Measurement: Sensors 19 (2022) 100061

Contents lists available at ScienceDirect

Measurement: Sensors

journal homepage: www.sciencedirect.com/journal/measurement-sensors

A survey on multisource heterogeneous urban sensor access and data management technologies

Fei Yang [a], Yixin Hua [a,*], Xiang Li [a], Zhenkai Yang [a], Xinkai Yu [a], Teng Fei [b]

[a] Institute of Surveying and Mapping, Information Engineering University, Zhengzhou, 450052, China
[b] School of Resource and Environmental Sciences, Wuhan University, 430079, Wuhan, China

ARTICLE INFO

Keywords:
Smart city
Urban sensor
Access
Data management
Computational burden
Energy consumption
Cybersecurity

ABSTRACT

Urban sensors are an important part of urban infrastructures and are usually heterogeneous. Urban sensors with different uses vary greatly in hardware structure, communication protocols, data formats, interaction modes, sampling frequencies, data accuracy and service quality, thus posing an enormous challenge to the unified integration and sharing of massive sensor information resources. Consequently, access and data management methods for these multisource heterogeneous urban sensors are extremely important. Additionally, multisource heterogeneous urban sensor access and data management technologies provide strong support for intelligent perception and scientific management at the city scale and can accelerate the construction of smart cities or digital twin cities with virtual reality features. We systematically summarize the related research on these technologies. First, we present a summary of the concepts and applications of urban sensors. Then, the research progress on multisource heterogeneous urban sensor access technologies is analysed in relation to communication protocols, data transmission formats, access standards, access technologies and data transmission technologies. Subsequently, the data management technologies for urban sensors are reviewed from the perspectives of data cleaning, data compression, data storage, data indexing and data querying. In addition, the challenges faced by the technologies above and corresponding feasible solutions are discussed from three aspects, namely, the integration of massive Internet of Things (IoT), computational burden and energy consumption and cybersecurity. Finally, a summary of this paper is given, and possible future development directions are analysed and discussed.

1. Introduction

Urban sensors play an irreplaceable role in meeting the demand of humans for ubiquitous intelligent perception in cities. With the development of the Internet of Things (IoT), the Internet and communication technologies, the number and types of urban sensors have increased rapidly in recent years [1], with the data volume exponentially exploding, and the functions of sensors have become increasingly broad to support smart cities and digital twin city construction [2]. Nowadays, urban sensors have been widely used in many fields, such as environmental monitoring, disaster forecasting, scientific research, military actions, intelligent transportation, industrial monitoring, and medical health [3,4], thus supporting the real-time and dynamic perception of urban activities and enabling resource sharing, comprehensive perception, precise control, intelligence clustering and rapid correspondence in smart city management. In addition, almost all the sensing tasks can be accomplished with the support of urban sensors, such as ubiquitous computing, which integrates computing, communication, and sensing [5,6]; social awareness computing, which emphasizes real-time perception and interactions in the real world [7-9]; context awareness, which highlights the expression of contextual information and the corresponding interactions [10]; urban computing, which focuses on urban perception and information services [11,12]; edge computing, which can be used for the real-time analysis and processing of big data on the data source side [13-15]; fog computing, which involves the migration of central cloud computing tasks to the device side to perform highly virtualized computing [16,17]; or digital twin city applications, which are characterized by the symbiosis of the physical world and the digital world [18,19]. Other relevant applications include digital twin battlefields [20], urban brains [21], digital China [22] and voluntary geographic information (VGI) [23]. To be clarified, the "urban sensors" we discuss in this paper are a subset of sensors that includes sensors that

* Corresponding author.
E-mail address: hy3gcdx@163.com (Y. Hua).

https://doi.org/10.1016/j.measen.2021.100061
Received 11 March 2021; Received in revised form 8 September 2021; Accepted 9 September 2021
Available online 29 September 2021

are widely distributed in various industries and city departments and that are used for the collection, storage, processing, and interoperability of urban perception information [24-26]. In some works, social media, or large social network platforms (Big SNSs, such as Twitter and Facebook) have been regarded as types of "virtual sensors" or "social sensors" [27-30], but they are not considered in this paper. In other words, the "urban sensors" in this paper only refer to physical sensors.

Urban sensor access and data management technologies are the key technologies in urban sensor networks and lay a solid foundation for the construction of urban sensor applications, such as smart cities or digital twin cities. In recent years, with different application needs, investment scales and planning levels, urban sensors differ greatly in their hardware structures, communication protocols, data formats, interaction modes, sampling frequencies, data accuracy levels and service quality, and urban sensor data have been distributed in various organizations or departments, which creates an enormous challenge related to the unified integration and sharing of massive sensor information resources for urban sensor applications. As a result, research on multisource heterogeneous urban sensor access and data management technologies has become an urgent need and still require further review. In detail, there are two aspects of technical problems that have a great influence on urban sensor application construction but have not been solved well and still need to be improved:

. Multisource heterogeneous urban sensor access technologies. These technologies help connect urban sensors or other perception devices with physical and network layers, and enable the data transmission of urban sensors between urban sensor nodes. Many aspects are required in multisource heterogeneous urban sensor access technologies to support urban applications better, and these aspects include communication protocols, data transmission formats, access standards, access technologies and data transmission technologies.
. Multisource heterogeneous urban sensor data management technologies. These technologies are developed for data cleaning, data compression, data storage, data indexing and data querying, respectively corresponding to the requirements for higher data quality, preferable data transmission efficiency, better data integration and quicker data querying response.

Many researchers have performed extensive research on multisource heterogeneous urban sensor access and data management· technologies and have made considerable progress, and there have been much literature review concerning the summary of these works. However, not all the aspects of the technologies needed in urban sensor applications have been concerned in previous reviews. We summarize the existing recent relevant studies and highlight their research gap in Table 1. From Table 1, it can be concluded that most of the existing surveys covered only some aspects of the technologies used in urban sensor applications, which still need to be updated based on the latest technological advances, and even some crucial aspects were not considered in them, such as data transmission formats and data cleaning technologies. In detail, what has been missing from previous surveys on the proposed domain are listed as follows:

. Urban sensor application platforms. Few of the existing studies only concerned few of the existing relevant platforms, which lacks sufficient credibility.
. Data transmission formats. The existing studies didn't consider the data transmission formats, which is important for urban sensor applications construction.
. A systematic survey of sensor access standards. Some important urban sensor access standards are missing from the existing studies, such as the OGC PUCK standard and the DICOM standard, which have been popular and widely used in urban sensor access in recent years.
. A uniform technical framework of access technologies based on customized adaptation layers. Although some sensor access technologies based on customized adaptation layers have been proposed recently, they were applied in different scenarios and not uniformed.
. The survey of urban sensor data cleaning technologies. Current studies didn't cover some needed technologies of multisource heterogeneous urban sensor data management, such as data cleaning technologies.
. The newest analysis of the main open challenges faced by the latest relevant technologies. This kind of analysis is needed as a guideline for future research to drive innovative development in the construction of urban sensor applications such as smart cities or digital twin cities.
. Research directions for future work. Possible directions for future work have not been well covered in current state-of-the-art research.
. The newest survey of the other relevant technologies, such as data transmission technologies and data compression technologies.

In contrast, this paper offers the newest analysis on the available technologies by reviewing studies mostly in the past five years on various aspects that support urban sensor applications. This study was motivated by the need for reliable and efficient technical solutions for the rapid development of urban sensor applications. Therefore, this paper aims at investigating all the latest technologies of multisource heterogeneous urban sensor access and data management, thus

Table 1
A summary of the existing recent relevant studies and a highlight of the research gap between these studies and the proposed survey. The contributions of the previous relevant surveys and the proposed survey can be more easily measured and compared from this table.

Contributions	Previous Surveys										The Proposed Survey
	[31]	[32]	[33]	[34]	[35]	[36]	[37]	[38]	[39]	[40]	
Applications	✔	✘	✔	✔	✔	✘	✔	✔	✔	✔	✔
Systems & Platforms	✘	✘	✘	✔	✔	✔	✔	✔	✔	✘	✔
Communication Protocols	✔	✔	✔	✘	✔	✔	✔	✘	✘	✔	✔
Data Transmission Formats	✘	✘	✘	✘	✘	✘	✘	✘	✘	✘	✔
Sensor Access Standards	✔	✔	✘	✔	✔	✘	✘	✘	✘	✘	✔
Sensor Access Technologies & Frameworks	✔	✔	✔	✔	✔	✔	✘	✔	✔	✔	✔
Routing Protocols	✘	✔	✘	✘	✔	✘	✔	✔	✔	✔	✔
Data Transmission and Communications	✔	✘	✔	✘	✔	✔	✔	✔	✔	✔	✔
Data Cleaning Technologies	✘	✘	✘	✘	✘	✘	✘	✘	✘	✘	✔
Data Compression Technologies	✘	✘	✔	✘	✘	✘	✔	✘	✔	✘	✔
Data Storage Technologies	✘	✘	✘	✔	✔	✘	✔	✔	✘	✘	✔
Data Indexing and Querying Technologies	✘	✘	✘	✔	✘	✘	✔	✔	✔	✘	✔
Device & Data Integration	✘	✘	✔	✘	✔	✔	✔	✔	✔	✔	✔
Cloud & Edge Computing	✔	✘	✘	✘	✘	✔	✔	✔	✘	✘	✔
Computational Burden and Energy Consumption	✔	✔	✔	✔	✔	✘	✔	✔	✔	✔	✔
Cybersecurity	✔	✔	✘	✔	✔	✔	✔	✔	✔	✘	✔

providing a more comprehensive insight into the requirements, platforms, recent technical advances and open research challenges of urban sensor applications for scholars and governors in fields related to the IoT and smart cities. Detailly, the new additions and contributions of this paper to previous surveys are listed as follows:

- The investigation and presentation of 26 typical urban sensor application systems and IoT platforms to make a summary and comparison of the most popular communication protocols and data transmission formats adopted in current urban applications. This comparative investigation can make it easier for scholars or developers to construct urban sensor applications when considering communication protocols and data transmission formats.
- Adding the survey of data transmission formats, which is important for urban sensor applications construction. The existing studies didn't consider data transmission formats, which is important for urban sensor applications construction. Instead, the most popular data transmission formats of urban sensor applications are contained in the proposed review paper to provide better choices for designers during the construction of urban sensor applications.
- The study of 4 most popular sensor access standards, which provide a uniform paradigm for urban sensor access. Compared with the existing surveys, in addition to making a more systematic survey of the SWE Standards and the IEEE 1451 Standard, we add a survey of the OGC PUCK Standard and the DICOM Standard, which is an effective complement to the previous works. Additionally, we make a comparative analysis of the standards. This analysis can make it more efficient to select sensor access standards for urban sensor applications construction.
- The proposal of a uniform technical framework of access technologies based on customized adaptation layers. This framework shows the basic principles and procedures of the existing access technologies, which serves as a guide for scholars or developers to build customized adaptation layers and connect urban sensors more easily.
- Adding the survey of urban sensor data cleaning technologies. data cleaning technologies play a key role in improving the quality of urban sensor data. This paper provides the relevant discussion for scholars or designers to improve the quality of data services.
- The newest analysis of 3 main open challenges faced by the existing relevant technologies. This analysis can serve as a guide for future research directions in the integration of massive IoT, the energy saving and the security protection of urban sensor applications.
- The discussion and suggestion of new research directions. We suggest the direction of the concerned technologies in urban sensor applications for solving specific problems as a technical guideline for future research.
- Other contributions: 1) The survey of the remaining technologies, such as data transmission technologies and data compression technologies, have also been updated based on the latest literature and works. 2) We adopt different kinds of effective methods to summarize existing recent relevant works, such as the thematic taxonomy method based on the key parameters. All these contributions help to complement and update the review of relevant literature.
- The improvement and extension of the existing body-of-the-knowledge on the relevant theory and technologies. Based on all the contributions above, the existing body-of-the-knowledge on the theory and technologies of urban sensor applications can be complemented and extended. Besides, this work will enable scholars a better understanding of what the most important technologies are and what will have to be done, thus providing them better insight for future research and the development of urban applications; on the other hand, it also provides more reference for governors in the future informatization construction of smart cities or digital twin cities, so that all the public services, such as policy making and emergency rescue can be more efficient and reliable to us.

To achieve the goals above, different kinds of methods have been used to summarize existing recent relevant works. In addition to the method of literature research that can be seen in most of the surveys, some other methods are also adopted in this paper. For example, when concerning multisource heterogeneous urban sensor access technologies, the method of case study is applied where we analyse the communication protocols and data transmission formats in section 2. In detail, firstly we select 26 typical urban sensor application systems and IoT platforms and make a survey of their application fields; then we deeply study the communication protocols and data transmission formats of these systems and platforms through a comprehensive analysis and test of the Application Programming Interface (API) services; finally, the results of the survey are summarized in Table 2. Additionally, we finish this survey through: 1) making much comparison between different related technologies, such as a comparative analysis of the power, frequency band and band width of the communication protocols in the data link layer in section 2.1.1, and a comparison between post-fusions and pre-fusions technologies in section 4.1.2; 2) some fieldwork in the smart city and IoT technology exhibitions; 3) participating some important academic conferences concerned in this area, such as the 2019 World Sensor Summit and the First Military IoT Innovation and Development Forum of China in 2019.

The remainder of the paper is structured as follows. The technologies used by multisource heterogeneous urban sensor access are analysed in section 2 from the aspects of communication protocols, data transmission formats, access standards and access technologies. Urban sensor data management technologies, including data cleaning, compression, storage, indexing and querying technologies, are discussed in section 3. Most of the technical challenges that are faced by these technologies and remain open are discussed in section 4. Finally, in the final section, we conclude the paper and discuss possible development trends with related technologies in the future.

2. Multisource heterogeneous urban sensor access technologies

The number and types of urban sensors have increased rapidly in recent years, with the data volume exponentially exploding. The variety of urban sensors in applications, structures, communication interfaces and data formats, etc. has led to the division of urban sensors into small different collections, thus posing an enormous challenge to the unified massive integration of massive sensor information resources. Therefore, how to realize the efficient multiple communication protocol conversion and adaption become the major challenge of urban sensor integration. To solve this problem, previous studies mainly focus on the manual conversion and adaption of communication protocols. However, with the increasingly urgent need for real-time urban perception, more efficient reliable solutions for urban sensor access are needed. As new efficient approaches to solving this problem, software development kits (SDKs) become popular in recent years.

2.1. Communication protocols and data transmission formats

2.1.1. Communication protocols

Communication protocols define standards that connect urban sensors or other perception devices with physical and network layers. As Fig. 1 shows, we list the most widely used communication protocols for urban sensor access, which are derived from 26 typical urban sensor application systems and IoT platforms shown in Table 2, and we regarded all the protocols as protocol stacks, i.e., protocol layers divided by functions, to obtain a five-tier protocol framework. We focus more on the data link layer and application layer in protocols, because. In Fig. 1, protocols corresponding to the data link layer are used to link devices in the device layer with the data link layer, and they can be divided into four types: application layer protocols, secure transmission layer protocols, network layer protocols and data link layer protocols. Protocols associated with the application layer are used to define communication

Table 2
Some mainstream urban sensor application systems and IoT platforms and their communication protocols and data transmission formats. These systems and platforms are mainly launched by developers from the U.S.A, China and Germany, from 2007 to 2017. Their communication protocols and transmission formats can be more easily compared from this table.

System or Platform Name	Affiliation	Launch Dates	Main Application Fields	Communication Protocols	Transmission Formats
Bosch IoT Suite[a]	Bosch	2008	Smart agriculture, home, industry and trade; Automobiles	LwM2M, OMA-DM, TR-069, Modebus, BLE, DECT ULE, EEBUS SPINE, BACnet, Home Connect, KNX, ONVIF, ZigBee, Z-Wave, APIs or self-defined SDK	JSON
Oracle IoT[b]	Oracle	2012	Remote asset maintenance, Industry 4.0, Smart manufacturing and logistics	MQTT, HTTP, APIs or self-defined SDK	JSON
ThingWorx IIoT[c]	PTC	2013	Aerospace and national defence, automotive, industrial machinery, smart manufacturing, oil and gas	MQTT, CoAP, APIs or self-defined SDK	Private data model and data formats
ArcGIS GeoEvent[d]	Esri	2013	Smart environment, transportation; public opinion monitoring	NMEA0183, HTTP, TCP/UDP, Web Socket	TXT, JSON, XML, RSS, Esri Feature
Watson IoT[e]	IBM	2014	Intelligent asset maintenance; Intelligent real estate and facility management	MQTT, TLS, HTTP, SMS, APIs or self-defined SDK	JSON
Predix Platform[f]	GE Digital	2014	Smart manufacturing, digital factories, oil and gas	BLE, BACnet, DECT ULE, EEBUS SPINE, Home Connect, KNX, Modbus, ONVIF, ZigBee, Z-Wave	JSON
Cisco IoT Cloud Connect[g]	Cisco	2014	Smart industry, cities, transportation, manufacturing, government departments	LTE, Lora, HTTP, MQTT,	JSON
AWS IoT[h]	Amazon	2015	Digital manufacturing	MQTT, HTTP, Web Socket, APIs or self-defined SDK	JSON
Azure IoT[i]	Microsoft	2015	Remote monitoring; Industrial IoT; Predictive maintenance; Equipment simulation	MQTT, Web Socket, AMQP, HTTP, APIs or self-defined SDK	JSON
Google Cloud IoT[j]	Google	2017	Smart retail, finance; Healthcare and life sciences, manufacturing, energy	MCU, Wi-Fi, HTTP, APIs or self-defined SDK	JSON, JWT
Alibaba Cloud IoT[k], Feifeng Platform[l]	Alibaba	2017	Edge computing; Smart manufacturing, cities, real estate, agriculture, parks	MQTT, HTTP, TCP, LoRa, NB-IoT, Wi-Fi, Ethernet, MCU, 2/3/4G, Modbus, ZigBee, KNX, Z-Wave, RS485, APIs or self-defined SDK	JSON
QQ IoT[m]	Tencent	2014	Smart homes, life	Lora, Wi-Fi, SOC, 2G, BLE, ZigBee, Z-Wave, APIs or self-defined SDK	DATAPOINT, JSON
WeChat IoT[n]	Tencent	2017	Smart cities, agriculture, buildings, environment and medical	Lora、Wi-Fi、2/3/4G、BLE、ZigBee, APIs or self-defined SDK	JSON、XML
IoT Connection Management Platform[o]	Huawei	2016	Smart homes, family, transportation, water, parking; Assisted driving	LwM2M, NB-IoT, Lora, 2/3/4/5G, Wi-Fi, ZigBee, MQTT, CoAP, Modbus, OPC-UA, TCP/UDP, HTTP, APIs or self-defined SDK	JSON, Binary stream
IoT Hub[p]	Baidu	2016	Smart marketing, customer service, offices, cities, finance, medical, manufacturing, water, energy	MQTT, CoAP, HTTP, Web Socket, GB-T32960, APIs or self-defined SDK	JSON
Jingdong IoT[q]	Jingdong	2015	Smart homes, health, security, driving and production	MQTT, Modbus, CoAP, Lora, BLE Mesh, Ethernet, OPC-UA, APIs or self-defined SDK	JSON
SuperMap iServer Streaming[r]	SuperMap	2012	Smart parks, retail, banks, catering, insurance, home appliances	Web Socket, HTTP, APIs or self-defined SDK	CSV, TXT, JSON, GeoJSON
Jizhi Cloud[s]	Gizwits	2007	Smart manufacturing, homes, industry and logistics; Electronic consumption	LwM2M, MCU, SOC, NB-IoT, Wi-Fi, Lora, BLE Mesh, APIs or self-defined SDK	JSON
FogCloud[t]	MXCHIP	2010	Smart communities, homes, hotels, pensions and agriculture	MQTT, HTTP, Micro USB, Wi-Fi, BLE, APIs or self-defined SDK	JSON
ONEnet[u]	China Mobile	2014	Smart cities, life, agriculture, health, campuses, finance and retail; Industrial IoT, Security laboratory	LwM2M, MQTT, HTTP, EDP, JT/T-808, NB-IoT, TCP, 2/3/4/5G, Wi-Fi, ZigBee, Lora, RS485, Modebus, APIs or self-defined SDK	JSON
CTWing[v]	China Telecom	2017	Smart water, gas, fire control, photovoltaic power, electric vehicles and animal husbandry, retail	NB-IoT, Modebus, MQTT, LwM2M, HTTP, TCP, JT/T-808, Wi-Fi, BLE, ZigBee, APIs or self-defined SDK	JSON
AbleCloud[w]	AbleCloud	2014	Smart homes, medical and health; Internet of Vehicles (IoV), Industrial IoT	MQTT, CoAP, TCP/UDP, HTTP, MCU, SOC, 2/3G, Wi-Fi, APIs or self-defined SDK	JSON, Binary stream
WISE-PaaS[x]	Wise2C	2015	Smart medical, logistics, factories, transportation energy, environment and manufacturing	MQTT, Modbus, CoAP, Lora, BLE Mesh, Ethernet, OPC-UA, APIs or self-defined SDK	SimpleJSON
Xiaomi IoT[y]	Xiaomi	2015	Smart homes, electrical appliances, wearable devices and travel	Wi-Fi, BLE, BLE Mesh, 2/3/4/5G, ZigBee, MCU, APIs or self-defined SDK	JSON
COSMOPlat[z]	Haier	2017	Smart life and home; Business administration, Industrial IoT, Industrial security	MQTT, HTTP, Web Socket, OPC-UA, Wi-Fi, BLE, ZigBee, HomeKit, APIs or self-defined SDK	JSON

