. Smartphones for Crowd Sensing.* Static sensing is traditional and mature but has node coverage, maintenance, and scalability issues. Mobile crowd sensing is more flexible, manageable, and scalable, especially when vast numbers of smartphones are used as sensing nodes in cities or towns. The fast increasing number of smartphone users, various inherent mobile applications, and exponential increasing capacity of 3G/4G networks lead to this new mobile sensing paradigm. Currently, smartphones are used as sensors for localization, personal/surrounding context recognition, traffic monitoring, and other daily life related applications. But, in the near future, other applications, such as environmental pollution detection, health care monitoring, and social life analysis, will generate large amount of sensing data. Unlike conventional sensor networks, mobile crowd sensing is more human related; therefore privacy and security should be carefully considered. Otherwise, smartphone users will be unwilling to share their devices and subsequent data with others. To the best of our knowledge, there is no mature platform for mobile crowd sensing and researchers are working in that direction. For example, researchers proposed Medusa [18], which can provide high-level abstractions for stages in completing crowd sensing tasks and a distributed system which can coordinate the execution of these tasks between smartphones and the cloud.

How to attract users to participate in projects of crowd sensing becomes a very important problem. Unlike conventional methods of constructing sensor networks, there is less support from institutions or organizations. The willingness of personal users decides the scale of mobile crowd sensing. In [19], two system models are proposed. The platform-centric

model is designed to award participating users who share information with others, and the user-centric model can help individuals to ask for a reserve price for their sensing service. The former is run as a Stackelberg game to maximize the utility of this platform and no user can improve its utility by deviating from the current strategy unilaterally. In this model, the total benefit for user is fixed and competition exists. The second model introduces a strategy in which users calculate their won cost and ask for prices. In this model, users receive payments which are not lower than their asked prices, if their prices are accepted. These two models normalize user behaviors in crowd sensing networks to protect users' benefits, in order to encourage individuals to join in sharing networks.

In the above two paragraphs, we introduced two popular applications in mobile crowd sensing. With the rapidly increasing number of smartphones, more and more research topics are developed, like strategy of data collection, mobile sensing performance, communication quality, privacy and security, energy efficiency, and other categories of applications. The fast development of mobile crowd sensing not only leads to a generation of vast amounts of data, but also requires fast and efficient data processing abilities. Science of big data can be one of mobile crowd sensing's fundamental research fields [20].

*2.1.3. Smartphones for Environment Monitoring.* Weather and environment monitoring are usually the responsibility of governments and some specific institutions. But if billions of mobile phones can be utilized for such jobs, more diversified and abundant information can be used to improve human's living conditions. Currently, combined with a cloud of supporting web services, large amount of smart mobile devices make such a distributed data collection infrastructure possible, though not immediately usable. An appropriate platform can be used in this field for further applications. Paper [21] proposed the Personal Environmental Impact Report (PEIR), a system that combines web and personal mobile techniques to inform users of environmental impact and exposure, which can help people make more informed and responsible decisions. PEIR is built on location tracing and GPS records that are sampled. Based on the GPS information, users' trips are predicted and environmental impact or exposure measurements are aggregated from each trip. This platform can be used for a number of applications, such as traffic condition measurement, environmental pollution monitoring, and vehicle emission estimating. Though only four applications were proposed by the authors, new models can be developed based on this platform and scalability, stability, performance, and usability are the foreseeable promising directions for this kind of platforms.

While the above paper [21] shows an example of platform building for environment monitoring using smartphones, [22] is a good instance to show a specialized application. Nericell is a system designed to make full use of mobile phone sensing components to provide rich sensing information about the road and traffic conditions. In this system, microphones, GSM radios, and GPS sensors are organized to detect potholes, bumps, braking, and honking. The large amounts

of mobile phones and the variety of information from each mobile device can guarantee an effective road and traffic condition detection without significant energy consumption. Unlike similar approaches which use meaningful digital information, Nericell also utilizes sharp changes of analog signals like acceleration alternation from accelerometers and then builds certain models to detect incontinuous vehicle running behaviors. This type of application largely enriches the utilization of smartphone sensors and shows a broader prospect of mobile sensing.

*2.1.4. Smartphones for Health Monitoring.* On-body sensing with small, inexpensive, and low-power sensors has led to series of research on human health monitoring. With the improvement of artificial intelligence and computing capability of mobile devices, machine learning has been applied to provide health suggestions by analyzing data acquired by sensors [23]. Mobile phones, as the “most frequently carried devices,” are the best human behavior monitor devices. Without buying expensive sensors or carrying additional heavy sensors, people can simply get their activities and health suggestions from their cell phones. Researchers have found that regular daily activity is important to people’s physical and psychological health, regardless of their static body conditions. Therefore, mobile phones can be the best choice over any other approaches if they are carefully utilized. Paper [24] introduces UbiFit Garden, a system that is designed to interpret and reflect on the data about people’s physical activities, and provides certain health information to users. This system is comprised of three parts: (i) a fitness device which uses 3D accelerometer and barometer to acquire and process data, (ii) an interactive application which runs on mobile phones to interact with users about practice activities, and (iii) a glanceable display that presents key information about the user’s physical activities and goal attainments. Though a special designed fitness device is used in this paper, the proposed technique can leverage the 3D accelerometers and barometers in smartphones as well. Based on this platform, a smartphone network can be built and people’s health information can be aggregated, compared, and analyzed by central servers; then, useful health suggestions are sent back to individuals’ smartphones based on machine learning or doctor suggestions (if certain health institutions are involved).

*2.1.5. Common Issue of Smartphone Related Applications.* In previous sections, we introduced different applications enabled by smartphones. One common research issue among the wide variety of applications that use smartphones as sensing data sources is power consumption. With the development of smartphones, more and more embedded devices and powerful processors are attached. Therefore, smartphones consume significantly more energy than the previous generation of cellular phones. A smartphone which never stops using its GPS, not to mention those applications which might combine GPS with other components, may run out of energy within several hours. So, for every newly developed application, power consumption is an unavoidable problem.

Crowd sensing with smartphones (and its advantages) is discussed in the previous subsection; for example, observing and measuring phenomena over a large area by collecting and sharing data is implied [25]. However, due to limited battery storage, smartphones usually cannot support nonstop sensing tasks. Thus, for every newly developed application, power consumption should be considered. This paper proposed a Mobile Publish/Subscribe (MoPS) middleware system which focuses on the requirements of mobile and resource-constrained environments with a goal of reducing overall energy consumption and building a general platform for mobile crowd sensing. The basic idea of MoPS is filtering out uninteresting data from mobile Internet-connected objects to avoid redundant information being transferred to the cloud. The filter method for sensor data depends on contexts before transmission. For example, a specific application is covered by multiple smartphones and only one needs to transfer data to the cloud.

