Research on key technologies of data processing in internet of things

Yangqing Zhu and Peiying Liang

Department of Internet of Things Engineering in Electronic and Information Engineering School, Foshan University, Teacher room 506, Foshan City, China.
Email: zhuyangqing_830516@163.com

Abstract. The data of Internet of things (IOT) has the characteristics of polymorphism, heterogeneous, large amount and processing real-time. The traditional structured and static batch processing method has not met the requirements of data processing of IOT. This paper studied a middleware that can integrate heterogeneous data of IOT, and integrated different data formats into a unified format. Designed a data processing model of IOT based on the Storm flow calculation architecture, integrated the existing Internet security technology to build the Internet security system of IOT data processing, which provided reference for the efficient transmission and processing of IOT data.

1. Introduction

Internet of things (IOT) can be taken as an important extension of Internet, are widely connected to the physical world by building the pervasive environment perception infrastructure, in the human world, the physical world and information world construction consists of a network, information and service ecosystem. In the traditional Internet era, information was shared among people. On the basis of the Internet, IOT further realize the information sharing of human, machine and things, to form a more close information flow ecosystem between the information world and physical world, and change the mode that people use the traditional Internet for information communication and resource sharing. [1]

IOT connects the digital world and physical world, the technological changes and challenges that it brings will be difficult to estimate. While relying on hardware devices, it needs more intelligent analysis and processing. Data processing is the most important part of the core technology of IOT. The large and real-time data into the system, how to transmit and process these massive, complex and real-time data to get the effective results, through the feedback allows users to perform intelligent management and control to the IOT objects, is the key technology of IOT. [2]

This paper analyzed the characteristics of IOT data, and researched some key technologies of IOT, such as data acquisition middleware, flow data processing mode and data security framework, provided a reference path for the construction of IOT data processing system.

2. Data Characteristics of IOT

2.1. Polymorphism and Heterogeneity of Data

There are a variety of sensors in wireless sensor networks (WSN), each of sensor has different function in different application systems. These sensors have different structures and different performances, and their collected data structures are different. In the Radio Frequency Identification (RFID) system, there are multiple RFID tags and multiple reader-writers. The micro computing devices in the machine to machine system are all different. Their data structures do not follow a unified pattern. The data in the IOT include text, images, audio, video and other multimedia data. There are also static data and dynamic
data. The diversity of data and the heterogeneity of the sensing model lead to the heterogeneity of data. This heterogeneity exacerbates the difficulty of data processing and software development. [3]

2.2. The Magnanimity of Data
The IOT is often a dynamic network formed by the combination and connection of several wireless recognized objects. The quantity of goods in a medium-sized supermarket can be as many as millions or even tens of millions pieces. In a supermarket RFID system, assume that 10 million items need to be tracked, read 10 times a day, each time 100 bytes, the amount of data per day will reach 10GB, each year will reach 3650GB. In the field of real-time monitoring such as ecological monitoring, wireless sensor networks need to record multimedia information from multiple nodes, and the amount of data is amazing, reaching more than 1TB per day. In addition, in some emergency monitoring systems, the data are generated in real-time, high-speed and continuously in the form of stream, which aggravates the mass of data.

2.3. Timeliness of Data
The thing state of being perceived may change rapidly. Therefore, regardless of the WSN or RFID system, data acquisition of IOT is carried out at any time, every certain period of time to send data to the server, data update soon, historical data is only for recording the development process of transactions, although you can backup, but because of the massive not long-term preservation [4]. Only new data can reflect the existing state of the things perceived by the system, so the response speed or response time of the system is the key to the reliability and practicability of the system. This requires that the software processing system of IOT must have sufficient speed of operation, otherwise it may lead to erroneous conclusions and even cause great losses.

