Challenges of Big Data based Cyber-Physical System

Wei Zhong[1, a], Lichen Zhang[2, b]

School of Guangdong University of Technology, Guangzhou 510000, China.

[a]1240500584@qq.com

Keywords: CPS,Big Data ,Modeling.

Abstract. With the development of science and technology, What we need in Cyber-Physical System （CPS） have became more and more complicated。Because the system seem more vast 、more sensors,and need to be more intelligent , the data and information be collectted, stored, processed also need even more. So big data were used to rapidly processing and storage for huge amounts of data in the CPS.However, there are many differences between big data and CPS. The combination of these two systems have different with general systems。Such as time and space model , data flow, modeling and so on.The paper will systematically elaborate the challenges and of the CPS besed on big data be faced and illustrate the characteristics of its modeling demand.

1. Introduction

Definition of big data has actually varies over time.today, Widely from capacity, speed and so-called '3 v (volume, velocity and variety)'.Capacity refers to the big data can handle, storage of great capacity of data, speed is refers to the big data can handle data faster, diversity refers to the big data can store、handle many different kinds of data.According to the different of specific requirements, and for it to add a v (value, variability, or virtual). That is to say, the information processed by big data must be valuable. There is a more complete and detail definition of big data in 2012: “Big data is huge, high speed and high diversity of information assets, the information assets need to be more efficient processing, in order to get the value to make corresponding decision, discover new information or the optimization process ". In fact, if a vast amounts of data collection combined with the existing technology can achieve data capture, analysis and management that can say that it is the big data.The application of big data efficiency would be profound.According to McCarthy institute concluded report, big data will give the medical industry in the United States, the European Union's public security management industry, retail industry in the United States, global manufacturing bring great impact. main application areas of big data are: business, social management, the study of science.

CPS is also the current academic research hot spot. we know the Internet of things is a typical CPS.Small to intelligent home appliances、robot, big to car networking, flight Cyber-Physical System(FCPS), and Internet of things,the CPS brings us a new technological revolution. CPS integrate computing, communication and sensing module, on the surrounding environment of sensing, data transmission, information processing, real-time control, and can through the man-machine interface with reliable, real-time and security cooperation.CPS fusion environmental sensing, embedded computing, network control and network communication at an organic whole, makes the equipment hava function of communication, computing, collaboration, accurate control and autonomy, etc.Now the CPS have a wide range of applications in smart city, smart grid, smart home and network appliances, robot assisted living, environmental monitoring, traffic system, and many other fields, the CPS technology affect the information interaction mode between the physical world

And information world.CPS based on big data did not meet the demand at present, with the development of big data, Real-time information storage, processing, analysis, as demand increases, making the system based on information fusion physics of big data has become a major new trend, making the CPS based on big data has become a hot research field.

2. The challenge of CPS based on big data

CPS based on big data consists of two main parts, one is big data, one is CPS.Since these two areas at the forefront of social development. The presence of their respective itself many research problems unresolved. Because of their unique， the combination of the two different attributes, resulting in the analysis and design, modeling, application of this system, there are some technical challenges. In the next, the paper will elaborate the challenge of CPS based on big data on these three directions.

1) *Big data integration:*The extensive existence of data make the data more and more scattered in different data management system, To facilitate data analysis required for the integration of data. Data integration look is not a new problem, but the data integration of the era of big data need a new demand, so it is faced with new challenges.
2) *Analysis of big data (analytics):*Mainly for structured data in the traditional sense of the data analysis, and has formed a whole set of effective system of analysis. First of all use database to store structured data, on the basis of building data warehouse, on-line analytical processing according to need to build a data cube (online analytical processing, OLAP), can be multiple dimensions of drill (drill down) or roll (roll - up) operation. For derived from the data at a deeper level of knowledge needs to promote the generation of data mining technology, and clustering, correlation analysis and a series of effective methods in practice.
3) *Big data privacy issues:*Privacy has been a problem long ago, the advent of the computer makes more and more data in the form of digital stored in a computer, the development of the Internet makes the data easier to produce and transmission, data privacy is more and more serious.
4) *Energy consumption of big data problem:*In energy prices, has been expanding its data center storage, high energy consumption has gradually become a major bottleneck restricting the development of big data rapidly. From small to large-scale data center cluster are faced with lower energy consumption, but has not caused enough attention, the relevant research results less. In the big data management system, the energy consumption is mainly composed of two parts: hardware and software for energy consumption of energy consumption, and both give priority to with hardware consumption. Ideally, the energy consumption of big data management system should be and is directly proportional to the system utilization. But the truth is not as expected, when the system utilization of 0 still have energy consumption.
5) *Big data processing and hardware together:*the development of the big data, to some extent, but it also caused a large number of different hardware architecture of coexisting situation. From the data integration to the data analysis, until the final data explanation, ease of use should be throughout the whole process of big data. The challenge of ease of use is embodied in two aspects: first, the data of the era of big data, analysis of more complex, the results are more diversified. Its complexity is far beyond the traditional relational database.
6) *Performance benchmark test (benchmark):*Relational database products succeed without represented by TPC series test benchmark. It is with these test datum, can accurately measure the performance of different database products, and the existing problems to improve.

3. The challenge of CPS based on big data

1) *Cyber - Physical integration under the heterogeneous environment :*The traditional physical system is controlled by differential equation and boundary conditions of continuous and dealing with problems, attention to detail factors that affects the system implementation;Traditional information system is built on the basis of the cognitive and discrete mathematics and other unstructured knowledge, is not sensitive to the continuity of time and space, usually only care about the realization of system functions. Physical and information fusion system involving physical components and information components at the same time, therefore, how to overcome these differences, the depth of the layer and physical layer implementation information coupling, is a primary issue in the study of the CPS.