Reference [26] focuses on how to save power from smartphones, presence services. The main idea of this paper is similar to MoPS. By analyzing a large mobile data challenge data set, smartphones learn and infer user presence status by using available context data to enable nonintrusive and energy-efficient maintenance automatically. Besides using the calendar or other settings as static grounds for status alternating, GPS, accelerometers, and microphones are applied to sense user’s behaviors. Whenever people enter an “unavailable” or another status in which it is not convenient for users to response to a real-time conversation, the presence service frequency is reduced. Since smartphones usually have a considerable number of present related applications, turning off presence service is an effective method to save power.

*2.2. Techniques for Smartphone Enabled Applications.* Smartphones, due to their vast number, wide coverage range, multiple embedded sensing components, significant computing ability, and convenient network accessing, are currently considered to be the largest sensing data source. The potential of embedded components (e.g., cameras, microphones, GPS, compresses, and accelerometers) is not yet well developed. Every combination or new application of these components can provide a brand new direction for mobile sensing. For example, utilizing microphones to detect vehicle horns can infer traffic conditions [22]. With the development of computing capabilities, every mobile phone can act as a high performance terminal, in which case cloud and parallel computing can be applied with the help of multiple network accessing ability like WiFi, 3G, Bluetooth, and so forth. Based on these hardware advantages of smartphones, various software designs and policies are proposed. These include information sharing tactics, data management, privacy preservation, and security protection. At the system level, scalability, robustness, and other requirements call for further research and novel techniques. On the other hand, techniques of studying smartphone sensing are highly diversified. Multiple existing data science techniques (e.g., cloud computing [27], data mining [28, 29]) have been applied in this field. In [27], an approach (called Pickle) was proposed

to prevent privacy leakage when applying cloud computing to collaborative learning for mobile sensing. Pickle perturbs the training data by premultiplying a private random matrix to train feature vector matrices. Since the private random matrix can be seen only at the user side, user's information is unavailable to cloud server or other participants after perturbing.

Data mining is considered as another frequently used technique to analyze smartphone sensing information. Various embedded sensing devices (e.g., cameras, microphones, accelerometers, light sensors, and GPS) generate abundant information to achieve innovative applications. When large amount of sensing data are aggregated together, data mining can be applied to extract useful and interesting information from them. The rapid growth of smartphone number shows great opportunity for data mining and introduces new challenges at the same time. Paper [30] (i) discusses the limitation and impact on applying data mining to mobile sensing in detail and (ii) introduces their solution: a method based on their wireless sensor data mining which is a smartphone-based sensor mining architecture. In this paper, the authors discussed issues which include the following: limited resources, scalability, real-time responsibility, granularity, configurability of polling rate, interactions with normal phone functions, conflicts with the needs of sensor mining, convenience for developers, self-learning ability, trade-offs between application scalability and limited resources, database management, I/O bottleneck of real-time transmission, parallelism requirements, pipelining requirements, programming language choice, algorithms for different application, secure connection/communication/storage, privacy control, trade-offs between sensing mining performance and energy/resources, and data compression (encoding).

Besides the above mature data analyzing sciences, other general or special purpose techniques are also developed. For example, [31] introduces a method which can utilize human-carried mobile phones to mule information from distributed sensors to other sensor nets.

*2.3. Other Applications.* Besides the smartphone enabled applications, wireless sensor networks [32–35] also enable a lot of applications. In this section, we introduce these applications including building energy management, pollution monitoring, and smart transportation systems.

*2.3.1. Building Energy Management.* Since sensor devices need to continuously collect data, energy management of sensor devices [36–38] is critical. On the other hand, utilizing sensors for building energy management [39–41] is an emergent application in sensor network community. As one of the most important research fields in the world of sensing, building energy management investigates energy consumption information in both space and time domains, by utilizing smart meters. The energy utility companies in the United States have deployed millions of “smart meters” in both residential and commercial buildings to better understand the electricity demand of consumers. This advanced metering infrastructure generates huge amount of data about the energy consumption of a customer at high granularity

(e.g., at second level). But the utility companies have been inefficient at getting maximum utilization from such a wealth of data. About 27% of the total electricity consumption in the USA is utilized for thermal conditioning (HVAC), that is, heating and cooling of premises in response to the outside temperature. One of the recent works [42] focused on building thermal profiles of residential energy users using smart meter data. Another paper [43] by the same authors leveraged the concept by building thermal profiles at both individual and group levels and applying them in a dynamic model for studying the thermal sensitivity in a given sample of users. Such profiles can also be utilized by the utility companies in their demand-response programs that focus on temperature-dependent consumption. The paper also analyzed the seasonal and time-of-day effects on thermal sensitivity at both individuals and their neighborhoods. Finally, it presented a methodology for aggregation of thermal profiles based on geographically homogeneous groups of users.

The rate at which data are being generated from the current electric microgrids and smart grids is tremendous. Efficient utilization of the generated real-time streaming sensor data remains a challenging task considering the sheer volume, complexity, and the rate of acquisition. Therefore, there is an urgent need to effectively manage and control such data via advanced processing, modelling, optimization, real-time forecasting, and analytics. There are internal factors (related to the grid) and external factors (e.g., weather, user behavior, and user economics) that affect the management of real-time data. Paper [44] proposes large-scale predictive analytics for real-time energy management by deploying a microgrid in a university campus aiming at maximizing its operational benefits. This particular environment was chosen due to the rich resources of cutting-edge analytics and high performance computing available for studying the huge and complex real-time data streams generated by the deployed microgrid. The proposed model aims at improving operational efficiency, lowering operating costs, and reducing the overall carbon footprint of the microgrid by using novel time series prediction algorithms.

Today, residential and commercial buildings are equipped with large number of different sensors and smart meters. These devices are primarily used as a mode of providing value added services by service providers and getting important feedback for customers on their usage patterns. But these devices can be used to make unwanted inferences about occupants and their behaviors. The research paper [45] explores this possibility of unwanted inferences (e.g., privacy) from the sensor data available to the utility companies. It attempts to infer answers to the following questions: (i) is a particular space occupied? (ii) how many people are there in that space? (iii) if that space is occupied, what are its occupants' identities? and (iv) which particular subspaces do they occupy? The paper focuses on inferences from two different types of sources: motion sensors (i.e., passive infrared sensors) installed by security companies and smart electric meters deployed by utility companies.

