Mechanical Systems and Signal Processing 138 (2020) 106564

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

Data management techniques for Internet of Things

Bassirou Diène [a], Joel J.P.C. Rodrigues [b,c,d], Ousmane Diallo [a,*], EL Hadji Malick Ndoye [a], Valery V. Korotaev [d]

[a] Department of Informatics, University of Assane Seck, Ziguinchor, Senegal
[b] Federal University of Piauí (UFPI), Teresina, PI, Brazil
[c] Instituto de Telecomunicações, Portugal
[d] ITMO University, St. Petersburg, Russia

ARTICLE INFO

Article history:
Received 25 May 2019
Received in revised form 6 November 2019
Accepted 11 December 2019

Keywords:
Data management
Internet of Things
IoT applications
Big data
Industrial Internet of Things

ABSTRACT

Internet of Things (IoT) is a network paradigm in which physical, digital, and virtual objects are equipped with identification, detection, networking, and processing functions to communicate with each other and with other devices and services on the Internet in order to perform the users' required tasks. Many IoT applications are provided to bring comfort and facilitate the human life. In addition, the application of IoT technologies in the automotive industry has given rise to the concept of Industrial Internet of Things (IIoT) which facilitated using of Cyber Physic Systems, in which machines and humans interact. Due to the diversity, heterogeneity, and large volume of data generated by these entities, the use of traditional database management systems is not suitable in general. In the design of IoT data management systems, many distinctive principles should be considered. These different principles allowed the proposal of several approaches for IoT data management. Some middleware or architecture-oriented solutions facilitate the integration of generated data. Other available solutions provide efficient storage and indexing structured and unstructured data as well as the support to the NoSQL language. Thus, this paper identifies the most relevant concepts of data management in IoT, surveys the current solutions proposed for IoT data management, discusses the most promising solutions, and identifies relevant open research issues on the topic providing guidelines for further contributions.

1. Introduction

In computing science's history, the Internet had become more present in people's lives in a very short time than any other technology. It had revolutionized the possibility of people's communication. Later, Internet involves the possibility of connecting machines, devices, software, and objects [1]. Through the use of the Transmission Control Protocol/Internet Protocol (TCP/IP) architecture, these objects can be connected a network. This allowed communication between them without human intervention. Most recently, a new concept called, Internet of Things (IoT), appears and it is considered the main evolution to the fourth generation of the Internet. IoT does not have a unambiguous and acceptable definition from users of the global community [2]. Only the term is defined by several entities associating academicians, researchers, practitioners, innovators, developers, and businessmen but its initial use was attributed to Kevin Ashton, an expert in digital innovation [3]. One of the best IoT definition considers it a dynamic and global network infrastructure, in which intelligent things (objects), subsystems

* Corresponding author.
E-mail address: o.diallo@univ-zig.sn (O. Diallo).

https://doi.org/10.1016/j.ymssp.2019.106564
0888-3270/

and individual physical and virtual entities, are identifiable, autonomous, and self-configurable [4,5]. This sort of network has generally a large number of nodes (things) widely distributed on a given area and communicating among themselves in order to interact with environment by exchanging sensed data, while reacting to events and triggering actions to control the physical world. Each network node is considered smart (given by the software that allows autonomous behaviors) with generally very limited resources, especially those of storage, processing and energy. This implies, among other problems, the reliability and validity of data, performance, security, the privacy.

Over the years, with the development of RFID technology, sensors and actuators, wireless mobile communication, embedded systems and cloud computing, IoT technologies offer several potentialities and enabled the development of applications that evolve around the Internet network. These applications are used to improve people live conditions in many areas. Indeed, applications are being developed for automotive industry with the concept of Industrial Internet of Things (IIoT) [6,7], smart environments, such as smart homes and offices, logistic and distribution systems, healthcare, surveillance and security, supply chain, manufacturing industry, etc. [8,9]. An illustration of an IoT applications architecture may be found in Fig. 2. These IoT applications not only enhance the comfort of human life, but they provide more controls to simplify work routine and personal tasks, as well as the benefit offered to the industry. Industrial IoT considered as a subset of IoT (Fig. 1) covers the domains of machine-to-machine (M2M) and industrial communication technologies with automation applications [10]. Thus, IIoT has allowed a better understanding of the manufacturing process for efficient and sustainable production. Furthermore, to increase productivity and efficiency through smart and remote management, manufacturers, utility companies, agriculture producers and healthcare providers have adopted the IIoT [11]. The diversity of devices that continually send data provides a large volume of geographically dispersed and heterogeneous data in real time. In this context, the study and design of databases management techniques that take into account all these specificities of IIoT-based applications are therefore necessary [12]. Traditionally, data management is a wide concept that refers to architectures, practices, and procedures for managing the data lifecycle requirements of system. In the context of IoT, data management should act as a layer between the objects and devices where the applications use data for analysis purposes and services [13].

Due to the very large amount of data and their distinctive characteristics [14], the use of traditional database systems would not be a better solution. Indeed, the design of IoT's data management systems is based on several principles. Depending on these different principles, several data management approaches have been proposed, such as middleware-based IoT oriented on data and sources, data storage, and indexing solutions, and IoT data schema support solutions. The middleware-based IoT approaches provide a middleware between objects and data storage locations [15,4]. Thus, the middleware facilitates data streams routing. IoT data can be generated quite rapidly, the volume of data can be huge and with a diversity of data types. In order to address these potential problems, data storage solutions propose data storage environments not only enabling efficient storing of massive IoT data, but also integrating both structured and unstructured data with the possibility for being accessed by NoSQL queries [16,17]. On the other hand, data indexing solutions enable search and discover data efficiently [18,19]. Some of these systems use hash tables with centralized indexing but others perform distributed indexing in dynamic IoT networks. As data volumes are diverse, solutions that define flexible schema support will ensure interoperability.

There are some published survey works that present different aspects of the data management in IoT. For example, in [4], the authors explore some proposed data management solutions for IoT. They give the distinctive design primitives that should be addressed in an IoT data management solution, and they present their data management framework for IoT. Padiya *et al.* address the challenges of managing huge interactive sensor data using Resource Description Framework (RDF) including in their experiments different RDF storage mechanisms, such as triple store, property table, vertically, and horizontally partitioned table, column store, and data aware hybrid storage [20]. The authors of [21] discuss and identify different aspects of data in IoT including the sources of data, data gathering, data processing, and the transmission devices. They also discuss challenges brought by the needs to manage vast quantities of heterogeneous data across heterogeneous systems. The storage of a huge quantity of generated data becomes more and more complex, and emerged the cloud IoT paradigm. Thus, Botta *et al.* survey the integration of Cloud and IoT discussing their complementarity, detailing what is

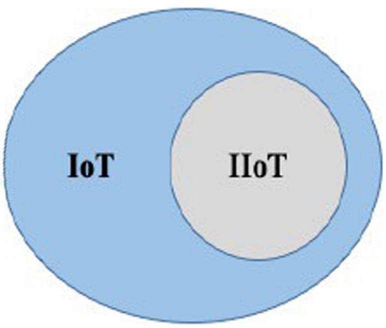

Fig. 1. Illustration of the relationship between Internet of Things (IoT) and Industrial Internet of Things (IIoT).

Fig. 2. Illustration of IoT applications (verticals).

Table 1
Comparison of the most relevant surveys on the topic.

Challenges	IoT applications and type of Data	IoT data management Taxonomy	Challenges	Open issues Discussion	Lessons learned
Abu-Elkheir et al. [4]	✓	✗	✓	✓	✗
Padiya et al. [20]	✗	✗	✓	✗	✗
Vongsingthong et al. [21]	✗	✗	✓	✓	✗
Botta et al. [22]	✗	✗	✓	✗	✗
This survey	✓	✓	✓	✓	✓

driving to their integration [22]. The comparison table below (Table 1) shows the contribution of this survey regarding other existing surveys on IoT data management.

This paper analyzes the state of the art of data management techniques for Internet of Things. Thus, the main contributions of this paper are the following:

- A deep study that shows how databases management techniques are adapted for IoT applications in order to improve people's live conditions in many areas by well managing the geographically dispersed and heterogeneous IoT data, in real time;
- A detailed presentation of various applications of IoT and an interesting study of the data types generated by IoT applications;
- A comparative study of data management in IoT and the traditional relational database management;
- A presentation, ranking, and discussion considering the most recent and relevant proposals of data management in IoT;
- Identification of open issues on IoT data management techniques suggesting further contributions in the topic.

The remainder of this paper is organized as follows. Section 2 details the most relevant verticals on IoT and Section 3 explores the data types for IoT applications presenting the essential conceptual features of IoT data management systems. In Section 4, the current and most relevant proposals on IoT data management are described. A deep and interesting discussion carried out on the different proposals of data management techniques for IoT, summarizing and classifying these proposals is addressed in Section 5. Moreover, relevant open research issues are identified. Finally, Section 6 presents IoT data requirements for processing and storage and Section 7 concludes the paper.

2. Internet of things verticals

Years ago, advances in Radio Frequency Identification (RFID) [23] and related technologies, such as Industrial Internet of Things (IIoT) were particularly applied to the supply chain applications of consumables and retail industries. Over the years, the potentialities offered by IIoT have allowed the development of several applications that revolve around it [2]. In many aspects of industries, IIoT has gained more and more space. Several applications are being developed for smart homes and offices, intelligent transportation systems, smart healthcare service industry (hospitals), smart businesses and industries, safer mining production, firefighting, etc. [2,24,25]. These different types of IoT applications are usually referred to Verticals, as illustrated in Fig. 2. In this section a deep study of various IoT application domains has been presented.

