A Survey on Internet-of-Things for Smart Cities and Urban Informatics

Abstract—

*Index Terms—*smart cities, Internet-of-Things

I. Introduction

II. Overview

We may describe a smart IoT system built on four different layers, including the sensing plane, network plane, data plane, and the application plane. The research challenges, related work, and future directions could be discussed in the following sections of the paper.

III. Research Problems

(Each topic: research challenges, existing technologies, future directions)

A. Application plane

By accumulating and analyzing the urban information collected from ubiquitous sensing devices in the city-scale, intelligent urban informatics can help the urban administrators better understand the metropolitan needs and challenges, as well improve urban citizens' awareness of city conditions in various aspects, e.g., the states of transportation systems and the surround environments, to facilitate people's daily life. Traditionally, people deploy some static instrumenting devices to continually sense urban city, and then provide the public with real-time urban informatics. For example, *Kyun* [1] monitors the real-time traffic queues via analyzing wireless signals of sensor nodes placed at roadsides. *CitySee* [2] is a real-time CO_2 monitoring system based on thousands of sensor nodes deployed in city-scale, which makes government authorities and the public be aware of precise carbon emission. Such static sensors based monitoring and sensing systems usually need to deploy substantial instruments, which will incur huge deployment and maintenance costs and thus result in limited coverage of urban areas. Recent studies turn to gather and mine urban information generated from advanced infrastructures existing in modern cities, e.g., public transportation systems and wireless cellular networks. Though these systems were originally designed for other purposes, they potentially carry some information about the urban dynamics, through which people can better understand the urban activities and phenomenons of urban cities. Taking the call detail records from cellular networks, *WHERE* [3] can model the human mobility, which is useful for urban planning in smart city. *mPat* [4] further enhances the human mobility model by involving data from public transportation systems, e.g., taxi, buses, and subways. The human mobility models are used to design efficient carpool systems [5] and last-miles services [6]

to improve the travel experiences of urban citizens. Besides, the data from advanced infrastructures, e.g., taxi trajectories, can provide people with knowledges about real-time traffic volumes [7], traffic conditions [8][9], road surface conditions [10], better driving routes [11], and many other applications and services [12], which largely strength people's perception about the infrastructures of urban city.

Compared to above static and passive data collection approaches, nowadays people begin to proactively collect data from urban city, and have designed various applications based on the emerging data collection paradigm called Crowdsourcing [13]. Crowdsourcing based systems for urban informatics leverages modern smartphones, which are equipped with rich sensors and powerful CPUs. The high penetration and rich availability of such smartphones make it possible to outsource the urban informatics sensing and data collection tasks to the ordinary people, who are both the information source provider and the major consumers of final output. Crowdsourcing plays an important role to realize intelligent transportation systems, and makes such kind of systems operate independent of any city authorities and deployments of special infrastructures. Zhou *et al.* [14] presents a system that can accurately predict the bus arrival time based on the limited crowdsourced information, e.g., cell tower signals, movement status and audio recordings, from the bus passengers. They also utilize the same information collected from the bus riders to derive the instant traffic map of the city [15]. By collaboratively sensing the traffic light status among drivers in a road network, *SignalGuru* [16] can predict traffic signal's future schedule and thus advise optimal speeds for drives when they head towards a signalized intersection for the fuel-saving purpose. *PocketParker* [17] is a parking lot availability prediction system that can recognize parking events via activity detection and analyze the parking lot status with a sophisticated parking lot availability model based on the amount of parking information crowdsourced from users. Besides the application in transportation, crowdsourcing the urban informatics tasks to the public can also achieve noise pollution monitoring of urban area [18], indoor localization [19], indoor navigation [20], air quality monitoring [21][22], and so on.

B. Data plane

With urban sensing in city-scale and urban data collected from various instrumenting devices, urban information and knowledge have to be analyzed and mined from those raw sensing data so as to support a variety of upper urban informatics applications. As a standard knowledge discovery procedure which includes data pre-processing, data management, and data fusion (or knowledge discovery), urban informatics data

analytics involves many techniques to convert the noisy, sparse, heterogeneous, and massive urban data to insightful and valuable urban information and knowledge. In the following, we will discuss the main research challenges involved in urban informatics data analysis and the existing available techniques.