In the current era of smart meters deployed by the utility companies, the rate at which data is being generated by such smart devices is immense. The consumers, who are

the key stakeholders of the energy usage data, are often not involved in the analysis of this data. There are no existing systems which (i) empower users with access controls and (ii) provide control and access of their energy usage data with high granularity. In [46], the authors propose a new system design which (i) offers cloud-based personal data and execution containers for persistent data storage and (ii) at the same time gives independence to consumers in choosing their analytic algorithms. In this system, the consumers can also utilize third party applications which analyze data in a privacy-preserving fashion. Finally, the containers can also be utilized for secure and private control of home appliances from any Internet-enabled device.

**2.3.2. Pollution Monitoring.** Urban air pollution is one of the growing concerns in major cities worldwide. Large amount of data in the form of air pollution maps helps health protection agencies in assessing air quality. Ultrafine particles (UFPs) are often neglected as atmospheric pollutants, due to their small contribution to the total particle mass. The authors in [47] try to understand the impact of these high spatial variability particles on human health by proposing a mobile measurement system for producing accurate UFP pollution maps with high spatiotemporal resolution. The static measurement systems are inefficient at measuring such kinds of highly spatial variability pollutants. Moreover, these systems have high acquisition and maintenance costs. To enable a large urban coverage, the proposed system has its 10 sensor nodes installed on top of public transport vehicles. It also utilizes land-use regression models for modeling pollution concentrations at locations not covered by the mobile sensor nodes.

**2.3.3. Smart Transportation System.** Today's modern cities are one of the major contributors to the generation of big data. The different mobile sensing devices as well as the city infrastructure sensors produce large amounts of data, which provide a wealth of information about their surroundings and can be utilized for improving the social lives of human beings. In the current scenario of more precise and pervasive sensing, lots of dynamic information about individual cars becomes available through car-to-car (C2C) and car-to-infrastructure (C2I) communication. Paper [48] dwells on the possible research area of dynamic infrastructure-to-car communication where dynamic information about vehicles is exploited. The main contribution of the paper is a model of a distributed intelligent speed adaptation system. The authors also provide a formal proof about the correct dissemination of speed limit information by such a system. This information is in the form of speed advice from traffic centers, traffic sign detectors, or obstacle detectors. The paper proposes a global control system, to be used by highway authorities, for considering incidents (such as accidents, construction sites, or traffic jams) which are well beyond the scope of sensor coverage of a local vehicle. The paper also identifies the safely operable bounds of such a system.

In [49], the authors present Context-Aware Platform using Integrated Mobile services (CAPIM) which is basically a platform enabling smart management of the large amount

of available contextual information. CAPIM focuses on collection and aggregation of context data (e.g., location, user's profile, and characteristics) through smart services offered by mobile devices like smartphones and tablet PCs that have multiple sensors. The platform supports collaborative environment by enabling its users to learn about their surroundings through sharing data without too much user interaction. The authors then present an intelligent transportation system that is designed on top of CAPIM, for improving the understanding of traffic related problems. Finally, they propose a solution called context-aware framework which deals with the efficient storage of context data on a larger scale.

### 3. Summary of Big Data Techniques

As discussed above, a lot of applications are in the urgent need of novel big data techniques. However, big data itself is a new data science. Currently, there is no mature architecture for it. Presently, some of the researchers in this field are devoting themselves to building general platforms, architectures, and analysis methodologies. The others are focusing on developing solutions for particular problems.

**3.1. Platform Development.** One of the significant features of sensing in future is "gigantism." Concepts like smart cities and IoT require vast number of sensors to work together under certain control policies. Conventional topologies, policies, architectures, and methods are no longer suitable. Platforms which can deal at city level, country level, or even world level with sensor data are in need.

In [4], the authors explored five key challenges, which all researchers will face in the field of future sensing in developing a city level sensing platform. The first challenge mentioned is crowd sourcing and collaboration. This is mainly about how to create a mature system from which users can get tangible benefits through sharing and using information. Current single-provider model no longer fits the requirement of future sensing but multiple-provider model is suffering from lack of structure and consistency. A mature platform must support operations for sharing, annotating, reusing, and analyzing data itself. The second challenge is heterogeneity and disparity. Sensing data in a city are distributed anywhere and it is impossible to aggregate them in one central location. Data collected by individuals under diverse regimes are different as a matter of course. An effective informatics system which can extract useful information from different data format is necessary. The third challenge is multiresolution and multiscale which relate to the fact that there is no unified standard for sensing so far. While data from different sources are aggregated for new applications, multiresolution is the first problem researchers are facing. Even worse, will the conclusion based on these resources lead to future ambiguity? The fourth challenge is data uncertainty and trustworthiness. Data from some sources may be wrongly calibrated or inaccurate due to sensing devices. Sensor system should be able to identify uncertainty and distinguish trustful information sources from others and ensure that users can

manage and get profits from different sources. The fifth challenge is model and decision making. The quality of analysis depends on data and leveraging weights of different data sources are key issues. Moreover, the costs of time and resources processing and analyzing large amounts of data are too high given that real-time decisions need to be made.

Paper [5] focuses on building cloud-based big data architecture for supporting sensor services. Data quality is key aspect of their system. The purpose of this paper is building a sensing infrastructure for federated sensor services paradigm. However, several design requirements must be considered. The first one is models for feed content and quality. A cloud network designed for federated sensor services should be able to satisfy customers' requirements in terms of both content and quality. The second is techniques for feed discovery, composition, and adaptation. Techniques for a federated sensor services' cloud should be able to adapt various environmental dynamics. The third is markup language. A semantics-rich markup language is required for user applications to express their feed requirements and feed providers. The fourth is massively scalable feed storage and analytics. A federated sensor service cloud should provide scalable storage and analytic services for feeds. The fifth is pricing models and service-level agreements (SLA). Benefits are incentives for users to join certain services. A federated sensor service cloud should be able to support real-time pricing model, based on service quality. And an effective SLA is critical for sensor data markets.

The authors of [17] proposed another model that is designed for wireless sensor networks to aggregate sensor data from various devices. Nowadays, a vast amount of mobile devices is connected to Internet and users can get access to sensing data by using user-friendly mobile applications anytime and anywhere. Then integration of all sorts of data through Internet is challenging. The proposed model in this paper fully utilizes existing infrastructures to aggregate, process, and distribute data. It can be considered as ubiquitous since it is designed for general data integration scenes. The whole model contains a REST Web service which relies on open standards such as Hypertext Transfer Protocol (HTTP) and Extensible Markup Language (XML) and a MySQL database to store information from mobile devices. Then, the data can be delivered to mobile clients in XML messages by HTTP servers.

**3.2. Data Processing Techniques.** Big data, just as its name implies, is a data science which cannot be easily processed using existing infrastructure or data processing methods. Currently, researchers are working in two directions to solve this problem. One is modifying and improving current infrastructures, for instance, strengthening processing abilities or optimizing computing structures, to handle data more efficiently. Another direction is developing new data management methods. Various techniques are applied in each direction and it is hard to categorize them precisely. So, we only introduce several representative papers in this section.