[a] Bosch IoT Suite. https://www.bosch-iot-suite.com/, 2008.
[b] Oracle IoT. https://www.oracle.com/internet-of-things/, 2012.
[c] ThingWorx IIoT. https://www.ptc.com/cn/products/thingworx, 2013.
[d] ArcGIS GeoEvent. https://www.esri.com/en-us/arcgis/products/, 2013.
[e] Watson IoT. https://www.ibm.com/cn-zh/cloud/internet-of-things, 2014.
[f] Predix Platform. https://www.ge.com/digital/iiot-platform, 2014.
[g] Cisco IoT Cloud Connect. https://www.cisco.com/, 2014.
[h] AWS IoT. https://aws.amazon.com/cn/iot-core/, 2015.
[i] Azure IoT Suite. https://azure.microsoft.com/zh-cn/overview/iot/,2015.
[j] Google Cloud IoT. https://cloud.google.com/solutions/iot/, 2017.

[k] Alibaba Cloud IoT. https://open.iot.aliyun.com, 2017.
[l] Feifeng Platform. https://www.cloudwalk.com/, 2017.
[m] QQ IoT. https://iot.open.qq.com/, 2014.
[n] Wechat IoT https://iot.weixin.qq.com/, 2017.
[o] IoT Connection Management Platform. https://developer.huaweicloud.com/, 2016.
[p] IoT Hub. https://cloud.baidu.com/, 2016.
[q] Jingdong IoT. http://devsmart.jd.com/, 2015.
[r] SuperMap iServer Streaming Service. http://www.supermapol.com/realspace/, 2012.
[s] Jizhi Cloud. https://www.gizwits.com/, 2007.
[t] FogCloud. https://v2.fogcloud.io/, 2010.
[u] ONEnet. https://open.iot.10086.cn/, 2014.
[v] CTWing. https://www.ctwing.cn/, 2017.
[w] AbleCloud. http://www.ablecloud.com/, 2014.
[x] WISE-PaaS. https://www.advantech.com.cn/, 2015.
[y] Xiaomi IoT. https://iot.mi.com/, 2015.
[z] COSMOPlat. https://www.cosmoplat.com/, 2017.

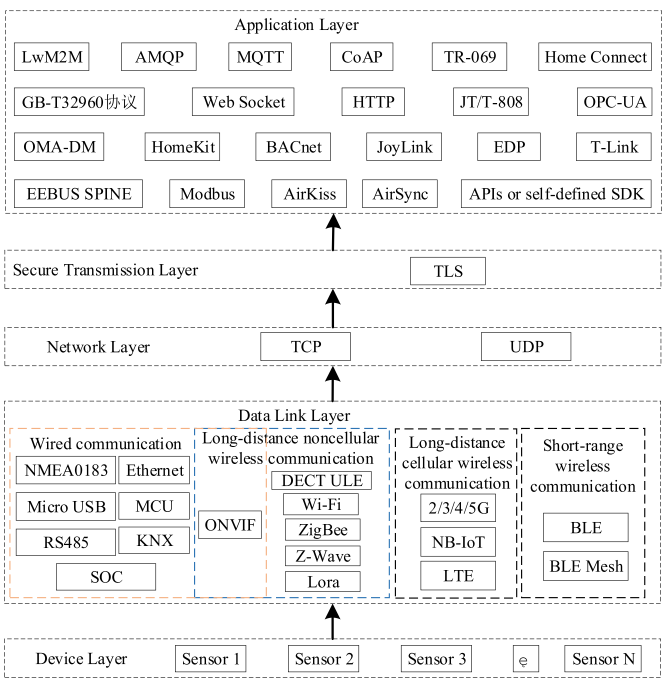

Fig. 1. Hierarchical architecture of communication protocols for urban sensor access, containing four types of protocols, i.e., application layer protocols, secure transmission layer protocols, network layer protocols and data link layer protocols.

standards that link urban sensors in the network layer with the application layer and are much closer to applications than are protocols in the data link layer. These protocols can, to some degree, reflect the IoT application or situation in various urban production and life activities. For example, the Home Connect protocol is often used in smart homes, and the JT/T-808 protocol is mainly used in communication involving road transportation vehicle satellite positioning systems.

For traditional urban sensor access, developers need to know the technological implementation details, such as those associated with data parsing, transmission protocols (e.g., TCP/UDP, Web Socket and SMTP), access process monitoring, and data analysis and data processing [41]. To address these problems, many urban sensor application systems and IoT platforms have achieved the integration of existing mainstream protocols and private protocols in recent years by encapsulating and implementing different kinds of SDKs for different kinds of protocols. An SDK consists of different types of APIs, and an API is essentially a real-time sensing data access function which realizes and encapsulates a certain kind of communication protocols. By calling these SDKs, developers don't need to know the underlying implementation details of the corresponding protocols any more. The APIs or SDKs of the communication protocols and data transmission formats can be divided into two types: those achieved by the integration or encapsulation of existing mainstream protocols or data formats and those achieved and owned by a single entity. We conducted a survey of the communication protocols and data transmission formats of 26 typical urban sensor application systems and IoT platforms, and the results are displayed in Table 2. When developers need to access urban sensors in certain communication protocols, they can look up the protocols in this table and select the most appropriate SDKs of the platforms.

In urban sensor networks, different kinds of sensors have huge differences in their power, frequency band and band width, which is determined by the communication protocols adopted and results in the diversity of sensor network deployment styles. Based on the surveyed results in Fig. 1 and combined with our latest research results, a comparative analysis of the power, frequency band and band width of the communication protocols in the data link layer, which is currently public and directly related to physical connection, is shown in Table 3 below.

In recent years, Fifth Generation (5G) technology has been applied in various fields of smart cities and has become increasingly mature. Compared with 4G technology, 5G technology mainly solves the problems of high-definition video transmission and high-speed transmission. However, at present it is difficult for 5G technology to support the deep integration and application of humans, machines and materials in information breadth, information speed and information depth. Much progress needs to be made in the current communication technology and its network architecture to support the business needs of future urban applications. Subsequently, the Beyond 5G (B5G) technology was proposed and quickly became a new research hotspot. As a more advanced technology, also as a new communication protocol, B5G will expand the communication space and improve the communication wisdom while continuing to accelerate the communication rate, to provide more reliable technical support for building ubiquitous and integrated information networks. Specifically, B5G uses a higher frequency band for its signal carrier, with a bandwidth of more than 1 Tbps. It further expands the communication space, from the current land coverage to multidomain and wide-area coverage in the ocean, sky, and into the space. Moreover, B5G can further strengthen and improve communication intelligence, which facilitates the evolution of intelligent information processing from the current single device to that from collaborative cross-domain multiple devices and multiple networks. In the technical architecture of B5G technology, there are at least three layers, i.e., the physical layer, the transmission layer, and the network layer [42]. Many technical problems need to be solved in the layers above, including the following aspects: 1) terahertz technology, and intelligent signal processing technologies based on deep learning in the physical layer; 2) unbalanced polarization transmission technology, polar code-based relay, multiple input and multiple output (MIMO) technologies in the transmission layer; and 3) full analytic B5G network architecture for the

Table 3
A comparative analysis of the power, frequency band and band width of the communication protocols in the data link layer, which is currently public and directly related to physical connection. The power, frequency band and band width of urban sensors can be more easily compared from this table by the protocols they adopt, and scientific decisions in deploying urban sensor networks can be more easily made for designers.

Protocol Types	Protocol Name	Corresponding Standards	Maker	Power/Frequency Band	Band Width
Wired Communication	Micro USB	USB Protocol	Compaq, Hewlett Packard, Intel, Lucent, Microsoft, NEC and Philips	USB2.0: ≤2.5 W USB3.0: ≤5 W USB3.1: ≤100 W USB3.2: ≤100 W	USB2.0: 480 Mbps USB3.0: 5 Gbps USB3.1: 10 Gbps USB3.2: 20 Gbps
	Ethernet	IEEE 802.3	DEC, Intel and Xerox	IEEE 802.3af: ≤12.95 W IEEE 802.3 at: ≤25.5 W	Ethernet: 10 Mbps, Fast Ethernet: 100 Mbps, Gigabit Ethernet: 1000 Mbps and 10G Ethernet: 10000 Mbps
	RS485	TIA/EIA-485-A	EIA and TIA	≤96 W	10 Mbps
Long-distance Non-cellular Wireless Communication	Wi-Fi	IEEE 802.11b	IEEE 802.1 Working Group	2.402 GHz–2.485 GHz	11 Mbps
	ZigBee	IEEE 802.15.4	ZigBee Alliance	868 MHz, 915 MHz and 2.4 GHz	20 kbps, 40 kbps, 250 kbps
	Z-Wave	Z-Wave, Z-Wave LR (2020)	Z-wave Alliance	900 MHz (ISM), 908.42 MHz(USA) and 868.42 MHz(Europe)	9.6 kbps, 40 kbps and 100 kbs
	Lora	LoRaWAN Specification V1.0	LoRa™ Alliance	EU433 MHz, CN470 MHz, EU868 MHz, AU915 MHz	0.3 kbps, 1.2 kbps, 2.4 kbps, 4.8 kbps, 9.6 kbps, 19.2 kbps
Long-distance Cellular Wireless Communication	2G	GSM and CDMA	3GPP	Up Link: 4 Frequency Bands including 880 MHz−890 MHz, Down Link: 4 Frequency Bands including 925 MHz−935 MHz	UL: 2.7 kbps, DL: 9.6 kbps
	3G	CDMA2000, WCDMA, TD-SCDMA, WiMAX	3GPP	Up Link: 6 Frequency Bands including 698 MHz–716 MHz, Down Link: 6 Frequency Bands including 870 MHz−880 MHz	UL: 384 kbps, DL: 3.6 Mbps
	4G	TDD and FDD	3GPP	Up Link: 42 Frequency Bands including 3600 MHz–3800 MHz, Down Link: 42 Frequency Bands including 734 MHz−746 MHz	UL: 50 Mbps, DL: 150 Mbps
	5G	3GPP Release.15 and 3GPP Release.16	3GPP	29 Frequency Bands in total, where 26 in 450 MHz-6 GHz,and 3 in 24250 MHz–52600 MHz	1 Gbps
	NB-IoT	3GPP Release.13 and 3GPP Release.14	3GPP	Up Link: 18 Frequency Bands including 703 MHz–748 MHz, Down Link: 18 Frequency Bands including 2110 MHz 2170 MHz	200 kbps
Short-range Wireless Communication	Blue Tooth	IEEE 802.15.1	SIG	2.4020 GHz–2.4835 GHz	721 kbps

integration of humans, machines and materials, mobile network architecture based on artificial intelligence, intelligently defined network architecture for cognitive enhancement and decision-making deduction in the network layer.

2.1.2. Data transmission formats

It can be concluded from Table 2 that most urban sensor application systems or IoT platforms have adopted JSON as their data transmission format. JavaScript Object Notation (JSON) is a text-based data storage and exchange format with a lightweight, understandable and independent programming language. In addition, SimpleJSON, GeoJSON, Extensible Markup Language (XML) and other formats have also been used to meet different transmission needs. SimpleJSON is a JSON-based format implemented in Python that is lighter than JSON for storage and provides faster loading and parsing speeds. GeoJSON is also JSON-based but has been mainly used for geospatial information exchange and feature-based data coding, such as for points, lines and polygons. JSON Web Token (JWT) is a user authentication encrypted format based on both JSON and Token. XML is a widely used data storage and transmission format that is recommended by W3C; with self-descriptiveness, its tags can be self-defined similar to those of JSON. Received signal strength (RSS) is an XML-based data sharing format that is often used by news and blog websites to distribute and aggregate web content. CSV and TXT are both text-based formats: CSV is a comma-separated format, and TXT is the most universal and cross-platform format. DATAPOINT is a unified data storage format for interdevice signals and statuses, and it is only used in the QQ IoT; in these applications, IDs and data organization methods are established for every type of sensor data. An Environmental Systems Research Institute (ESRI) feature is similar to a GeoJSON feature and is an abstraction of entities from the real world. A binary stream is also frequently used in communication, and it consists of 0 and 1 values that are easy for computers to process and transmit.

As discussed, due to the diversity of application scenarios, investment scales, business focuses, etc., different platforms, companies and organizations have taken advantage of different communication protocols and data transmission formats in urban sensor access, resulting in many “information islands”, which make it difficult to share data. To address this problem, SWE, IEEE 1451, Programmable Underwater Connector with Knowledge (PUCK), DICOM and other standards have been formulated by various institutions and organizations for accessing multisource heterogeneous city sensors, and many researchers have studied related technologies based on these standards or subsequent improved versions. These technologies, standards and their pros and cons will be summarized and discussed below.

2.2. Access standards

As discussed above, due to the diversity of application scenarios, investment scales, business focuses, etc., different platforms, companies and organizations have taken advantage of different communication protocols and data transmission formats in urban sensor access, resulting in many “information islands”, which make it difficult to share urban sensor resources on a larger scale.

To address this problem, access standards were formulated by various institutions and organizations for connecting multisource heterogeneous urban sensors, and these standards include the IEEE 1451 standards, the SWE standards, the PUCK standard and the DICOM standard. 1) The IEEE 1451 standards were firstly formulated to define the standards of different kinds of wired or wireless communication protocols to connect sensors in the device layer into the network layer.

Therefore, these standards focus more on defining and standardizing the physical connection protocols between the device layers and the network layers. 2) SWE standards define a uniform technical framework for large-scale protocol conversion, adaption and data transmission between the device layer, service layer and application layer. Compared with the IEEE 1451 standards, these standards define a logical architecture for urban sensor access, where all the sensors and sensor data are regarded as sensor resources; then the resources are modeled and published as different services for further applications. 3) With the significant growth of the underwater wireless sensor networks (UWSNs) and medical wireless sensor networks (MWSNs) in the world market in recent years, a diverse range of new application requirements have contributed to the pressing constraints on the underwater or medical sensor access. The Programmable Underwater Connector with Knowledge (PUCK) and DICOM standards are formulated to resolve these problems. Compared with the former two standards, these standards are more applied in specific fields like underwater or medical industries.

Based on the standards above, many access technologies have been proposed for multisource heterogeneous urban sensors access, and we define them as access technologies based on standards. To date, the most widely used access technologies based on standards are those based on the IEEE 1451 standards and the SWE standards, but they both have been proved to be error prone or inefficient. To solve this problem, access technologies based on customized adaptation layers are proposed as more reliable and efficient solutions.

2.2.1. The IEEE 1451 standard

The IEEE 1451 standard, also known as the Networked Smart Sensor Interface Standard, was jointly formulated by the National Institute of Standards and Technology (NIST) and the Sensing Technology Committee of the IEEE Instrumentation and Measurement Society (IMS) with the goal of developing a software and hardware connection scheme for smart sensors and control networks or to support other network technologies. IEEE 1451 is a protocol family based on the design of the electronic data sheet (TEDS) [43]; these sheets are used to establish software models for networked smart sensors and make the sensor modules compatible and plug-and-play ready. The Meta-TEDS in the IEEE 1451.2 protocol establishes a globally unique description identifier universal unique identifier description (UUID) for each urban sensor, which is significant for sensor modelling; the corresponding data type structure is illustrated in Table 4.

2.2.2. SWE standards

The Open Geospatial Consortium (OGC) has focused on developing open-interface standards and associated encoding standards and best practices that enable developers to create information systems that can easily exchange geospatial information and instructions with other information systems. As a subset of these standards, sensor web enablement (SWE) standards, which consist of information models and service interfaces, were designed to provide multisource and heterogeneous sensor information resources for the application layer and service layer. The SWE framework is relatively simple, as shown in Fig. 2. The information models are composed of a sensor model language (SensorML), SWE common data models, observations and measurements, and the service interfaces are composed of SWE service models, sensor observation services, sensor planning services and sensor components.