1) Data quality.: In urban informatics systems, there are different types of sensing devices, which may be statically deployed, e.g., sensor networks, or movable in large scale like smartphones in crowdsourcing based data collection, and are with heterogeneous capability and facing various complicated environments. As a result, the gathered sensing data are usually in poor quality along with various data issues like noise, sparsity, and heterogeneity. It is thus necessary to carefully pre-process these raw data with some advanced techniques so as to preserve the insight and valuable information in raw data. On one hand, we have to filter the noisy or unreliable data to guarantee the accuracy of urban informatics; on the other hand, we need to recovery the complete big picture of the urban informatics to analyze the urban phenomenons without detail missing.

The data quality of urban informatics sensing is mainly caused by the capability of sensing devices, e.g., in low power, or the complexity of sensing environment, thus it is essential to remove those noise or erroneous data by borrowing some advanced techniques from database domain. At the sensing device end, different techniques, e.g., outlier detection [23] and data filtering [24], should be involved to detect and filter the unreliable data, so as to improve the data uploading efficiency. At the server end, advanced techniques like data cleaning [25] could be further employed to eliminate the inaccurate data. Besides the unreliable data issue, urban informatics systems will always have the data sparsity problem due to the sparse and dynamical characteristics of distributed sensing network. For example, there may exist data missing due to sensor failure in static sensor networks, while in the crowdsourcing scenario, it is impossible for the volunteers to participant in the data collection at each point of interest at any time. As a result, it turns to be challenging to derive the complete urban informatics observations at the server end. Data sparsity is a general challenge that has been studied for years in many computing domains, and some typical techniques, such as collaborative filtering [26], singular value decomposition [27], and tensor decomposition [28], can be used for reference in the urban informatics system design. Recently, a promising technique named compressive sensing becomes a powerful tool for the data sparsity issue, which can recover the global signal even with a small number of measurements if the signal of interest is sparse or compressible [29]. The compressive sensing theory fits perfectly with the sensing network design whose sensing end is less capable than the server end, and thus have been applied in various urban informatics applications, e.g., traffic estimation [8]. Compressive sensing makes global recovery available even with sparse data mainly due to the correlation among data, and can achieve much better performance if such correlation is well exploited. Wu *et al.* [30] propose a compressive crowdsensing framework that explicitly derives the data correlation and enable compressive sensing in various crowdsourcing scenarios.

2) Big data management.: The data generated in urban city is an typical example in big data era, and possess the three characteristics of big data, i.e., volume, variety, and velocity. There exist different urban data, e.g., data from transportation system, environment monitoring, and social networks, and they are generated in large volumes at a high speed, e.g., taxis equipped with GPS devices usually report one status message every minute and there are tens of thousands of taxis in a city. How to store, manage, and analyze these massive, heterogeneous, and noisy urban data is a big challenge.

The well-known programming model, i.e., MapReduce [31], is widely used to process the big data for high performance computing in a parallel and distributed manner. The Hadoop Distributed File System (HDFS) [32] can store vary large data set reliably and provide a framework for the analysis and transmission of these data using MapReduce paradigm. Building on the MapReduce concept and taking the HDFS as the storage base, various tools for big data managing and processing have been proposed. To simplify the MapReduce programming and quicken the query on the large data set, Hive [33] built on Hadoop is a warehousing solution which supports queries expressed in a SQL-like declarative language. Similarly, Pig [34] is a high-level dataflow system and can support SQL-style high-level data manipulation constructs as well. To further optimize the performance of MapReduce on those applications that reuse a working data set across multiple parallel operations, Zaharia *et al.* propose a new framework Spark [35], which keeps data in memory for operation and introduces an abstraction named resilient distributed datasets (RDDs) [36] to retain the scalability and fault tolerance of MapReduce. Building on the Spark framework, GraphX [37] can efficiently express graph computation to achieve data-parallel and graph-parallel, so as to enable some advanced machine learning and data mining algorithms. Above frameworks and tools immensely improve our capability to handle the massive and heterogeneous data gathered from urban city. For example, by making use of these tools for huge transportation data management and analysis, Zhang *et al.* have proposed a series of applications to support urban informatics, e.g., carpool service [5], last-mile service [6], and human mobility modeling [4], and so on.