2.1. IoT in manufacturing industry

With advances noted in embedded systems, industrial manufacturing sector has evolved considerably by supporting the development of IoT producing intelligent products. The industrial sector not only provides intelligent equipment, but also uses IoT technology itself [26]. Thus, to show the effectiveness of IoT in the reconfiguration of industrial manufacturing, fully dynamic networks of cooperating equipment are often created. For example, remote monitoring applications can increase visibility by integrating wireless location sensors into wearable products and devices [27]. This will allow continuous people and property monitoring, such as stocks and lines of available materials. Sensors integrated in the production machines allow also monitoring of each machine along the production chain. The German initiative "Industry 4.0" (where 4.0 means the fourth industrial revolution) shows the importance of adopting IoT in industrial manufacturing. This concept arises when the IoT paradigm is merged with the Cyber-Physical Systems (CPSs) idea [28]. It promotes the integration of production processes, IoT technology and cyber-physical systems in a so-called smart factory. In these factories, the products communicate with their environment, with other products, machines and users [26]. In addition, the intelligent systems of flower shops based on distributed embedded devices are proposed. These devices focus behavioral and intelligent programming on a small number of monolithic computing resources, associated with a large number of devices without many utilities. Intelligence and behavior are adapted and programmed individually for each application.

2.2. IoT in supply chain

With the development of technologies used in IoT, it is possible to track the location and conditions of an object throughout its production cycle and supply chain. Today, these technologies are still used to detect diversions into illicit markets as well as the introduction of counterfeit products. This ensures the complete integrity of the product supply chain. Complete integrity that includes the physical integrity, the integrity of the means of transport and the integrity of the product and all its subcomponents [8]. Sensors are used to ensure that the product is not exposed to environmental conditions potentially damaging, for example if the temperature or the shock levels have never exceeded thresholds of deterioration. The integrity of the means of transport can be ensured, i.e., the product has never been in an area where it should not be. For example, dangerous articles cannot be transported to densely populated or ecologically sensitive areas given the serious consequences of an accident. The integrity of a product and all its sub-components according to the means of production must also be ensured. It must be ensured that a final product has not exceeded certain levels of gas emissions and traces of carbon. IoT technologies can enable recording of all the emissions generated during the production and in the transport of each sub-component through a multi-level of a complex supply chain. In general, new technologies are mostly used to ensure the full compliance of all the supply chain participants against a set of agreed rules: legal regulations, internal policies, and service level agreements.

2.3. IoT in energy

The use of modern technologies associated with the concepts of IoT and Internet of Services (IoS) [29] has driven to a paradigm change with new concepts and innovative applications. The current trend is to provide more dynamic services and market-based infrastructure where efficiency and energy savings can be better addressed through interactive distribution networks. One example is the creation of Advanced Metering Infrastructure (AMI) in the energy field. These infrastructures refer to systems that collect, measure, and analyze the use of energy from advanced equipment, such as electricity, gas, and water meters. These devices are generally referred to as smart meters. They are wireless sensors that connect to each other, but can also cooperate with sensor networks at home to transmit their data. In this perspective, Kamilaris et al. [30] designed and implemented a prototype of smart gateway for a smart meter, beginning by illustrating the application of the IoT architecture monitoring and controlling home energy consumption [31]. They contributed to an IoT realization by integrating real-world objects, such as wireless sensor networks, embedded devices, and home appliances with other

Internet content. In addition, their paper describes two paths for integrating devices to the Internet, first using REST Web services [32] that orients the integration on the progress of embedded computing and, secondly, an intelligent gateway-based approach to peripheral devices with limited resources [33].

2.4. IoT in healthcare

IoT is proposed to improve the quality of life and well-being by automating the basic tasks that humans should perform. In this sense, monitoring and decision making can be managed specifically by machines [2]. One of the main applications of IoT in the medical domain is for assisted living scenarios. Sensors are integrated into the health monitoring equipment used by patients. Thus, the information collected by these sensors is available on the Internet for physicians, family members, and other caretakers to improve treatment and responsiveness. In addition, IoT devices can be used monitor patients monitoring drugs and to evaluate the risk of new drugs in case of allergic reactions and negative interactions. In this perspective, the authors of [34] proposed the application of IoT in the medical system, especially in clinical care. Similarly, at the Jena University Hospital in Germany, RFID technology and IoT are used to optimize both logistical processes and to provide better care, to follow the equipment of patients, especially newborns, and drugs [23]. This benefit both medical providers in the form of savings as well as for patients since false treatments can be avoided. RFID technology applied to blood management can effectively achieve patient identification, reduce blood contamination and improve data collection effectively. In addition to RFID, sensors are deployed in patient rooms to capture patient data and transfer it to health professionals or hospital staff.

2.5. IoT in smart environments

The application of the IoT paradigm to an environmental context is of particular interest as it responds to the need of many national governments to adopt ICT solutions in the management of public affairs, thus realizing the so-called Smart City concept [35]. This paradigm in the Smart City context may bring a number of benefits in the management and optimization of traditional public services, such as, water supply, lighting, surveillance and maintenance of public areas, salubrity of hospitals and school [36].

2.5.1. Smart water supply

Smart cities can monitor the water supply to ensure adequate access for resident and business needs [37]. Wireless sensor networks provide technology for cities to monitoring water pumping systems more precisely and to discover their greater risk of water loss [2]. Cities that address water leakage issues with sensor technology have produced more savings on their investments. The system can report pipe flow measurement data regularly, and can also send alerts if water usage exceeds an estimated normal flow. This allows a smart city to determine the location of leaking pipes and repairs in priority based on the amount of water loss that can be avoided [38,39]. In [11], Thames Water the largest provider of water and waste-water services in the UK, anticipates equipment failures and provide fast response to critical situations such as leaks or inclement weather events. This is possible with the use of sensors, as well as the real-time data acquisition and analytics. More than 100,000 smart meters are installed in London and the company's goal is to cover more customers (currently 15 million) with smart meters by 2030. With their 4200 leaks detected on customer pipes this program has already saved 930,000 L of water per day across London [6].

2.5.2. Smart public buildings

The maintenance of the historical buildings in a city requires continuous monitoring of the current conditions of each building and identification of the most vulnerable areas for external agents. A complete characterization of the environmental conditions of buildings can be achieved by installing vibration and deformation sensors for building stress monitoring, atmospheric agent sensors in the vicinity to monitor pollution levels, and temperature and humidity sensors [40]. These data collection is a distributed database of structural integrity measures for buildings. This will reduce the need for costly periodic structural testing by human operators and enable targeted and proactive maintenance and recovery actions. It will also be possible to combine vibration and seismic readings to better study and understand the impact of light earthquakes on urban buildings.

2.5.3. Smart waste management

The most common problem in modern large cities is waste management because of both the cost of service and the problem of the storage of waste in landfills. With IoT, the integration and use of smart waste containers can enable the detection of load levels and the optimization of collector truck routes. This will reduce the cost of managing waste collection and improve the quality of recycling [41,42].

2.5.4. Smart homes and offices

Different electronic gadgets revolve around people. These include microwave ovens, refrigerators, heaters, air conditioners, fans, and lamps. Actuators and sensors can be installed in these devices in order to use energy efficiently and also to add more comfort in life. These sensors can measure the temperature it makes outside, and even can determine the occupants inside the rooms and so can control the amount of heat, freshness and light flow. These efforts can help to minimize the cost of energy and increase the economy [2].

2.5.5. Automation and salubrity of public building

The salubrity monitoring of environment and energy consumption in public buildings (administration offices, schools, museums, etc.) is another important application of IoT technologies. Using simple different types of sensors and actuators that control temperature, lights, temperature, it is possible to enhance the level of comfort of the persons that live in these environments. This may also have a positive return in terms of productivity, while reducing the costs for heating/cooling [43].

2.6. IoT in transport and distribution

The transport and distribution sector is among the first to adopt IIoT. Simple applications developed on vehicles allow remote monitoring and facilitate the location and training of fleets, trucks, and vans [44–46]. Logging driver hours, speed and driving behavior can improve safety and simplify compliance. Vehicles like buses and trains with sensors-equipped roads and rails can provide accurate information to the driver for better and safer travel. With the use of assisted driving, it would be possible to find a good route with information in priority on traffic jams and incidents. Similarly, in the new smart car parks, sensors are installed to detect arrival and departure of vehicles. This provides extensive parking management solutions that help motorists to save time and fuel [47,48] and to have a good traffic management [49]. A significant contribution to congestion comes from motorists seeking accessible parking spaces [2]. Providing accurate information about parking spaces provides a better flow of traffic, and it will also help reservation of parking spaces directly from the vehicle [50]. This would also reduce CO_2 emissions.

In the context of a company, information on the vehicle carrying the products, information related to the type and status of the products can be integrated to provide accurate information on delivery time, delivery delays, and mistakes [51,45]. The implementation of IoT in retail chain monitoring has several advantages: RFID and Near Field Communication (NFC) can be used for monitoring each link in the supply chain, ranging from the details of the commodity, purchase of raw materials, production, transportation and storage, product sales and after-sales services. IoT will track the inventory in the warehouse, so the stock can be reloaded at the right time for a continuous sale, which will reduce the waiting time of the customer resulting in customer satisfaction and increased sales [52]. Customers can also be regularly informed about the delivery times of the scheduled products. Even the containers themselves are now equipped with tables of limited capacity, which open possibilities for tracking and organizing articles in a logistic chain of storage and delivery [27].

3. Background and basic concepts of data management

Data is one of the most valuable aspects of IoT. Thus, their management is a very important role and becomes a key research topic. Several works have been done to allow the processing and analysis of data from IoT elements, especially in RFID, Wireless Sensor Networks (WSNs), and other related technologies. With the variety and diversity of the data sources, several types of data can be identified.

3.1. Data types in IoT

In IoT environments, data comes from different kind of devices and represents billions of objects. IoT data can be classified into several categories (see Fig. 3) [53]. Some data is discrete or continuous, others are generated automatically, and several others introduced by users.

3.1.1. Radio Frequency identification (RFID) data

Radio Frequency Identification refers to identification and tracking using radio waves. RFID tags can be inserted into objects and used to transmit and receive data. The RFID tag is based in an integrated circuit that can store data and an antenna to receive and transmit signals. Tags can be miniaturized to few millimeters in length and width, thus enabling their ubiquitous use for everyday objects. A tag is activated by radio waves emitted from an RFID reader. The reader communicates wirelessly with the tag. Once activated, the tag sends data stored in its memory relating to the item back to the reader. RFID tags can be active or passive. Active RFID tags contain a battery and can operate autonomously. Passive RFID tags are only activated when they receive a radio wave sent by an RFID reader. The technology is now used in many areas, such as passports, livestock tracking, road tolling, supply chain management, logistics, stock control, and healthcare.

