

Preliminary Study of Network System Reliability Based on Big Data

Chao Shen^{1, a}, Hongzhong Deng^{1, b}

¹College of Information and Management, National University of Defense Technology, Changsha, 410073, China.

^astevenshen1203@sina.com, ^bhongzhongdeng@vip.tom.com

Keywords: Reliability Big Data; Analysis Process of Reliability Big Data; Reliability Data Application Cases

Abstract. Complex network system has characteristics of numerous units, complex structure, and dynamic changes, which brings many problems for the study of network reliability. However, the big data of network system reliability and big data analysis technology provide more data and thought for supporting traditional reliability research. This paper explores the source, collecting process, preprocessing of the big data of network system reliability and research idea of big data analysis, meanwhile introduces research field to which big data of network system reliability relates at present.

Introduction

With the innovation of science and technology and the development of society, the scale and structure of the modern network system and the platform is becoming larger, whose units and connection mode is more complex, then the research problem of reliability is becoming more prominent. Network system is complex systems with particular structure and function, composed of many nodes (equipment, personnel, etc.), which is connected by edges (material, information, energy). Modern network system, such as communication network, traffic network, electric power network, has certain forming principle and task requires. It has characteristics of numerous units, complex structure, dynamic changes, etc. Traditional research of network system reliability usually have two kinds of thought: The one kind is that describing, simplifying and calculating network with series-parallel way. The other kind is that doing multiple simulations for network system if people have suitable condition.

In recent years, information technology and computer network technology develops quickly and technology and services such as cloud computing, Internet, mobile phone constantly improve. Meanwhile, data scale of many kinds of systems has explosive growth. Big Data has become a hot issue in today's social studies[1]. Huge amounts of data acquisition in network system reliability and utilization become more convenient with the help of big data technology and analysis method. The big data hidden behind nodes, edges and the connection relations in the network system has a broad application space. Laws and properties revealed in big data play an important role in aspects such as exploring network system reliability problems in key stage, data correlation in models, prediction of network system reliability, which provide more data and thought for supporting traditional reliability research.

This paper discusses the related contents and process of the study of network reliability based on the big data. Specific structure is as follows: the second part expounds the basic content of network reliability and some big data examples of network system reliability. In the third part the paper analyzes collection of the network reliability data. In the fourth part, starting from the concept of big data analysis, this paper introduces the general methods of big data analysis, and put forward some corresponding research idea combing with the network reliability data. The fifth part lists the field of study that the big data of network reliability is involved and applied preliminarily. The paper ends with a summary in the sixth part.

Description of Network Reliability and Its Reliability Big Data

Network Reliability Definition. In the process of the development of the network system reliability research, according to different network characteristics and research target, the standard of the definition of network reliability is not uniform. For example, in the communication network, according to the connotation of basic reliability and mission reliability, the network reliability is defined as the ability that complex network keeps entire network connectivity or realizes communication function under prescribed conditions and within the stipulated time. For other network systems, refers to the ability that transmits the material, information and energy completely and correctly in the network for certain time range of users' expectations under prescribed conditions and within the stipulated time.

Network Reliability Measure Parameters. The research of reliability problems is based on the analysis and evaluation of reliability measure parameters. At present, compared with the relatively perfect reliability measure parameter system, network reliability measure parameters have not yet formed a consistent view. But it always stresses that an important standard of reliability evaluation is to meet the needs of users in most of the definition. Therefore, by analyzing the inherent requirement of reliability and some existing research results, this paper briefly summaries some relevant parameters of the network system reliability as the following three aspects.