3) Multi-dimensional data fusion.: Considering the inherent multi-dimensional nature of urban informatics, viewing the data across multi-categories and substantial data fusion on dimensions is critical to guarantee the best information retrieval accuracy. Instead of being independent, information on different dimensions is usually highly related to each other. Separate data observation on individual dimension though may still provide some useful information, this kind of approach fails to exploit the relationship of various data categories and thus are not able to deeply mine the status of urban city. How to characterize the relationship among different data dimensions and how to make the best use of such relationship depends on the application characteristics. How to design an efficient multi-dimensional data fusion is still a challenging problem in the urban informatics domain.

There are three major approaches to achieve the data fusion across various data sources as summarized in [38]. Firstly,

we can treat different data sources equally and put them together to extract the overall feature vector. Secondly, we can also use data at different stages, and thus view them as the output of different phases in urban dynamics. Thirdly, we can feed data into different parts of one model or feed each data source to different models, and then manipulate the final model. There exist some tangible examples in the literature for multi-dimensional data fusion. Zheng *et al.* [22] propose a system called *U-Air* to infer real-time and fine-grained air quality information throughout the city using a variety of data sources. Specifically, they take the road network topology and information about point of interests like shopping mall as input to train an artificial neural network (ANN) for modeling the spatial correlation between air quality and different locations. They also take the meteorology and traffic data as input to train a linear-chain conditional random field (CRF) for modeling the temporal correlation among air quality in a location. They finally manipulate the ANN and CRF models by a weighting strategy to achieve accurate results. Zhang *et al.* [39] propose a framework named *coMobile* to infer real-time human mobility based on the modeling of multi-source data. They feed different data sources, e.g., cellphone data, taxi data, bus data, subway data, and etc., into different models to train several single-view models, e.g., cellphone-view and transportation-view of human mobility, and then integrate them together to obtain an multi-view model by solving a convex optimization problem. The derived human mobility is proved better than single-view modeling approach.

C. Network plane

network infrastructure in city, wireless communications, green networking

D. Sensing plane

Compared to the static wireless sensor networks and complex CPS systems, crowdsourcing based urban informatics systems make the urban sensing and data collection feasible in large scale and independent of special infrastructures. However, how to design an efficient and reliable crowdsourcing based data collection approach is non-trivial due to various factors. In the following, we present the main research challenges of crowdsourcing based urban sensing and data collection, as well discuss the existing works.

1) Energy efficient sensing.: Given the wide range of available sensing capabilities (accelerometer, magnetic sensors, cameras, microphones, gyroscope, etc.), smartphones are the essential sensing resources in crowdsourcing based applications. Due to limited energy density and battery size, however, the energy issue of smartphone is usually one major concern when used for sensing and data collection. Configured with more powerful hardware and more complex software, the energy drain of smartphones becomes even more serious. As a result, carefully consideration of both the sensing accuracy and energy efficiency is critical to satisfy the crowdsourcing application requirements and placing affordable overhead that encourage user participation.