3.1.2. Descriptive data about Objects, Processes, and systems

Much of the power of IoT will come from data or metadata that will be recorded on the participating objects, processes, and systems. Metadata is data about data and is essential to enable users to find and access the appropriate data. There are research questions on how the data will be stored, represented, and validated, to ensure maximum efficiency and non-repudiation in its retrieval and update. Objects will need to be self-describing and have the ability to report on dynamic char-

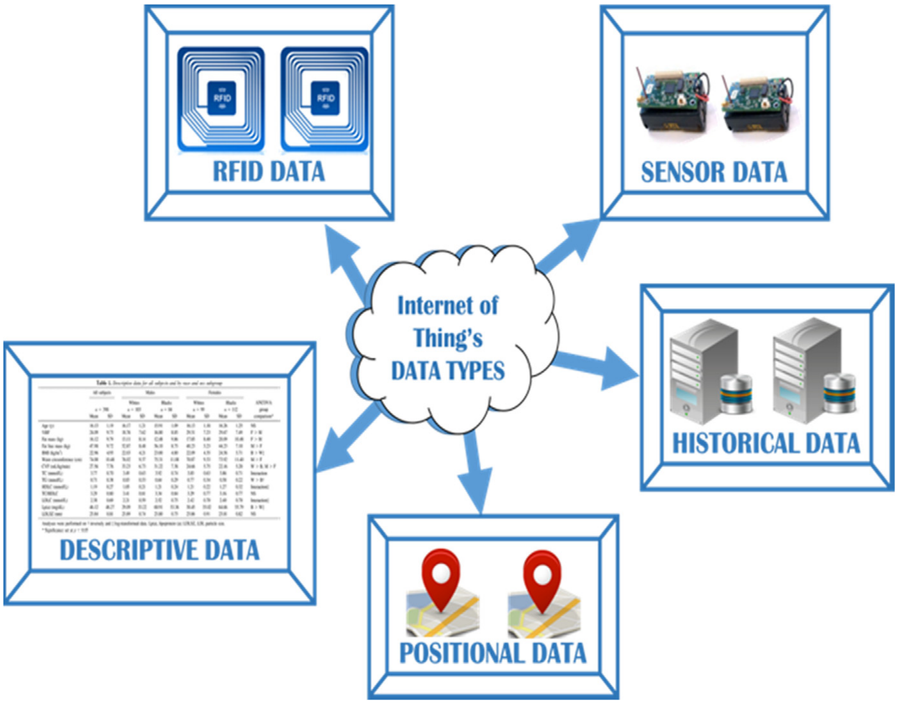

Fig. 3. Illustration of data types in IoT.

acteristics to maximize sharing. Not only data about basic objects, but data about processes and systems will also need to be stored. Systems and processes can be regarded as special types of objects, although objects of a more complex nature than basic objects. It is important to store data about objects, processes, and systems, so that users know how to take advantage of the services and facilities offered by IoT. IoT services will allow users to see in which processes and systems an object participate. Metadata about processes and self-describing data, together with good indexing systems, will be helpful. An interesting research area will be the development of suitable representation schemes to capture different types of objects and their metadata to maximize usage and usability.

3.1.3. Descriptive data about Objects, Processes, and systems

One of the way by which data is received in IoT systems is through WSNs. Advances in electronics have made it relatively easy to set up these networks for monitoring all sorts of environmental phenomena (e.g., weather, temperature, and noise). Decisions need to be made on how frequently data should be captured, for instance, continuously, at regular intervals, or only when queried. Questions arise on how can be ensured and obtained a representative sample in an efficient manner and how much data captured should be archived. Sensors, cloud, and grid technology have made it possible to capture vast amounts of data very quickly, but querying and mining these data can be problematic, particularly when the analysis must be achieved in real-time.

3.1.4. Positioning data

Positioning data provides location of a particular tagged object either within a Global Positioning System (GPS) or within a local positioning system. GPS is implemented with multiple satellites sending signals to a controlling unit from which objects can ascertain their position through triangulation. Local positioning systems work in a similar way, with smaller coverage. Some examples of local technologies are cellular base stations, Wi-Fi access points, and TV towers. Local positioning systems can be used in collaboration with GPS or, sometimes, instead of GPS. They can be used inside buildings or heavily built-up areas. In smaller areas, such as a room inside a house, positional data can come from locally placed sensors and transmitters. Multiple sensors send signals to a smart object, which can then work out its location or the location of a collaborating object. Positional data will play an important role in IoT, given that its components may be static or mobile. A very fine-grained positioning remains a challenge and is being actively addressed by researchers.

3.1.5. Historical data

Petabytes (and more) of data will be captured by sensors in IoT systems. Such data may require to be stored. As time passes data becomes historical. Volume becomes a challenge. Application-oriented design decisions need to be made on how the data should be retained and which data should be retained. Some will be kept in active data warehouses for frequent querying, some may be needed less frequently and may be archived in less accessible structures. In both cases, time needs to be captured alongside data, along with location information, ownership, and capture method. Issues in data archiving, such as, data loss, inaccurate recording, missing information, and dependence on obsolete technology have been recognized for quite some time. The database community has offered some solutions toward successful data archiving. Such solutions may be adapted to IoT.

3.2. Data management systems in IoT vs traditional database management systems

Traditionally, data management is a large concept that refers to architectures, practices, and procedures for meeting the data lifecycle requirements of a system. In the context of IoT, data management will serve as a layer between objects and devices generating data and the applications that use them for analysis and services [13]. Due to the diversity, heterogeneity, and large amount of data generated by objects [17], the use of traditional relational database management systems would not generally be suitable for many reasons. First, in traditional database systems, the mass of data is collected from finite and predefined sources and then stored in scalar form according to strict rules of relationship normalization. In IoT, while there is a large and growing number of data sources with sensors, RFIDs, embedded systems, and so on. Secondly, traditional data management systems handle storage, retrieval, and updating of basic data items, records, and files. In the context of IoT, data management systems must take into account online data while providing storage, logging and auditing capabilities for online analysis. Unlike occasional queries and updates that are sent to traditional database management systems (DBMS), data in IoT is sent continuously from multitude objects to data warehouses, and queries are more frequent and with more requirements. Third, in IoT, obtaining a strict relational database schema and a practice of relational normalization can be relaxed in favor of unstructured and more flexible forms of structures that adapt to different types of data and to different complex queries. However, traditional database management systems have many aspects that can be used by IoT. These include the use of remote storage at the object layer, support for unstructured data, relaxation of ACID properties to ensure consistency, availability, and integration of data. Taking into account the transparency requirements of traditional DBMS, in IoT data management systems is based on principles that determine the logical and physical structure of data management solutions for IoT.

4. Current proposals of data management systems in IoT

Data management systems design addresses the architecture of data management systems and how data should be stored and archived. Processing elements deal with access and data processing in storage systems. As part of the design of data management systems several approaches are adopted, such as the following and illustrated by the taxonomy in Fig. 4:

- the deployment of an IoT middleware centered on sources and on data,
- the establishment of a schema support adapted to IoT data
- the designing of an efficient storage and indexing platform

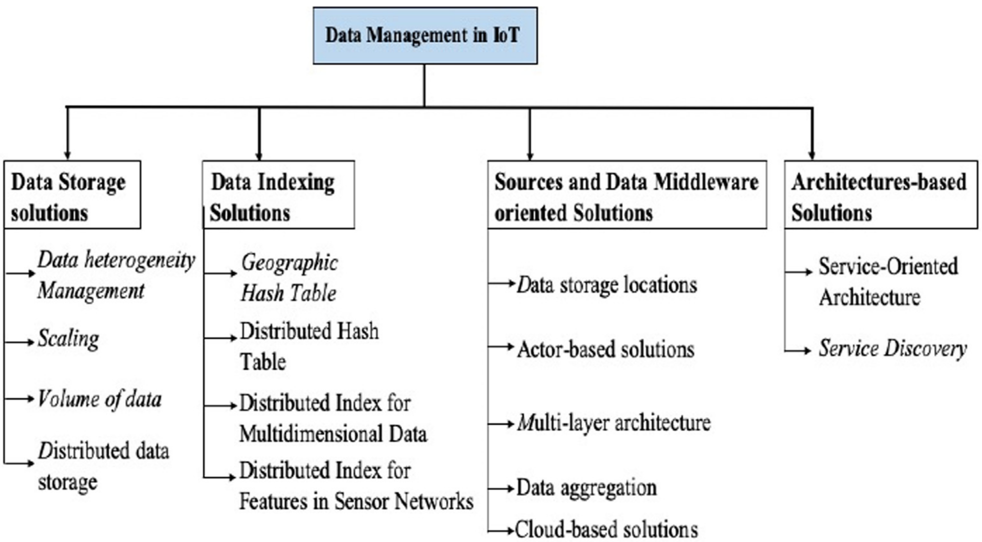

Fig. 4. Taxonomy of Data management solutions in Internet of Things.

4.1. Data storage solutions

The devices involved in IoT systems (ubiquitous sensors, RFID readers, etc.) can generate data rapidly. Furthermore, because the data can increase speedily and the volume is very large, data storage solutions for IoT data must not only be able to store massive data efficiently but also support scalability. Moreover, the IoT data collected come from several different sources and can be based on various structured and unstructured data. Data storage mechanisms must have the ability to deal with heterogeneous data resources. Thus, storage becomes a very important challenge in data management for IoT. In IoT context, if all the data generated at certain time could be stored in a permanent storage system, obsolescence would be noted and new data would not be available. This problem is solved by streaming and real-time databases which principles are widely adapted especially in sensor networks [54,55]. Only, the design of these database systems takes into account the centralized storage locations, whereas this is not favorable for the distribution and evolution of data volumes generated by IoT sources. Thus, designing a system for IoT, therefore, requires much more effective updating and access mechanisms [56,57].

To address the potential problems of enormous IoT data volume, Jiang et al. [16] propose a data storage framework that not only allows efficient storage, but also integrates both structured and unstructured data. In their solution (see Fig. 5), if data coming from IoT applications are structured, then they will be managed by a database management module that takes into account relational databases and NoSQL databases [58]. On the other hand, if data are unstructured, they will be taken into account by a file repository module.