In [50], the authors introduce a well designed sensor network (RACNet) that can be used for monitoring data

center's environmental conditions. RACNet is a large-scale sensor network for high-fidelity data center environmental monitoring. The sensor nodes of this network are custom-made. And the protocol applied here is a congestion control policy called Wireless Reliable Acquisition Protocol (WRAP), which is developed by leveraging frequency and time multiplexing. The experimental results show that RACNet can improve the data center's safety and energy efficiency. WRAP is the most important part in RACNet for reliable wireless data acquisition. It inherits advantages from both distributed and centralized data collection policies. A distributed system will suffer channel contention which eventually leads to packet losses due to lack of coordination, especially under high network load, while a centralized data collection system requires additional communication load from or to the gateway, especially when the number of nodes in a network is large. The square increasing control information load adds a great burden to the large-scale sensing network. As a hybrid approach, WRAP transfers tokens, which can be passed one by one through distributed nodes, to exchange authority of sending control information. Thus, tokens can avoid being passed to interflow contention which may lead to congestion and packet loss.

In [51], the authors propose prediction models to improve geometric monitoring framework. These models provide significant communication savings ranging from two to three orders of magnitude, compared to the transmission cost of the original monitoring framework. Multiple predictor models are proved to fit this kind of large-scale monitoring network. Actually, the concepts of the predictor models proposed in this paper have existed for a long time, but applying them to significantly reduce the communication burden is the key idea of building a big data sensing network. If the current infrastructure cannot afford the impact of rapid growing data volume, there is a need to improve or redesign current systems for higher computing abilities or data throughput.

Paper [52] introduces a data management method that is designed for data query processing. Packets sent by sensors usually lack time information, and even timestamps are embedded. Query processing is still challenging due to the infinite amount of sensor data. Conventional model-based query processing approaches mostly employ the relational data model on top of modeled segments of sensor data. MapReduce is applied in the cloud era to have time series stored in key value stores. In this paper, the authors proposed KVI-index, which combines the advantages of key value stores and the MapReduce parallel computing together, to dynamically accommodate new sensor data segments efficiently.

Opportunistic sensing is another new approach which exploits sensing capabilities of mobile devices. It can be applied as tactics to enlarge mobile sensing scales without additional investments. Paper [53] describes a framework for fully distributed opportunistic sensing which can perform recruitment and collect data. Profile-cast and opportunistic geocast are used for recruitment. An original version of profile-cast aims at reaching nodes which match a certain target profile, but the recruitment also needs to reach the nodes that match only a part of the target profile.

Based on opportunistic geocast, geodissemination which calculates EVR for the buildings in the traces, instead of for the hexagonal cells, achieves better performance when recruiting nodes. Similar to the recruiting case, data collection aims to reach any of the nodes that match the target profile, since sensing nodes are usually greatly out of sync.

Another way of dealing with large amount of data is compression. Different compressing algorithms suit different application scenes. Paper [54] introduces GAMPS, a compressing method which processes sensing data before they are aggregated in data center for mining. Though the compressing method is not lossless, maximum error is acceptable compared to the significant profits. Two key ideas are proposed in this paper. One is dynamically compressing data in a group which contains related signals, and the other is considering different amplitudes of signals and reconstructing the joint signal within the maximum allowed reconstruction error bound. Besides these two compressing methods, GAMPS maintains an index so that several important queries can be issued directly from compressed data.

The authors of [55] worked on a data set which is relatively “big.” In this realm of wireless sensing, nodes with deployed devices are usually inexpensive and have limited computing ability, energy, bandwidth, and storage space. In this kind of sensing networks, there are new challenges in data processing and dissemination. Though the total amount of data is not that large, compared to the limitation of sensor nodes, novel techniques are still required to improve the networks’ data processing capabilities. The method proposed in this paper compresses data streams from different sensors based on the historical information they carried. Though not lossless, the compressing algorithm in this paper has a lower compressing error ratio than conventional methods. The method is designed to find correlation and redundancy from measured information of the same sensors. A base signal is extracted based on the difference of correlation signals which are from real measurement features. These measurement features are used to encode signals as well. The proposed algorithm is not restricted to particular sensing application scenario. So it can be applied to any data set in which correlation and redundancy exist.

Sensing in the future will grow in size with no doubt, and large amount of data can be aggregated in many physical systems over time. But since these series usually exhibit various behaviors, it is challenging to build one static model to analyze them efficiently and benefit from the growth of data. In [56], a dynamic model which integrates multiple existing models is proposed. It selects suitable models for different series based on their extracted features. In the feature extraction techniques which are used for individual time series, both linear and nonlinear methods are applied. The main idea known as “trajectory mining” is used to model the evolution path of time series in the feature space. This paper shows that combining and improving current techniques is a convenient way to solve the upcoming sensing data problems.

*3.3. Techniques for Specific Problems.* The increasing scope of applications of the wireless sensor networks is producing data at an extremely higher rate than before. The sudden inconsistencies of data, or outliers, often affect applications which heavily rely on timely and reliable sensory data. Current approaches to identifying outlier values introduce an overwhelming communication overhead which limits their practical implementations. The researcher of [57] proposes Tunable Approximate Computation of Outliers (TACO), an outlier detection framework that trades bandwidth for accuracy. TACO supports various similarity measures such as the cosine similarity, the correlation coefficient, and the Jaccard coefficient. It involves two levels of hashing mechanisms. The first level deals with dimensional reduction using locality sensitive hashing. The second level of hashing comes into picture during the intracluster communication phase. TACO also employs a boosting process for improving its accuracy. The TACO’s novel load balancing and comparison pruning mechanisms ensure reduced processing and communication load at clusterheads, resulting in a more uniform, intracluster power consumption. Therefore, TACO can prolong unhindered network operations.

Recently, the wide-area shared sensing has been the center of attraction. Different from a typical wireless sensing application, it has certain characteristics such as a relatively diverse set of queries (e.g., Max/Min, Sum, Uniform Samples, Quantiles, Top-k readings, frequent readings, and push-based data collection). There are several reasons for using the push-based data collection technique, for example, large number of geographically dispersed sensors, substantial high query rate to the shared sensor compared to the data collection or reporting frequency of the sensor, and occasional connectivity of some sensors (e.g., once per hour) for data reporting purposes. These reasons make it unfeasible to use pull-based data collection at query time. The portals usually outsource data collection and query processing tasks to the third parties, called aggregators who provide data aggregation services. Such an outsourced aggregation model faces key security challenges such as the fact that aggregators can be untrusted, compromised, or even malicious. Thus the correctness of answers provided by aggregators should be verified to prevent incorrect query answers.