There have been some system-level tools to manage the rare energy of smartphones. *eDoctor* [40] diagnoses abnormal battery drain via identifying abnormal App behavior. Pathak *et al.* [41] proposes a fine-grained power modeling method by tracing the system calls of applications, and thus enable the accurate online energy estimation. Besides careful energy management, it is also critical for smartphone to intelligently make the best decision about urban informatics sensing and data collection in accordance to the instant environment and user conditions. Context awareness is the key for such adaptive sensing, and there exists many valuable research efforts. *IODetector* [42] is a fast, efficient, and lightweight generic service for indoor outdoor detection, and can benefit plenty of location-based sensing and data collection tasks. A^3 [43] is an accurate and automatic attitude detector, which derives the gesture of smartphone and thus ensure the data collection quality. Besides above context awareness, there are many other aspects of context awareness desired in urban informatics applications, e.g., logical localization (e.g., restaurant, bookstore, etc.) [44], user activity awareness (e.g., walking, jogging, cycling, etc.) [45], transportation mode awareness (e.g., vehicle, subways, etc.) [46], and so on. To perform energy efficient context detection, *ACE* [47] dynamically learns relationships among different context attributes and supports continuous context awareness while mitigating sensing costs. *PCS* [48] performs crowdsourcing data collection by exploiting smartphone App opportunities, such as placing a call or using applications, to lower the energy overhead of user participation. Context awareness could help decide when to sense and collect the data, while it is equally important to decide what kinds of data to collect given the urban informatics requirements. This is because though different sensors may delivery similar information while they may have distinct energy consumption profiles. For example, both the GPS and cellular sensor can collect information to infer the location while GPS consume much more energy than the cellular sensor. For example, the work in [14] adopts using cellular sensor rather than the energy-hungry GPS to collect data for bus location inference. Energy-friendly sensors of smartphones are more preferable in the urban informatics application design on the condition that they can collect the same useful information.

2) Sensing privacy.: Due to the involvement of human in crowdsourcing based sensing and data collection, a natural requirement in crowdsourcing based system is to guarantee the data sensing privacy. As participants' smartphones serve as the major sensing probes, which may sense and contain the privacy data of users, such as location and personal information, the protection of user privacy is essential to encourage the participation. Given the characteristics of distributed, dynamical, large-scale, and heterogeneous in crowdsourcing based urban informatics system,s however, it becomes more critical and challenging for the protection of sensing privacy.

To guarantee the participants' privacy in crowsouring, on one hand, it is important for the data sensing techniques to avoid collecting privacy sensitive data for the same sensing purpose. On the other hand, it is necessary to avoid revealing the privacy sensitive data, once these data have been collected for utility purpose, to other parts. *ipShield* [49] is a context-

aware privacy enforcement framework that allows users to define the safe shared sensor data, which are impossible to be used for privacy inference. [50] proposes a privacy-aware framework for spatial crowdsourcing to protect the location privacy of participants via exploiting differential privacy technique, which allows interactions of a center database only by means of aggregate (e.g., sum, count) queries. *PPTD* [51] is a cloud-enabled privacy-preserving framework for weighted aggregation based truth discovery. By exploiting homomorphic cryptosystem, *PPTD* can not only protect the users' sensitive sensor data, but also their reliability scores, which could also be used for private personal information inference, derived by truth discovery approaches. It is non-trivial to achieve fully complete privacy-preserving as it calls for efforts of many parties involved in the crowdsourcing procedure. At the user end, the participant should carefully define the safe sharable sensor data to guarantee her own privacy. At the server end, the urban informatics system should enable the authentication of participants and only let truthful and reliable participants join in the crowdsourcing tasks [52]. For example in the mobile Healthcare emergence scenario, the medical users should be enabled to decide who can participate in the crowdsourcing computing to assist in processing her private personal health information via an efficient user-centric privacy access control to minimize the privacy disclosure [53].

3) Incentive mechanisms.: Despite the rich potential of crowdsourcing based urban informatics applications, people are increasingly aware that their effectiveness critically depend on the willingness of participants from user engagement to their compliance with the crowdsourced tasks. While participating in a crowdsourcing campaign, the smartphone users consume their own resources such as battery and computing power, and meanwhile undertake the potential risks of privacy disclosure. For the sakes of attracting more participants for the scale and quality requirements of urban data collection, and as well compensating the costs of crowdsourcing volunteers for their resource and time investments, it is necessary and essential to design an effective incentive mechanisms for the crowdsourcing based urban informatics systems.