Table 4
Data structure of UUIDs in the Meta-TEDS of the IEEE 1451.2 protocols.

Number	Description	Digit Places
1	Location field	42
2	Manufacturer field	4
3	Date field	12
4	Time field	22

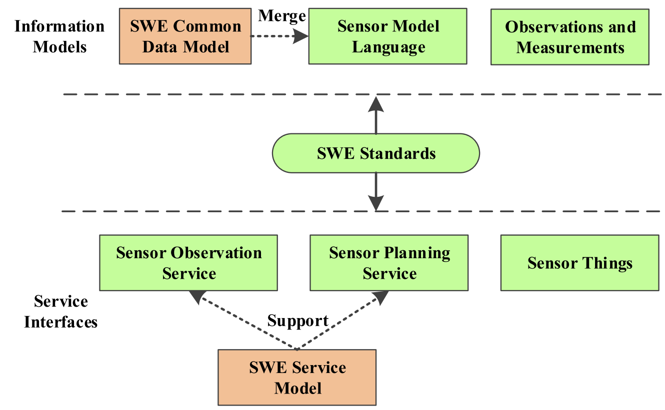

Fig. 2. The architecture of SWE standards, consisting of two parts, i.e. information models and service interfaces. Information models are proposed to build theoretical and abstract models for sensors or their observations and measurements; service interfaces are designed and developed to define standards for sensor observation, sensor planning, and sensor access into service layers.

2.2.3. The PUCK standard

The PUCK standard was derived from the PUCK device [44] designed and developed by the Monterey Bay Aquarium Research Institute (MBARI) in the United States; this device stores sensor metadata, drivers and other information and is built into or externally connected to sensors to bind the metadata and hardware. PUCK is primarily applied in marine perception, and the corresponding sensor access method was later developed into a standard and adopted by the OGC. The PUCK standard can be used by any sensor network consisting of RS232 or Ethernet-connected devices and mainly focuses on solving association and matching problems for software components and hardware devices in traditional sensor networks; thus, manual installation and configuration are required. This approach can be used to extract metadata or other information from SensorML files, TEDS tables and device drivers for the automatic identification, configuration and operation of sensors to provide standard interfaces and establish plug-and-play components.

2.2.4. The DICOM standard

The DICOM (Digital Imaging and Communication in Medicine) standard was first jointly formulated by the National Electrical Manufacturers Association (NEMA) and the American College of Radiology (ACR) with the goal of standardizing the transmission, storage, retrieval, printing, processing and display of medical digital images from different medical equipment so that image acquisition devices, image archiving and communication systems, printers and workstations from different manufacturers in the medical field can be integrated [45]. The most recent version of the DICOM standard contains 20 parts, providing various functions from data organization to object definition in the physical layer, from information exchange to network communication in the communication layer, and from network configuration in the service layer to program hosting in the application layer, thus forming a complete standard system.

2.2.5. Comparison and analysis

Based on the analysis above, a comparative analysis of the four kinds of sensor access standards is shown in Table 5 below. SWE standards are the most widely used standards for sensor network services; however, linking urban sensors in a physical layer into a network layer based on the SWE standards requires considerable adaptations from bottom sensor protocols to SWE protocols, and this approach is error prone and inefficient. In addition, the SWE standards do not describe where and how these adaptations are performed, resulting in an "interoperability gap" between the physical layer and the network layer. The TEDS of IEEE 1451, which cannot provide additional description information for virtual sensors, is primarily designed for physical layer sensor access; as

Table 5
A comparative analysis of the four most popular data access standards. The makers, applications and characteristics including advantages and disadvantages of these standards can be more easily compared from this table. This analysis can make it more efficient to select sensor access standards for urban sensor applications construction.

Sensor Data Access Standards	Makers	Applications	Characteristics	
			Advantages	Disadvantages
SWE Standards	OGC	Suitable for different scenarios of urban sensor access	The most widely used standards with highly uniformed technical architectures	Considerable adaptations required; interoperability gap between the physical layer and the network layer; error prone and inefficient
The IEEE 1451 Standard	NIST, IMS	Suitable for physical urban sensor access	Significant for sensor modelling with UUIDs; compatible and plug-and-play	Unsuitable for virtual sensor access
The PUCK Standard	MBARI	Suitable for wired urban sensor networks consisting of RS232 or Ethernet-connected devices	High reliability and correctness in connection	Much manual installation and configuration required
The DICOM Standard	NEMA, ACR	Suitable for medical urban sensor access	Providing a complete standard system of medical sensor access	Unable to connect sensors in non-medical fields

a result, it is unable to provide heterogeneous urban sensor access. The OGC PUCK standard provides a hardware solution, and the hardware structure of urban sensors needs to be changed to achieve adaptive access. Therefore, this standard is used in applications with few sensors and is not as popular as the previous two standards. The DICOM standard is mainly suitable for medical sensor access and medical image integration due to its object-oriented characteristics and openness, but it is unable to connect sensors in other fields.

2.3. Access technologies

As discussed above, due to the diversity of application scenarios, investment scales, business focuses, etc., different platforms, companies and organizations have taken advantage of different communication protocols and data transmission formats in urban sensor access, resulting in many "information islands", which make it difficult to share urban sensor resources on a larger scale.

To address this problem, access standards were formulated by various institutions and organizations for connecting multisource heterogeneous urban sensors, and these standards include the IEEE 1451 standards, the SWE standards, the PUCK standard and the DICOM standard. 1) The IEEE 1451 standards were firstly formulated to define the standards of different kinds of wired or wireless communication protocols to connect sensors in the device layer into the network layer. Therefore, these standards focus more on defining and standardizing the physical connection protocols between the device layers and the network layers. 2) SWE standards define a uniform technical framework for large-scale protocol conversion, adaption and data transmission between the device layer, service layer and application layer. Compared with the IEEE 1451 standards, these standards define a logical architecture for urban sensor access, where all the sensors and sensor data are regarded as sensor resources; then the resources are modeled and published as different services for further applications. 3) With the significant growth of the underwater wireless sensor networks (UWSNs) and medical wireless sensor networks (MWSNs) in the world market in recent years, a diverse range of new application requirements have contributed to the pressing constraints on the underwater or medical sensor access. The Programmable Underwater Connector with Knowledge (PUCK) and DICOM standards are formulated to resolve these problems. Compared with the former two standards, these standards are more applied in specific fields like underwater or medical industries.

Based on the standards above, many access technologies have been proposed for multisource heterogeneous urban sensors access, and we define them as access technologies based on standards. To date, the most widely used access technologies based on standards are those based on the IEEE 1451 standards and the SWE standards, but they both have been proved to be error prone or inefficient [62]. To solve this problem, access technologies based on customized adaptation layers are proposed as more reliable and efficient solutions.

2.3.1. Access technologies based on standards

(1) Access Technologies Based on The SWE Standards

The essence of technologies based on SWE standards is to connect urban sensors to the service layer or application layer via the implementation or extension of SWE standards, and the process can be simplified, as shown in Fig. 3. Sensors in the device layer are converted into objects defined in SensorML and later registered to services such as the sensor observation services (SOS). Then, all sensor information can be made available in the service layer, where users from the application layer can call the interfaces with various functions and sensor resources [46-48].

The geospatial sensor web (GSW), as a cyber-physical infrastructure for geoscience research applications, is a typical implementation of SWE standards; this model supersedes the previous models of geoscience research based on ground experiments or sensors and improves the cost and efficiency of geoscience perception and analysis [49–54]. Xie and Liu [55] designed a technical scheme to connect heterogeneous sensor observation data to the SOS and encompass the processes of sensor registration, access network selection, data reception and analysis, insertion document generation and data insertion. With regard to the extension of SWE standards, Broering et al. [56] further extended the SensorML approach and designed a sensor bus model to combine the physical layer with the service layer. In addition, a heterogeneous sensor

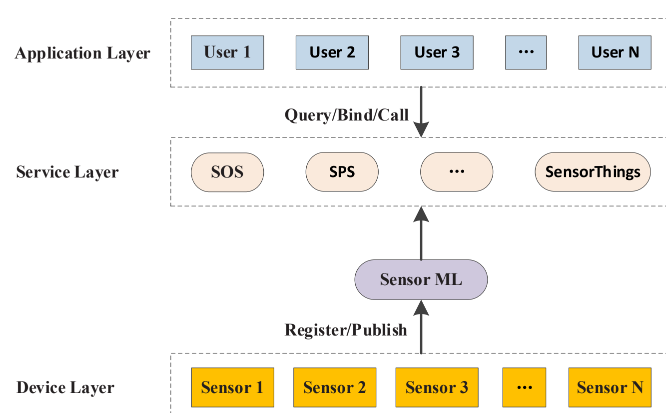

Fig. 3. Urban sensor access based on the SWE standards. Sensors in the device layer are converted into objects defined in SensorML and later registered to services such as the SOS, SPS and SensorThings. Then, all sensor information can be made available in the service layer, where users from the application layer can call the interfaces with various functions and sensor resources.

information model (HSIM) [57] was designed to achieve automatic analysis and dynamic adaptation by connecting heterogeneous sensing data to SWE standards. The IoT-plug-and-play (PNP) architecture [44] supports the automatic registration process from embedded devices to the SensorThings Application Programming Interface (API) service, thus enabling seamless resource integration for the device layer and the gateway layer.

(2) Access Technologies Based on Non-SWE Standards

Wang et al. [57] designed a sensor spreadsheet suitable for remote environmental monitoring; notably, a transmitter interface module, a wireless transmitter interface module and a network adapter were designed, and an environmental monitoring system was developed. Pinto et al. [58] designed an electric bicycle sensor network sharing architecture, that has good interoperability, connectivity, scalability and plug-and-play ready. The PUCK device realizes the binding of the sensor hardware and its metadata [59] The Evolved Packet Core (EPC) sensor network [60,61] was developed based on the EPC network infrastructure standards, which can effectively provide a standardized wireless sensor network (WSN)/radio-frequency identification (RFID) integration framework to support sensor data sharing.

2.3.2. Access technologies based on customized adaptation layers

To overcome the shortcomings of SWE standards, access technologies based on customized adaptation layers have been proposed to construct customized dynamic adapters for urban sensor access. As Fig. 4 shows, a dynamic adapter cluster layer is established for automatic protocol conversion and data modelling between the device layer and the service layer in Fig. 3.

Li et al. [62] proposed a custom adapter technology based on the Unified Sensor Data Interface (USDI) model to build a "virtual instrument layer" to communicate between the sensor layer and the service layer, with the objective of performing automatic sensor data structure analysis according to the relevant data transmission protocols to automatically achieve protocol conversion and dynamic adapter generation. The SentinAir System [63] can be used to evaluate the performance of any type of sensor; it provides configurability and adaptive sensing functions with high flexibility and availability and is compatible with multiple WSNs. With the goal of smart environmental gas monitoring, João et al. [64] designed and implemented a PnP adapter, which is integrated with IoT middleware to form an adaptation layer, thus enabling storage of multisource and heterogeneous context sensing data and allowing Android-based apps to warn users about their surroundings.

2.4. Data transmission technologies

Data transmission efficiency has become a major challenge for urban sensor networks. The crucial factors that affect data transmission efficiency in urban sensor networks are the statuses of sensor nodes and the network topologies that reflect the deployment mode of these nodes. When the normal operation of these nodes is ensured, the network topology that characterizes the way of a network's organization has become the main factor affecting the urban sensor data transmission efficiency.

To date, many researchers have performed extensive research on the optimization algorithms of network topologies. Traditional urban sensor network topologies can be divided into three types: tree-based topology, multi-path-based topology and hybrid topology combining the previous two topologies [65-69]. With the increasing demand for real-time information perception in cities, the traditional urban sensor network topology algorithms cannot meet the application requirements any more. On the other hand, inefficient data transmission usually means higher energy consumption.

To resolve these problems, urban sensor network topology and data transmission optimization algorithms based on intelligent algorithms have been proposed. The intelligent algorithms mainly include machine learning algorithms (e.g., deep learning and reinforcement learning algorithms) and other algorithms (e.g., the fog-based delay-sensitive algorithms, genetic algorithms, and ant colony algorithms).

2.4.1. Urban sensor network topologies based on traditional algorithms

In tree-based topologies, all nodes communicate with each other through a single path. This helps to minimize the transmission cost, but is susceptible to data loss and node failures: when a node failure occurs, the data of the corresponding subtree will be lost. Multipath-based topologies provide important support for a message propagating through multiple paths until the message reaches the base station. In such a case, when a message in one path is lost, it can be successfully transmitted through other paths. As a result, when compared with tree-based topologies that are sensitive to a single path, multipath-based topologies can effectively avoid data transmission failures. However, it is at the cost of higher communication cost and data redundancy. Hybrid approaches are based on the above two topologies, in which tree-based topologies are used to organize reliable nodes with stable communication links in an urban sensor network, while the other nodes are organized based on multipath approaches.

With the goal of balancing node energy in the transmission process of WSNs, Sun et al. proposed a multislot allocation WSN data transmission

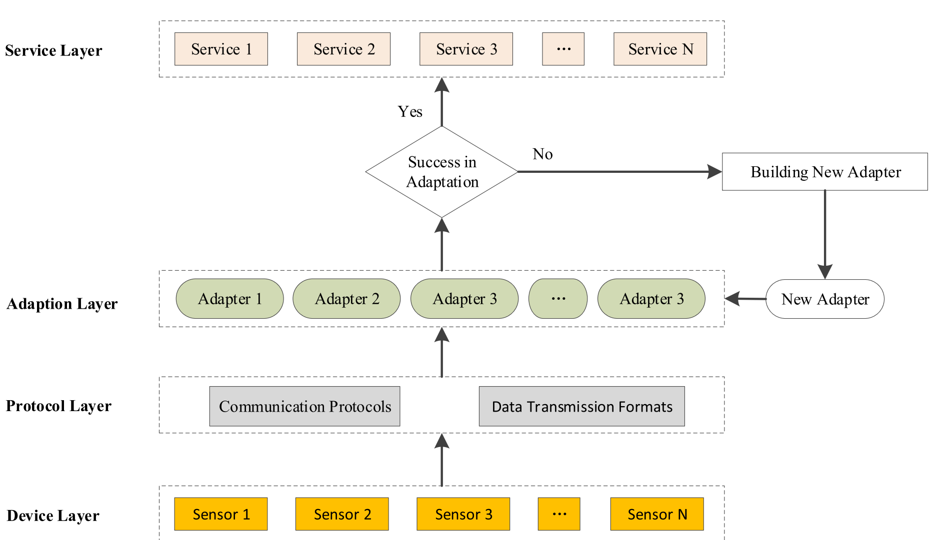

Fig. 4. Urban sensor access based on a customized adaptation layer. A dynamic adapter cluster layer, also named the adaptation layer, is established for automatic protocol conversion and data modelling between the device layer and the service layer. The adaptation layer can automatically check the communication protocols and data transmission formats between a sensor and all the adapters. If any of the adapters matches successfully, it proves to be effective. Otherwise, a new adapter will not be built until the adaptation succeeds.

algorithm based on dynamic tree-based topologies [70], and analysed the data transmission mode by establishing a tree link model with time slot requirements, where the relationships between parents and child are used in tree-based topologies to enable nodes to perform frame time slot allocation based on time slot requirements, giving a sequence pattern of receiving time slots and a sequence pattern of sending time slots, allowing nodes to be more flexible to receive data packets sent by other nodes in a channel with less interference, reduce the waste of time slots and improve channel utilization efficiency. Compared with the WSN life cycle extension algorithm based on data transmission optimization, and the reliable data transmission algorithm based on energy perception and time slot allocation, the energy delivery efficiency and the length of life of the network based on this algorithm are significantly improved.

2.4.2. Urban sensor network topologies based on intelligent algorithms

With the rise and continuous evolution of deep learning and reinforcement learning technologies in recent years, research on sensor network topology and data transmission optimization algorithms based on these technologies has become a hot topic [71–73]. Meng et al. proposed a self-organized WSN topology optimization algorithm based on deep reinforcement learning for self-organized energy-saving WSNs [71]. A unified search framework is defined, and deep neural networks are used to guide the Monte Carlo Tree Search (MCTS) to facilitate the simulation; the learning of the neural network is strengthened by the results of the tree-based search; the results of multiple simulation experiments show that the algorithm has better performance than the heuristic algorithm, and can adapt to the environment and network changes without looping execution of the algorithm from the beginning. In addition, the current related research also includes the optimization of sensor data transmission based on fog-based delay-sensitive algorithms, genetic algorithms, and ant colony algorithms [74–77], which have also improved the quality and effectiveness of sensor network data transmission to varying degrees.

3. Multisource heterogeneous urban sensor data management technologies

With the further application of location-based services (LBSs) and various types of urban sensors being used for space and ground observation, large volumes of heterogeneous and spatiotemporal sensor data have been generated [78–80]; such data are difficult for relational database management systems (RDBMSs) to manage. Facing this challenge requires improvements to relational database performance or the construction of spatiotemporal indices and query rules in nonrelational databases (NoSQL) or streaming databases. Moreover, because of the low battery, network outage, calibration error, mechanical failure and other external issues, the sensor data obtained through the access technologies above are often noisy, multidimensional and dynamic and need to be cleaned and compressed before storage.

3.1. Data cleaning technologies

Most of the applications in smart cities require abundant real-time or near real-time urban sensor data. However, incorrect sensor data values are often unavoidable, due to many factors such as network failures, discharged batteries or vandalism of low-cost sensors from humans or animals. To address these problems, data cleaning technologies are needed to detect and correct erroneous sensor data values.

As traditional solutions, the most common data cleaning technologies can be summarized as eight aspects, namely deleting multiple columns, changing data types, converting categorical variables to numeric variables, checking for missing data, deleting strings in columns, deleting spaces in columns, connecting two columns with strings (with conditions), and converting timestamps (from string format to datetime format). However, when dealing with large volumes of sensor data with multidimensional and spatiotemporal attributes, these solutions become less efficient.

To meet this challenge, model-based cleaning methods are proposed and used in recent years, and sensor data are cleaned based on certain mathematical models. The essence of the model-based cleaning method is to use an established model to estimate the most likely sensor value and then compare the difference between the estimated value and the actual measured value to detect and process abnormal values. At present, the most frequently used sensor data cleaning models are respectively regression models, probability models, and outlier detection models. The advantages of the model-based cleaning methods can be concluded as two aspects: 1) outliers of sensor data values can be located more quickly based on the mathematical models; 2) model-based data cleaning technologies can take advantage of the spatiotemporal correlations between two related data streams to infer the outliers of each other.