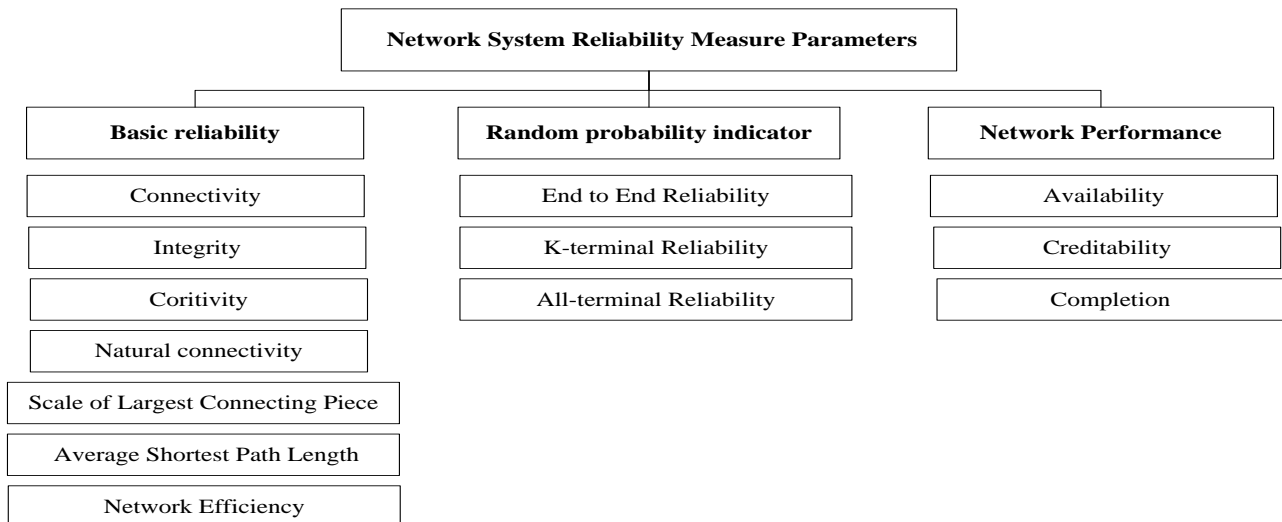


Fig. 1 Network System Reliability Measure Parameters

Traditional Calculation Method. Network reliability calculation method plays an important role in the study of network reliability analysis that is mainly divided into two categories: analytic method and simulation method. Among them, the analytical method is divided into the precise and approximate calculation, which includes state-space techniques, inclusion-exclusion principle, sum of disjoint products, factor analysis approach, stochastic process method, simulation method, bound method, the graph transformation method and other methods. These methods can completely solve the reliability calculation problem of series-parallel network, and simplify problem of network which is not series-parallel. For network with huge scale and complex structure, the computer simulation plays an important role in reliability calculation, including Petri net method, neural network, MonteCarlo simulation method and other simulation methods.

Reliability Big Data Samples. With the advent of the era of big data, numerous research fields with the aid of big data application have made a big breakthrough. The paper considers that modern network system has large number of component unit, various types and dynamic evolution feature and it has rich data resources[2] [3]. Network system is equipped with sensors and smart chip, providing convenient conditions for data collection and transmission. These data resources can provide data support for the traditional network reliability research. The following network system

preliminarily has the ability to provide reliability big data, and has certain economic value and social effect.

Table 1 Samples of network reliability big data

Network System	Information Provided by Big Data
Medical Network	Huge amounts of working parameters, energy consumption and use, etc.
Wind Power Network	Physical characteristic information of Equipment or component, etc.
Space Network	System working parameters, status, environment, information, etc.
Traffic Network	Temperature, pressure, geographical position information of equipment.

Collection and Preprocessing of Reliability Big Data of Network

The Source of Reliability Big Data. Considering the characteristic of network structure and function elements and using other big data application cases for reference, this paper argues that the network system reliability data should cover the following aspects of content respectively from the internal and external of network :