There are many research efforts focusing on the incentive mechanisms recently due to the popularity of crowsourcing based data collection paradigm. The incentive types can be subdivided into *intrinsic* motivations, which can satisfy individual via entertainment or services, and *extrinsic* motivations, which attract participants via credits or even monetary payment [54]. For the former case, it requires the crowdsourcing applications to be mutually benefit for both the application itself and the participants. For example, in the bus arrival time prediction service [14] and the noise monitoring application [18], the participants of crowdsourcing are both information providers and information consumers. Such kinds of crowdsourcing applications are born to attract participants, while there are still lots of crowdsourcing applications needing external incentives for encouraging more participants. There exists some incentive mechanisms designed for the latter incentive scenario. Yang *et al.* [55] have considered two types of incentive mechanisms models: platform-centric inventive mechanisms and user-centric inventive mechanisms. For the

platform-centric model where the platform leads the reward assignments among users, they make use of gaming theory to maximize the overall utility. For the user-centric model, they adopt an auction-based mechanisms. Zhao *et al.* [56] consider a new crowdsourcing scenario where participants are dynamical available and propose an online incentive mechanisms which can select a subset of participants to maximize the utility for a given budget constraint by modeling the problem as an online auction. Li *et al.* [57] propose two incentive mechanisms with or without an online trusted third part, both of which can simultaneously address the privacy and incentive problems in crowdsourcing based systems. Rather than increasing the data size, Kawajiri *et al.* [58] have proposed the steered crowdsourcing framework with incentive design, which controls incentives by introducing gamifications to guarantee data quality and the final quality of service. There are many other valuable incentive mechanisms, and interested readers are referred to a survey paper in [59].

IV. CROSS-PLANE DESIGN

- emphasizing on the cross-layer design, and the interaction of different layers in the system.
- research consideration for the smart system as a whole

V. FUTURE DIRECTIONS

A. Future direction of crowdsourcing

A holistic crowdsourcing based data collection scheme, which includes the considerations about energy efficiency, privacy protection and incentive mechanisms, is desired for the accurate, efficient, and reliable urban informatics systems. As crowdsourcing is loose-controlled data sensing and collection paradigm, the data generated via a crowdsourcing method is nonuniformly distributed in the geographical space. We may have more data than we need at some places while we have no input at some other locations. Rather than simply assigning participants to collect data at these locations without crowdsourced data, it is more preferable to mine the spatial correlation among locations firstly and then predict the most valuable barren locations, whose data can improve the overall informatics accuracy the most, based on current available data. Such a feedback based adaptive sensing can not only determine the target incentive locations in advance to meet the timeliness of urban informatics systems, but also achieve the maximum benefit, e.g., accuracy improvement, with the minimum incentive budget. Besides, for the privacy issue in crowdsourcing, it is important to study how the same type of data can be of different privacy sensitivity in different environment and conditions. Such a differential design can not only protect the privacy of participants according to the context, but also enrich the available crowdsourcing data.

VI. CONCLUSIONS

REFERENCES

[1] R. Sen, A. Maurya, B. Raman, R. Mehta, R. Kalyanaraman, N. Vankadhara, S. Roy, and P. Sharma, "Kyun queue: a sensor network system to monitor road traffic queues," in *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems (SenSys)*. ACM, 2012, pp. 127–140.

[2] X. Mao, X. Miao, Y. He, X.-Y. Li, and Y. Liu, "Citysee: Urban co2 monitoring with sensors," in *Proceedings of IEEE INFOCOM*. IEEE, 2012, pp. 1611–1619.

[3] S. Isaacman, R. Becker, R. Cáceres, M. Martonosi, J. Rowland, A. Varshavsky, and W. Willinger, "Human mobility modeling at metropolitan scales," in *Proceedings of the 10th international conference on Mobile systems, applications, and services (MobiSys)*. ACM, 2012, pp. 239–252.

[4] D. Zhang, J. Huang, Y. Li, F. Zhang, C. Xu, and T. He, "Exploring human mobility with multi-source data at extremely large metropolitan scales," in *Proceedings of the 20th annual international conference on Mobile computing and networking (MobiCom)*. ACM, 2014, pp. 201–212.

[5] D. Zhang, Y. Li, F. Zhang, M. Lu, Y. Liu, and T. He, "coride: carpool service with a win-win fare model for large-scale taxicab networks," in *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems (SenSys)*. ACM, 2013, p. 9.

[6] D. Zhang, J. Zhao, F. Zhang, R. Jiang, and T. He, "Feeder: supporting last-mile transit with extreme-scale urban infrastructure data," in *Proceedings of the 14th International Conference on Information Processing in Sensor Networks (IPSN)*. ACM, 2015, pp. 226–237.