Currently, there is a need to maximize the overall value of the collected data, subject to resource constraints, in a particular class of sensor networks that focus on the reliable collection of high-resolution signals. The main characteristic of such systems is that the collected data is more than the amount of data that can be delivered to the base station, due to the severe limitations on radio bandwidth and energy. These systems also cannot utilize the in-network data aggregation due to the high data rates and raw signals requirement. Moreover, applications look for the most “interesting” signals rather than wasting resources on “uninteresting” signals. Some examples of sensor network applications where high-resolution signals are needed from low-power wireless sensor nodes include monitoring acoustic, seismic and vibration waveforms in bridges, industrial equipment, volcanoes, and animal habitats. The researchers in [58] present Lance, a system that aims at providing value-driven bandwidth and

energy management framework for high-data-rate sensor networks. Lance uses cost estimators to predict the energy cost for reliably downloading each Application Data Unit from the network. It also utilizes user-supplied policy modules for decoupling resource allocation mechanisms from application-specific policies, allowing the system to be tailored to a broad range of applications.

*3.4. Security and Privacy Preserving Techniques.* In this field, researchers have investigated secure network protocols [59, 60] and privacy-preserving techniques [61, 62]. The design and evaluation of large-scale urban sensing networks often utilize mobility traces of people. There is a growing privacy concern about the public availabilities of such real user traces. The reason that the synthetic movement models produce inaccurate traces in network design is leading to increasing efforts towards having real-world participants in such systems. The effectiveness of some cloaking techniques, such as introducing noise or reducing the resolution of the recorded data, in protecting privacy of the real-world users is not known. Hence, the side information or the information about the whereabouts of the participants (victims) in public spaces can be obtained by an adversary over an extended period of time. The researchers in [63] analyze, both theoretically and experimentally, the ways in which an attack can be carried out by an adversary either through direct observations or indirect information sources based on the huge amounts of publicized data about real user traces available on either consolidated data portals or websites. The results indicate that it may lead to potential privacy breach. The researchers of [64] present SECOA, the first unified framework with a family of optimally secured (i.e., no false positive/negative) protocols. SECOA supports a large set of aggregations with *Most Popular Readings* and *Frequent Readings* aggregation in a secure aggregation scheme. SECOA also utilizes RSA encryption in one-way chains for aggressive optimization to reduce computation overhead.

The amount of data that smartphones are generating is huge with the help of various embedded sensors. The need for classification of data naturally arises. The researchers in [61] explore an entirely new way of building robust classifiers through collaborative learning where users contribute sensor data as training samples such as audio clips. Such learning enables user diversity; thus it helps train a model to robustly recognize the environment the user is in. The employment of cloud computing platform for classifier construction raises privacy concern on submitted samples. The authors propose Pickle, a new approach to privacy-preserving collaborative learning. It encourages user's participation by ensuring privacy of the contributed training samples. Pickle also boasts many desirable properties such as high accuracy, independent user operation, tuning the level of privacy, and robustness to poisoning attacks.

There is a growing privacy concern on the large number of applications available on the Apple iPhone App Store that are accessing private user information without user's consent. The private user information can be user's location, address book, music, photos, and unique identifiers such as IMEI number, UDID, and Wi-Fi MAC addresses. The

incorporation of free applications from untrusted developers who rely on third party advertisement frameworks as a source of income often leads to access of private information by these advertisement frameworks when a particular user installs such an application. The authors in [65] compare the other leading mobile OS platform Android with Apple iOS. Android puts the responsibility of reviewing app permissions on users at the time of download while iOS checks apps before including them on App Store. But due to the recent cases of private data leakage because of some applications on iOS, there has been a public outcry in general. The authors propose the ProtectMyPrivacy system which detects access to private information by apps at runtime. The unique feature of this system is its crowdsourced recommendation engine which provides app privacy recommendations based on collected and analyzed user protection decisions.

In today's era, where mobile devices such as smartphones and PDAs are ever-growing in terms of sensing, computation, storage, and communication capabilities, huge amounts of data are being generated by such devices very rapidly. People now are active data contributors instead of being just passive data users as was the case several years ago. People-centric urban sensing is one of the promising fields in this new direction which supports urban-scale distributed data collection, analysis, and sharing. But the privacy concerns in such a system result in user reluctance for participation in contributing personal data. For example, a study on relationship between air quality and public health requires researchers to obtain people's health data such as heart rates, blood pressure levels, and weights for some aggregate statistics. But most of people will not provide their personal data unless they assure that their data will not be misused to invade their privacy. The researchers in [62] propose PriSense, a privacy-preserving data aggregation solution in people-centric urban sensing. PriSense consists of two main components: one for dealing with additive aggregation functions and the other for nonadditive aggregation functions. It utilizes the concept of data slicing and mixing. It can support different functions such as Sum, Average, Variance, Count, Max/Min, Median, Histogram, and Percentile with accurate aggregation results. The level of user privacy can be increased substantially by tuning threshold number of colluding users and aggregation servers.

#### 4. Future Research Directions

With the development of sensing techniques and rapid growth of sensing devices (e.g., smartphones and tablets) large amount of sensing data will be generated and, thus, big data has become a hot topic. However, big data is a relatively new concept in the world of data sciences. The future research directions of big data in sensing have a lot of challenges and also great opportunities for researchers.

Mature infrastructures for sensing data generation, collection, classification, analysis, and processing are desired. For now, several key network techniques [66, 67] can be applied to build this kind of general purpose infrastructures. Cloud computing and parallel structure are essential techniques to build high performance platforms. Grid or

stream computing and relevant programming models beyond Hadoop/MapReduce and STORM can be used to define basic architectures of future sensing. Currently, sensor networks are usually restricted to small regions. They are commonly developed and maintained by individuals, labs, or certain groups. However, sensor networks in the future should be at the town or city level, or even world level. They are expected to be maintained by large companies, institutions, or governments. Data will be aggregated and distributed in different methods to all potential users. Therefore, large profits will be gained during the data sharing process. Smartphone sensing is the forerunner of building such large-scale networks and it is one of the top concerned topics in this research field. Mobile sensing will lead this field in the coming future. Therefore, existing localization techniques [68, 69] should be improved to support mobile sensing.

Based on certain infrastructures, data management methods will bloom. But other data sciences have been introduced to solve problems in the world of big data, such as data mining, crowd sourcing, techniques on data base, data management, security and privacy, data protection and integrity, data storage, machine learning, and neural networks. Currently, researchers are focusing on data management performance based on existing techniques. But in the future, with the development of sensing infrastructure, high performance data management methods will flourish. These data management methods include (i) different optimization techniques which improve data analysis ability, (ii) compression methods which condense data values, and (iii) searching approaches which extract useful information from database.