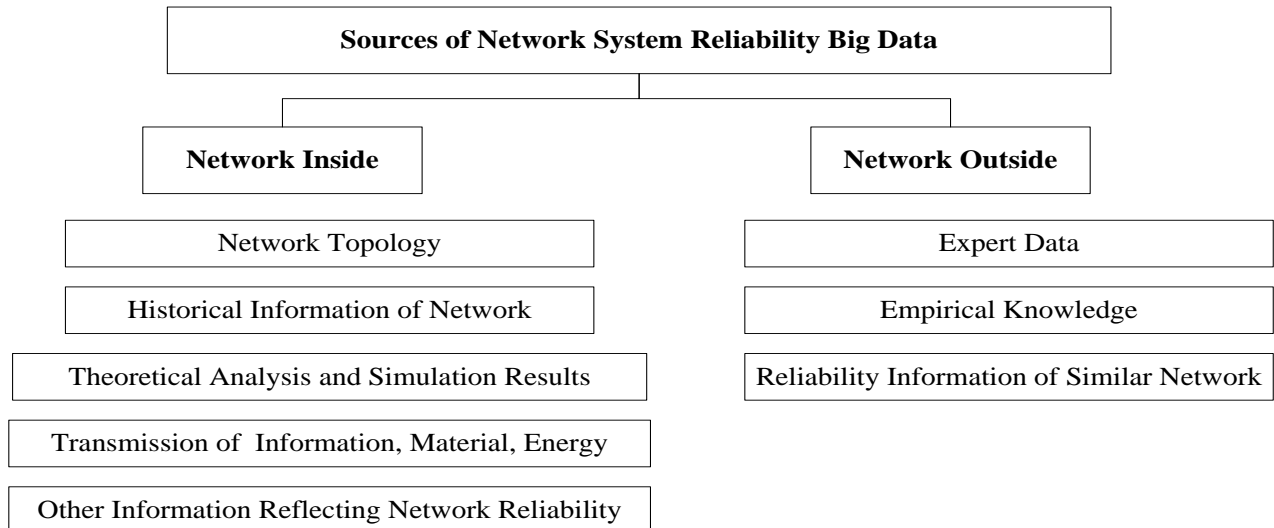


Fig. 2 Sources of Network System Reliability Big Data

Actually in a broad sense, the reliability data of network system should also include reliability data of product or equipment, namely, all kinds of data which reflects product reliability level and state in the reliability work and activity in each stage of product life cycle. These data can well describe the reliability parameters of the node layer. The traditional system reliability analysis method in general can be handled on the basis of these data. The following is the sources of reliability of the product data[4]:

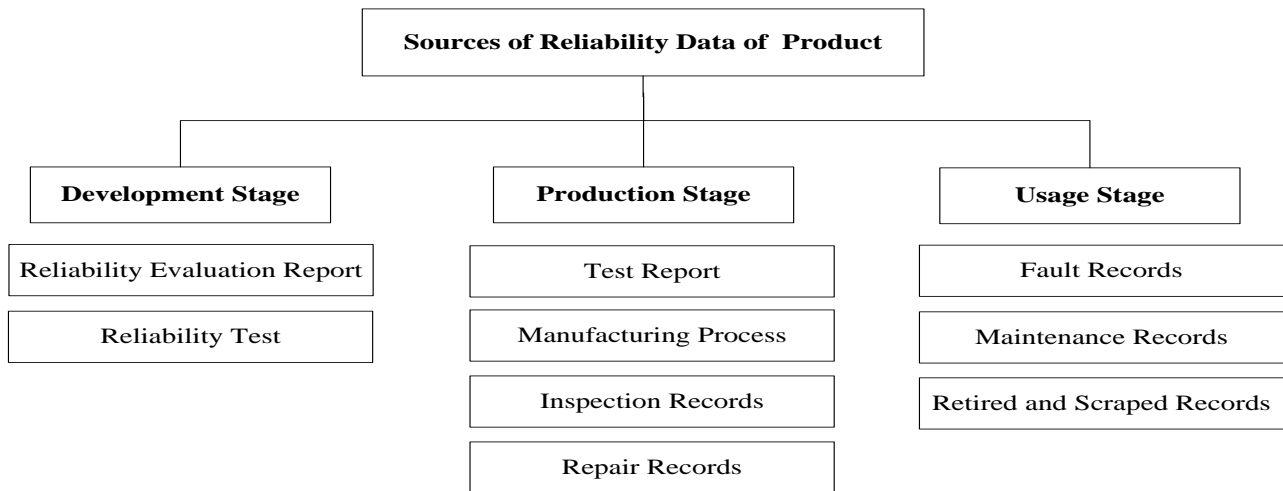


Fig. 3 Sources of Network System Reliability Big Data

The Collection of Reliability Big Data.