[7] J. Aslam, S. Lim, X. Pan, and D. Rus, "City-scale traffic estimation from a roving sensor network," in *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems (SenSys)*. ACM, 2012, pp. 141–154.

[8] Y. Zhu, Z. Li, H. Zhu, M. Li, and Q. Zhang, "A compressive sensing approach to urban traffic estimation with probe vehicles," *IEEE Transactions on Mobile Computing (TMC)*, vol. 12, no. 11, pp. 2289–2302, 2013.

[9] B. Yang, C. Guo, and C. S. Jensen, "Travel cost inference from sparse, spatio temporally correlated time series using markov models," *Proceedings of the VLDB Endowment*, vol. 6, no. 9, pp. 769–780, 2013.

[10] J. Eriksson, L. Girod, B. Hull, R. Newton, S. Madden, and H. Balakrishnan, "The pothole patrol: using a mobile sensor network for road surface monitoring," in *Proceedings of the 6th international conference on Mobile systems, applications, and services (MobiSys)*. ACM, 2008, pp. 29–39.

[11] J. Yuan, Y. Zheng, X. Xie, and G. Sun, "Driving with knowledge from the physical world," in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD)*. ACM, 2011, pp. 316–324.

[12] Y. Zheng, "Trajectory data mining: an overview," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 6, no. 3, p. 29, 2015.

[13] J. Howe, "The rise of crowdsourcing," *Wired magazine*, vol. 14, no. 6, pp. 1–4, 2006.

[14] P. Zhou, Y. Zheng, and M. Li, "How long to wait?: predicting bus arrival time with mobile phone based participatory sensing," in *Proceedings of the 10th international conference on Mobile systems, applications, and services (MobiSys)*. ACM, 2012.

[15] P. Zhou, S. Jiang, and M. Li, "Urban traffic monitoring with the help of bus riders," in *The 35th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2015.

[16] E. Koukoumidis, L.-S. Peh, and M. R. Martonosi, "Signalguru: leveraging mobile phones for collaborative traffic signal schedule advisory," in *Proceedings of the 9th international conference on Mobile systems, applications, and services (MobiSys)*. ACM, 2011, pp. 127–140.

[17] A. Nandugudi, T. Ki, C. Nuessle, and G. Challen, "Pocketparker: Pocketsourcing parking lot availability," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*. ACM, 2014, pp. 963–973.

[18] R. K. Rana, C. T. Chou, S. S. Kanhere, N. Bulusu, and W. Hu, "Ear-phone: an end-to-end participatory urban noise mapping system," in *Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. ACM, 2010, pp. 105–116.

[19] Z. Yang, C. Wu, and Y. Liu, "Locating in fingerprint space: wireless indoor localization with little human intervention," in *Proceedings of the 18th annual international conference on Mobile computing and networking (MobiCom)*. ACM, 2012, pp. 269–280.

[20] Y. Zheng, G. Shen, L. Li, C. Zhao, M. Li, and F. Zhao, "Travi-navi: Self-deployable indoor navigation system," in *Proceedings of the 20th annual international conference on Mobile computing and networking (MobiCom)*. ACM, 2014, pp. 471–482.

[21] D. Hasenfratz, O. Saukh, S. Sturzenegger, and L. Thiele, "Participatory air pollution monitoring using smartphones," *Mobile Sensing*, 2012.

[22] Y. Zheng, F. Liu, and H.-P. Hsieh, "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 (KDD)*. ACM, 2013, pp. 1436–1444.

[23] Y. Zhang, N. Meratnia, and P. Havinga, "Outlier detection techniques for wireless sensor networks: A survey," *IEEE Communications Surveys & Tutorials*, vol. 12, no. 2, pp. 159–170, 2010.

[24] A. G. Parameswaran, H. Garcia-Molina, H. Park, N. Polyzotis, A. Ramesh, and J. Widom, "Crowdscreen: Algorithms for filtering data with humans," in *Proceedings of the VLDB Endowment*. ACM, 2010, pp. 361–372.

[25] E. Rahm and H. H. Do, "Data cleaning: Problems and current approaches," *IEEE Data Eng. Bull.*, vol. 23, no. 4, pp. 3–13, 2000.