3.1.1. Sensor data cleaning based on regression models

Regression models are mainly used to quantitatively analyse and describe the correlation between two or more variables. The principle of cleaning sensor data based on a regression model is that sensor data essentially sample values observed by sensors in the real world. The same sampling process is often continuous, and different sampling processes may be correlated; thus, a functional model between two different observed variables (e.g., sensor values and time) can be built and used for inferring a sensor value at a certain time. There are currently two common regression models, namely, the polynomial regression model and Chebyshev regression model [81,82].

3.1.2. Sensor data cleaning based on probability models

Urban sensor data can also be cleaned based on probability models, and the corresponding process is shown in Fig. 5 [83]. At time $t_i = 6$, the depicted probability model uses the previous values $v_{2j}, \ldots, v_{5j}$ in the sliding window to infer a distribution (e.g., Gaussian distribution) that these values may obey; then, the expected value at time $t_i = 6$, such as the mean of the distribution $\bar{v}_{6j}$, is considered the inferred sensor value for sensor *j*. As shown in Fig. 5, the anomaly detector checks that the value $\bar{v}_{6j}$ is within a reasonable accuracy range, and if so, the sensor value $\bar{v}_{6j}$ is considered normal. By repeating the same process at time $t_i = 7$, the sensor value $\bar{v}_{7j}$ is deemed abnormal.

Ma'arif et al. used the Kalman filter to detect the noise or dirty values of sensors, and inferred the missing values of sensor data [84,85]; Verner used the long short-term memory (LSTM) model to detect and classify the possible dirty values in sensor data [86]; A classifier based on the Bayesian network model and the linear regression model was established to identify outliers in sensor data by Szewcyzk et al. [87,88]. To obtain good performance and high-quality services in medical IoT systems, a method to infer missing or wrong values in RFID data based on the genetic algorithm adaptive neuro-fuzzy inference system genetic algorithm (ANFIS-GA) and the particle swarm optimization algorithm ANFIS particle swarm optimization (ANFIS-PSO) was proposed by Turabieh et al. [89]. Since the inferred values of sensor data can be used for quantitative quality and accuracy analysis of raw sensor data, almost all mathematical models used to infer sensor values can be used for quality analysis and evaluation of sensor data [90–93].

3.1.3. Sensor data cleaning based on outlier detection models

Some researchers have proposed outlier detection models to clean sensor data because abnormal values are usually quite different from other normal values. Templ and Yoon et al. performed a comprehensive classification of data outlier detection methods [94,95]. Some outlier detection methods have been specifically proposed for sensor data. For example, Aya and Van et al. comprehensively reviewed the outlier detection technologies used in sensor network applications [96–98]. Sathe et al. studied the correlations among variables, extended Jaccard coefficients, and approximation methods based on regression models in

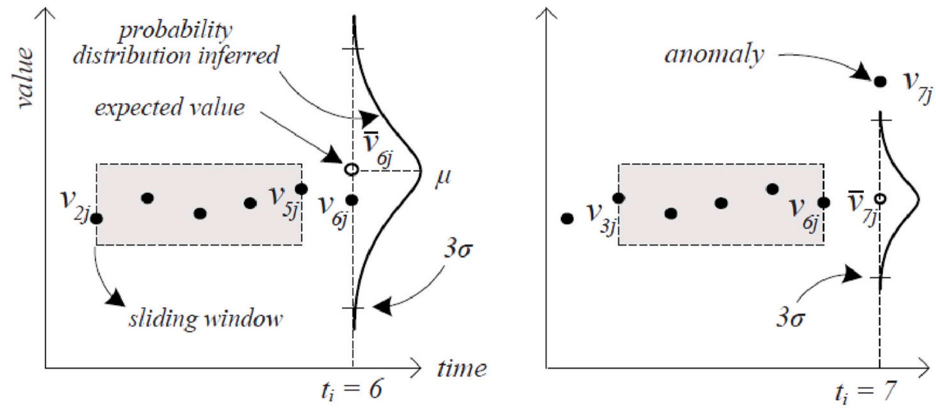

Fig. 5. The process of sensor data cleaning based on a probability model.

the process of sensor data cleaning [99]. In addition, methods based on histograms or nonparametric sequence models were applied to identify outliers in sensor data [100–102]. Abid developed a new density-based spatial clustering of applications with noise (DBSCAN) algorithm to detect outliers in real-time sensor data [103].

3.1.4. Comparison and analysis

Based on the surveyed results above, a comparative analysis of the three kinds of sensor data cleaning technologies is shown in Table 6 below. The three kinds of sensor data cleaning technologies are all essentially model-based methods, but the models are different. Regression model-based data cleaning technology mainly uses a polynomial regression model, such as the Chebyshev regression model, which establishes a function-based relationship among time, regression coefficients and a constant term to express a sensor value, similar to a mathematical approximation. Therefore, the model is usually built based on a series of historical data. Probability model-based data cleaning technology also relies on historical sensor data but focuses on finding a probability distribution that all the sensor data may obey to determine whether a value is normal or abnormal by checking whether it resides within a reasonable accuracy range. Outlier detection model-based data cleaning technologies require comparatively large amounts of data, and many models, including those above and others, such as histogram, distance-based, density-based and triangular mesh surface models, can be used.

Table 6
A comparative analysis of the four most popular data access standards. The makers, applications and characteristics including advantages and disadvantages of these standards can be more easily compared from this table. This analysis can make it more efficient to select sensor access standards for urban sensor applications construction.

Sensor Data Cleaning Technologies	Models	Principles	Application Scenarios
Regression model-based data cleaning technology	Regression Models, such as the polynomial regression model and Chebyshev regression model	Building a functional model (e.g., the Chebyshev regression model) between two different observed variables (e.g., sensor values and time) of the sample values as sensor data, by which a sensor value at a certain time can be inferred	The observed historical sequence of the sample values as sensor data can be simulated by a linear polynomial or high-degree polynomial
Probability model-based data cleaning technology	Probabilistic Models, such as the Kalman filter	Finding a probability distribution that all the historical sensor data may obey to determine whether a value is normal or abnormal by checking whether it resides within a reasonable accuracy range	The observed historical sequence of the sample values as sensor data tends to obey a certain probability distribution
Outlier detection model-based data cleaning technology	Outlier detection models, including the models above and others, such as histogram, distance-based, density-based and triangular mesh surface models	Cleaning sensor data based on the outlier detection methods, such as correlation, extended Jaccard coefficients and regression-based approximation	Outlier detection models are always helpful for sensor data cleaning, especially when the former technologies don't work

3.2. Data compression technologies

Many sensors, such as Global Position System (GPS) devices, RFIDs, thermometers, hygrometers, accelerometers and microphones, have been embedded in mobile or fixed devices such as cars, scanners, and phones in cities, thereby forming massive and dynamic sensor data networks with high processing, storage, querying and analysis expenses. Therefore, it is necessary to compress the sensor data to a reasonable volume while ensuring usability. Data compression efficiency is the eternal theme of data compression technologies.

To date, there have been many effective data compression algorithms, such as the quicklz algorithm and the snappy algorithm of Google. However, these algorithms are more suitable for non-spatiotemporal data; when dealing with spatiotemporal sensor data, the efficiency of these algorithms cannot be guaranteed because the sensor data with multidimensional and spatiotemporal attributes are often high-dimensional and more heterogeneous than non-spatiotemporal sensor data.

To address these problems, model-based compression methods are proposed as new reliable solutions in recent years, and sensor data are compressed based on efficient mathematical models. The most widely used models for sensor data compression mainly include data segmentation models, piecewise approximation models, and related data stream compression models. Compared with the previous data compression technologies, model-based data compression technologies can take advantage of the spatiotemporal correlations between two related data streams to compress the data within a certain error norm. In addition, orthogonal transformation models have been used as new efficient solutions for reducing the dimensionality of sensor data to reduce the amount of storage space.

3.2.1. Data segmentation models

Sensor data segmentation, which is based on segmentation approximation technology, involves segmenting a data stream containing a large amount of data. The current sensor data segmentation models can be divided into three types, namely, sliding window segmentation models, bottom-up segmentation models and top-down segmentation models [104–107]. The differences among the three models are associated with the order of segmentation processing. The sliding window model can be used for the online processing of sensor data, and it employs a look-ahead method. The other two methods are more efficient than the sliding window model but need to scan the entire data stream in advance, thus limiting online and real-time processing. The monoid tree aggregator (MTA) [108] seamlessly integrates the online processing capability of the sliding window model and bottom-up segmentation, thus providing a window aggregation and sliding strategy.

3.2.2. Piecewise approximation models

As the most commonly used data flow approximation approach, piecewise linear approximation [109] uses a piecewise linear function model to approximate each segment after sensor data are divided. In the simplest form of the model, all piecewise functions are constants, and this scenario is called piecewise constant approximation (PCA). PCA is widely used because it often satisfies the approximate accuracy requirements of data stream segmentation, and the constant value of each segment can be replaced with the starting value [110], average value [111] or median value of the segment [112], among which the median is most commonly used. In addition, a simplified form of the piecewise linear approximation model named the linear filter has also been widely used [113].

3.2.3. Related data stream compression models

Some studies have used the correlations between different data streams or data segments to compress sensor data. For example, Farias et al. proposed a related data flow model based on a random-restart steep climbing algorithm [114] to dynamically analyse the correlations between different sensor data segments; then, a polynomial time approximation algorithm was used to collaboratively compress these data streams within the maximum error range. The implementation of the model is divided into three stages. In the first stage, the similarities between the different data streams are identified in the fusion window. In the second stage, sets of two data streams are aggregated based on the similarity results. Finally, the data streams are compressed. Lin et al. proposed an adaptive linear vector quantization model (ALVQ) [115], which uses the spatial correlations between different data streams and reference signals to compress sensor data. A linear regression model can be applied to obtain the linear correlations between different data segments, which can effectively improve the compression accuracy and greatly reduce bandwidth consumption. Albuquerque proposed an adaptive fuzzy learning linear vector quantization model (AFLVQ) [116] that improved the ALVQ model, and sensing data from three-axis accelerometers and gyroscopes were used to perform experiments on the classification and compression of time series of human activity data. The classification accuracy and smoothness of learning convergence were much better than those of the ALVQ model.

3.2.4. Orthogonal transformation models

Orthogonal transformation models are often used for dimensionality reduction with time series data. The basic concept involves treating a sensor data stream of a given length as a finite sequence of real coefficients; this stream is then mapped to an N-dimensional Euclidean space via a transformation function, and a subset of orthogonal transform coefficients is selected as a feature set to form a feature space. Finally, some coefficients in the feature set are selected as approximate values in the original data stream. Two kinds of orthogonal transforms have been used for dimensionality reduction of sensor data, namely discrete Fourier transform and discrete wavelet transform [117–123].

3.3. Data storage technologies

Urban sensor data have been significantly multisource and heterogeneous in recent years. With the rapid emergence of many cheap and smart urban sensors, sensor data storage has become a major challenge for smart city applications. Urban sensor data are large in scale, diverse in type, and complex in structure, and most data have multidimensional and spatiotemporal attributes. Urban sensor data can be divided into structured data, semistructured data (e.g., JSON or XML), and unstructured data (e.g., audio or video) categories. According to the location where urban sensor data are stored, the current storage methods include local storage, centralized storage and distributed storage [124]. Local storage involves storing sensor data on a storage unit that is tied to the sensor. Centralized storage refers to the transmission of collected sensor data from each node in a network to the same data centre. Distributed storage refers to storing sensor data at each node or sink node in a network based on distributed technology. Among the three methods above, local storage has considerable limitations due to the storage capacity limitations and performance constraints of sensors, as well as other factors that limit data sharing. Therefore, centralized storage and distributed storage are the most widely used sensor data storage methods.

From the perspective of databases, current solutions to the sensor data storage problems and their pros and cons can be summarized as three aspects below: 1) Relational database storage. Relational databases such as Oracle, MySQL, and PostgreSQL, which store data in relational tables, are very mature and reliable. The tables are associated through foreign keys, so they are characterized by high consistency and availability. However, as the volume of urban sensor data has rapidly increased in recent years and many data types are now recorded, according to the CAP law (consistency, availability, and partition tolerance of a database) [125], relational databases have issues associated with horizontal expansion. 2) Nonrelational database storage. Nonrelational databases such as Cassandra, HBase and MongoDB, which provide high speeds and effective scalability, can address the problem by weakening the consistency or availability and enhancing the partitioning capability. Notably, network partitioning can effectively solve the above problems [126]. 3) Real-time database storage. Real-time databases such as Apache Strom, ThinkDB and PI have also become very popular and are mainly used to support the rapid writing, storage and querying of large amounts of measurement data in industrial scenarios and real-time feedback control. These approaches have notable advantages in storing time-efficient data [127], but some scalability problems remain.

To resolve these problems, Ding et al. proposed a database cluster system framework called IoT-ClusterDB [128] for managing massive sensor sampling datasets from the IoT to address the problems of both traditional distributed storage and cloud data management systems. The architecture of IoT-ClusterDB can be summarized as shown in Fig. 6. The sampling data used by IoT-ClusterDB are organized as “atomic monitoring objects”, and all data for the same object are organized in a temporal sequence. IoT-ClusterDB considers the massiveness, heterogeneity, spatiotemporal sensitivity and dynamics of sensor data, thus providing good data writing and query processing performance (see Fig. 7).

Relational databases usually have limitations in scalability, availability and parallel reading and writing, especially in big data management. In contrast, nonrelational databases have good performance in these features, but they do not support querying operations based on SQL well and the querying performance often depends on the patterns of their row key design [129–132]. As a result, DeCloud-RealBase [133], LaUD-MS [134], IOTMDB [135] and other storage frameworks [136] based on NoSQL have been designed and implemented to solve the problems above.

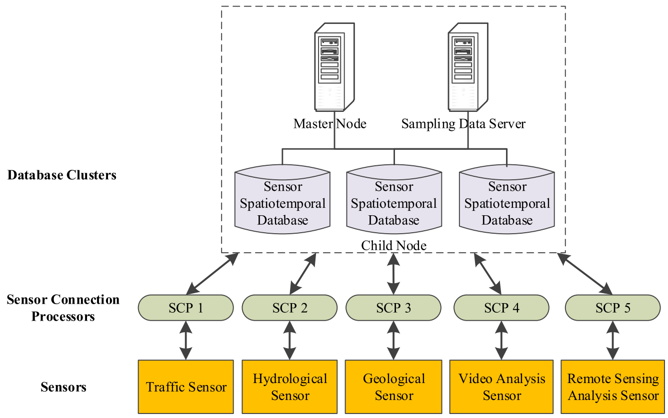

Fig. 6. The architecture of IoT-ClusterDB [128]. The sampling data are organized as "atomic monitoring objects", and all data for the same object are organized in a temporal sequence. IoT-ClusterDB considers the massiveness, heterogeneity, spatiotemporal sensitivity and dynamics of sensor data, thus providing good data writing and query processing performance.

3.4. Data indexing and querying technologies

3.4.1. Data indexing technologies

All data indices are constructed to increase the efficiency of data query responses, and the differences among indices are determined by the data or objects being indexed. With the rapid increases in the use of urban sensors and the availability of big data, an increasing contradiction has formed in smart city applications regarding the query efficiency and the complexity and diversification of sensor data. Research on sensor data indices has progressed in recent years [137–150], and indices can be classified into six types, as shown in Table 7.

As shown in Table 7, the efficiency of a sensor data query depends on the structure of the data index. The data indexing technologies listed in Table 4 are all suitable for urban sensor data queries. However, because urban sensor data may be diverse and uncertain, different indices are only suitable for specific kinds of sensor data, and the choice should be determined according to the actual application requirements. Taking the most common GPS trajectory data as an example, data queries can be performed based on a tree index [151], a hash index [152], or a combined hash and tree index such as B+ tree, 3DR tree or B* tree [153, 154]. Therefore, it is reasonable to construct an urban sensor data index in three steps. The first step involves selecting a suitable index method considering the characteristics of the queried data; the second step is to comprehensively evaluate the index based on the construction time and query efficiency; and the final step is to select the best data index.

3.4.2. Data querying technologies

Traditional data management is mainly for static data, and the corresponding data queries are generally predefined; the optimization goal of a query is to minimize the I/O and CPU costs of the system. Sensor data queries are dynamic and often required in real time, so the best query will provide high-efficiency and high-accuracy results [155–158]. The research progress related to sensor data query technologies can be summarized in the following four categories.

(1) In-Network Queries

In-network query technology has been developed by Cougar [159], TinyDB [160] and TiNA [161] for query processing involving sensor network data. These methods can be implemented by constructing an overlay network with a sensor based on a tree structure (e.g., a semantic routing tree(SRT)) and then performing queries. The overlay network is used to query the results of the root nodes gathered from the leaf nodes, thus improving the efficiency of sensor value aggregation and query processing.

(2) Distributed Queries

The distributed query technologies currently used in sensor networks are mainly based on the research results from the computer field, where the complexity of communication calculations, the implementation mechanisms of indices in various data flow models, achieving a load balance, query energy consumption and unstable or missing query data are common problems [162]. Balazinska proposed a global sensor network management framework [163], which is a typical distributed sensor data management and application system architecture. One of the key issues with this framework is performing efficient distributed indexing and querying, and it is necessary to unify the sensor-related syntax and semantics in various existing standards, such as SWE and IEEE 1451, to realize automatic query processing tasks. However, considering the mobility of sensors, nodes and users and the multidimensional characteristics of urban sensor data, it is currently difficult to achieve such an architecture. Other frameworks, such as IrisNet [164], share similar ideas with the framework above. Additionally, distributed hash tables have also been widely used in sensor data management in recent years [165].

(3) Centralized Queries

Centralized indices and queries are suitable for sensor data managed in a centralized manner. The current related research involves three types of centralized queries: symbol-based queries, semantic state-based queries and event-based queries.