The fundamental goal of a database management model is combining multiple databases and unifying access interfaces. The main approaches used in this model are object-entity mapping and query adapting method. Object-Entity mapping classifies objects in real world to entities in databases enabling developers to operate data in databases as they operate the objects in real world. This mapping of different databases is performed by abstracting structures of the collected data. The query adapting module is deployed with many adaptors enabling the processing of some operations such as the joining operations not supported by NoSQL databases. A file repository model which extend the Hadoop Distributed File System (HDFS) [59] allows unstructured data files management. HDFS is also wrapped to implement the features required by a file repository. This model uses an Electronic Product Code (EPC) of devices and a generation timestamp to identify the data generated by a device in a given time. Every IoT data file is named with a string considering the EPC. The timestamp and the model store the timestamp and EPC in the corresponding file's name.

Li et al. proposes a centralized data storage management system for massive and heterogeneous IoT data [17]. This solution uses an architecture called IOTMDB and is based on NoSQL. NoSQL systems are successfully applied to solve the storage and management problems of massive IoT data. The data provided by applications is collected and sent to IOTMDB system (see Fig. 7). In this architecture, data is received by a data reception node and it is stored in a slave node. They also propose storage strategies on the expression and organization of IoT data uniformly. In their approach, data of various application objects are represented in the IOTMDB system in the form of a collection of *SampleRecords*. SampleRecord being basic unit of the data stored in IOTMDB and is composed of a set of *SampleElement*. SampleElement is the smallest unit of IOTMDB and is defined as follows:

$$< sampleElement > = < key, value >$$

where the *key* belongs to String equals to the name of the *value* and the value belongs to String and Number is the actual sample value.

Fig. 6 shows the data expression in IOTMDB and illustrates that there is static and dynamic information parts. Static information is the invariable information about the object, such as object ID, the field the object belongas to, etc. Dynamic infor-

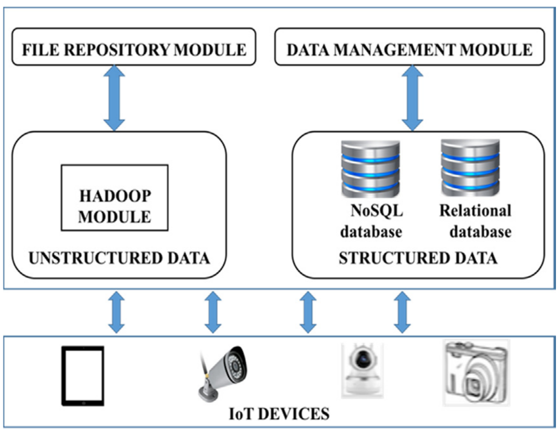

Fig. 5. Internal architecture of data management in the framework, without layers for access and services management.

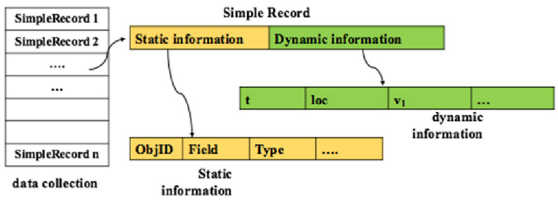

Fig. 6. Data expression in IOTMDB.

mation is sampling values which describe the state change of the object, such as the sampling time, location, speed, or ozone concentration, etc. Comparing a collection to a table in a RDBMS system, a *SampleRecord* can accordingly be seen as a row in the table. It is defined as follows:

$$SampleRecord = ((objID,\ field,\ type), (t,\ loc,\ (v_i)_{i=1}^{n})$$

where *objID* belonging to String is the ID of the object; *field* belonging to String is field of the object belonging to, such as agriculture and logistics, *t* belonging to Instant is the time when the data was sampled; *loc* belonging to Point is the location where the object was when it was sampled, v belonging to String and Number is the value of the lightweight data.

The authors of [60] propose the use of a custom data management infrastructure that can be adapted to the various IoT requirements. This infrastructure supports robustness and reliability by managing node failure and unreliable communication through the use of distributed data storage inspired by the RAID system as a data storage layer that resides between a device and the customizable Data Base Management System (DBMS).

The work of [18] proposes to dispense to a conventional vision of captured data in a form of stream filtered and continuously aggregated towards a vision centered on the storage. In this case, sensors are equipped with flash memories and embedded systems for storing the data locally on the nodes thus contributing to reduce the communication costs when the processing of requests is performed on a node. A database management system, called StoneDB, takes data from locally stored sensors to create a database that can support archived queries and even data mining tasks. The transmission is taken into account only to send requests to the database and when the results are returned to the initiator of the request.

4.1.1. Lessons learned

It is noted that the speed of growing massive data generated by connected objects arises storage problems that cannot be managed by traditional data management systems. These data stores in scalable form while respecting the requirements of a predefined scheme. This is not the case in IoT where data sources are numerous and data are available in different forms and formats. Thus, IoT data storage requires platforms and storage mechanisms that can take into account this data diversity. Researchers have proposed storage-oriented management solutions that are struggling to scale up and are often centralized. This centralization occurs high response times and does not favor the least transmission delays. These are solutions that do not offer the replication of data to cope with possible failures but also to make them more available to meet the simultaneous demands of several user applications. Moreover, due to the increasing requirement in computational algorithms and knowledge from the applications, and the intrinsic resources scarcity of real world active small objects, such as limited storage, processing and energy capacity, the data storage solutions for IoT should be aware. Among other problems, reliability and validity of information, and performance should carefully addressed.

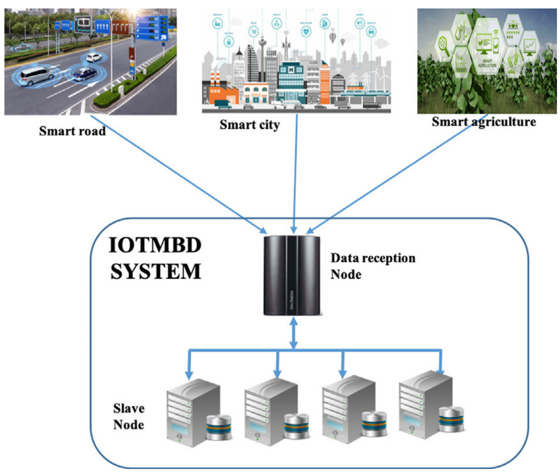

Fig. 7. Illustration of the IOTMBD architecture.

4.2. Data indexing solutions

To perform efficient and fast research and discovery of data and services in an IoT network, continuous scanning of all the connected devices appears to be inefficient and computationally intensive. In fact, efficient, scalable, and effective distributed indexing solutions are required to support the discovery and access of data in large scale IoT networks [61,62].

Many available frameworks and IoT indexing solutions rely on the use of predefined resource links or centralized data locations. However, having multiple indices used for different extraction devices is essential for dynamic IoT applications. Thus, to achieve this dynamic and distributed indexing, the models or mechanisms include sliding window indexing, multi-granule indexing, wave indices, and temporal index model [63].

In [18], dynamic indexing is proposed for frequently polled data. Intelligent aging mechanisms are used to manipulate the limited storage space on sensors, putting access on the least useful or least accessed data. In addition, aggregation and summaries of data can be used to further optimize storage capacity. Gateways that perform intelligent space-time caching and indexing result summaries are used to logically unify query views from heterogeneous sensor platforms. Distributed indexing techniques of IoT resources can also be adopted. To achieve this goal the Geographic Hash Table (GHT) is used [19]. The core step in GHT is the hashing of a *key k* into geographic coordinates where *k* is an event type name (e.g., high temperature). In GHT a *key-value* pair (*value* is the data location) is stored at a node in the vicinity of the location to which its key hashes. This allows GHT to group nodes with the same type of information together, although these nodes may be distant. To assure both persistence and consistency when nodes fail or move, GHT uses a novel perimeter refresh protocol. This protocol replicates stored data for key k at nodes around the location to which k hashes, and ensures that one node is chosen consistently as the home node for that k, so that all storage requests and queries for k can be routed to that node. Greenstein et al. [64] proposes a solution that extends GHT by introducing a Distributed Index for Features in Sensor Networks (DIFS) in order to support interval requests. In their approaches, indexes are constructed as a tree in which each node stores information on a number of values in a geographic area. Thus, non-root nodes in the structure can have multiple parents. However, the index (DIFS) is likely to have a distance sensitivity problem if some of the parent nodes of a child node in the tree are located far away in different geographical areas. Thus, a Distributed Index for Multidimensional Data is described in [65]. The work is based on dividing the entire sensor field into partitions (i.e., areas) and preserving the locality of the data by hashing the multi-attribute events into geographic areas. This allows the construction of a multidimensional search tree in which each geographic area is represented by multi-attribute events to support the routing of multidimensional queries.

Another distributed indexing solution is proposed in [66]. The indexing mechanism assumes that there is a one-to-one relationship between nodes (resources and services) and their indexes. IoT also needs mechanisms for discovering and querying resources and their data. Thus, different IoT discovery services have been listed in [67]. Most of these discovery service solutions are centralized [68] and others offer limited functionality. The use of Distributed Hash Table [69] is a solution. It makes it possible to locate the distributed data efficiently.

Fathy et al. proposed a novel distributed and efficient indexing mechanism in a hierarchical distributed network to allow discovery of IoT data [70]. In their architecture, Query Processing (QP) module receives user queries and forwards them to Discovery Service (DSs) module as shown in Fig. 8. DSs components are responsible for routing and locating a set of related GateWays (GWs) that might have a resource providing a response to the requested query, or DS can search historical data that are archived in IRs. Indexing IoT resources is performed on a set of distributed gateways. Each gateway represents a cluster (and its centroid) and has a direct access to the IoT resources that belong to this cluster. Each gateway represents a cluster (and its centroid) and has a direct access to the IoT resources that belong to this cluster [71].

In their semi-distributed approach, there are aggregated representations of the connected resources to the lower layer (i.e. gateways) at the upper layers (i.e. DSs) with less overhead. The resource attributes and values are accessed at the resource level, and the gateway has only referenced to its connected resources. However, a set of different types of the resources connected to each gateway is sent along with its cluster centroid to the higher-level DS that is linked to that gateway. The "Fully distributed" approach is an extension of the semi-distributed where a tree structure per cluster which has n children is built, where n is the number of types per cluster. While each gateway has a tree structure of its types if DS receive a user query, the gateway is not going to search sequentially to find a connected resource with the requested attributes, but it will only access the resources with a given type (temperature). This mechanism takes into count the change in device locations and availability. To face this, the membership of a new resource in a cluster is predicted by selecting a gateway whose distance from its centroid to the location of the new data item (resource) is minimized [17].