2) *component abstraction:*CPS component abstraction is a big difficulty in the study of the CPS. Main facing how to "ignore" different applications and different platforms of types of knowledge and the technical features, will be mixed with the interaction process of decoupling, from dynamic space-time, dissimilar, in the field of multidimensional mixture of different information and quantitative problems such as physical device of abstraction. In addition to further optimize the existing software engineering related theory, also can consider combining GIS and multidimensional data mining technology, through the analysis of the existing data and extract for component identification and abstract useful information. Can also be combined with graph theory and complex network technology, through the information abstract for the "node" and "edge" in the network, the characteristics of information is analyzed, and through a complex network of cluster methods of benefit and community division, realize the mixed messages and the extraction of components.

3) *Autonomy and autonomous coordination:*The emergence of large-scale CPS will be provided to decision makers and the generalUsers with vast amounts of information and much higher than the level of demand of a large number of can be controlled by electronic equipment. This may be beyond people's ability to control.CPS, therefore, should be able to digest the information and knowledge, and dynamically control the ability of each device independently. Only meet the autonomy, in order to realize the global optimization performance of the CPS system. Therefore, to solve the components in the system of independent interaction and resource coordination scheduling problem.

4)*Real-time:*In the existing embedded system and network control technology, the real-time performance is also a research focus. In the CPS environment for real-time operating more demanding, not only to improve the accuracy of real-time system, also want to improve the compatibility of the system, in order to have in the different platforms and different space system real-time response ability. How to wide-area implement the real-time control of the system under the scope of time and space, and how to under the premise of without sacrificing performance of the system and resources to improve system real-time performance and predictability, is difficult to realize the CPS technology.

5) *System safety and reliability:*Safe and reliable is a leading indicator of large complex system. In the CPS environment, information and physical interaction between components is more convenient than the original network communication structure is frequent, and the network users even intelligent component position also enjoy more equal freedom. In addition, the physical components and object-oriented software components to the standard of security also have essentially different, the traditional pattern of a single thread and method call based will no Longer apply.

6)*Verifiability:*Using what technology and index each of the components in the CPS and methods of performance measurement, validation and optimization is a big problem in the design of the system. Such as how to calculation module of the code in the consumption statistics and assessment, how to measure the energy loss in the physical device in the network environment, etc. The CPS validation is still a lack of general approved standard. Can consider scheduling ability from the system, the system energy consumption, speed, the system memory usage, deadlock and privacy aspects of conducting research. Combined with the existing modeling simulation and verification technology, to verify this model and method of the CPS and assessment. At the same time, also can make full use of and improve the existing various human-computer interaction system and complex system simulation platform, such as the university of Virginia and Berkeley university and other research institutions have adopted Macro Lab as the foundation of the CPS programming and verification environment, has been a good experimental effect.

4. Data fusion system modeling of physical characteristics

1) *Dynamic continuous:I*nformation physical define data fusion system is real-time and continuous, so you need to specific method to describe the continuity of the physical world, namely the discrete data of continuous.Such as train on the track to run through the sensor of the discrete data is continuous, so as to ensure stable operation system.

*2) Architecture modeling:*CPS architecture modeling need from two aspects of system's hard software on the system modeling design, architecture design is the basis of the CPS modeling design. hardware architecture, hardware architecture is the basis of CPS, also is the important part of the underlying model.System commonly used hardware mainly includes: the processor, memory, bus, etc.Therefore the design of the system hardware model need through physical modeling language for information fusion system rationalization modeling artifacts related to hardware. the software architecture, software architecture description system composition of each model, through the formal expressions of the child how to integration, interaction between model and how to communicate.

*3) Behavior modeling:*Behavior modeling is divided into two types: the interaction; the execution behavior.Interaction is mainly refers to physical information fusion information interaction between tasks in the system.Which occur concurrent data due to the big data system, require the system to support the mutex and synchronization operation, in the different priority tasks concurrent sharing resources, is likely to lead to priority inversion and deadlock, so as to make the error system.Execution behavior is the key step in the real-time system model, including real-time task scheduling and the system modeling and so on.

*4) Trouble shooting modeling:*Big physical data system and the information fusion system in running cycle to avoid errors or fault, the fault as defined by big data systems more the norm, so when the system failure occurs, should set up the system how to handle errors or failures, how to make the system failure occurs, the influence of the minimum of the whole system, and conform to the requirements of the CPS modeling.

5. Summary

The CPS Based on Big Data is the CPS with the needs of The Times development.It’s product of data scale explosive growth. Eager to through the Internet, big data and cloud computing technology to solve the problems in the data analysis, data processing, data storage.This paper expounds on big data and technical challenges that the CPS have faced and we needed to do to response.In this paper，author describe the big data based CPS modeling requirements.

References

[1]. Agrawal R,Srikant R.Privacy preserving data mining[C],Proc of SIGMOD2000,NewYork,ACM，2000,439-450.

[2]. Dwork C. Differential privacy[C] ,Proc of ICALP, 20006. Berlin : Springer,2006;1-12.

[3]. Harder T,Hudlet V, Ou Y,et al.Energy efficiency is not enough, energy proportionality is needed![]C,Proc of DASFAA 2011. Berlin; Springer,2011;226-239.

[4]. Baru C,Bhandarkar M,Nambiar R, et al.Introducting the big data benchmarking community[EB/OL],Proc of XLDB ,2012.[2012-10-02].

[5]. GILLES, Olivier.HUGUES, Jérôme. Expressing and enforcing user-defined constraints of AADL models. In: Proceedings of the 5th UML and AADL Workshop (UML and AADL 2010).