With the development of data infrastructures and data management methods, it is foreseeable that sensing in the future will step into every corner of this world, for example, smart grids [70–72]. Then more security and privacy problems will arise. Without solving security problems, techniques may introduce damages instead of profits. Currently, researchers are mostly focusing on privacy leakages and user data protection. However, with the development of sensing infrastructures and data management techniques, more and more sensing data will flood. Then the sensor network itself can be a target of attackers, just like Internet. Current sensor packets are usually not encrypted and a single node which runs the same protocols can decode information from the network or even inject attacker's malicious information. To address this problem, we need encryption which leads to additional burden to sensor nodes and may impact energy efficiency of sensor networks. How to protect sensing information efficiently is a promising direction.

Applications and research methods are inseparably interconnected. Various and innumerable applications might be developed based on people's needs as determined by the big data collected, processed, and analyzed over time. Though, currently, smartphones enabled applications are the most popular applications in the sensing world, other sensing applications (such as monitoring systems, remote sensing, and sustainable computing) are also promising directions to be investigated in the future.

## 5. Conclusion

In this survey paper, we introduced research circumstances of big data in the field of sensing. We first introduce different applications that deal with big sensing data and then summarize techniques used to solve the big sensing data problems. Finally, we propose some future research directions. A large number of platforms which have the capacity for sensing at the city level are still in the designing concept stage, but a lot of research methods have been proposed. Though most of them are based on existing data processing and management techniques, they are still very useful. Mobile sensing and smartphone applications are still considered as the most popular topic. Researchers will dedicate themselves to smartphone applications in the near future because it is the most mature large-scale sensor network so far.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgment

This work is supported by the NSF Grant CNS-1503590.

## References

- [1] D. Laney, "3D data management: controlling data volume, velocity and variety," *Application Delivery Strategies*, 2001.
- [2] IEEE BigData 2013, <http://cci.drexel.edu/bigdata/bigdata2013/>.
- [3] F. Khafa and C. Dobre, "Intelligent services for big data science," *Future Generation Computer Systems*, vol. 37, pp. 267–281, 2014.
- [4] C. Wu, D. Silva, O. Tsinalis et al., "Building a generic platform for big sensor data application," in *Proceedings of the IEEE International Conference on Big Data*, pp. 94–102, Silicon Valley, Calif, USA, October 2013.
- [5] L. Ramaswamy, V. Lawson, and S. V. Gogineni, "Towards a quality-centric big data architecture for federated sensor services," in *Proceedings of the IEEE International Congress on Big Data (BigData '13)*, pp. 86–93, July 2013.
- [6] F. Mattern and C. Floerkemeier, "That 'internet of things' thing, in the real world things matter more than ideas," *RFID Journal*, 2009.
- [7] D. Georgakopoulos, A. Zaslavsky, and C. Perera, "Sensing as a service and big data," in *Proceedings of the International Conference on Advances in Cloud Computing (ACC '12)*, Bangalore, India, July 2012.
- [8] X. Yang, W. Song, and D. De, "LiveWeb: a sensorweb portal for sensing the world in real-time," *Tsinghua Science and Technology*, vol. 16, no. 5, pp. 491–504, 2011.
- [9] J.-F. Créteaux, W. Jelinski, S. Calmant et al., "Sols: a lake database to monitor in the near real time water level and storage variations from remote sensing data," *Advances in Space Research*, vol. 47, no. 9, pp. 1497–1507, 2011.
- [10] N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. T. Campbell, "A survey of mobile phone sensing," *IEEE Communications Magazine*, vol. 48, no. 9, pp. 140–150, 2010.