4.2.1. Lessons learned

To facilitate the discovery and search data indexing mechanisms are proposed. Dynamic indexing solutions based on geographic hash functions are first introduced. Unfortunately, they only support binary events (i.e., an event occurs or not) and exact queries. After, other distributed indexing techniques based on the features of network objects are also studied. However, advanced routing algorithms are resource-intensive and not scalable with large-scale networks. Thus, efficient indexing mechanisms based on distributed hash tables are cited in this study. However, they support exact matching only for a given key. This poses scalability problems and efficient index updates in dynamic IoT networks. Only one can see that it is necessary that these solutions be improved.

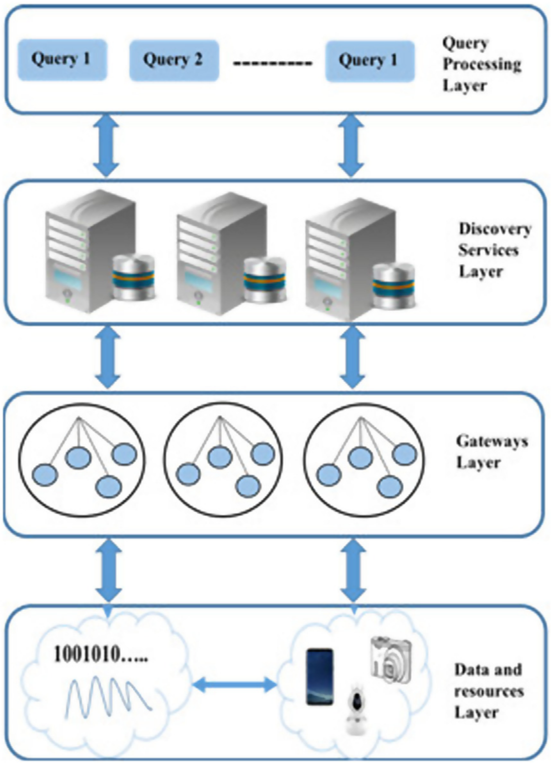

Fig. 8. Architecture of Distributed indexing (Indexing is done in the gateway layer).

4.3. Sources and data middleware oriented solutions

Internet of Things lacks a unified standardization to allow communication and integration between heterogeneous sensory devices, WSNs, and other mobile communication networks [72]. Integration of ubiquitous things (sensory devices and actuation services) into the Web can be direct where the objects must be IP-enabled or indirect in which a proxy (i.e. smart gateway or middleware) is used to provide uniform access to the Web [73,74]. Middleware hides the complexity and heterogeneity of underlying networks by providing common interfaces and protocol compatibility that allow network-enabled devices to share and publish their data and services seamlessly on a global network [75]. A middleware facilitates application development and deployment on different platforms. There is often a trade-off between the generality of middleware and domain-specific applications that can be built on top of them [76–78]. However, a separation between the generic functionality of middleware and unique features of each application is required.

In IoT, some middleware are either complex or domain specific [79–81]. Others are inflexible; they focus on centralized approach, in which collected sensor data is stored in a centralized database [82]. Moreover, conventional middleware are typically complex; they include communication, energy, and memory overhead [76,83]. The idea is to have a data-centric middleware between the network of communication-centric objects and storage-centric data storage locations. This middleware will give more possibilities for the discovery and access to data sources but also the analysis of heterogeneous data distributed on several sites [84]. This middleware will also facilitate the large-scale routing of large amounts of data flowing between the object network and the data storage locations.

Abu-Elkheir et al. [4] proposed a framework for IoT data management that is more compatible with the IoT data lifecycle and addresses the design principles of an IoT data management system. Their framework is a middleware-oriented approach centered on data and sources. This system uses the concepts of federated database management systems to ensure the independence of independent IoT subsystems. The system has a multi-layered architecture that is able to respond to real-time and archival query, analysis and service requirements. It is also a flexible architecture since it ensures arrivals and departures of objects in the system.

Aberer et al. proposed GSN (Global Sensor Networks), a service-based IoT middleware, that aims to provide a uniform platform for flexible integration, sharing and deployment of heterogeneous IoT devices. The central concept is the virtual sensor abstraction, which enables users/developers to declaratively specify XML-based deployment descriptors to deploy a sensor. This make possible to mask heterogeneity of the network and facilitate the data aggregation [15]. The central element of their architecture is the use of a virtual sensor, called Sink Node. This architecture of GSN follows the same container architecture as in J2EE where each container can host multiple virtual sensors and the container provides functionalities for lifecycle management of the sensors which includes persistency, security, notification, resource pooling and event processing. The input to the virtual sensor is one or more data streams which are processed according to the XML specification. These include the sampling rate of the data, the type and location of the data stream, the persistency of the data, the output structure of the data, and the SQL processing logic for the data stream. Furthermore, GSN architecture provides an SQL-based database that stores all the raw sensor data if the permanent storage attribute of the virtual sensor is specified as "true" in the XML specification. In addition, each virtual sensor contains a key-value pair which can be registered and discovered in GSN.

4.3.1. Lesson learned

In this section one can see the need to have a middleware in an IoT system. It should also be noted that the middleware-based approaches proposed for the development of innovative IoT applications and services that were not imagined so far are satisfactory. This section has also enabled to provide a more in-depth analysis of the capabilities of an IoT middleware in terms of the abstraction provided by the middleware for connecting and accessing the physical devices, and the services offered for flexible composition of IoT devices and services. Only, these solutions oriented-middleware are not facilitating the ease of connecting IoT devices and interpreting collected data from them. This is compounded by the fact that each IoT middleware advocates programming abstraction and architecture for accessing and connecting to IoT devices. Furthermore, these proposed approaches have a problem of centralization of query processing and the ability to support all types of applications. Scalability arises for solutions proposing a federation system of a set of data sources and the development of an IoT middleware that will be available in the cloud as well as on the edge to better control latency and support diverse applications.

4.4. Architectures-based solutions

The infrastructure of IoT is becoming increasingly important and it is now possible to integrate it into the real-world environment. For example, its integration into enterprise systems will surely change the way of designing, deploying, and using services [85]. For the integration in enterprise and real-world services, several efforts have been explored. To ensure interoperability around systems, some works have opted for the application of the Service-Oriented Architecture (SOA) concept, in particular the Web services standards on devices. This implementation in devices has many advantages in terms of end-to-end integration and programming, thus reducing the need for gateways and translation between components. This will allow direct orchestration of services running on devices. Since real-life services are in very dynamic environments where devices and associated services are degrading and reappearing constantly, there is a need to dynamically and automatically discover devices and services as well as manage them effectively.

In [86], Spiess et al. proposed an architecture that effectively integrates an IoT infrastructure with enterprise services. They contributed to a service discovery process for real-world services initiated in [32]. Their process known as Real-World Service Discovery and Previsioning Process (RSDPP), is a tool that developers can use to discover real-life services to include in their composite applications. Since most services are offered by embedded services with limited processing capabilities, there is a need for a service-oriented mechanism that does not generate too much overhead. Thus, to ensure a minimal overhead to provide the functionality of embedded devices as a service, the authors use Device Profile for Web Services (DPWS) [87] and its dynamic network discovery mechanism. Opportunities are also offered that external sources can be used to better formulate queries. These can be simple keyword searches and taking into account the user's context where a requirement to have support for a contextual and dynamic search. Thus, faced with this requirement, their system includes the context of development and aligns it with the context extracted from real-life services.

Guinard et al. [88] provides an appropriate process and system architecture that enables developers and business process designers to query, select, and use real-world service execution instances (i.e., services performed on physical devices) dynamically or even deploy new ones on demand, all in the context of real-world composite business applications [84,15].

4.4.1. Lessons learned

Since IoT systems often interact with other processes or systems, the proposal for architectures that facilitate interaction and data exchange becomes very useful and important. This section presented integration solutions that adapt to the user context and take into account various types of requests. This is made possible with the use of Service-Oriented Architecture. Since the environments with which IoT applications interact evolves continuously, new services or devices may come online anytime with unavailability of existing devices. It is therefore essential to be able to discover or query for compatible services at the right time and in the right place, both at design time as well as at runtime. Thus, it would also be possible to provide semantic service discovery systems that goes beyond discovery of IP addresses of the IoT devices.

5. Discussion and open issues

Internet of Things has experienced a rapid growth through the number of developed applications. In recent years, the most relevant research in IoT focuses on how to manage the data produced by the various objects. Data management is a very important concept for IoT applications and building data management techniques that take into account the specifics characteristics of this kind of application becomes a big challenge. Indeed, traditional database management systems are not generally well suitable because the very distinctive characteristics of data. While efforts are being made to address IoT data management issues, there are still many challenges to overcome for better solutions.

Table 2 summarizes different data management proposals for IoT applications; their advantages and limitations; the type of evaluation (experiments, simulation, or in a real environment). Moreover, this table highlights the classification of each proposal into specific categories. When some storage solutions take into account the heterogeneity of data (structured or unstructured), others do not take it into account. This table also shows that indexing solutions allow efficient and effective research. In the category of indexing approach simulation is adopted, no solution was done in a real environment. It is shown also in this table the effectiveness of middleware-based approaches in the processing of temporal and real-time queries by presenting unified access interfaces.