A symbol-based query for sensor data was proposed as part of the

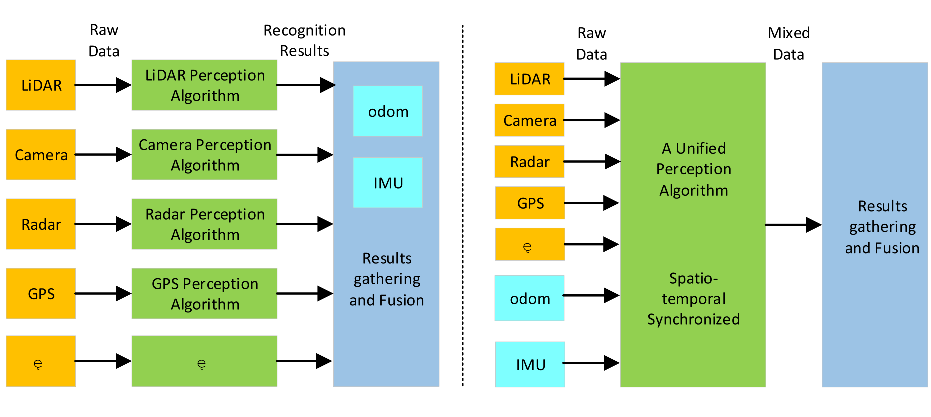

Fig. 7. A comparison between post-fusions and pre-fusions.

Table 7
Sensor data index classification.

Classification Standard	Index Type	Interpretation
Index Structure	Tree Index	Including the AVL Tree, B + Tree, B* Tree, R + Tree, R*Tree, T Tree, T*Tree and other tree indices such as I2B + tree for efficient time series data querying, IR-Tree for graph document querying, M-Tree for efficient big data querying and QDR-Tree for complex-space keyword querying
	Hash Index	Achieved by defining a hash function, mapping the keyword to a certain address set, and directly calculating the target address through the obtained function. This approach is efficient and widely used in accurate query functions, but it performs poorly in range queries and can cause data skew issues.
	Bitmap Index	Mapping the index column value to the bitmap array based on mapping functions, where each bit in the array indicates whether the keyword exists
	Grid Index	Dividing an object based on a grid, after which the subdivided grid is coded and the index is constructed. The grid can be regular or irregular and is suitable for both two-dimensional and three-dimensional space. A two-dimensional grid is usually a four-corner grid (rectangular) or a hexagonal grid, and a three-dimensional grid can be a Wulff net, an adaptive grid, or a regular polyhedral grid.
Query Requirements	Single Column Index	An index built based on a single attribute and suitable for single-attribute queries
	Composite Index	Building an index based on multiple attribute columns in sequence and suitable for multi-attribute queries
	Join-Index	Records the physical location of the connections in an index with a pre-connection operation, which is likely to occupy a large amount of memory and as much space as the data table itself
	Self-defined Index	An index built based on the calculation result of a function or formula, with logical conditions processed more than twice
Optimization Strategies	Inverted Index	Also known as the keyword-document reverse index, where an attribute value and the address of a record associated with the attribute value are saved and the record location is determined by the attribute value. This approach is commonly used in search engines, where each index item is called an inverted item.
	Partial Index	An index built based on part of the attribute columns and proposed by the Turing award winner M. Stonebraker, with low time and memory costs.
Skiplist Index		An index with a data structure is derived from a tree index, the Skiplist index is slightly inferior to the AVL tree in query performance but superior to tree indices in concurrency; additionally, the insertion and deletion performance is superior with this approach, such as for multi-core jump tables, writing optimization jump tables, jump graphs and jump networks. This index is currently used in multiple open-source databases such as Redis and LevelDB
Distributed Indices		Indices for distributed storage and querying, such as the non-primary key index CG-Index for cloud storage systems, the TI for efficient distributed queries of Weibo, the real-time index LSII for Weibo data, the Searchindex for the Alibaba Cloud and the HIndex for Huawei's HBase
Other Indices		Indices created according to specific application requirements, such as clustered, non-clustered, dense, sparse, and multi-level indices, among others.

FunctionDB project [166]. Specifically, the variables associated with the query conditions are regarded as independent variables, and the target variables of the query are regarded as dependent variables. First, the independent variables are filtered through a set of identifiable symbols in the database to obtain an approximate query interval, which is then interpolated within certain time intervals. The value of a dependent variable at each interpolation point is calculated, and a query is performed to return a dependent variable value range and the corresponding independent variable value range that meet the query conditions.

A semantic state-based sensor data query approach was proposed in the model-based index structure (MIST) framework [167]. The basic idea is that a semantic state implicitly present in the sensor data may comply with a user's query needs. For example, for the temperature of a room, the measured value is usually characterized by "temperature too low", "temperature low", "moderate temperature", "temperature high" and "temperature too high" rather than "<0 °C″, "0–20 °C″, "20–30 °C″, "30–40 °C″ and ">40 °C″, which is easier for users to understand. The query adopts the hidden Markov model to extract the semantic state from the sensor data, and each query returns this state to the user.

As an important type of query in sensor networks, event-based queries are used to query possible events in the continuous monitoring data sets obtained by sensors, and these events may include forest fires, traffic accidents or criminal behaviour [168]. In addition, event-based queries also support user registration, and the results can be used to monitor a single event or record a series of important events.

(4) Spatiotemporal Queries

Time and space are two crucial factors in the analysis and presentation of urban sensor data. The spatiotemporal characteristics of sensor data are often embodied as the spatiotemporal characteristics of the sensors and the observed sensor data. To achieve interoperability among sensor networks, middleware is often used for model conversion in distributed queries. However, the representation ability of middleware is limited, model conversion often requires additional computational costs, and long-term historical data cannot be queried. Therefore, centralized queries have become popular in recent years. Ding et al. proposed [128] a spatiotemporal database model for sensors, and it provided a distributed global index and global query functionalities by constructing a distributed global key B+ tree index, GFTKB + -tree, for each leaf node and a distributed global spatiotemporal R-tree index, GSTR-tree. This model queries only the key sampling data from the sensors to support queries based on spatiotemporal logic conditions, thus improving the query efficiency. In addition, a distributed composite spatiotemporal index named VegaIndexer was proposed to query large-scale spatiotemporal sensor datasets [169]. VegaIndexer is a cloud platform-based distributed spatiotemporal index that consists of a series of global and local indices. Later, a multiversion distributed enhanced R+ tree (MDR+) algorithm was proposed to accelerate data retrieval and improve the spatiotemporal query efficiency.

4. Discussion

In Sections 2 and 3, the general realization process and main technologies of multi-source heterogeneous urban sensor access and data management are analysed. With the acceleration of smart city and digital twin city construction, urban sensor application systems and IoT platforms are supposed to be more ubiquitous, intelligent, and contextualized. More challenges have been faced by these technologies, which can be summarized in three aspects: 1) the integration of massive IoT; 2) computational burden and energy consumption; and 3) cybersecurity. An analysis of these challenges, with feasible solutions and a comparison of related technologies, are presented in the following section.

4.1. The integration of massive IoT

The integration of massive IoT is a key factor in comprehensive and accurate information-assisted decision-making to realize the intelligence of smart city management. To achieve this goal, large-scale information from different groups, networks, and application systems is collected and integrated into a whole, which enables enterprises and people to better formulate strategies and back-trace historical information. The integration of massive IoT can be divided into two types: device integration and data integration.

4.1.1. Device integration

Device integration refers to the integration of devices in the same networks or among different networks. The two most critical issues are physical interface and protocol adaptation and cybersecurity in integration. 1) physical interface and protocol adaptation. Access standards and technologies can provide technical support for device integration when faced with multiple communication protocols introduced in Table 2 and Fig. 1. To date, there have been no unified standards or technical architectures for device integration, regardless of the physical layer and the network layer, and it can be predicted that this will continue to be a challenge in the future. Easy access to the Internet may give us some insights: by defining extensive and unified APIs for interoperability in the same networks or among different networks, we can easily avoid over-attention in physical interfaces and protocol adaptation. In addition, many platform-as-a-service (PaaS) platforms have been developed, as listed in Table 2, and large-scale device integration can be achieved via these solutions. 2) As the cybersecurity issue is also involved in data integration, it will be discussed in a uniform way in the end.

4.1.2. Data integration

Data integration is a more challenging issue than device integration in the integration of massive IoT. Although many efforts have been made in recent years to facilitate the integration and sharing of IoT data among different systems, many problems still need to be solved, mainly including 1) data complexity and redundancy; 2) uneven data quality; 3) inconsistent data formats; and 4) data security issues. The data cleaning technologies and data compression technologies surveyed above can be used to effectively reduce data complexity, remove redundant data, correct erroneous data and compress correlated data streams. However, these technologies have been mainly applied to deal with similar data or correlated data, while they are far from suitable for processing data, which has large differences in formats and structure. There have been many kinds of data formats in current urban sensor application systems, and the data transmission formats listed in Table 2 are only a subset of them. Data fusion technologies are a feasible solution to these problems at present and in the future. The data fusion technologies of urban sensor networks will be analysed in the following section, and the cybersecurity issue will be discussed in a uniform way in the end.

The data fusion technologies in urban sensor networks process and combine the data from different kinds of sensors or networks, to form a unified and consistent interpretation of the observed environment and obtain more information for prediction. Urban sensor data fusion technologies have been applied in many fields, such as military decisions, pattern recognition, remote sensing image fusion and medical image processing. The existing urban sensor data fusion technologies can be classified into five types by the methods adopted [170–172]: 1) technologies based on probability statistics methods, such as the weighted average method, Bayesian estimation and Kalman filtering; 2) technologies based on logical reasoning methods, such as DS evidence theory and fuzzy logic; 3) technologies based on neural networks; 4) technologies based on feature extraction, such as principal component analysis (PCA), linear discriminant analysis (LDA) and independent component analysis (ICA) methods; 5) technologies based on search algorithms, such as genetic algorithms and particle swarm optimization.

On the other hand, considering the fusion order, the existing urban sensor data fusion technologies almost belong to post-fusions. In a post-fusion, each sensor independently collects and processes the generated data, and then all these data are transported to the main processor, where the data fusion is finally accomplished. As a result, each sensor needs to execute a perception algorithm in post-fusion, which leads to much energy consumption and response delay in application systems [173]. In comparison, each sensor in a pre-fusion still independently collects data but does not execute perception algorithms any more. Instead, a unified perception algorithm is developed, in which all the collected data are fused in the primitive layer, and the fused data can be regarded as a supersensor. In this way, all the data streams will maintain spatiotemporal synchronization with each other. Compared with post-fusions, pre-fusions have at least two obvious advantages: 1) the energy consumption and response delay of application systems from pre-fusion will be much less than those of post-fusions; 2) much invalid and useless data filtered out in post-fusions can be integrated with other sensor data in pre-fusions, which helps to improve the perception comprehensiveness and the information utilization. As a result, pre-fusion technologies will be predictably popular in the future.

4.2. Computational burden and energy consumption

Computational burden and energy consumption are two other important issues in urban sensor networks. How to reach a balance between the two factors has become an important challenge for urban sensor network designers. Fundamentally, this challenge derives from the dialectical relations between computational burden and energy consumption: 1) the two factors jointly influence the overall cost of urban sensor networks; 2) the two factors are mutually related and restricted from each other. For example, as the computational burden mainly comes from the quantity of data to be transmitted over the network, the kind of processing to perform locally on the node and the complexity of data management technologies, the overall cost of urban sensor networks will increase with the increase in any of the three indices above.

The technical architecture formed by the technologies surveyed in this paper, from urban sensor access technologies to urban sensor data management technologies, is essentially a cloud computing architecture; thus, all the technologies will produce computational burden, while the energy consumption mainly comes from data access, data processing on the node and data transmission. The computational burden and energy consumption of these technologies are analysed below.

4.2.1. Computational burden and energy consumption in sensor data access

Urban sensor data access technologies involve two major categories, namely the technologies based on standards and those based on customized adaptation layers. In access technologies based on standards, most of the common sensors and their data can be connected into the network layer or the application layer via the collaborative use of the IEEE1451 standard and SWE standards, without effort to redefine the conversion rules for physical interfaces and transmission protocols or to develop related architectures. Of course, for new types of sensors that are not defined in these standards, manual adaptation and conversion are also required. In contrast, in access technologies based on customized adaptation layers, the building of the customized adaptation layers needs to define the conversion rules and finish the interface adaptation for all kinds of urban sensors. A good design of the adaptation layers can make it easier for new types of sensors to be connected. Therefore, in the sensor adaptation and access phase, the computational burden and energy consumption from the access technologies based on customized adaptation layers are greater than those from the access technologies based on standards. However, in the later expansion and maintenance phase, a good design of the adaptation layers can be more effective and cause less computational burden and energy consumption.

4.2.2. Computational burden and energy consumption in sensor data processing and transmission

When deploying urban sensor networks, there are two factors that mainly affect computing burden and energy consumption: the power supply mode and data processing and transmission.

1) The power supply mode. Most of the current urban sensor networks are wireless sensor networks, which use batteries as their power source, and energy constraints are still a major bottleneck. To extend the life of a network, minimizing the energy consumption is essential, especially when the network is deployed in harsh or inaccessible environments, such as in the scenarios of wildlife tracking or habit monitoring. To meet this challenge, wireless charging technologies have been invented and have rapidly become popular in recent years. However, wireless charging is more expensive than wired charging, and the time consumption of wireless charging is often greater. Therefore, it is not entirely applicable for urban sensor networks yet. In the future, more solutions for energy supply can be found, such as the energy of wind power, solar energy and tidal power.
2) Computational burden and energy consumption in data processing and transmission. Most components on a sensor node, such as the sensing unit or the CPU, always spend substantial power in an active state, including sending, receiving and listening, and bring less power consumption when sleeping into an inactive state. For example, the energy consumption of a radio in different statuses is compared in Table 8 [174]. Analogous to radios, sensors often spend more energy when carrying out tasks such as data collection, processing, and communication than when they are sleeping. As a result, except for necessary transmission tasks, most of the nodes keep listening even if there is no transmission task in the network, thus causing considerable unnecessary energy waste. In addition, the base station, also called the sink node, is usually responsible for data aggregation and analytic calculation, thus facing greater data storage and query pressure and spending more energy. Furthermore, the closer a node is to the sink node, the more data needs to be processed and transmitted on this node.

To meet the challenges above, early optimization algorithms improved the sleeping/waking mechanism by setting a fixed time window, according to which each sensor works regularly [175]. When there is no task, the sensor changes into the sleeping status. In this way, much energy can be saved. Nevertheless, it is often difficult for a sleeping node to obtain the working status of the neighbouring nodes in real time. In an extreme situation, an energy black hole is formed when a large number of nodes are asleep, and the network becomes paralyzed. Therefore, how to reduce energy consumption as much as possible while ensuring its availability becomes a new challenge for designers. Some improved routing algorithms were proposed to avoid too many nodes sleeping by setting an uncertain working time window. However, these algorithms rely too much on information sharing between nodes, and the energy consumption is often larger than that of algorithms based on fixed windows [176].

Table 8
Power consumption of a radio in different statuses. Radios often spend more energy when they are in an active status, such as data collection, processing, and communication, than when they are in an inactive status, such as sleeping.

Radio Status	Power
Sending	60 mW
Receiving	45 mW
Listening	45 mW
Sleeping	90 umW

The energy consumption of data transmission is also an important part of the total cost of urban sensor networks. Many routing algorithms have been designed to optimize the solutions. The purpose of the routing algorithm is to converge the data to the central node through the most time-saving and shortest path, while most of the data converge through the nodes with the most energy. The key point of this solution is the optimization of sensor network topologies.

Among the three data transmission technologies based on network topologies in section 2, 1) in tree-based topologies, all nodes communicate with each other through a single path, and each node only sends messages to its parent node, which helps to minimize the transmission cost. 2) In multipath-based topologies, a message is propagated through multiple paths to ensure successful data transmission, which is at the cost of higher communication cost and data redundancy. 3) Hybrid approaches are based on the above two topologies, in which tree-based topologies are used to organize reliable nodes with stable communication links in an urban sensor network, while the other nodes are organized based on multipath approaches. Therefore, from the perspective of energy consumption, tree-based topology algorithms have the lowest computational burden and energy consumption, and multipath-based topology algorithms have the highest computational burden and energy consumption, while the computational burden and energy consumption of the hybrid topology algorithms are in the middle.

4.2.3. Computational burden in sensor data management

The computational burden of the data cleaning technologies, data compression technologies and data storage technologies in section 3 can be measured by their computational complexity, where the more complex a technology is, the more computational burden it has. For example, considering the data cleaning technologies, when the historical observation data series of sensors can be simulated by linear polynomial or multivariate polynomial, the polynomial regression model is suitable for sensor data cleaning; and when the requirements for the similarity between the simulation results and the actual observation values are strict, the polynomial regression model with high degrees is more suitable, When the requirements are not that strict, the constant regression model or the linear regression model can ensure the effect of data cleaning, where the threshold value of the difference between the actual observation value and the simulated value needs to be selected according to specific application requirements. When the observed historical data tend to obey a certain probability distribution, the technologies based on probability models are more suitable for data cleaning, and the selection of confidence intervals needs to be based on the specific application requirements. When the former two technologies cannot meet the application requirements, technologies based on histograms, triangular mesh surfaces and other outlier detection models may provide a useful supplement. The above three data cleaning technologies are all desirable, and their selection mainly depends on specific application scenarios. The comparison of the computing burden between them can be transformed into the comparison of their computational complexity.

Another example is the comparison of the urban sensor data management technologies in section 3. At present, the two main data management technologies for urban sensor application systems or IoT platforms are centralized storage and distributed storage, respectively. In centralized storage, the data are stored on the local nodes of the sensor networks, but limited storage and computational resources make it impossible for this kind of network to serve long-term observations. In distributed storage, all related data can be stored together, but the disadvantage is that to balance the cost of all the, complex algorithms are required to determine where the data should be stored, which also causes considerable energy consumption.

4.2.4. In comparison with edge computing

As mentioned above, the technical architecture formed by the technologies surveyed in this paper is essentially a cloud computing architecture, where most of the data access tasks are finished in sensor networks and the data management services are finished on cloud servers. In contrast, as a technology that has become popular in recent years, edge computing finishes data processing and calculation on the object side. Suppose that there are urban sensor networks based on edge computing; almost all the data access and data management tasks will be accomplished in each sensor or on each node. It will be easier for edge computing to keep data flow localized and distributed to make the most important applications or key services closer to users. Compared with cloud computing and edge computing, the service response delays in edge computing will be reduced to a smaller degree by avoiding frequent cloud access. In addition, as most data processing and calculations in edge computing are performed locally on the sensor nodes, the network security and user privacy can be more easily protected. However, under the existing technical conditions, the power supply and storage and computational resources of each node are very limited, especially in wireless sensor networks, resulting in the fact that edge computing is not as popular as cloud computing at present. However, with the continuous innovation of the IoT and sensor technology, it is foreseeable that edge computing will be more widely used in all kinds of applications.