The Demand of Data Collection. In general, the requirements of the network reliability big data is decided by the need of reliability analysis in the network system for of different operational stage. In other words, it is necessary to analyze the purpose of obtaining the data and clear the specific use of data.

The Principle of Data Collection. The quantity and quality of the data itself had a great influence on analysis and application of reliability big data. So in order to obtain the high quality of the data in the process of data collection, researchers shall meet the truthfulness, continuity and completeness of the three basic principles.

The Procedure of Data Collection. Big data collection process generally has the following steps: researchers should analyze the demand and be clear of content and purpose of collection of network big data; researchers should determine the source of data, by analyzing the background of reliability problem and make clear channels of data acquisition; researchers should establish a standardized database and use multiple databases to receive and store the data from outside and carry out simple queries and processing tasks.

The Preprocessing of Reliability Big Data. The premise of Big data analysis is not only obtaining data of high quality, but also is the effective management of data. And data preprocessing is one of the important parts. Data preprocessing is the last step before researchers use and analyze reliability big data. Its' aim is to convert disorganized data and inconsistent structure to meet the requirements of the reliability analysis. In complex network system, the data of reliability behind network grows exponentially due to a lot of reasons. It is not practical to manage huge amounts of data artificially or in traditional way for most of people. Researcher must use the corresponding tools and instruments. Big data preprocessing mainly includes the data cleaning and data integration.

Data cleaning: due to subjective or objective reasons, the quality of the data in network system has several problems such as incomplete data, disorder data, noise, which need to use the corresponding data cleaning technology. It extracts reliability big data from the original data. In particular, it does need quickly effective management techniques and algorithm in order obtain high-quality data when data scale is too big.

Data integration: researcher should merge multiple data sources from different channels, put in a data store and remove redundant data. For large data source, data storage infrastructure includes cloud storage or distributed file storage, etc.

The Analysis of Reliability Big Data of Network

The Concept of Big Data Analysis. In general, big data analysis refers to obtaining intelligent and valuable information by analyzing the specific problems and background of huge amounts of data and information. And these big data attributes, including the number, speed, diversity and so on presents the complexity of big data. So big data analysis method is particularly important in the field of big data, which can decide whether information is of value.

The Content of the Big Data Analysis and General Method. Big data analysis focuses on two aspects of content: one is big data and analysis techniques; the second is combination of them. Therefore, big data analysis mainly involves data mining algorithm, visualization analysis, predictive analysis, network data mining, natural language processing, etc.

Among them, the data mining algorithm is one of the important theoretical basis for big data analysis[5]. In big data analysis, data mining algorithm mainly includes the following aspects: Classification algorithm: decision tree, random forest, neural network and support vector machine; Clustering algorithm, system hierarchical clustering method, k-means algorithm, BIRCH algorithm, clustering algorithm based on density, etc; Association rule mining: Apriori algorithm, Carma algorithm, etc; High-dimensional regression: model selection, generalized linear model regression, compression of high-dimensional regression coefficient; Graphic models: graphic classification, graphic clustering, graphic pattern matching, etc; Bayesian classification: naive bayes and bayesian network; Social network analysis: community mining algorithms and so on; Model of natural language and text mining: vector space model, statistical language model, and the LDA model, etc.

The Train of thought of the Big Data Analysis. It can be found that network system and reliability big data has its characteristic. This paper argues that the network system reliability research mainly has two problems: the one is difficult to establish a network system reliability model due to extensive data sources and complex network structure; the other is difficult to extracting reliability information from large working parameters and working indicators of network system. However, it can be regarded as an effective way to solve the reliability problem by extracting characteristics of data, analyzing the data evolution trend and finding the correlation among data. Therefore, this paper provides the corresponding reliability big data analysis thought based on the data characteristics, trends and correlation.