[26] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," *Advances in artificial intelligence*, vol. 2009, p. 4, 2009.

[27] Y. Zhang, M. Roughan, W. Willinger, and L. Qiu, "Spatio-temporal compressive sensing and internet traffic matrices," in *ACM SIGCOMM Computer Communication Review*, vol. 39, no. 4. ACM, 2009, pp. 267–278.

[28] J. C. Ho, J. Ghosh, and J. Sun, "Marble: high-throughput phenotyping from electronic health records via sparse nonnegative tensor factorization," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD)*. ACM, 2014, pp. 115–124.

[29] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory (TIT)*, vol. 52, no. 4, pp. 1289–1306, 2006.

[30] L. Xu, X. Hao, N. D. Lane, X. Liu, and T. Moscibroda, "More with less: lowering user burden in mobile crowdsourcing through compressive sensing," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*. ACM, 2015, pp. 659–670.

[31] J. Dean and S. Ghemawat, "Mapreduce: simplified data processing on large clusters," *OSDI*, 2004.

[32] K. Shvachko, H. Kuang, S. Radia, and R. Chansler, "The hadoop distributed file system," in *IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST)*. IEEE, 2010, pp. 1–10.

[33] A. Thusoo, J. S. Sarma, N. Jain, Z. Shao, P. Chakka, S. Anthony, H. Liu, P. Wyckoff, and R. Murthy, "Hive: a warehousing solution over a mapreduce framework," *Proceedings of the VLDB Endowment*, vol. 2, no. 2, pp. 1626–1629, 2009.

[34] A. F. Gates, O. Natkovich, S. Chopra, P. Kamath, S. M. Narayanamurthy, C. Olston, B. Reed, S. Srinivasan, and U. Srivastava, "Building a high-level dataflow system on top of map-reduce: the pig experience," *Proceedings of the VLDB Endowment*, vol. 2, no. 2, pp. 1414–1425, 2009.

[35] M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker, and I. Stoica, "Spark: cluster computing with working sets," in *Proceedings of the 2nd USENIX conference on Hot topics in cloud computing*, vol. 10, 2010, p. 10.

[36] M. Zaharia, M. Chowdhury, T. Das, A. Dave, J. Ma, M. McCauley, M. J. Franklin, S. Shenker, and I. Stoica, "Resilient distributed datasets: A fault-tolerant abstraction for in-memory cluster computing," in *Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation (NSDI)*. USENIX Association, 2012, pp. 2–2.

[37] R. S. Xin, J. E. Gonzalez, M. J. Franklin, and I. Stoica, "Graphx: A resilient distributed graph system on spark," in *First International Workshop on Graph Data Management Experiences and Systems*. ACM, 2013, p. 2.

[38] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban computing: concepts, methodologies, and applications," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 5, no. 3, p. 38, 2014.

[39] D. Zhang, J. Zhao, F. Zhang, and T. He, "comobile: Real-time human mobility modeling at urban scale using multi-view learning," 2015.

[40] X. Ma, P. Huang, X. Jin, P. Wang, S. Park, D. Shen, Y. Zhou, L. K. Saul, and G. M. Voelker, "edoctor: Automatically diagnosing abnormal battery drain issues on smartphones," in *10th USENIX Symposium on Networked Systems Design and Implementation (NSDI)*, vol. 13, 2013, pp. 57–70.

[41] A. Pathak, Y. C. Hu, M. Zhang, P. Bahl, and Y.-M. Wang, "Fine-grained power modeling for smartphones using system call tracing," in *Proceedings of the sixth conference on Computer systems*. ACM, 2011, pp. 153–168.

[42] P. Zhou, Y. Zheng, Z. Li, M. Li, and G. Shen, "IODetector: A generic service for indoor outdoor detection," in *Proceedings of the 10th ACM conference on embedded network sensor systems (SenSys)*. ACM, 2012, pp. 113–126.

[43] P. Zhou, M. Li, and G. Shen, "Use it free: Instantly knowing your phone attitude," in *Proceedings of the 20th annual international conference on*

Mobile computing and networking (MobiCom).   ACM, 2014, pp. 605–616.