2.4. Data Security and Transmission Reliability
When the IOT interacts with data, it may face malicious attacks from outside users, thus reducing the probability of success of network data transmission. With the increase of malicious behaviour, IOT interactive data outage probability will increase, which will also lead to increased data retransmission, thereby increasing the burden of communication network, and may even lead to paralysis of the network. At the same time, malicious eavesdropping may also exist, which will destroy the security of data transmission. In a word, the malicious attack and wiretapping behaviours of the external users not only reduce the reliability of the data transmission of IOT, but also deteriorate the security of IOT information. [5]

3. The Key Technologies of Data Processing in IOT

3.1. Middleware Technology
Service-Oriented Architecture (SOA) provides a better solution for heterogeneous information fusion in the IOT data processing. Using the SOA architecture to build IOT applications, can better integrate multiple sources and a wide range of services. [6]

3.1.1. Service-Oriented Application Architecture Description. SOA is a component model that uses different functional services of related applications and uses the interfaces and contracts defined between these services to build the application. The interface is defined in a neutral manner, which should be independent of the hardware platform, the operating system, and the programming language that implements the service. This allows services to interact in a unified manner in a wide variety of systems. Services are the foundation of SOA, so that components can interact directly and efficiently with application systems and software agents. In the SOA, a typical business operation involves a number of different components that typically reflect the needs of the underlying business process in an event driven or asynchronous way. In the context of IOT, traditional and emerging resources are open on the Internet in the form of services. Therefore, the research of data fusion application technology based on SOA has very important application value [7]. The SOA architecture usually consists of five main parts:
- Consumers. Obtain information from producers, at the same time, processing information.
• Application. Providing application interfaces or interoperable services that can be called from one another to different programs.
• Service. A task entity that implements certain specified actions.
• Service runtime environment support. The relevant service support functions required by the SOA application, such as the data interface of devices in the IOT, etc.
• Service provider. An interface entity that provides the specified function using related components.

Figure 1. Middleware Architecture Diagram

Figure 1 mainly contains data acquisition unit and data format standard service unit. The middleware was between the perceived hardware layer and the service application layer, realized information exchange and management with front-end hardware sensing device, automatic acquired the original information of hardware perception, refined the effective information, finished application of information exchange with the upper application and provided a standard format of data calls for the application software, completely got rid of the restrictions that reader-writer device of non-standardized protocol brought development, maintenance and extension.

3.1.2. Middleware Design of IOT Based on SOA. By comparing the characteristics of traditional Internet and data in IOT, the design, abstract and integration services of service oriented architecture middleware are used to be compatible with various types of data and protocols. Therefore, this paper presents the basic framework of IOT applications based on SOA, as shown in figure 1.

3.2. Flow Computing Technology
In view of the massive and real-time characteristics of the IOT data, the traditional query, statistics and model calculation based on relational database cannot meet the requirements. With the development of Web 2.0, large data processing technology developed rapidly, such as Google GFS, MapReduce, Hadoop platform, HDFS, MapReduce and so on, but these technologies are only suitable for dealing with batch static data. In recent years, stream computing has provided a very good model for data computing in IOT. [8]

Stream computing is a real-time calculation of stream data, and can be applied to a variety of scenarios, such as Web services, Machine Translation, IOT data processing, and so on. Take the sensor monitoring as an example, the atmospheric PM2.5 concentration can be real-time monitored by placing the sensor PM2.5 in the atmosphere, the real-time monitoring data will be transmitted back to the data center, the monitoring system real-timely analyze the return data, to predict the air quality change trend, if the air quality to affect human health level in the next period of time, it start the emergency response mechanism.

Message processing is the foundation of real-time computing for real-time systems that need to process a large number of messages. The core of message processing is how to do not lose data in the
process of message processing, and can make the whole processing system have good expansibility so as to be able to deal with the bigger message flow. [9]

At present, stream computing platform can be divided into three categories, the first category is commercial grade stream computing platform, such as IBM InfoSphere Streams, IBM StreamBase, the second is the open source stream computing framework, such as Twitter Storm, Yahoo S4, the third is the company occupied stream computing framework, such as Facebook Puma, Baidu DStream, etc.

In the IOT data processing system, the open source computing platform Storm is a very good solution. [10]

3.2.1. The Design Idea of Storm. Storm main includes the components design idea of Streams, Spouts, Bolts, Stream Groupings and Topology. Streams refers to an infinite sequence of Tuple. Spouts refers to reading stream data from the outside and continuing to send Tuple. Bolts refers to encapsulation of Tuple processing procedures, including filtering, aggregation, query and other operations. The function of Stream Groupings is to transfer Tuple between two components. Topology is a conversion diagram of stream data [11]. Figure 2 shows the flow diagram of the Topology of the Storm. In Figure 2, the stream data that Spout read is grouped by the Grouping component according to certain rules, such as Shuffle Grouping, Fields Grouping, All Grouping, Global Grouping, and Direct Grouping and so on. The tuple after grouped is send to the corresponding Bolt component for processing. The processed tuple data can be sent to the next Bolt to continue processing or to feed back to the user. Some processing results can be permanently stored on distributed disks for further data mining in the future.