After the detailed analysis of the most recent and relevant data management proposals for IoT, the following open issues can be identified and suggested:

- Most of the indexing mechanisms presented in this study effectively address the issues of indexing and discovering IoT objects. Some of them rely on using pre-defined resource link or centralized data repositories which make them not scalable. Others distributed solutions based on Distributed Hash Table or Graphic Hash Table are proposed, however, scalability and efficient update of their indexes in dynamic IoT networks are their real drawbacks. Another problem with these solutions is that they only support the processing of exact queries. Thus, they can be revised and improved to provide a dynamic in-network distributed mechanism with an efficient indexing model that takes into account the resolution of both exact queries and distributed queries.
- Regarding Table 2, there is approximately no proposal based on real-time database techniques, which should be useful for IoT applications.
- Some management solutions used in large enterprises traditionally adopt highly coupled SOAs with all difficulties of doing tests and resulting in high maintenance costs. The emergence of microservices can be an alternative solution to SOAs, which orchestrates highly scalable and loosely coupled cloud services.
- IoT is generally characterized by small and smart real world, digital and virtual objects, widely distributed on a given area, with very limited resources, especially those of storage, processing, and energy. While Cloud computing has virtually unlimited capabilities in terms of storage and processing power. The design of a cloud-based framework that takes into account the various particular characteristics of IoT applications can well meet the requirements to a generalized database management system to handle various types of applications and user needs. Another important open research issue is the database security for IoT applications. In fact, object data is always shared across the network, ensuring security and privacy becomes important challenges. First, many IoT applications are provided to bring comfort and facilitate the life and human being, so it is question about data from the daily life of human beings. Generally, the choice of the appropriate cryptographic methods depends on the processing capability of interconnected smart objects, then it can be difficult to have unified solution for all Internet of Things. Therefore, the design of a security mechanism adaptable to various IoT applications could be interesting. Second, most of traditional security protocols are designed for fixed topologies. Whereas with applications based on mobile smart objects, the mobility of the devices influences the objects network topology, therefore leads to many problems in secure routing protocols.
- Several centralized storage solutions were studied in this work. Since the volume of IoT data is generally very large and can increase rapidly, a data storage approach for IoT data must not only be able to store massive data efficiently but also support horizontal scaling. This approach must also avoid failure, such as bottlenecks. So, the proposal of a distributed storage mechanism would be a good idea. Moreover, IoT data considered various unstructured and structured data, data storage components are expected to have the ability to deal with heterogeneous data resources. Most of the proposals studied in this research work interrogates data through SQL or NoSQL queries. Some have even proposed unified access interface allowing the use of both languages. Modules are developed with few methods for managing queries from applications. Since NoSQL does not support certain types of queries, such as complex joins, the proposal of mechanisms that take into account this type of query would be a challenge in a distributed and heterogeneous environment integrating structured and unstructured data. The possibility of integrating the various data offered by the proposed frameworks. Since no data representation schema is defined beforehand, the architecture of Web services such as Service-Oriented Architecture (SOA) is used to facilitate the processing of user queries.
- The studied proposals did not evaluate the impact of the dynamicity of the network on the performance of data and resources indexing and discovery. Thus, the proposal of a fully distributed framework for efficient, flexible and scalable management of IoT data is required, ensuring node mobility and unreliable node connection or disconnection.
- The integration of Cloud computing and IoT has provided good results for data-centric storage in datacenters. Generally, IoT applications require relatively low throughput per node and capacity is not a major concern. The need to connect a very large number of devices to the Internet at low cost, with limited hardware capabilities and energy resources make latency, energy efficiency, cost, and the reliability more desired features. Thus, Fog Computing has been proposed with challenges on how to place data or applications on network devices constituting the fog layer.
- The vast amount of IoT data provided makes that associated applications are considered as Big Data Applications (BDA). These applications are designed by considering efficiency objectives in terms of storage, in-network data movement, energy consumption, privacy, security and real-time knowledge availability. The approach used to help maximize value

Table 2
Comparison of the most relevant solution types available in the literature and their advantages and limits.

Papers	Proposal approach	Advantages and basic characteristics	Limits	Queries expression	Types of evaluation
[15]	Middleware	• Flexible architecture • Heterogeneity and distributed query • Integration data from multiple sensors • Temporal processing	• Virtual node problem	SQL	Experiment
[4]	Middleware-oriented framework	• Real-time and offline processing modes • Comprehensive IoT data management solution • Federated, data and source-centric approach	• Security and confidentiality • Heterogeneous data integrity	–	Experiment
[16]	Storage framework	• integration structured and unstructured data • unified access interface to different databases • Hadoop extension • distributed file repository	• little adaptors for NoSQL DBs • few features of IoT data	SQL	Experiment
[17]	Storage system	• based system on NoSQL • storage of massive and heterogeneous data • ontology for data sharing • query syntaxes-based NoSQL • two-tier B + tree index and temporal-spatial queries	• Data/access balancing techniques • Data compression techniques • The maximize value of massive IoT data	NoSQL	Experiment
[19]	Indexing System	• Geographic Hash Table ***(key, value)*** • Data Centred Storage method • Perimeter Refresh Protocol: data replication • Geographic Hierarchy: load distribution through GPSR geographic routing algorithm	• non-uniform data distribution • possibility that used keys point outside the network prior knowledge of node positions	–	Analytic verification and simulation
[64]	Distributed indexing system	• research on temporal data • failures management of index nodes • event and interval searches • load balancing communication • Simple Quad Tree Approach Geographically Bounded Hash	• indexing on several attributes of data • non-scalability system if the distribution changes continuously • no degree tolerance for transactions	–	Simulation
[70]	Distributed indexing system	• effective discover and indexing of real-life data • distributed indexing in hierarchical networks • machine learning algorithm • supports exact search queries • thematic and spatial attributes	• degradation of performance in a very dynamic network • No support of approximate queries	–	Experiment
[89]	Distributed storage system	• Scalable structured and semi-structured data Management • flexible, high-performance solution • data indexed using row and column names	• no support of a fully relational database	–	Experiment
[86]	Service oriented Architecture	• integration of IoT in enterprise services • event-based approach internal communication is based on web services	• scalability	–	Implementation and test

creation for enterprise in terms of operating cost for these Big Data systems is to give priority for devices and systems that near to data sources. This will allow the maximum data collection, filtration and processing before data arrive in cloud computing systems. Thus, the operational costs of data storage and the use of cloud services will be minimized. To achieve these objectives, the orchestration of BDA using concentric computing systems that provide computational and storage support via different devices and systems would be favorable. These concentric computing systems will also ensure real-time or near real-time intelligence at the endpoints, IoT devices, and other data sources in IIoT systems.

6. IoT data requirements for processing and storage

Today, with the accelerated growth of IoT equipment, large amounts of information are generated (Big Data concept). The evolution of these types of equipment and different data sources has brought a great diversity of types of data that are in different formats (structured, unstructured, semi-structured and mixed data). When IoT devices interact, the amount of data generated becomes so large. The speed of generation of IoT data compared to traditional transaction processing is totally different because the devices capture data continuously. This speed of production of IoT data increases exponentially. These high data volumes, combined with increasing data production speed, as well as an increased variety of data, generate a large amount of raw data that requires a very favorable analytical treatment to create added value. This also imposes requirements for data storage and processing for dynamic adaptation to the data format.

Although, the need to quickly process these increasingly diverse, complex and less structured data and to store them efficiently becomes a very strong requirement for existing IoT data management solutions, it would be fundamentally necessary to upgrade or propose distributed architectures with massively parallel processing that will meet these requirements. In this case, two types of costs must be considered, the cost of communication and the cost of treatment. Since the former is often much higher than the latter, these proposals will have to minimize communication costs while satisfying the storage and application data requirements. They will also need to consider bandwidth and latency which are two major network features that may affect the cost of communication.

A survey has been done in these aspects in order to have an understanding of the contributions of the existing literature. This thorough analysis has given an in-depth insight of the existing developments of IoT data management techniques, which further helps to extract the open gaps and limitations in these existing literatures to put forward several open solutions research areas. Some research has relied on the integration of IoT and Cloud Computing to provide resource management and heavy computing potential for applications. However, the major constraint noted is that IoT's performance will depend exclusively on the capacity of data centers in the cloud to satisfy the requirements of connected devices and their applications in real time. Some applications, such as healthcare monitoring systems, smart cities, smart grids, ad-hoc vehicle networks (VANETs), etc., require real-time computing and minimal response time. In this context, the use of fog computing will solve the problem of quality of service (QoS) management and dynamic adaptability. In addition, Table 3 summarizes some requirements for IoT data. Solutions using future technologies are proposed to deal with the detected problems.

7. Conclusion

The development of RFID, sensors, actuators, mobile, and embedded systems has been allowed the rapid grow of Internet of Things (IoT). This has made easy the development of several applications around it, especially in automotive industry with the concept of IIoT, where objects are connected and generate a huge amount of heterogeneous data. So, data management of these applications becomes a key research topic.

The data management systems used in traditional databases are not generally suitable for IoT data given their specificities. These data are different in their data structures, volume, accessing methods, and some others aspects. Then, the research community has provided new IoT data management methods, such as middleware or architecture-oriented solutions to facilitate the integration of generated data, efficient storage and indexing methods of structured and unstructured data, as well as supports for NoSQL languages. These methods can be categorized around three main principles which are data collection, data management system design, and data processing.

Table 3
Summary of open research challenges in data management and their perceived solutions.

IoT data Requirements	Issues	Causes	Solutions
Storage	- Heterogeneous data sources - Data offloading	- low storage capacity of IoT devices - Massive data production	- Storage in network devices - Using edge computing - Using fog computing
Scalability	- Node connexion/disconnexionIntegration of IoT in enterprise services	- Energy constraints - Devices mobility	- Using microservices - Fault management algorithm
Efficiency	- High energy utilization - Resources-constraints - Device-overloading	- Massive and continuous data generation and device operations - On-device data management	- Enabling energy and computation-efficient algorithms for IoT data management in IoT

This work has deeply reviewed the various solutions for managing IoT data available in the literature. To this end, after a detailed presentation of the various IoT verticals, the basic concepts of data management for IoT were identified, and an analysis of the oldest to the most recent and relevant proposals in this topic has been performed. This can provide a good foundation for researchers and practitioners who are interested to gain an insight into the data management techniques for IoT. To conclude, some open issues regarding the direction of future contributions were identified and lessons learned were well detailed. This study has also shown that proposed storage approaches have many advantages but have some limitations in managing the quality of service (response time) and storage due to bandwidth or storage problems.

In future works, one intends to propose mechanisms using future technologies such as Fog computing to improve data management. To do this, strategies for placing data in network devices can be proposed.

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.