- [11] J. Laurila, D. Gatica-Perez, I. Aad et al., "The mobile data challenge: big data for mobile computing research," in *Proceedings of the Mobile Data Challenge Workshop (MDC '12)*, June 2012.
- [12] J. K. Laurila, D. Gatica-Perez, I. Aad et al., "From big smartphone data to worldwide research: the Mobile Data Challenge," *Pervasive and Mobile Computing*, vol. 9, no. 6, pp. 752–771, 2013.
- [13] A. Sheth, C. C. Aggarwal, and N. Ashish, "The internet of things: a survey from the data-centric perspective," in *Managing and Mining Sensor Data*, chapter 12, pp. 383–428, Springer, New York, NY, USA, 2013.
- [14] Z. Li, M. Li, J. Wang, and Z. Cao, "Ubiquitous data collection for mobile users in wireless sensor networks," in *Proceedings of the IEEE INFOCOM*, pp. 2246–2254, IEEE, Shanghai, China, April 2011.
- [15] D. Tracey and C. Sreenan, "A holistic architecture for the internet of things, sensing services and big data," in *Proceedings of the 13th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid '13)*, pp. 546–553, Delft, The Netherlands, May 2013.
- [16] L. Atzori, A. Iera, and G. Morabito, "From 'smart objects' to 'social objects': the next evolutionary step of the internet of things," *IEEE Communications Magazine*, vol. 52, no. 1, pp. 97–105, 2014.
- [17] A. G. F. Elias, J. J. P. C. Rodrigues, L. M. L. Oliveira, and B. B. Zarpelão, "A ubiquitous model for wireless sensor networks monitoring," in *Proceedings of the 6th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS '12)*, pp. 835–839, Palermo, Italy, July 2012.
- [18] T. L. Porta, R. Govindan, M.-R. Ra, and B. Liu, "Medusa: a programming framework for crowd-sensing applications," in *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services (MobiSys '12)*, pp. 337–350, 2012.
- [19] X. Fang, J. Tang, D. Yang, and G. Xue, "Crowdsourcing to smartphones: incentive mechanism design for mobile phone sensing," in *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking (MobiCom '12)*, pp. 173–184, August 2012.
- [20] L. Lenzini, V. Luconi, A. Vecchio, A. Faggiani, and E. Gregori, "Lessons learned from the design, implementation, and management of a smartphone-based crowdsourcing system," in *Proceedings of the 1st International Workshop on Sensing and Big Data Mining (SenseMine '13)*, pp. 1–6, Roma, Italy, November 2013.
- [21] M. Mun, S. Reddy, K. Shilton et al., "PEIR, the personal environmental impact report, as a platform for participatory sensing systems research," in *Proceedings of the 7th ACM International Conference on Mobile Systems, Applications, and Services (MobiSys '09)*, pp. 55–68, June 2009.
- [22] P. Mohan, V. N. Padmanabhan, and R. Ramjee, "Nericell: Rich monitoring of road and traffic conditions using mobile smartphones," in *Proceedings of the 6th ACM Conference on Embedded Networked Sensor Systems (SenSys '08)*, pp. 323–336, November 2008.
- [23] W. Qin, J. Zhang, B. Li, and L. Sun, "Discovering human presence activities with smartphones using nonintrusive wi-fi sniffer sensors: the big data prospective," *International Journal of Distributed Sensor Networks*, vol. 2013, Article ID 927940, 12 pages, 2013.
- [24] T. Toscos, M. Y. Chen, J. Froehlich et al., "Activity sensing in the wild: a field trial of ubifit garden," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1797–1806, April 2008.
- [25] K. Pripuzić, I. P. Žarko, and A. Antonić, "Publish/subscribe middleware for energy-efficient mobile crowdsensing," in *Proceedings of the ACM Conference on Ubiquitous Computing (UbiComp '13)*, pp. 1099–1110, Zurich, Switzerland, September 2013.
- [26] A. Antonic, I. P. Zarko, and D. Jakobovic, "Inferring presence status on smartphones: the big data perspective," in *Proceedings of the 18th IEEE Symposium on Computers and Communications (ISCC '13)*, pp. 600–605, July 2013.
- [27] B. Liu, Y. Jiang, F. Sha, and R. Govindan, "Cloud-enabled privacy-preserving collaborative learning for mobile sensing," in *Proceedings of the 10th ACM Conference on Embedded Networked Sensor Systems (SenSys '12)*, pp. 57–70, November 2012.
- [28] G. M. Weiss and J. W. Lockhart, "Identifying user traits by mining smart phone accelerometer data," in *Proceedings of the 5th International Workshop on Knowledge Discovery from Sensor Data (SensorKDD '11)*, pp. 61–69, ACM, August 2011.
- [29] J. W. Lockhart and G. M. Weiss, "A comparison of alternative client/server architectures for ubiquitous mobile sensor-based applications," in *Proceedings of the 14th International Conference on Ubiquitous Computing (UbiComp '12)*, pp. 721–724, September 2012.
- [30] J. C. Xue, S. T. Gallagher, A. B. Grosner, T. T. Pulickal, J. W. Lockhart, and G. M. Weiss, "Design considerations for the WISDM smart phone-based sensor mining architecture," in *Proceedings of the 5th International Workshop on Knowledge Discovery from Sensor Data (SensorKDD '11)*, pp. 25–33, 2011.
- [31] U. Park and J. Heidemann, "Data muling with mobile phones for sensor networks," in *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems (SenSys '11)*, pp. 162–175, November 2011.
- [32] L. Oliveira, J. Rodrigues, A. Elias, and B. Zarpelo, "Ubiquitous monitoring solution for Wireless Sensor Networks with push notifications and end-to-end connectivity," *Mobile Information Systems*, vol. 10, no. 1, pp. 19–35, 2014.
- [33] O. Diallo, J. Rodrigues, M. Sene, and J. Lloret, "Distributed database management techniques for wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 2, pp. 604–620, 2015.
- [34] L. D. P. Mendes, J. J. P. C. Rodrigues, J. Lloret, and S. Sendra, "Cross-layer dynamic admission control for cloud-based multimedia sensor networks," *IEEE Systems Journal*, vol. 8, no. 1, pp. 235–246, 2014.
- [35] S. Ullah, J. Rodrigues, F. Khan, C. Verikoukis, and Z. Zhu, "Protocols and architectures for next-generation wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2014, Article ID 705470, 3 pages, 2014.
- [36] T. Zhu, Z. Zhong, T. He, and Z.-L. Zhang, "Energy-synchronized computing for sustainable sensor networks," *Ad Hoc Networks*, vol. 11, no. 4, pp. 1392–1404, 2013.
- [37] Y. Gu, L. He, T. Zhu, and T. He, "Achieving energy-synchronized communication in energy-harvesting wireless sensor networks," *ACM Transactions on Embedded Computing Systems*, vol. 13, no. 2, article 68, 2014.
- [38] T. Zhu, Y. Gu, T. He, and Z. Zhang, "Achieving long-term operation with a capacitor-driven energy storage and sharing network," *ACM Transactions on Sensor Networks (TOSN)*, vol. 8, no. 4, article 32, 2012.