[44] M. Azizyan, I. Constandache, and R. Roy Choudhury, "Surroundsense: mobile phone localization via ambience fingerprinting," in *Proceedings of the 15th annual international conference on Mobile computing and networking (MobiCom)*.   ACM, 2009, pp. 261–272.

[45] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," *ACM SigKDD Explorations Newsletter*, vol. 12, no. 2, pp. 74–82, 2011.

[46] S. Hemminki, P. Nurmi, and S. Tarkoma, "Accelerometer-based transportation mode detection on smartphones," in *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems (SenSys)*.   ACM, 2013, p. 13.

[47] S. Nath, "ACE: exploiting correlation for energy-efficient and continuous context sensing," in *Proceedings of the 10th international conference on Mobile systems, applications, and services (MobiSys)*.   ACM, 2012, pp. 29–42.

[48] N. D. Lane, Y. Chon, L. Zhou, Y. Zhang, F. Li, D. Kim, G. Ding, F. Zhao, and H. Cha, "Piggyback crowdsensing (pcs): energy efficient crowdsourcing of mobile sensor data by exploiting smartphone app opportunities," in *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems (SenSys)*.   ACM, 2013, p. 7.

[49] S. Chakraborty, C. Shen, K. R. Raghavan, Y. Shoukry, M. Millar, and M. Srivastava, "ipshield: a framework for enforcing context-aware privacy," in *11th USENIX Symposium on Networked Systems Design and Implementation (NSDI)*.   USENIX Association, 2014, pp. 143–156.

[50] H. To, G. Ghinita, and C. Shahabi, "A framework for protecting worker location privacy in spatial crowdsourcing," *Proceedings of the VLDB Endowment*, vol. 7, no. 10, pp. 919–930, 2014.

[51] C. Miao, W. Jiang, L. Su, Y. Li, S. Guo, Z. Qin, H. Xiao, J. Gao, and K. Ren, "Cloud-enabled privacy-preserving truth discovery in crowd sensing systems," in *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems (SenSys)*.   ACM, 2015, pp. 183–196.

[52] K. Yang, K. Zhang, J. Ren, and X. Shen, "Security and privacy in mobile crowdsourcing networks: challenges and opportunities," *IEEE Communications Magazine*, vol. 53, no. 8, pp. 75–81, 2015.

[53] R. Lu, X. Lin, and X. Shen, "SPOC: A secure and privacy-preserving opportunistic computing framework for mobile-healthcare emergency," *IEEE Transactions on Parallel and Distributed Systems (TPDS)*, vol. 24, no. 3, pp. 614–624, 2013.

[54] J. P. Rula, V. Navda, F. E. Bustamante, R. Bhagwan, and S. Guha, "No one-size fits all: Towards a principled approach for incentives in mobile crowdsourcing," in *Proceedings of the 15th Workshop on Mobile Computing Systems and Applications*.   ACM, 2014, p. 3.

[55] D. Yang, G. Xue, X. Fang, and J. Tang, "Crowdsourcing to smartphones: incentive mechanism design for mobile phone sensing," in *Proceedings of the 18th annual international conference on Mobile computing and networking (MobiCom)*.   ACM, 2012, pp. 173–184.

[56] D. Zhao, X.-Y. Li, and H. Ma, "How to crowdsource tasks truthfully without sacrificing utility: Online incentive mechanisms with budget constraint," in *Proceedings of IEEE INFOCOM*.   IEEE, 2014, pp. 1213–1221.

[57] Q. Li and G. Cao, "Providing privacy-aware incentives in mobile sensing systems," *IEEE Transactions on Mobile Computing (TMC)*, 2015.

[58] R. Kawajiri, M. Shimosaka, and H. Kahima, "Steered crowdsensing: incentive design towards quality-oriented place-centric crowdsensing," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*.   ACM, 2014, pp. 691–701.

[59] X. Zhang, Z. Yang, W. Sun, Y. Liu, S. Tang, K. Xing, and X. Mao, "Incentives for mobile crowd sensing: A survey," *IEEE Communications Surveys & Tutorials*, 2015.