Acknowledgements

This study was supported by National Funding from the FCT – Fundação para a Ciência e a Tecnologia through the UID/EEA/50008/2019 Project; by the Government of the Russian Federation, Grant 08-08; by Brazilian National Council for Scientific and Technological Development (CNPq) via Grant No. 309335/2017-5; and by the Non Linear Analysis, Geometry and Applications Project sponsored Simons Foundation.

References

[1] E. Biddlecombe, UN Predicts "Internet of Things", (2009).
[2] R. Parashar, A. Khan, A survey: the internet of things, 4 (2016) 7.
[3] Michael Chui, Markus Löffler, Roger Roberts, The Internet of Things, Httptcindymediaorgfilesbigbrotherinternet-Spy-Allpdf. (n.d.).
[4] M. Abu-Elkheir, M. Hayajneh, N. Ali, Data management for the internet of things: design primitives and solution, Sensors 13 (2013) 15582–15612, https://doi.org/10.3390/s131115582.
[5] R. Aggarwal, M.L. Das, RFID security in the context of internet of things, in: Proc. First Int. Conf. Secur. Internet Things, ACM, 2012, pp. 51–56.
[6] E. Sisinni, A. Saifullah, S. Han, U. Jennehag, M. Gidlund, Industrial internet of things: challenges, opportunities, and directions, IEEE Trans. Ind. Inform. 14 (2018) 4724–4734.
[7] C.H. Cella, G.W. Duffy, J.P. McGuckin, Methods and systems for the industrial internet of things, 2019.
[8] S. Haller, S. Karnouskos, C. Schroth, The internet of things in an enterprise context, in: Future Internet Symp., Springer, 2008, pp. 14–28.
[9] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, M. Ayyash, Internet of Things: a survey on enabling technologies, protocols, and applications, IEEE Commun. Surv. Tutor. 17 (2015) 2347–2376, https://doi.org/10.1109/COMST.2015.2444095.
[10] H.-N. Dai, H. Wang, G. Xu, J. Wan, M. Imran, Big data analytics for manufacturing internet of things: opportunities, challenges and enabling technologies, Enterp. Inf. Syst. (2019) 1–25.
[11] Thames Water: A Strategic Move into Predictive Planning, Wipro Digit. (n.d.). https://wiprodigital.com/cases/progressive-metering-a-utilitys-strategic-move-into-predictive-planning/ (accessed July 27, 2019).
[12] L.D. Xu, W. He, S. Li, Internet of Things in industries: a survey, IEEE Trans. Ind. Inform. 10 (2014) 2233–2243, https://doi.org/10.1109/TII.2014.2300753.
[13] G. Pujolle, An autonomic-oriented architecture for the internet of things, in: Mod. Comput. 2006 JVA06 IEEE John Vincent Atanasoff 2006 Int. Symp. On, IEEE, 2006: pp. 163–168.
[14] Z. Ding, J. Xu, Q. Yang, SeaCloudDM: a database cluster framework for managing and querying massive heterogeneous sensor sampling data, J. Supercomput. 66 (2013) 1260–1284, https://doi.org/10.1007/s11227-012-0762-1.
[15] K. Aberer, M. Hauswirth, A. Salehi, Infrastructure for data processing in large-scale interconnected sensor networks, in: 2007 Int. Conf. Mob. Data Manag., IEEE, 2007: pp. 198–205. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4417143 (accessed November 30, 2016).
[16] Lihong Jiang, Li Da Xu, Hongming Cai, Zuhai Jiang, Fenglin Bu, Boyi Xu, An IoT-Oriented Data Storage Framework in Cloud Computing Platform, IEEE Trans. Ind. Inform. 10 (2014) 1443–1451. doi: 10.1109/TII.2014.2306384.
[17] T. Li, Y. Liu, Y. Tian, S. Shen, W. Mao, A Storage Solution for Massive IoT Data Based on NoSQL, in: IEEE, 2012: pp. 50–57. doi: 10.1109/GreenCom.2012.18.
[18] Y. Diao, D. Ganesan, G. Mathur, P.J. Shenoy, Rethinking Data Management for Storage-centric Sensor Networks., in: CIDR, 2007: pp. 22–31.
[19] S. Ratnasamy, B. Karp, L. Yin, F. Yu, D. Estrin, R. Govindan, S. Shenker, GHT: a geographic hash table for data-centric storage, in: Proc. 1st ACM Int. Workshop Wirel. Sens. Netw. Appl., ACM, 2002: pp. 78–87.
[20] T. Padiya, M. Bhise, P. Rajkotiya, Data management for Internet of Things, in: Reg. 10 Symp. TENSYMP 2015 IEEE, IEEE, 2015: pp. 62–65.
[21] S. Vongsingthong, S. Smanchat, A Review of Data Management in Internet of Things, (n.d.) 26.
[22] A. Botta, W. de Donato, V. Persico, A. Pescapé, Integration of Cloud computing and Internet of Things: a survey, Future Gener. Comput. Syst. 56 (2016) 684–700, https://doi.org/10.1016/j.future.2015.09.021.
[23] T.I. Cerlinca, C. Turcu, C. Turcu, M. Cerlinca, RFID-based information system for patients and medical staff identification and tracking, in, Sustain. Radio Freq. Identif. Solut., InTech (2010).
[24] E. Ahmed, I. Yaqoob, A. Gani, M. Imran, M. Guizani, Internet-of-things-based smart environments: state of the art, taxonomy, and open research challenges, IEEE Wirel. Commun. 23 (2016) 10–16, https://doi.org/10.1109/MWC.2016.7721736.
[25] L. Atzori, A. Iera, G. Morabito, The internet of things: a survey, Comput. Netw. 54 (2010) 2787–2805.
[26] M. Brettel, N. Friederichsen, M. Keller, M. Rosenberg, How virtualization, decentralization and network building change the manufacturing landscape: an industry 4.0 perspective, Int. J. Mech. Ind. Sci. Eng. 8 (2014) 37–44.
[27] A.K.R. Venkatapathy, A. Riesner, M. Roidl, J. Emmerich, M. ten Hompel, PhyNode: An intelligent, cyber-physical system with energy neutral operation for PhyNetLab, in: Smart SysTech 2015 Eur. Conf. Smart Objects Syst. Technol. Proc. Of, VDE, 2015: pp. 1–8.
[28] M. Wollschlaeger, T. Sauter, J. Jasperneite, The Future of Industrial Communication: Automation Networks in the Era of the Internet of Things and Industry 4.0, IEEE Ind. Electron. Mag. 11 (2017) 17–27. doi: 10.1109/MIE.2017.2649104.
[29] L. Da Xu, W. He, S. Li, Internet of things in industries: a survey, IEEE Trans. Ind. Inform. 10 (2014) 2233–2243.
[30] A. Kamilaris, V. Trifa, D. Guinard, Building web-based infrastructures for smart meters, in: Proc. First Workshop Energy Aware. Conserv. Pervasive Appl. Hels. Finl., Citeseer, 2010: p. 16.