4.3. Cybersecurity

At present and in the future, cybersecurity is another important challenge faced by device integration, data integration and data fusion on urban sensor application systems and IoT platforms. Furthermore, cybersecurity is also a challenge for urban sensor data transmission. In this section, an analysis of cybersecurity technologies in urban sensor applications is presented below. Finally, we take RFID as an example to make a related discussion because cybersecurity has again been a great challenge in the case of sensor nodes and short-range transmissions in recent years.

4.3.1. Cybersecurity technologies in urban sensor applications

One of the most popular ideas of traditional solutions for cybersecurity in sensor networks is based on data partitioning. In detail, sensor data are first divided into data segments before being transmitted from one endpoint to another; then, the segments are assembled and recovered based on certain algorithms when they reach the endpoint. The shortcomings of these methods are obvious: pure data partitioning will bring more data redundancy and energy cost; more importantly, if disinformation from attackers is added into the data, the original information will be damaged and become unavailable.

To cope with these challenges, different kinds of data encryption technologies have been proposed and applied in sensor data transmission or data protection after being integrated. There have been many data encryption technologies thus far, such as homomorphic encryption, data perturbation, secure multi-party calculation, modular arithmetic, camouflage arithmetic, and complex number arithmetic. Both homomorphic encryption and secure multi-party calculation involve complex encryption and decryption operations, which will greatly increase energy consumption when used for sensor data transmission. Among the remaining technologies, data perturbation is the most widely used technology, with the idea of adding a perturbation sequence to the original data sequence, and the data will be recovered at the endpoint. However, simple data disturbance will increase the amount of data transmitted in the sensor network. In addition, spatiotemporal context information based on the correlations between different sensor data streams can be utilized by new means of cybersecurity privacy attacks, such as principal component analysis and spectrum filtering technologies. To solve this problem, data fusion technologies are used to add noise with a known distribution to the original data while simultaneously reducing the data volume during the transmission process [177], and these technologies are likely to be widely used in future cybersecurity protection of sensor applications.

4.3.2. Cybersecurity technologies in RFID applications

In addition, in some short-distance communication applications, cybersecurity is still a principal challenge, and we will take RFID technology as an example below. RFID has been widely used in many fields, such as the retail industry, automatic vehicle identification, logistics and other fields. However, the tags used by RFID technology can be easily tracked without user awareness. In this way, much private information of the users, such as their shopping habits and even the relationships with others, can be acquired. When contending with these challenges, the current common solutions include: 1) adding a locking and unlocking mechanism into the RFID tags, and users can lock the tags with a password when necessary; 2) using a blocking tag, so that when two RFID tags send different signals to the RFID reader at the same time, a broadcast conflict will occur and prevent the reader from deciphering any response and prevent the attacker from attacking other's privacy with fake RFID; 3) other privacy protection technologies, such as k-anonymity and t-closeness algorithms. As popular solutions for spatiotemporal data privacy protection, location privacy protection technologies have been proposed in recent years, such as the technology based on differential privacy strategies, which builds a multilevel location information tree model and use the Laplace scheme to add noises to accessing frequency of the selected data, and thus greatly improve the security, privacy, and applicability of sensor data [178]. As urban sensor data will be more dynamic, more heterogeneous and more spatiotemporally complicated in the future, location privacy protection technologies can provide powerful technical supports for future urban sensor applications.

5. Conclusions

As an important part of urban infrastructures, urban sensors have been widely used in many urban applications to dynamically sense urban activities in real time, thus facilitating the precise, scientific and intelligent management of smart cities and providing important physical support for the construction of digital twin cities. As a result, the development of multisource heterogeneous urban sensor access and data management technologies is significant for the realization of new smart cities and digital twin cities in the future. Researchers worldwide have performed extensive research on these technologies and have achieved notable results that provide a basis for follow-up research. Even though there have been a lot of achievements in urban sensor applications and also much literature review concerning the summary of these works, the technologies is are continuously upgraded and the review of them must be further pro-moted. Most of the existing surveys covered only some aspects of the technologies used in urban sensor applications, which still need to be updated based on the latest technological advances, and even some crucial aspects were not considered in them.

This paper has reviewed the current work in multisource heterogeneous urban sensor access and data management technologies. We have reviewed various relevant approaches, models, technologies, and architectures that have been applied in smart cities and other urban sensor applications in recent years. Most of the approaches, models, technologies, and architectures are still in the development stage. After making a more comprehensive survey of the latest relevant achievements by updating and complementing the current literature review, we provide an extensive study of the existing body-of-the-knowledge on the theory and technologies of urban sensor applications. For example, most of the existing surveys didn't cover data compression technologies as important technologies in urban sensor applications, and even data transmission formats and data cleaning technologies, as a crucial aspect of sensor access technologies, has not been considered. In contrast, we have made a consolidated review of all the necessary technologies and identified existing research issues and challenges. On one hand, this work can provide better insight for scholars for future research and the development of urban applications like IoT platforms; on the other hand, it also provides more reference for governors in the future informatization construction of smart cities or digital twin cities, so that all the public services, such as policy making and emergency rescue can be more efficient and reliable to us.

Additionally, we suggest the research directions of the technologies concerned in this paper as a technical guideline for future research, which are listed as follows:

- Research directions of multisource heterogeneous urban sensor access technologies:
- Research directions of urban sensor access standards. 1) The SWE standards are currently the most widely used among the urban sensor access standards, and they will likely be used in many applications in the future. In addition, other standards, such as the IEEE 1451 standards and the PUCK standard, are mainly designed for embedded device access from the physical layer to the service layer, thereby providing supplements for SWE standards. 2) Sensor access standards should be unified. Many application systems and IoT platforms have achieved their APIs or SDKs, but there is still much difference between the technologies. To connect and manage urban sensors in the worldwide sensor web in the future, the existing standards should be improved and applied in a more uniform way for the integration of these systems or platforms. 3) With the development of IoT and artificial intelligence (AI), more smart urban sensors will come into being. When accessing these smart sensors, automatic reverse control should be adequately considered to realize smart urban sensor networks. Therefore, more intelligent protocols and standards should be developed and formulated to provide such technical supports.
- Research directions of urban sensor access technologies. 1) The problem of large-scale protocol conversion and adaption is still a major challenge for urban sensor access technologies, and this problem should be addressed through the construction of adaptation layers or the continuous development of open interfaces with the service layer in a unified manner. 2) Moreover, the integration of nonstandard access technologies and their incorporation into a unified sensor access technology standard system is a feasible solution. 3) With the increasing number of smart urban sensors in the future, there will be more urgent need for real-time interaction among humans, devices, and environments, and edge computing technologies are likely to be utilized. As a result, automatic sensor access and control will be more considered in the current technologies.
- Research directions of urban sensor data transmission technologies. 1) As mature solutions of urban sensor network topologies, hybrid approaches use tree-based topologies to organize reliable nodes with stable communication links in an urban sensor network, and the other nodes are organized based on multipath approaches. As a result, it reaches a good balance between efficiency, reliability and energy consumption and thus will likely be continuously further applied in urban sensor network deployments. 2) With the development of intelligent algorithms, especially the machine learning algorithms, urban sensor data transmission will be more efficient owing to the continuous optimization of urban sensor network topologies.
- Research directions of multisource heterogeneous urban sensor data management technologies:
- Research directions of urban sensor data cleaning and compression technologies. 1) Data quality is a constant issue in sensor data management. The current data cleaning and compression technologies for urban sensors are mostly model based. Regression models, probability models and outlier detection models are often applied in sensor data cleaning, and researchers may seek to improve the modelling efficiency or and effectiveness of data cleaning in the future. 2) Data segmentation, piecewise ap-proximation, data stream compression and orthogonal transformation models are currently used in sensor data compression, and future research may focus on comparing the effectiveness of existing models and developing new mathematical models. For example, the time series point data, such as GPS trajectory data, in some sensor data sets are a type of vector data, so some models in geosciences, such as the Douglas-Puck algorithm, can be introduced for data compression.
- Research directions of urban sensor data storage technologies. 1) For urban sensor data storage, a unified solution has not been identified, but certain principles need to be followed. For example, semi-structured and unstructured sensor data are often stored in non-relational databases, and structured data are commonly stored in relational databases. 2) Furthermore, graph databases have become popular in recent years and support efficient relational data storage and operations. Therefore, the storage of spatiotemporal relational data from urban sensors based on graph databases is a possible research hotspot in the future.
- Research directions of urban sensor data query technologies. 1) The current data query technologies used by urban sensors are mainly based on big data indices, especially tree indices and hash indices, and the choice of index should be determined according to the actual requirements of urban sensor applications. Because the current cloud-based data management systems do not effectively support spatiotemporal logic queries and constrained attribute queries, the exploration of cloud-based data management systems that support spatiotemporal data queries involving urban sensors may continue. Moreover, the complex query technologies used in large-scale distributed sensor data management systems remain problematic and need to be further addressed. 2) Hybrid sensor data indices and the indices based on space-filling curves will continue to be widely used because of their high efficiency, and future research should focus more on the performance and optimization of the indices in distributed urban sensor data management systems and real-time urban sensor applications.

Author Contributions

Conceptualization, Fei Yang; methodology, Fei Yang, Yixin Hua, Xiang Li, Zhenkai Yang, Xinkai Yu and Teng Fei; software, Fei Yang, Xiang Li and Teng Fei; validation, Zhenkai Yang and Xinkai Yu; investigation, Fei Yang and Teng Fei; writing—original draft preparation, Fei Yang, Yixin Hua, Xiang Li, Zhenkai Yang, Xinkai Yu and Teng Fei; writing—review and editing, Fei Yang, Zhenkai Yang and Xinkai Yu; supervision, Fei Yang; funding acquisition, Fei Yang and Yixin Hua. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by the project Pan-Spatial Information System and Intelligent Facility Management which is financially supported by the National Key Research and Development Project, grant number 2016YFB0502300.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] R. Frank, Understanding Smart Sensors, Artech House, Boston, UK, 2013, pp. 1–14.

[2] S. Zhong, Thinking and Development of Intelligent Society, Aerospace Science and Industry Corp., Beijing, China, Tech. Rep., Dec. 2019.

[3] B.R. Ruiz, O.M. Muñoz, Using multivariate outliers from smartphone sensor data to detect physical barriers while walking in urban areas, Technologies 8 (4) (Aug. 2020) 51–58, https://doi.org/10.3390/technologies8040058.

[4] F. Gao, M.I. Ali, E. Curry, et al., Automated discovery and integration of semantic urban data streams: the ACEIS middleware, Future Generat. Comput. Syst. 76 (Jan. 2017) 561–581, https://doi.org/10.1016/j.future.2017.03.002.

[5] J.G.P. Rodrigues, A. Aguiar, J. Barros, Sense My City: Crowdsourcing an Urban Sensor, Computer Science, 2014 arXiv preprint arXiv:1412.2070.

[6] E. Morris, X. Liu, A. Manwar, et al., Application of distributed urban sensor networks for actionable air quality data, ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences 6 (Oct. 2020) 119–126, https://doi.org/10.5194/isprs-annals-VI-4-W2-2020-119-2020.

[7] O. Lock, T. Bednarz, C. Pettit, "HoloCity–exploring the Use of Augmented Reality Cityscapes for Collaborative Understanding of High-Volume Urban Sensor Data," in the 17th International Conference on Virtual-Reality Continuum and its Applications in Industry, VRCAI, Brisbane, Australia, Nov. 2019, pp. 1–2, https://doi.org/10.1145/3359997.3365734.

[8] S. Wang, M.J. Paul, M. Dredze, Social media as a sensor of air quality and public response in China, J. Med. Internet Res. 17 (3) (Mar. 2015) 125–138, https://doi.org/10.2196/jmir.3875.

[9] H.W. Chun, Y.L. Tsai, Social sensor: an analysis tool for social media, Int. J. Electron. Commer. Stud. 7 (1) (Jul. 2016) 77–94, https://doi.org/10.7903/ijecs.1411.

[10] A. Suprem, C. Pu, Assed: A framework for identifying physical events through adaptive social sensor data filtering, in: Proceedings of the 13th ACM International Conference on Distributed and Event-Based Systems (DEBS), Jun. 2019, pp. 115–126, https://doi.org/10.1145/3328905.3329510. Darmstadt, Germany.

[11] P.K. Samatha, M. Rafi, Expelling information of events from critical public space using social sensor big data, International Journal of Trend in Scientific Research and Development 3 (5) (Aug. 2019) 445–448.

[12] M. Weiser, The computer for the 21st century, Nature 265 (3) (Sep. 1991) 3–11, https://doi.org/10.1145/329124.329126.

[13] M. Weiser, Hot topics-ubiquitous computing, Computer 26 (10) (Oct. 1993) 71–72, https://doi.org/10.1109/2.237456.

[14] Z.W. Yu, Z.Y. Yu, X.S. Zhou, Socially aware computing, Chin. J. Comput. 35 (1) (Jan. 2012) 16–26, https://doi.org/10.3724/SP.J.1016.2012.00016.

[15] Y. Liu, K. Wang, H. Guo, et al., Social-aware computing based congestion control in delay tolerant networks, Mobile Network. Appl. 22 (2) (Apr. 2017) 174–185, https://doi.org/10.1007/s11036-016-0759-8.

[16] H. Flores, R. Sharma, D. Ferreira, et al., Social-aware device-to-device communication: a contribution for edge and fog computing?, in: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UBICOMP), Sep. 2016, pp. 1466–1471, https://doi.org/10.1145/2968219.2968589. Heidelberg, Germany.

[17] D. Silva, R.C. Sofia, "A Discussion on Context-Awareness to Better Support the IoT Cloud/edge Continuum," IEEE Access, Oct. 2020 arXiv preprint arXiv: 2010.04563.

[18] Y. Zheng, L. Capra, O. Wolfson, et al., Urban computing: concepts, methodologies, and applications, ACM Transactions on Intelligent Systems & Technology 5 (3) (Sep. 2014) 1–55, https://doi.org/10.1145/2629592.

[19] Y. Zheng, Urban Computing, MIT Press, Cambridge, UK, 2019, pp. 3–41.

[20] S.I. Candanedo, R.S. Alonso, Ó. García, et al., Edge computing, IoT and social computing in smart energy scenarios, Sensors 19 (15) (Jun. 2019) 3353, https://doi.org/10.3390/s19153353.

[21] M. Satyanarayanan, The emergence of edge computing, Computer. 50 (1) (Jan. 2017) 30–39, https://doi.org/10.1109/MC.2017.9.

[22] W. Yu, F. Liang, X. He, et al., A survey on the edge computing for the Internet of Things, IEEE Access 29 (6) (Nov. 2017) 6900–6919, https://doi.org/10.1109/ACCESS.2017.2778504.

[23] J. Wu, M. Dong, K. Ota, et al., FCSS: fog computing based content-aware filtering for security services in information centric social network, IEEE Transactions on Emerging Topics in Computing 7 (4) (Sep. 2017) 553–564, https://doi.org/10.1109/TETC.2017.2747158.

[24] L. Liu, Z. Chang, X. Guo, Socially aware dynamic computation offloading scheme for fog computing system with energy harvesting devices, IEEE Internet of Things Journal 51 (3) (Mar. 2018) 1869–1879, https://doi.org/10.1109/JIOT.2018.2816682.

[25] M. Farsi, A. Daneshkhah, F.A. Hosseinian, et al., Digital Twin Technologies and Smart Cities, Springer, Berlin, Germany, 2020, pp. 1–217.

[26] A. Francisco, N. Mohammadi, J.E. Taylor, Smart city digital twin–enabled energy management: toward real-time urban building energy benchmarking, J. Manag. Eng. 36 (2) (Mar. 2020) 1–11, https://doi.org/10.1061/(ASCE)ME.1943-5479.0000741.

[27] C. Wythe, N. Fisher, S. Bobrov, et al., Digital Twins to Computer Vision: A Rapid Path to Augmented Reality Object Detection on the Battlefield, Presented at IT2EC 2020 Annual Meeting. [Online]. Available: https://www.itec.co.uk/_media/libraries/draft-abstracts–slides/65-Chuck-Wythe.pdf.

[28] L. Feng, L.Y. Fang, S. Yong, City brain, a new architecture of smart city based on the internet brain, in: IEEE 22nd International Conference on Computer Supported Cooperative Work in Design, CSCWD, Nanjing, China, Sep. 2018, pp. 624–629, https://doi.org/10.1109/CSCWD.2018.8465164.

[29] L. Xia, Improved heuristic data management and protection algorithm for Digital China cultural datasets, ACM Transactions on Asian and Low-Resource Language Information Processing 32 (5) (2020) 13–22.

[30] J. Severinsen, M. Roiste, F. Reitsma, et al., VGTrust: measuring trust for volunteered geographic information, Int. J. Geogr. Inf. Sci. 33 (8) (Feb. 2019) 1683–1701, https://doi.org/10.1080/13658816.2019.1572893.

[31] A. Kayes, R. Kalaria, I.H. Sarker, et al., A survey of context-aware access control mechanisms for cloud and fog networks: taxonomy and open research issues, Sensors 20 (9) (Apr. 2020) 2464, https://doi.org/10.3390/s20092464.

[32] F. Afroz, R. Braun, Energy-efficient MAC protocols for wireless sensor networks: a survey, Int. J. Sens. Netw. 32 (3) (Feb. 2020) 150–173, https://doi.org/10.1504/IJSNET.2020.105563.

[33] I. Tomić, J.A. McCann, A survey of potential security issues in existing wireless sensor network protocols, IEEE Internet of Things Journal 4 (6) (Dec. 2017) 1910–1923, https://doi.org/10.1109/JIOT.2017.2749883.

[34] P.D. Nguyen, L. Kim, Sensor system: a survey of sensor type, ad Hoc network topology and energy harvesting techniques, Electronics 10 (2) (Jan. 2021) 219, https://doi.org/10.3390/electronics10020219.

[35] J.Y. Gong, X. Zhang, L.G. Xiang, N.C. Chen, Progress and applications for integrated sensing and intelligent decision in smart city, Acta Geod. Cartogr. Sinica 48 (12) (Nov. 2019) 1482–1497, https://doi.org/10.1080/13658816.2019.1572893.

[36] S. Fattah, A. Gani, I. Ahmedy, M.Y.I. Idris, I.A. Targio Hashem, A survey on underwater wireless sensor networks: requirements, taxonomy, recent advances, and open research challenges, Sensors 20 (18) (Sep. 2020) 5393, https://doi.org/10.3390/s20185393.