- [39] T. Zhu, A. Mishra, D. Irwin, N. Sharma, P. Shenoy, and D. Towsley, "The case for efficient renewable energy management in smart homes," in *Proceedings of the 3rd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings (BuildSys '11)*, pp. 67–72, ACM, Seattle, Wash, USA, November 2011.
- [40] N. Sharma, J. Gummeson, D. Irwin, T. Zhu, and P. Shenoy, "Leveraging weather forecasts in energy harvesting sensor systems," in *Proceedings of the IEEE Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON '14)*, 2014.
- [41] Z. Huang, H. Luo, D. Skoda, T. Zhu, and Y. Gu, "E-Sketch: Gathering large-scale energy consumption data based on consumption patterns," in *Proceedings of the IEEE International Conference on Big Data (Big Data '14)*, pp. 656–665, Washington, DC, USA, October 2014.
- [42] A. Albert and R. Rajagopal, "Thermal profiling of residential energy use," *IEEE Transactions on Power Systems*, vol. 30, no. 2, pp. 602–611, 2014.
- [43] A. Albert and R. Rajagopal, "Building dynamic thermal profiles of energy consumption for individuals and neighborhoods," in *Proceedings of the IEEE International Conference on Big Data*, pp. 723–728, October 2013.
- [44] N. Balac, T. Sipes, N. Wolter, K. Nunes, B. Sinkovits, and H. Karimabadi, "Large Scale predictive analytics for real-time energy management," in *Proceedings of the IEEE International Conference on Big Data, Big Data*, pp. 657–664, October 2013.
- [45] L. Yang, K. Ting, and M. B. Srivastava, "Inferring occupancy from opportunistically available sensor data," in *Proceedings of the 12th IEEE International Conference on Pervasive Computing and Communications (PerCom '14)*, pp. 60–68, March 2014.
- [46] R. P. Singh, S. Keshav, and T. Brecht, "A cloud-based consumer-centric architecture for energy data analytics," in *Proceedings of the 4th ACM International Conference on Future Energy Systems (e-Energy '13)*, pp. 63–74, Berkeley, Calif, USA, May 2013.
- [47] D. Hasenfratz, O. Saukh, C. Walsler, C. Hueglin, M. Fierz, and L. Thiele, "Pushing the spatio-temporal resolution limit of urban air pollution maps," in *Proceedings of the 12th IEEE International Conference on Pervasive Computing and Communications (PerCom '14)*, pp. 69–77, March 2014.
- [48] S. Mitsch, S. M. Loos, and A. Platzer, "Towards formal verification of freeway traffic control," in *Proceedings of the IEEE/ACM 3rd International Conference on Cyber-Physical Systems (ICCPs '12)*, pp. 171–180, April 2012.
- [49] C. Dobre and F. Khafa, "Intelligent services for Big data science," *Future Generation Computer Systems*, vol. 37, pp. 267–281, 2014.
- [50] C.-J. M. Liang, J. Liu, L. Luo, A. Terzis, and F. Zhao, "RACNet: a high-fidelity data center sensing network," in *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems (SenSys '09)*, pp. 15–28, November 2009.
- [51] N. Giatrakos, A. Deligiannakis, M. Garofalakis, I. Sharfman, and A. Schuster, "Prediction-based geometric monitoring over distributed data streams," in *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD '12)*, pp. 265–276, May 2012.
- [52] T. G. Papaioannou, T. Guo, and K. Aberer, "Model-view sensor data management in the cloud," in *Proceedings of the IEEE International Conference on Big Data*, pp. 282–290, October 2013.
- [53] G. Benincasa, G. S. Tuncay, and A. Helmy, "Participant recruitment and data collection framework for opportunistic sensing: a comparative analysis," in *Proceedings of the 19th Annual International Conference on Mobile Computing and Networking (MobiCom '13)*, pp. 191–194, October 2013.
- [54] S. Gandhi, S. Nath, S. Suri, and J. Liu, "GAMPS: compressing multi sensor data by grouping and amplitude scaling," in *International Conference on Management of Data and 28th Symposium on Principles of Database Systems (SIGMOD-PODS '09)*, pp. 771–784, July 2009.
- [55] N. Roussopoulos, A. Deligiannakis, and Y. Kotidis, "Compressing historical information in sensor networks," in *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD '04)*, pp. 527–538, ACM, Paris, France, June 2004.
- [56] A. Sharma, G. Jiang, H. Xiong, B. Liu, and H. Chen, "Modeling heterogeneous time series dynamics to profile big sensor data in complex physical systems," in *Proceedings of the IEEE International Conference on Big Data*, pp. 631–638, October 2013.
- [57] N. Giatrakos, Y. Kotidis, A. Deligiannakis, V. Vassalos, and Y. Theodoridis, "TACO: tunable approximate computation of outliers in wireless sensor networks," in *Proceedings of the International Conference on Management of Data (SIGMOD '10)*, pp. 279–290, Indianapolis, Ind, USA, June 2010.
- [58] G. Werner-Allen, S. Dawson-Haggerty, and M. Welsh, "Lance: optimizing high-resolution signal collection in wireless sensor networks," in *Proceedings of the 6th ACM Conference on Embedded Networked Sensor Systems (SenSys '08)*, pp. 169–182, Raleigh, NC, USA, November 2008.
- [59] T. Zhu, S. Xiao, Y. Ping, D. Towsley, and W. Gong, "A secure energy routing mechanism for sharing renewable energy in smart microgrid," in *Proceedings of the IEEE 2nd International Conference on Smart Grid Communications (SmartGridComm '11)*, pp. 143–148, October 2011.
- [60] P. Yi, T. Zhu, Q. Zhang, Y. Wu, and J. Li, "Green firewall: An energy-efficient intrusion prevention mechanism in wireless sensor network," in *Proceedings of the IEEE Global Communications Conference (GLOBECOM '12)*, pp. 3037–3042, December 2012.
- [61] B. Liu, Y. Jiang, F. Sha, and R. Govindan, "Cloud-enabled privacy-preserving collaborative learning for mobile sensing," in *Proceedings of the 10th ACM Conference on Embedded Networked Sensor Systems (SenSys '12)*, pp. 57–70, ACM, Toronto, Canada, November 2012.
- [62] J. Shi, R. Zhang, Y. Liu, and Y. Zhang, "PriSense: privacy-preserving data aggregation in people-centric urban sensing systems," in *Proceedings of the IEEE INFOCOM*, San Diego, Calif, USA, March 2010.
- [63] C. Y. T. Ma, D. K. Y. Yau, N. K. Yip, and N. S. V. Rao, "Privacy vulnerability of published anonymous mobility traces," *IEEE/ACM Transactions on Networking*, vol. 21, no. 3, pp. 720–733, 2013.
- [64] S. Nath, H. Yu, and H. Chan, "Secure outsourced aggregation via oneway chains," in *Proceedings of the ACM International Conference on Management of Data (SIGMOD '09)*, Providence, RI, USA, June-July 2009.
- [65] Y. Agarwal and M. Hall, "ProtectMyPrivacy: detecting and mitigating privacy leaks on iOS devices using crowdsourcing," in *Proceedings of the 11th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '13)*, pp. 97–109, Taipei, Taiwan, June 2013.
- [66] J. Jun, L. Cheng, L. He, Y. Gu, and T. Zhu, "Exploiting sender-based link Correlation in wireless sensor networks," in *Proceedings of the IEEE 22nd International Conference on*

- Network Protocols (ICNP '14)*, pp. 445–455, Raleigh, NC, USA, October 2014.
- [67] Z. Zhou, M. Xie, T. Zhu et al., “EEP2P: an energy-efficient and economy-efficient P2P network protocol,” in *Proceedings of the International Green Computing Conference (IGCC '14)*, pp. 1–6, Dallas, Tex, USA, November 2014.
- [68] Z. Zhong, T. Zhu, D. Wang, and T. He, “Tracking with unreliable node sequences,” in *Proceedings of the 28th IEEE Conference on Computer Communications (INFOCOM '09)*, pp. 1215–1223, April 2009.
- [69] Q. Zhang, Z. Zhou, W. Xu et al., “Fingerprint-free tracking with dynamic enhanced field division,” in *Proceedings of the IEEE Conference on Computer Communications (INFOCOM '15)*, Hong Kong, April-May 2015.
- [70] W. Zhong, Z. Huang, T. Zhu et al., “iDES: incentive-driven distributed energy sharing in sustainable microgrids,” in *Proceedings of the International Green Computing Conference (IGCC '14)*, pp. 1–10, IEEE, Dallas, Tex, USA, November 2014.
- [71] L. He, Y. Gu, T. Zhu, C. Liu, and K. G. Shin, “SHARE: SoH-aware reconfiguration to enhance deliverable capacity of large-scale battery packs,” in *Proceedings of the ACM/IEEE 6th International Conference on Cyber-Physical Systems (ICCPS '15)*, pp. 169–178, Seattle, Wash, USA, April 2015.
- [72] Z. Huang, D. Corrigan, T. Zhu, H. Luo, X. Zhan, and Y. Gu, “Exploring power-voltage relationship for distributed peak demand flattening in microgrids,” in *Proceedings of the ACM/IEEE 6th International Conference on Cyber-Physical Systems (ICCPS '15)*, 2015.



**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