Extracting Characteristics of Data. For large network system or platform, reliability problems research is focused on implementation of the key stages and key link of network tasks under the premise of not affecting the network function. At this point, the related parameters influence the reliability problems tend to be clear. And the network system stores the parameters of relevant historical data in large amounts. Through statistical analysis and data mining, the main factors and its threshold range influencing the reliability issues can be decided. Compared to the current operation of the network systems, researcher can identify whether network system is healthy and find possible reliability problems.

Analyzing Data Evolution Trend. With the aid of experience knowledge, consulting experts, and comparing similar measures of the network system, researcher can master network system operation mechanism. By drawing curves, detecting abnormal point, time series analysis methods, researchers can analyze the related reliability parameters and historical information or data and infer the evolution trend of the data of the next stage in order to predict reliability problems.

Finding Correlation Among Data. Facing the reliability big data, researchers usually can not identify what data is valuable and effective quickly. But among network system running parameters, the subsystem and the data, it still can be excavated from valuable information a mass of clutter data through the analysis of the correlation of data, which reflects the deep problems. When going deep into research, application of correlation method also need to be constantly updated.

Application Fields about Network Reliability Big Data.

Based on analyzing inherent law of big data and supplemented by the corresponding data mining method, big data and its analysis technology has been gradually involved and even applied to the network system reliability study now. Among them, the more important applications are systems health management, prediction of remaining life, condition-based maintenance and early warning of reliability problems.

Systems Health Management. Health management systems is one of the most widely used method for reliability big data. Sensor technology can easily capture big data of network system reliability and people can detect abnormal or bad state of the system by analyzing these data. When people find network system goes wrong by testing, people usually adopt corresponding operation to make the system achieve security level.

Prediction of Remaining Life. Even if the network system is in normal working condition, people often also need to predict the rest of the life of system. By establishing a suitable model, researchers can predict the remaining life of main subsystem and key components by using reliability big data in order to make a regular maintenance program and avoid serious consequences caused by the network operation failure.

Condition-Based Maintenance. Generally, large network system in operation will inevitably have some problems. It is lack of efficiency to maintain that plans is based on the time schedule. However, it is gradually replaced by condition-based maintenance. The latter can timely maintain, according to the network system's need. This program has the advantages of high reliability and low cost.

Early Warning of Reliability Problems. As is known to all, the design of the network system model is different from the real situation. The latter may occur unexpected problems. The corresponding detection algorithm has been used to detect early problems of subsystem and component. Because working environment and usage of each part is different, so reliability of the same subsystem and component may be different. The application of reliability big data greatly improved the ability of early detection of reliability problems.

Summary

This paper preliminarily studies related content of network reliability based on big data. First of all, this paper summarizes the basic research content and method of network reliability, and introduces some sources of reliability big data from actual network. Second, our discussion analyzes the source of network reliability big data collection in detail and illustrates the demand, principle and general process of reliability. At the same time this paper stresses the importance of data preprocessing. Then this paper summarizes the general methods of big data analysis, put forward train of thought of the network reliability big data based on the characteristics of reliability big data. Finally, our examples introduce the application of big data and its' analysis techniques in reliability fields, including systems health management, prediction of remaining life, condition-based maintenance and early warning of reliability problems, etc. The paper provides some study basis for big data analysis in network reliability. But there are some deficiencies, the research content also need further validation and improvement in all kinds of specific network systems.

References

- [1] Li Guojie, Cheng Xueqi, Research status and scientific thinking of big data. Bulletin of Chinese Academy of Sciences, 2012, 27(6): 647—657(in Chinese)
- [2] Meeker, W.Q., Hong, Y. (2014). Reliability Meets Big Data: Opportunities and Challenges. Quality Engineering, 26:102–116.
- [3] Parker, P.A., Discussion of “Reliability Meets Big Data: Opportunities and Challenges”. Quality Engineering, 2013. 26(1): p. 117-120.

- [4] Zhao Yu, Data Analysis of Reliability, National Defense Industry Press, Beijing, 2011(in Chinese).
- [5] Wang Xing, Big Data Analysis: Methods and Applications, Tsinghua University Press, Beijing, 2013(in Chinese).