[37] H.M. Chen, H.L. Shi, M. Li, L. Cui, Service middleware for internet of things: challenges and approaches, Chin. J. Comput. 40 (8) (Aug. 2017) 1725–1749, https://doi.org/10.11897/SP.J.1016.2017.01725.

[38] B. Diène, J.J. Rodrigues, O. Diallo, E.H.M. Ndoye, V.V. Korotaev, Data management techniques for internet of things, Mech. Syst. Signal Process. 138 (4) (Apr. 2020) 106564, https://doi.org/10.1016/j.ymssp.

[39] A. Karim, A. Siddiqa, Z. Safdar, et al., Big data management in participatory sensing: issues, trends and future directions, Future Generat. Comput. Syst. 107 (6) (Jun. 2020) 1683–1701, https://doi.org/10.1016/j.future.2017.10.007.

[40] S.M. Chowdhury, A. Hossain, Different energy saving schemes in wireless sensor networks: a survey, Wireless Pers. Commun. 56 (3) (Mar. 2020) 114, https://doi.org/10.1080/13658816.2019.1572893.

[41] F. Ding, IoT Open Platforms: Platform Architecture, Key Technologies and Typical Applications, Publication House of Electronics Industry, Bei Jing, 2018, pp. 19–60.

[42] P. Zhang, B5G: ubiquitous fusion information networks, ZTE technology journal 25 (1) (Feb. 2019) 55–62, https://doi.org/10.12142/ZTETJ.201901009.

[43] E.Y. Song, M. Burns, A. Pandey, et al., IEEE 1451 smart sensor digital twin federation for IoT/CPS research, in: 2019 IEEE Sensors Applications Symposium (SAS). Porto, Portuga, May. 2019, pp. 1–6, https://doi.org/10.1109/SAS.2019.8706111.

[44] C.Y. Huang, H.H. Chen, An automatic embedded device registration procedure based on the OGC Sensor Things API, Sensors 19 (3) (Jan. 2019) 495, https://doi.org/10.3390/s19030495.

[45] Z. Wang, Q. Li, Y. Wang, et al., Medical protocol security: DICOM vulnerability mining based on fuzzing technology, in: Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security (CCS), the United Kingdom, Nov, London, 2019, pp. 2549–2551, https://doi.org/10.1145/3319535.3363253.

[46] B. Ventura, A. Vianello, D. Frisinghelli, et al., A methodology for heterogeneous sensor data organization and near real-time data sharing by adopting OGC SWE Standards, ISPRS Int. J. Geo-Inf. 8 (4) (Apr. 2019) 167, https://doi.org/10.3390/ijgi8040167.

[47] D.S.H. Van, J. Moßgraber, S. Grellet, et al., An environmental sensor data suite using the OGC Sensor Things API, in: International Symposium on Environmental Software Systems (ISESS), Jan. 2020, pp. 228–241, https://doi.org/10.1007/978-3-030-39815-6_22. Tokyo, Japan.

[48] C.C. Chung, C.Y. Huang, C.R. Guan, et al., Applying OGC sensor web enablement standards to develop a TDR multi-functional measurement model, Sensors 19 (9) (Sep. 2019) 4070–4078, https://doi.org/10.3390/s19194070.

[49] N. Chen, X. Wang, J. Gong, Design and implementation of geospatial sensor web information public service platform, Journal of Geo-Information Science 15 (6) (Jan. 2013) 887, https://doi.org/10.3724/SP.J.1047.2013.00887.

[50] J. Li, N. Chen, Geospatial sensor web resource management system for smart city: design and implementation, in: 14th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, May. 2014, pp. 819–827, https://doi.org/10.1109/CCGrid.2014.70. Chicago, IL, USA.

[51] X. Zhang, N. Chen, Z. Chen, et al., Geospatial sensor web: a cyber-physical infrastructure for geoscience research and application, Earth Sci. Rev. 185 (Oct. 2018) 684–703, https://doi.org/10.1016/j.earscirev.2018.07.006.

[52] X. Meng, F. Wang, Y. Xie, et al., An ontology-driven approach for integrating intelligence to manage human and ecological health risks in the geospatial sensor web, Sensors 8 (11) (Oct. 2018) 3619, https://doi.org/10.3390/s18113619.

[53] C. Xiao, N. Chen, X. Wang, et al., A semantic registry method using sensor metadata ontology to manage heterogeneous sensor information in the geospatial sensor web, ISPRS Int. J. Geo-Inf. 5 (5) (May. 2016) 63, https://doi.org/10.3390/ijgi5050063.

[54] D. Chen, X. Zhang, N. Chen, et al., "Geospatial Sensor Web adaptor for integrating diverse internet of things protocols within smart city," ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences 4 (2020) 115–121, https://doi.org/10.5194/isprs-annals-V-4-2020-115-2020, to be published.
[55] X.G. Xie, X.F. Liu, Heterogeneous sensor data access and management in Sensor Observation Service, Electronic Measurement Technology 39 (7) (Jul. 2016) 154–159, https://doi.org/10.19651/j.cnki.emt.2016.07.033.
[56] A. Broering, T. Foerster, S. Jirka, et al., Sensor bus: an intermediary layer for linking geosensors and the sensor web, in: Proceedings of the 1st International Conference and Exhibition on Computing for Geospatial Research & Application, Jun. 2010, pp. 1–8, https://doi.org/10.1145/1823854.1823870. Southampton, the UK.
[57] Z.L. Wang, F.L. Bu, Transducer and microsystem technologies, Transducer and Microsystem Technologies 34 (6) (Apr. 2015) 13–16, https://doi.org/10.13873/J.1000-9787(2015)06-0013-04.
[58] R. Pinto, S.A. Espírito, V. Paciello, Proposal of a sustainable e-bike sharing infrastructure based on the IEEE 1451 standard, in: The 45th Annual Conference of the IEEE Industrial Electronics Society, IECON, Lisbon, Portugal, Dec. 2019, pp. 5538–5543, https://doi.org/10.1109/IECON.2019.8927149.
[59] S. Lin, F. Lyu, H. Nie, An automatic instrument integration scheme for interoperable ocean observatories, Sensors 20 (7) (Apr. 2020) 1990, https://doi.org/10.3390/s20071990.
[60] X. Xuan, K. Li, Efficient anti-collision algorithm for RFID EPC generation-2 protocol based on continuous detection, Int. J. Wireless Inf. Network 27 (1) (Oct. 2020) 133–143, https://doi.org/10.1007/s10776-019-00465-4.
[61] V. Cherneva, J.L. Trahan, Serial-dependency grouping-proof protocol for RFID EPC Gen2 tags, IEEE Journal of Radio Frequency Identification 4 (2) (Oct. 2020) 159–169, https://doi.org/10.1007/s10776-019-00465-4.
[62] X. Li, R.S. Huang, A dynamic adapter method supporting sensor 'plug and play', Transducer and Microsystem Technologies 33 (7) (Apr. 2014) 20–23, https://doi.org/10.13873/j.1000-97872014.07.029.
[63] D. Suriano, G. Cassano, M. Penza, Design and development of a flexible, plug-and-play, cost-effective tool for on-field evaluation of gas sensors, Journal of Sensors 8 (1) (Aug. 2020) 1–20, https://doi.org/10.1155/2020/8812025.
[64] J.B. Gomes, J.J. Rodrigues, R.A. Rabêlo, et al., A novel Internet of things-based plug-and-play multi-gas sensor for environmental monitoring, Transactions on Emerging Telecommunications Technologies 4 (14) (Apr. 2020) 3967, https://doi.org/10.1002/ett.3967.
[65] S.L. Deshpande, Tree-based approaches for improving energy efficiency and life time of wireless sensor networks (WSN): a survey and future scope for research, in: Inventive Communication and Computational Technologies (ICICCT), Namakkal, India, Jan 2020, pp. 583–590, https://doi.org/10.1007/978-981-15-0146-3_55.
[66] Z. Hong, R. Wang, N. Wang, "A tree-based topology construction algorithm with probability distribution and competition in the same layer for wireless sensor network," Peer-to-Peer Networking and Applications, vol. 10, no. 3, pp. 658-669, doi: 10.1007/s12083-016-0514-8.
[67] S.M. Chowdhury, A. Hossain, Different energy saving schemes in wireless sensor networks: a survey, Wireless Pers. Commun. 114 (May. 2020) 2043–2062, https://doi.org/10.1007/s11277-020-07461-5.
[68] E.P. Gilbert, K. Baskaran, E.B. Rajsingh, et al., Trust aware nature inspired optimized routing in clustered wireless sensor networks, Int. J. Bio-Inspired Comput. 14 (2) (Aug. 2019) 103–113, https://doi.org/10.1504/IJBIC.2019.101637.
[69] B.D. Deebak, F. Turjman, A hybrid secure routing and monitoring mechanism in IoT-based wireless sensor networks, Ad Hoc Netw. 97 (Feb. 2020) 1–36, https://doi.org/10.1016/j.adhoc.2019.102022.
[70] L. Sun, X.Z. Song, Multi-slot allocation data transmission algorithm based on dynamic tree topology for wireless sensor network, J. Comput. Appl. 35 (10) (Oct. 2015) 2858–2862, https://doi.org/10.1504/IJBIC.2019.101637.
[71] X. Meng, H. Inaltekin, B. Krongold, Deep reinforcement learning-based topology optimization for self-organized wireless sensor networks, in: IEEE Global Communications Conference, GLOBECOM, Hawaii, USA, Dec. 2019, pp. 1–6, https://doi.org/10.1109/GLOBECOM38437.2019.9014179.
[72] H. Chang, J. Feng, C. Duan, Reinforcement learning-based data forwarding in underwater wireless sensor networks with passive mobility, Sensors 19 (2) (Oct. 2019) 256, https://doi.org/10.3390/s19020256.
[73] K. Nakashima, S. Kamiya, K. Ohtsu, et al., Deep reinforcement learning-based channel allocation for wireless lans with graph convolutional networks, IEEE Access 8 (Oct. 2020) 31823–31834, https://doi.org/10.1109/ACCESS.2020.2973140.
[74] S.E. Bouzid, Y. Seresstou, K. Raoof, et al., Multi-objective optimization of wireless network approach based on genetic algorithm, IEEE Access 8 (1) (Jun. 2020) 793–814, https://doi.org/10.1109/ACCESS.2020.2999157.
[75] S. Rani, S.H. Ahmed, R. Rastogi, Dynamic clustering approach based on wireless sensor networks genetic algorithm for IoT applications, Wireless Network 26 (7) (Jul. 2019) 2307–2316, https://doi.org/10.1007/s11276-019-02083-7.
[76] K.C. Chu, D.J. Horng, K.C. Chang, Numerical optimization of the energy consumption for wireless sensor networks based on an improved ant colony algorithm, IEEE Access 7 (Jul. 2019) 105562–105571, https://doi.org/10.1109/ACCESS.2019.2930408.
[77] A. Mubarakali, A.D. Durai, M. Alshehri, et al., Fog-based delay-sensitive data transmission algorithm for data forwarding and storage in cloud environment for multimedia applications, Big Data 14 (Jul. 2020) 1–9, https://doi.org/10.1089/big.2020.0090.16.
[78] N. Kabisch, P. Selsam, T. Kirsten, et al., A multi-sensor and multi-temporal remote sensing approach to detect land cover change dynamics in heterogeneous urban landscapes, Ecol. Indicat. 99 (Apr. 2019) 273–282, https://doi.org/10.1016/j.ecolind.2018.12.033.
[79] G. Veer, A. Ebert, N. Gershon, et al., "Visualizing information retrieved from (large) WHAT networks," in IFIP Conference on Human-Computer Interaction, INTERACT). Paphos, Cyprus (Aug. 2019) 735–740, https://doi.org/10.1109/IECON.2019.8927149.
[80] W. Li, M. Batty, M.F. Goodchild, Real-time GIS for smart cities, Int. J. Geogr. Inf. Sci. 9 (Oct. 2019) 311–324, https://doi.org/10.1080/13658816.2019.1673397.
[81] C.A. Charu, Managing and Mining Sensor Data, Kluwer Academic Publishers, Boston, 2013, pp. 1–533.
[82] H.Y. Teh, L.A.W. Kempa, I. Kevin, et al., Sensor data quality: a systematic review, Journal of Big Data 7 (11) (Feb. 2020) 11–18, https://doi.org/10.1186/s40537-020-0285-1.
[83] S. Sathe, H. Jeung, K. Aberer, Creating probabilistic databases from imprecise time-series data, in: The IEEE International Conference on Data Engineering (ICDE), Apr. 2011, pp. 327–338, https://doi.org/10.1109/ICDE.2011.5767838. Hannover, Germany.
[84] A. Ma'arif, I. Iswanto, A.A. Nuryono, et al., Kalman filter for noise reducer on sensor readings, Signal and Image Processing Letters 1 (2) (Feb. 2019) 50–61, https://doi.org/10.31763/simple.v1i2.2.
[85] S. Park, M.S. Gil, H. Im, Y.S. Moon, Measurement noise recommendation for efficient Kalman filtering over a large amount of sensor data, Sensors 19 (5) (Mar. 2019) 1168, https://doi.org/10.3390/s19051168.
[86] A. Verner, LSTM Networks for Detection and Classification of Anomalies in Raw Sensor Data, Ph.D. dissertation, College of Engineering and Computing, Florida, USA, 2019.
[87] S. Szewcyzk, K. Dwan, B. Minor, et al., Annotating smart environment sensor data for activity learning, Technology and Health Care 17 (3) (Jul. 2009) 161–169, https://doi.org/10.3233/THC-2009-0546, 2009.
[88] J. Ahn, E. Ko, E.Y. Kim, Highway traffic flow prediction using support vector regression and Bayesian classifier, in: International Conference on Big Data and Smart Computing (BigComp), Mar. 2016, pp. 239–244, https://doi.org/10.1109/BIGCOMP.2016.7425919. Hong Kong, China.
[89] H. Turabieh, M. Mafarja, S. Mirjalili, Dynamic adaptive network-based fuzzy inference system (D-ANFIS) for the imputation of missing data for Internet of medical Things applications, IEEE Internet of Things Journal 6 (6) (Dec. 2019) 9316–9325, https://doi.org/10.1109/JIOT.2019.2926321.
[90] Y. Liu, Z. Li, Z. Liu, et al., Real-time arm skeleton tracking and gesture inference tolerant to missing wearable sensors, in: Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services (Mobisys), the United States, New York, Jun. 2019, pp. 287–299, https://doi.org/10.1145/3307334.3326109.
[91] M. Gölz, M. Muma, T. Halme, et al., Spatial inference in sensor networks using multiple hypothesis testing and Bayesian clustering, in: 2019 27th European Signal Processing Conference (EUSIPCO), A Coruña, Spain, Sep. 2019, pp. 1–5, https://doi.org/10.23919/EUSIPCO.2019.8902986.
[92] R.R. Swain, P.M. Khilar, S.K. Bhoi, Underlying and persistence fault diagnosis in wireless sensor networks using majority neighbors coordination approach, Wireless Pers. Commun. 111 (2) (Oct. 2019) 763–798, https://doi.org/10.1007/s11277-019-06884-z.
[93] J.L. Hicks, T. Althoff, P. Kuhar, et al., Best practices for analyzing large-scale health data from wearables and smartphone apps, NPJ Digital Medicine 45 (2) (Jun. 2019) 1–2, https://doi.org/10.1038/s41746-019-0121-1.
[94] M. Templ, J. Gussenbauer, P. Filzmoser, Evaluation of robust outlier detection methods for zero-inflated complex data, J. Appl. Stat. 47 (7) (Sep. 2019) 1144–1167, https://doi.org/10.1080/02664763.2019.1671961.
[95] S. Yoon, J.G. Lee, B.S. Lee, NETS: extremely fast outlier detection from a data stream via set-based processing, Proceedings of the VLDB Endowment 12 (11) (Jul. 2019) 1303–1315, https://doi.org/10.14778/3342263.3342269.
[96] A. Aya, G. Oussama, M. Abdulfattah, Outlier detection approaches for wireless sensor networks: a survey, Comput. Network. 129 (1) (Dec. 2017) 319–333, https://doi.org/10.1016/j.comnet.2017.10.007.
[97] W.F. Van, Y. Wang, A. Khojandi, et al., Real-time sensor anomaly detection and identification in automated vehicles, IEEE Trans. Intell. Transport. Syst. 21 (3) (Apr. 2019) 1264–1276, https://doi.org/10.1109/TITS.2019.2906038.
[98] M. Zhang, X. Li, L. Wang, An adaptive outlier detection and processing approach towards time series sensor data, IEEE Access 7 (Dec. 2019) 192–212, https://doi.org/10.1109/ACCESS.2019.2957602.
[99] S. Sathe, T.G. Papaioannou, H. Jeung, et al., "A Survey of Model-Based Sensor Data Acquisition and Management," in Managing and Mining Sensor Data, 2013, pp. 9–50, https://doi.org/10.1007/978-1-4614-6309-2_2. Boston, MA.
[100] N. Paulauskas, A. Baskys, Application of histogram-based outlier scores to detect computer network anomalies, Electronics 8 (11) (Dec. 2019) 1251, https://doi.org/10.3390/electronics8111251.
[101] N. Nesa, T. Ghosh, I. Banerjee, Non-parametric sequence-based learning approach for outlier detection in IoT, Future Generat. Comput. Syst. 82 (1) (Dec. 2018) 412–421, https://doi.org/10.3390/electronics8111251.
[102] Y. Zhou, H. Zou, R. Arghandeh, et al., Non-parametric outliers detection in multiple time series a case study: power grid data analysis, in: Thirty-Second AAAI Conference on Artificial Intelligence, AAAI, 2018, pp. 201–210. Hawaii, USA.
[103] A. Abid, A. Kachouri, A. Mahfoudhi, Outlier detection for wireless sensor networks using density-based clustering approach, IET Wirel. Sens. Syst. 7 (4) (Aug. 2017) 83–90, https://doi.org/10.1049/iet-wss.2016.0044.

[104] F. Grützmacher, B. Beichler, A. Hein, et al., Time and memory efficient online piecewise linear approximation of sensor signals, Sensors 18 (6) (May. 2018) 1672, https://doi.org/10.3390/s18061672.

[105] X. Bin, Q. Li, Y. Wang, P. Wang, EPLA-DSTree: Extending piecewise linear approximation on a dynamic segmentation tree index in sensor-cloud systems, IEEE Access 7 (4) (May. 2020) 83–90, https://doi.org/10.3390/s18061672.

[106] M. Gao, G.Y. Shi, Ship spatiotemporal key feature point online extraction based on AIS multi-sensor data using an improved sliding window algorithm, Sensors 19 (12) (May. 2019) 2706, https://doi.org/10.3390/s19122706.