3.2.2. Data Processing Flow of IOT Based on Storm. Figure 3 shows the data processing flow diagram of IOT based on Storm. In Figure 3, the actual IOT devices, such as a variety of sensors, cameras, RFID and other devices, they generated data can be processed into a unified data format specification by Fusion Middleware, and then transferred to the Storm data processing framework. The data after processing by stream data processing framework, a part directly in the text, table or the graphical form can be displayed to the user, some data can be saved to the HBase, ElephantDB or other non-structured database, to provide data source for data mining in the future.
3.3. Data Security Technology

The current Internet security technology can be applied to the IOT. Figure 4 showed the data security transmission processing flow diagram of IOT. After a large number of sensors collect data, first obtain the corresponding digital summaries of these data, and then digitally sign them. The digital abstract and digital signature of the sensor data and the data itself constitute the original data packet of IOT. The original data packet of IOT can be transmitted by the network transmission environment (Internet, satellite communication network), the data need to be carried out data integrity verification and reliability verification in the data processing center. Their validation can be accomplished by digital abstraction techniques and digital signature techniques of the PKI security mechanism. The IOT sensor data that can be passed the security verification in the data center needs to be cleaned and filtered operation data, the valuable data can be preserved, and those not valuable can be filtered out. After cleaning and filtering, the data will be formed two types of data, one is the file type data, and the other is the data type of the database. The type of database data will be stored in the HBase database, and then protect data security using the security policy of database itself, the HBase database will eventually be stored in HDFS. All IOT data in the process of data migration, users can carry out certificate them with security authentication file type (PKI security) or the type of database security at any time. [12]

4. Summary

In some extent, the research of this paper can eliminate the difference of heterogeneous data sources of IOT and provide users with standard and unified data processing interface. Storm streaming data architecture can effectively deal with real-time and large amounts of data in the IOT, and is an effective solution for data processing in the IOT system, such as sensor networks and RFID system.
Comprehensive use of the current Internet security measures to solve the IOT data transmission and processing security issues, can protect the IOT data efficient transmission and security.

In the future work, can combine block chain technology to further study the data transmission and data security system of IOT.

5. Acknowledgments
This study was supported by the innovation project of Guangdong Provincial Education Department (No.2014GXJK109) and Foshan University laboratory open innovation fund project.

6. References
[1] Zuquan Weng and Qi Zhang 2014 China. Internet of Things Technology 6 88
[2] Xiaofang Wu 2017 China. Telecom World 1 1
[3] Yingxin Hu, Xinning Ma and Lijuan Zheng 2014 China. Internet of Things Technology 4 79
[4] Ye Tian, Bo Yuan and Tingli Li 2016 China. Acta Electronic Sinica 2 247
[5] Chuankun Wu 2016 China. Journal of Cryptologic Research 2 40
[6] Jiuyun Xu, Huanxing Chi, Dan Jiang, Han Xiao, Hongxia Zhang and Jiashu Guo 2017 China. Journal of Chinese Computer Systems 1 24
[7] Xingbang Lei and Jun Fang 2016 China. Computer Science 6 68
[8] Ziyu Lin 2015 Principles and Application of Big Data Technology (Beijing: Posts & Telecom Press) p 151
[9] Cheng Zhao and Jinlei Jiang 2016 China. ZTE Technology Journal 2 14
[10] Qianjin Zhang 2016 China. Journal of Changzhou Institute of Technology 6 30
[11] Minmin Chen, Xinchun Wang and Fengxian Huang 2015 Storm technology insider and big data practices (Beijing: Posts & Telecom Press) p 33
[12] Dan Wang, Wenbing Zhao and Zhiming Ding 2017 China. Journal of Beijing University of Technology 3 335