[31] M.-W. Ryu, J. Kim, S.-S. Lee, M.-H. Song, Survey on Internet of Things, SmartCR 2 (2012) 195–202.
[32] D. Guinard, M. Fischer, V. Trifa, Sharing using social networks in a composable web of things, in: Pervasive Comput. Commun. Workshop PERCOM Workshop 2010 8th IEEE Int. Conf. On, IEEE, 2010: pp. 702–707. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5470524 (accessed April 16, 2016).
[33] S. Chatziefthymiou, S.K. Goudos, A survey of IoT: Architecture, Applications and Future Vision, LAP LAMBERT Academic Publishing, 2017.
[34] Lu. Dongxin, Tao Liu, The application of IOT in medical system, IEEE (2011) 272–275, https://doi.org/10.1109/ITiME.2011.6130831.
[35] H. Schaffers, N. Komninos, M. Pallot, B. Trousse, M. Nilsson, A. Oliveira, Smart cities and the future internet: Towards cooperation frameworks for open innovation, in: Future Internet Assem., Springer, 2011: pp. 431–446.
[36] M. Dohler, I. Vilajosana, X. Vilajosana, J. Llosa, Smart cities: An action plan, in: Proc Barc. Smart Cities Congr. Barc. Spain, 2011: pp. 1–6.
[37] A. Zanella, N. Bui, A. Castellani, L. Vangelista, M. Zorzi, Internet of things for smart cities, IEEE Internet Things J. 1 (2014) 22–32.
[38] M. Sharifnejad, M. Shari, M. Ghiasabadi, S. Beheshti, A survey on wireless sensor networks security, in: 4th Int. Conf. Sci. Electron. Technol. Inf. Telecommun. SEIT Tunis. March 2007, Citeseer, 2007.
[39] H.T.L. de Paula, João B. A. Gomes, L.F.T. Affonso, Ricardo A. L. Rabêlo, Joel J. P. C. Rodrigues, An IoT-based Water Monitoring System for Smart Buildings, Secure Dependable Softw. Defin. Netw. Sustain. Smart Communities SecSDN 2019in Conjunction IEEE Int. Conf. Commun. IEEE ICC 2019 Shanghai China. (2019).
[40] J.P. Lynch, K.J. Loh, A summary review of wireless sensors and sensor networks for structural health monitoring, Shock Vib. Dig. 38 (2006) 91–130.
[41] K. Pardini, J. Rodrigues, S. Kozlov, N. Kumar, V. Furtado, IoT-based solid waste management solutions: a survey, J. Sens. Actuator Netw. 8 (2019) 5.
[42] T. Nuortio, J. Kytöjoki, H. Niska, O. Bräysy, Improved route planning and scheduling of waste collection and transport, Expert Syst. Appl. 30 (2006) 223–232.
[43] W. Kastner, G. Neugschwandtner, S. Soucek, H.M. Newman, Communication systems for building automation and control, Proc. IEEE 93 (2005) 1178–1203.
[44] B. Harpham, How the Internet of Things is changing healthcare and transportation, CIO. (2015). https://www.cio.com/article/2981481/healthcare/how-the-internet-of-things-is-changing-healthcare-and-transportation.html (accessed September 21, 2018).
[45] X.-Y. Liu, M.-Y. Wu, Vehicular CPS: an application of IoT in vehicular networks, Jisuanji Yingyong J. Comput. Appl. 32 (2012) 900–904.
[46] Z. Ning, J. Huang, X. Wang, Mobile Edge Computing-enabled Internet of Vehicles: Toward Energy-Efficient Scheduling, (n.d.) 8.
[47] I. Benkhelifa, N. Nouali-Taboudjemat, Sensor Technologies for Disaster Management Information Systems, CERIST, 2013.
[48] S. Lee, D. Yoon, A. Ghosh, Intelligent parking lot application using wireless sensor networks., in: CTS, 2008: pp. 48–57.
[49] X. Wang, Z. Ning, L. Wang, Offloading in internet of vehicles: a fog-enabled real-time traffic management system, IEEE Trans. Ind. Inform. 14 (2018) 4568–4578, https://doi.org/10.1109/TII.2018.2816590.
[50] X. Li, W. Shu, M. Li, H.-Y. Huang, P.-E. Luo, M.-Y. Wu, Performance evaluation of vehicle-based mobile sensor networks for traffic monitoring, IEEE Trans. Veh. Technol. 58 (2009) 1647–1653.
[51] J. Sung, T.S. Lopez, D. Kim, The EPC sensor network for RFID and WSN integration infrastructure, in: Pervasive Comput. Commun. Workshop 2007 PerCom Workshop 07 Fifth Annu. IEEE Int. Conf. On, IEEE, 2007, pp. 618–621.
[52] H. Aly, M. Elmogy, S. Barakat, Big data on internet of things: applications, architecture, technologies, techniques, and future directions, Int. J. Comput. Sci. Eng. 4 (2015) 300–313.
[53] J. Cooper, A. James, Challenges for database management in the internet of things, IETE Tech. Rev. 26 (2009) 320–329.
[54] H. Kopetz, Real-time systems: design principles for distributed embedded applications, Springer Science & Business Media, 2011.
[55] S.-H. Han, Y.-K. Kim, An Architecture of Real-Time, Historical Database System for Industrial Process Control and Monitoring, in: Comput. Netw. Syst. Ind. Eng. CNSI 2011 First ACISJNU Int. Conf. On, IEEE, 2011: pp. 31–34.
[56] Q. Zheng, X. Bi, An improved concurrency control algorithm for distributed real-time database, in: Adv. Manag. Sci. ICAMS 2010 IEEE Int. Conf. On, IEEE, 2010: pp. 364–367.
[57] Z.C. Liu, D.S. Lin, Y.Z. Ning, Embedded Real-Time Database System Concurrency Control Protocol AC-Based OCC-FV, in: Appl. Mech. Mater., Trans Tech Publ, 2013: pp. 1402–1406.
[58] J. Han, E. Haihong, G. Le, J. Du, Survey on NoSQL database, in: Pervasive Comput. Appl. ICPCA 2011 6th Int. Conf. On, IEEE, 2011: pp. 363–366.
[59] K. Shvachko, H. Kuang, S. Radia, R. Chansler, The hadoop distributed file system, in: Mass Storage Syst. Technol. MSST 2010 IEEE 26th Symp. On, Ieee, 2010: pp. 1–10.
[60] N. Siegmund, M. Rosenmüller, G. Moritz, G. Saake, D. Timmermann, Towards robust data storage in wireless sensor networks, IETE Tech. Rev. 26 (2009) 335–340.
[61] A. Gani, A. Siddiqa, S. Shamshirband, F. Hanum, A survey on indexing techniques for big data: taxonomy and performance evaluation, Knowl. Inf. Syst. 46 (2016) 241–284.
[62] Y. Ma, J. Rao, W. Hu, X. Meng, X. Han, Y. Zhang, Y. Chai, C. Liu, An efficient index for massive IOT data in cloud environment, in: Proc. 21st ACM Int. Conf. Inf. Knowl. Manag., ACM, 2012: pp. 2129–2133.
[63] M. Kholghi, M. Keyvanpour, Comparative evaluation of data stream indexing models, ArXiv Prepr. ArXiv12080684. (2012).
[64] B. Greenstein, S. Ratnasamy, S. Shenker, R. Govindan, D. Estrin, DIFS: a distributed index for features in sensor networks, Ad Hoc Netw. 1 (2003) 333–349.
[65] X. Li, Y.J. Kim, R. Govindan, W. Hong, Multi-dimensional range queries in sensor networks, in: Proc. 1st Int. Conf. Embed. Networked Sens. Syst., ACM, 2003: pp. 63–75.
[66] P. Liu, N. Kong, Y. Tian, X. Lee, B. Yan, A dht-based discovery service for rfid network, in: Internet Things IThings 2014 IEEE Int. Conf. Green Comput. Commun. GreenCom IEEE Cyber Phys. Soc. Comput. CPSCom IEEE, IEEE, 2014: pp. 344–347.
[67] S. Evdokimov, B. Fabian, S. Kunz, N. Schoenemann, Comparison of discovery service architectures for the internet of things, in: Sens. Netw. Ubiquitous Trust. Comput. SUTC 2010 IEEE Int. Conf. On, IEEE, 2010: pp. 237–244.
[68] E. Polytarchos, S. Eliakis, D. Bochtis, K. Pramatari, Evaluating discovery services architectures in the context of the internet of things, in: Unique Radio Innov. 21st Century, Springer, 2011: pp. 203–227.
[69] F. Paganelli, D. Parlanti, A DHT-based discovery service for the Internet of Things, J. Comput. Netw. Commun. 2012 (2012).
[70] Y. Fathy, P. Barnaghi, S. Enshaeifar, R. Tafazolli, A distributed in-network indexing mechanism for the Internet of Things, in: 2016 IEEE 3rd World Forum Internet Things WF-IoT, IEEE, Reston, VA, USA, 2016: pp. 585–590. https://doi.org/10.1109/WF-IoT.2016.7845472.
[71] L. Ding, P. Shi, B. Liu, The clustering of internet, internet of things and social network, in: Knowl. Acquis. Model. KAM 2010 3rd Int. Symp. On, IEEE, 2010: pp. 417–420. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5646274 (accessed April 16, 2016).
[72] Q. Zhu, R. Wang, Q. Chen, Y. Liu, W. Qin, Iot gateway: Bridgingwireless sensor networks into internet of things, in: Embed. Ubiquitous Comput. EUC 2010 IEEEIFIP 8th Int. Conf. On, Ieee, 2010: pp. 347–352.
[73] D. Zeng, S. Guo, Z. Cheng, The web of things: a survey, JCM 6 (2011) 424–438.
[74] M.A.A. da Cruz, J.J.P.C. Rodrigues, J. Al-Muhtadi, V.V. Korotaev, V.H.C. de Albuquerque, A reference model for internet of things middleware, IEEE Internet Things J. 5 (2018) 871–883, https://doi.org/10.1109/JIOT.2018.2796561.
[75] Soma Bandyopadhyay, Munmun Sengupta, Souvik Maiti, Subhajit Dutta, role of middleware for internet of things: a study, Int. J. Comput. Sci. Eng. Surv. 2 (2011) 94–105, https://doi.org/10.5121/ijcses.2011.2307.
[76] Y. Yu, B. Krishnamachari, V.K. Prasanna, Issues in designing middleware for wireless sensor networks, IEEE Netw. 18 (2004) 15–21.
[77] J. Gubbi, R. Buyya, S. Marusic, M. Palaniswami, Internet of Things (IoT): a vision, architectural elements, and future directions, Future Gener. Comput. Syst. 29 (2013) 1645–1660.
[78] A. Farahzadi, P. Shams, J. Rezazadeh, R. Farahbakhsh, Middleware technologies for cloud of things: a survey, Digit. Commun. Netw. 4 (2018) 176–188.

[79] A. Kamilaris, A. Pitsillides, V. Trifa, The smart home meets the web of things, Int. J. Ad Hoc Ubiquitous Comput. 7 (2011) 145–154.

[80] M.A. da Cruz, J.J. Rodrigues, A.K. Sangaiah, J. Al-Muhtadi, V. Korotaev, Performance evaluation of IoT middleware, J. Netw. Comput. Appl. 109 (2018) 53–65.

[81] A.H.H. Ngu, M. Gutierrez, V. Metsis, S. Nepal, M.Z. Sheng, IoT middleware: a survey on issues and enabling technologies 1 1 IEEE Internet Things J. (2016), https://doi.org/10.1109/JIOT.2016.2615180.

[82] M.A. Razzaque, M. Milojevic-Jevric, A. Palade, S. Clarke, Middleware for internet of things: a survey, IEEE Internet Things J. 3 (2015) 70–95.

[83] M.A. Chaqfeh, N. Mohamed, Challenges in middleware solutions for the internet of things, in: 2012 Int. Conf. Collab. Technol. Syst. CTS, IEEE, 2012: pp. 21–26.

[84] G. Chen, M. Li, D. Kotz, Data-centric middleware for context-aware pervasive computing, Pervasive Mob. Comput. 4 (2008) 216–253.

[85] D. Uckelmann, M. Harrison, F. Michahelles, An architectural approach towards the future internet of things, Springer, 2011. http://link.springer.com/chapter/10.1007/978-3-642-19157-2_1 (accessed April 12, 2016).

[86] P. Spiess, S. Karnouskos, D. Guinard, D. Savio, O. Baecker, L.M.S. De Souza, V. Trifa, SOA-based integration of the internet of things in enterprise services, in: Web Serv. 2009 ICWS 2009 IEEE Int. Conf. On, IEEE, 2009: pp. 968–975.

[87] F. Jammes, A. Mensch, H. Smit, Service-oriented device communications using the devices profile for web services, in: Proc. 3rd Int. Workshop Middlew. Pervasive Ad-Hoc Comput., ACM, 2005: pp. 1–8.

[88] D. Guinard, V. Trifa, S. Karnouskos, P. Spiess, D. Savio, Interacting with the soa-based internet of things: discovery, query, selection, and on-demand provisioning of web services, IEEE Trans. Serv. Comput. 3 (2010) 223–235.

[89] F. Chang, J. Dean, S. Ghemawat, W.C. Hsieh, D.A. Wallach, M. Burrows, T. Chandra, A. Fikes, R.E. Gruber, Bigtable: a distributed storage system for structured data, ACM Trans. Comput. Syst. TOCS 26 (2008) 4.