[107] J. Brito, K. Demirkaya, B. Etienne, et al., Efficient approximate query answering over sensor data with deterministic error guarantees, Rem. Sens. 12 (12) (Jul. 2017) 204–210, arXiv:1707.01414.

[108] Á. Villalba, J.L. Berral, D. Carrera, Constant-time sliding window framework with reduced memory footprint and efficient bulk evictions, IEEE Trans. Parallel Distr. Syst. 30 (3) (Mar. 2019) 486–500, https://doi.org/10.1109/TPDS.2018.2868960.

[109] T. Sato, Y. Takano, R. Miyashiro, Piecewise-linear approximation for feature subset selection in a sequential logit model, J. Oper. Res. Soc. Jpn. 60 (1) (Feb. 2017) 1–14, https://doi.org/10.15807/jorsj.60.1.

[110] J. Gao, W. Ji, L. Zhang, et al., Fast piecewise polynomial fitting of time-series data for streaming computing, IEEE Access 8 (43) (Feb. 2020) 43764–43775, https://doi.org/10.1109/ACCESS.2020.2976494.

[111] Y. Yu, Y. Zhu, D. Wan, et al., A novel symbolic aggregate approximation for time series, in: International Conference on Ubiquitous Information Management and Communication (ICUIMC), May. 2019, pp. 805–822, https://doi.org/10.1007/978-3-030-19063-7_65. Phuket, Thailand.

[112] H.Y. Peng, J. Cun, J. Ming, et al., A continuous segmentation algorithm for streaming time series, Collab. Comput.: Networking, Applications and Worksharing 201 (Jul. 2017) 140–151, https://doi.org/10.1007/978-3-319-59288-6_13.

[113] J.W. Lin, S.W. Liao, F.Y. Leu, Sensor data compression using bounded error piecewise linear approximation with resolution reduction, Energies 12 (13) (Jun. 2019) 2523, https://doi.org/10.3390/en12132523.

[114] C.M. Farias, L. Pirmez, F.C. Delicato, et al., A multi-sensor data fusion algorithm using the hidden correlations in multi-application wireless sensor data streams, in: IEEE 14th International Conference on Networking, Sensing and Control (ICNSC), Southern Italy, Calabria, 2017, pp. 1–7, https://doi.org/10.1109/ICNSC.2017.8000074.

[115] S. Lin, V. Kalogeraki, S. Gunopulos, Online information compression in sensor networks, in: IEEE International Conference on Communications (ICC), 2006, pp. 1–6, https://doi.org/10.1504/IJSNET.2006.012038. Providence, America.

[116] R.F. Albuquerque, P.D. Oliveira, A.P. "Braga, Adaptive fuzzy learning vector quantization (AFLVQ) for time series classification, in: North American Fuzzy Information Processing Society Annual Conference (NAFIPS), Cancun, Mexico, 2018, pp. 385–397, https://doi.org/10.1007/978-3-319-95312-0_33.

[117] X. Xu, R. Ansari, A. Khokhar, Parallel nonuniform discrete fourier transform (P-NDFT) over a random wireless sensor network, IEEE Trans. Parallel Distr. Syst. 28 (12) (May. 2017) 3615–3625, https://doi.org/10.1109/TPDS.2017.2706690.

[118] X. Liao, K. Li, J. Yin, Separable data hiding in encrypted image based on compressive sensing and discrete fourier transform, Multimed. Tool. Appl. 76 (Oct. 2017) 20739–20753, https://doi.org/10.1007/s11042-016-3971-4.

[119] B. Su, X. Ding, H. Wang, et al., Discriminative dimensionality reduction for multi-dimensional sequences, IEEE Trans. Pattern Anal. Mach. Intell. 40 (1) (Jan. 2018) 77–91, https://doi.org/10.1109/TPAMI.2017.2665545.

[120] H. Shen, G. Bai, J. Wang, et al., WGM: wavelet-based gamma model for video traffic in wireless multi-hop networks, Wireless Pers. Commun. 107 (Mar. 2019) 565–587, https://doi.org/10.1007/s11277-019-06289-y.

[121] R. Banerjee, S. Chatterjee, S.D. Bit, Performance of a partial discrete wavelet transform based path merging compression technique for wireless multimedia sensor networks, Wireless Pers. Commun. 104 (Jan. 2019) 57–71, https://doi.org/10.1007/s11277-018-6008-7.

[122] H.A. Hakami, Wavelet Based Multimedia Data Compression Techniques, " Ph.D. dissertation. University of Technology, Sydney, 2019.

[123] R. Banerjee, P. Roy, K. Layek, S.D. Bit, An energy efficient video compression using differential inter frame partial discrete wavelet transform for WMSN, in: IEEE Annual India Conference (INDICON), Dec. 2016, pp. 1–6, https://doi.org/10.1109/INDICON.2016.7838873. Bangalore, India.

[124] T.L. Li, Y. Liu, T. Tian, et al., A storage solution for massive IoT data based on NoSQL, in: The IEEE International Conference on Green Computing & Communications, Nov. 2012, pp. 50–57, https://doi.org/10.1109/GreenCom.2012.18. Besancon, France.

[125] E.A. Brewer, Towards robust distributed systems, Portland, Oregon, USA, in: Symposium on Principles of Distributed Computing, Jul. 2000, pp. 1–12. doi: [Online]. Available: https://www.itec.co.uk/_media/libraries/draft-abstracts–slides/65-Chuck-Wythe.pdf.

[126] M. Marinov, G. Georgiev, E. Popova, NoSQL approach for sensor data storage and retrieval, in: 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Jul. 2018, pp. 1427–1432, https://doi.org/10.23919/MIPRO.2018.8400257. Opatija, Croatia.

[127] K. Ferencz, J. Domokos, IoT sensor data acquisition and storage system using raspberry PI and Apache Cassandra, in: International IEEE Conference and Workshop in Óbuda on Electrical and Power Engineering, CANDO-EPE, Budapest, Hungary, Jul. 2018, pp. 143–146, https://doi.org/10.1109/CANDO-EPE.2018.8601139.

[128] Z.M. Ding, X. Gao, A database cluster system framework for managing massive sensor sampling data in the internet of things, Chin. J. Comput. 35 (6) (Jun. 2012) 1175–1191, https://doi.org/10.3724/SP.J.1016.2012.01175.

[129] S. Venkatraman, K. Fahd, S. Kaspi, et al., SQL versus NoSQL movement with big data analytics, Int. J. Inf. Technol. Comput. Sci. 12 (Dec. 2016) 59–66, https://doi.org/10.5815/ijitcs.2016.12.07.

[130] S. Agarwal, K.S. Rajan, Analyzing the performance of NoSQL vs. SQL databases for spatial and aggregate queries, in: Free and Open Source Software for Geospatial Conference Proceedings (FOSS4G), Sep. 2017, pp. 1–10, https://doi.org/10.7275/R5736P26. Boston, USA.

[131] Patent US, in: D. Yanovsky, T. Namoradze, V.D. Miloslavskaya, et al. (Eds.), Distributed Storage System Data Management and Security, vol. 10, 2020, Aug 4. *735*,137.

[132] Patent US, in: M. Kan, T. Torii (Eds.), Distributed Storage System and Distributed Storage Method, vol. 9, 2016, May 17, pp. 342–574.

[133] T. Lu, J. Fang, C. Liu, A unified storage and query optimization framework for sensor data, in: Web Information System & Application Conference, Feb. 2016, pp. 229–234, https://doi.org/10.1109/WISA.2015.13. Jinan, China.

[134] C. Huang, Q. Huang, D. Wang, Stochastic configuration networks based adaptive storage replica management for power big data processing, IEEE Transactions on Industrial Informatics 16 (1) (Jan. 2020) 373–383, https://doi.org/10.1109/TII.2019.2919268.

[135] D. Ousmane, D. Philippe, T. Ciprian, Data management mechanisms for Internet of Things: a position paper, in: International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, Dec. 2019, pp. 1214–1219, https://doi.org/10.1109/CSCI49370.2019.00228.

[136] M. Sharma, V.D. Sharma, M.M. Bundele, Performance analysis of RDBMS and NoSQL databases: PostgreSQL, MongoDB and Neo4j, in: International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE), Nov. 2018, pp. 1–5, https://doi.org/10.1109/ICRAIE.2018.8710439. Brisbane, Australia.

[137] W. Ge, Research on Techniques and Systems for Index and Query Optimization of Big Data, Ph.D. thesis, Dept. CST, Nanjing University, Nan Jing, 2019.

[138] E. Carneiro, A.V. Carvalho, M.A. Oliveira, I2B+ tree: interval B+ tree variant towards fast indexing of time-dependent data, in: 2020 15th Iberian Conference on Information Systems and Technologies (CISTI), Jun. 2020, pp. 1–7, https://doi.org/10.23919/CISTI49556.2020.9140897. Seville, Spain.

[139] H. Plieninger, Developing and applying IR-tree models: guidelines, caveats, and an extension to multiple groups, Organ. Res. Methods 4 (1) (Apr. 2020) 1–17, https://doi.org/10.1177/1094428120911096.

[140] J.M. Donkó, G. Szalai, G. Gombos, A. Kiss, An implementation of the M-tree index structure for PostgreSQL using GiST, in: IEEE 15th International Scientific Conference on Informatics (ISCI), Nov. 2019, pp. 189–194, https://doi.org/10.1109/Informatics47936.2019.9119265. Poprad, Slovakia.

[141] X. Zang, P. Hao, X. Gao, et al., QDR-Tree: An efficient index scheme for complex spatial keyword query, in: International Conference on Database and Expert Systems Applications (DEXA), Aug. 2018, pp. 390–404, https://doi.org/10.1007/978-3-319-98809-2_24. Bratislava, Slovakia.

[142] X. Guan, C. Bo, Z. Li, et al., ST-hash: An efficient spatiotemporal index for massive trajectory data in a NoSQL database, in: 25th International Conference on Geoinformatics (ICG), Aug. 2017, pp. 1–7, https://doi.org/10.1109/GEOINFORMATICS.2017.8090927. Buffalo, USA.

[143] X. Lu, X.A. Xu, Research on time series query method based on linear hash index, DEStech Transactions on Engineering and Technology Research 12 (7) (Nov. 2017) 482–486, https://doi.org/10.12783/dtetr/mcee2017/15798.

[144] B. Yildiz, Optimizing bitmap index encoding for high performance queries, Concurrency Comput. Pract. Ex. 7 (Sep. 2020) 1–9, https://doi.org/10.1002/cpe.5943.

[145] X.S. Zhao, J. Ben, W.B. Sun, et al., Overview of the research progress in the earth tessellation grid, Acta Geod. Cartogr. Sinica 45 (S1) (Dec. 2016) 1–14.

[146] X.Y. Huang, D.H. Zhao, W. Shao, et al., Research on method of battlefield situation representation based on multiplex information cell, J. Syst. Simul. 30 (9) (Jan. 2019) 3288–3292, https://doi.org/10.16182/j.issn1004731x.joss.201809008.

[147] J.C. Chuang, Y.C. Hu, C.M. Chen, et al., Joint index coding and reversible data hiding methods for color image quantization, Multimed. Tool. Appl. 78 (Nov. 2019) 35537–35558, https://doi.org/10.1007/s11042-019-08193-w.

[148] T. Cieslewski, D. Scaramuzza, Efficient decentralized visual place recognition using a distributed inverted index, IEEE Robotics and Automation Letters 2 (2) (Apr. 2017) 640–647, https://doi.org/10.1109/LRA.2017.2650153.

[149] M. Stonebraker, The case for partial indexes, ACM SIGMOD Record 18 (4) (Dec. 1989) 4–11, https://doi.org/10.1145/74120.74121.

[150] H. Wang, A. Belhassena, Parallel trajectory search based on distributed index, Inf. Sci. 388 (May. 2017) 62–83, https://doi.org/10.1016/j.ins.2017.01.016.

[151] R. Cai, Z. Lu, L. Wang, et al., DITIR: distributed index for high throughput trajectory insertion and real-time temporal range query, Proceedings of the VLDB Endowment 10 (12) (May. 2017) 1865–1868, https://doi.org/10.14778/3137765.3137795.

[152] X. Wei, Research of Mixed Index of Hash, B+, 3DR and B*, M.S. thesis, China University of Geosciences, Bei Jing, 2017.

[153] L.G. Xiang, M. Gao, D.H. Wang, et al., Geohash-Trees: an adaptive index which can organize large-scale trajectories, Geomatics Inf. Sci. Wuhan Univ. 44 (3) (May. 2019) 436–442, https://doi.org/10.13203/j.whugis20160523.

[154] J. Gong, S.N. Ke, Q. Zhu, An efficient trajectory data index integrating R-tree, Hash and B*-tree, Acta Geod. Cartogr. Sinica 44 (5) (May. 2015) 570–577, https://doi.org/10.11947/j.AGCS.2015.20130520.

[155] A.S. Elqusy, S.E. Essa, A. Sayed, Key management techniques in wireless sensor networks, Int. J. Netw. Secur. Appl. 6 (6) (Nov. 2014) 49–63, https://doi.org/10.5121/ijnsa.2014.6605.

[156] P. Sangat, S.M. Indrawan, D. Taniar, Sensor data management in the cloud: data storage, data ingestion, and data retrieval, Concurrency Comput. Pract. Ex. 2 (Nov. 2017) 1–10, https://doi.org/10.1002/cpe.4354.

[157] K. Mekki, W. Derigent, E. Rondeau, et al., In-network data storage protocols for wireless sensor networks: a state-of-the-art survey, Int. J. Distributed Sens. Netw. 15 (4) (Nov. 2019) 1–22, https://doi.org/10.1177/1550147719832487.

[158] J.B. Liang, F.Q. Ma, C. Jiang, Research progress on data query technology in dynamic wireless sensor networks, Computer Science 46 (11) (Aug 2019) 41–48, https://doi.org/10.1177/1550147719832487.

[159] Y. Yao, J. Gehrke, Query processing in sensor networks, in: First Biennial Conference on Innovative Data Systems Research (CIDR), 2003, pp. 1–12, https://doi.org/10.1109/MPRV.2004.1269131. Asilomar, USA.

[160] D. Abadi, S. Madden, W. Lindner, Sensor network integration with streaming database systems, in: Data Stream Management (DSM), 2016, pp. 409–428, https://doi.org/10.1007/978-3-540-28608-0. Berlin, Heidelberg, Germany.

[161] A.K. Qurabat, A.K. Idrees, Two level data aggregation protocol for prolonging lifetime of periodic sensor networks, Wireless Network 25 (Feb. 2019) 3623–3641, https://doi.org/10.1007/s11276-019-01957-0.

[162] A.K. Nayak, B.S. Mishra, H. Das, Computational Intelligence in Sensor Networks, 2019, pp. 1–12. Berling, Germany: Springer.

[163] M. Balazinska, A. Deshpande, M.J. Franklin, et al., Data management in the worldwide sensor web, IEEE Pervasive Computing 6 (2) (Jun. 2007) 30–40.

[164] C. Granell, A. Kamilaris, A. Kotsev, et al., Internet of things, in: Manual of Digital Earth, 2020, p. 852. Singapore: Springer.

[165] F. Yang, F.X. Li, H.L. Yu, et al., A hybrid peer-to-peer lookup service algorithm on distributed hash table, J. Software 3 (Mar. 2007) 714–721, https://doi.org/10.1109/MPRV.2007.27.

[166] K.A. Schmid, A. Züfle, Representative query answers on uncertain data, in: Proceedings of the 16th International Symposium on Spatial and Temporal Databases (SSTD), 2019, pp. 140–149, https://doi.org/10.1145/3340964.3340974. Vienna, Austria.

[167] A. Bhattacharya, A. Meka, A. Singh, MIST: distributed indexing and querying in sensor networks using statistical models, in: International Conference on Very Large Data Bases, University of Vienna, Austria, 2007, pp. 854–865.

[168] J.P. Yu, Y.P. Lin, Real-time query processing for sensor networks based on ant algorithm, J. Software 21 (3) (Mar. 2010) 473–489, https://doi.org/10.3724/SP.J.1001.2010.03413.

[169] Y. Zhong, J. Fang, X. Zhao, VegaIndexer: A Distributed composite index scheme for big spatiotemporal sensor data on cloud, in: IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Jul. 2013, pp. 1713–1716, https://doi.org/10.1109/IGARSS.2013.6723126. Melbourne, VIC, Australia.

[170] H.W. Durrant, T.C. Henderson, Multisensor Data Fusion, Springer handbook of robotic, Berling, Germany: Springer, 2016, pp. 867–896.

[171] F. Xiao, Multi-sensor data fusion based on the belief divergence measure of evidences and the belief entropy, Inf. Fusion 46 (Mar. 2019) 23–32, https://doi.org/10.1016/j.inffus.2018.04.003.

[172] Y. Song, Y. Deng, A new method to measure the divergence in evidential sensor data fusion, Int. J. Distributed Sens. Netw. 15 (4) (Apr. 2019) 1–8, https://doi.org/10.1177/1550147719841295.

[173] M. Muzammal, R. Talat, A.H. Sodhro, et al., A multi-sensor data fusion enabled ensemble approach for medical data from body sensor networks, Inf. Fusion 53 (Jan. 2020) 155–164, https://doi.org/10.1016/j.inffus.2019.06.021.

[174] M.S. Kerdabadi, R. Ghazizadeh, H. Farrokhi, et al., Energy consumption minimization and throughput improvement in cognitive radio networks by joint optimization of detection threshold, sensing time and user selection, Wireless Network 25 (4) (May. 2019) 2065–2079, https://doi.org/10.1007/s11276-018-1797-x.

[175] S. Zhang, Research on Balanced Energy Consumption Based on Energy-Consuming Model of WSNs, North China University of Technology, China, Beijing, 2017, pp. 12–17.

[176] Y. Wei, G. Duan, L. Tao, et al., New energy vessel routing and optimization for marine debris collection under uncertain environment, J. Phys. Conf. 1903 (1) (Jan. 2020) 12057.

[177] J. Wang, Y.H. Xu, L. Li, Privacy protection method for low energy consumption supporting integrity verification data fusion, J. Jilin Univ. (Eng. Technol. Ed.) (May. 2021) 1–9, https://doi.org/10.13229/j.cnki.jdxbgxb20210112.

[178] C. Yin, J. Xi, R. Sun, et al., Location privacy protection based on differential privacy strategy for big data in industrial internet of things, IEEE Transactions on Industrial Informatics 14 (8) (Aug. 2018) 3628–3636, https://doi.org/10.1109/TII.2017.2773646.