Intelligent Automation And Soft Computing, 2019
Copyright © 2019, TSI® Press
Vol. 25, no. 3, 561–572
https://doi.org/10.31209/2019.100000111

A New Rockburst Experiment Data Compression Storage Algorithm Based on Big Data Technology

Yu Zhang[1,2], Yan-Ge Wang[1], Yan-Ping Bai[3], Yong-Zhen Li[1,4], Zhao-Yong Lv[5], Hong-Wei Ding[6]

[1]School of Electrical and Information Engineering & Beijing Key Laboratory of Intelligent Processing for Building Big Data, Beijing University of Civil Engineering and Architecture, China;
[2]State Key Laboratory in China for GeoMechanics and Deep Underground Engineering (Beijing), China University of Mining & Technology, China;
[3]College of Management, Capital Normal University, China;
[4]School of Software, Beijing University of Technology, China;
[5]Computing Center, Beijing University of Civil Engineering and Architecture, China;
[6]China IPPR International Engineering CO., LTO, China

ABSTRACT
Rockburst phenomenon is a kind of phenomenon that the rock is out and ejected because the mineral was dug out, and the original force balance was destroyed in the process of mineral exploitation. From 2007, GeoLab (abbreviation of State Key Laboratory in China for GeoMechanics and Deep Underground Engineering) had made a series of important achievements in rockburst. Up to now, GeoLab's rockburst experiment data is reached 800T, and these data may occupy about 2PB hard disk space after analyzed. At this ratio, GeoLab need to buy a new hard disk to save all these data every 46 hours rockburst experiment. Since there is not enough hard disk space to save all these data, GeoLab had to slow down the speed of do rockburst experiment and only analyzed about 4 percent of the data. We call this phenomenon a dilemma for data storage. This hindered the research process of rockburst phenomenon. We proposed a structure to obtain data from a cloud platform based on big data technology. And basing on this we analyzed the distribution characteristics of rockburst experiment data, data frequency and data frequency domain. And a new rockburst experiment data compression storage algorithm (NDCS) based on big data technology and cloud platform was proposed. Then we compared NDCS with WinRAR and BDSS by occupied disk space, compress ratio and consuming time. Theoretical analysis and experiments show that NDCS has the best performance of all three algorithms. NDCS is the most suitable data compression storage algorithm for rockburst, and it has successfully solved the data storage dilemma in rockburst experiment.

KEY WORDS: Big Data; Cloud Platform; Compression Storage Algorithm; Rockburst

1 INTRODUCTION

THE rockburst is a phenomenon in which the rock is ejected because the mineral was dug out and the original force balance was destroyed in the process of mineral exploitation, and predicting it is one of the worldwide hard problems. In 2007, He Manchao et.al.(2007) successfully reproduced the process of rockburst indoors. Since then, scientific research about rockburst indoors has been possible. This is the foundation of the rockburst research of He Manchao, who is the sponsor of GeoLab (the abbreviation for the State Key Laboratory in China for GeoMechanics and Deep Underground Engineering). According to Google Scholar, this paper was cited 149 times in May 2018. In recent years, GeoLab has made a series of important achievements in rockburst as follows. Fracture angles were estimated by Gong Weili et.al. （2015） by using pressure analysis with the Mohr-cycle established. They found that the Mohr-cycle can perfectly represent the stress status of rockburst, whether in the situation of static loading or in the

CONTACT Yan-Ping Bai myscholar@sina.com

situation of dynamic unloading. It has the same good effect on the excavation surface and near the surface. He Manchao et.al.(2015) performed nearly 1000 rockburst tests and obtained a large amount of rockburst experimental data. They stored these data in an SQL (Structured Query Language, SQL) database and used traditional data mining technologies to analyze them. They developed a predictive model for the maximum stress (σRB) and risk index (IRB) in rockburst.He Manchao et.al.(2014) reviewed rockburst research in China and proposed some definitions and classifications in the rock domain, especially in the soft rock domain. For this purpose, they performed numerous experiments and accumulated numerous rockburst experiment data. Li, D et.al.(2014) performed rockburst experiments using a three-axial rockburst experiment system with four unloading rates and investigated the damage evolution from the cracked fracture surface and the characteristics of fragments. The characteristics include the amount of fragments, qualities of fragments, lengths of fragments, widths of fragments and thicknesses of fragments. These experiments show that the damage level is reduced with the decrease of the unloading rate. Sun X et al.(2017) used infrared thermography and acoustic emissions to experimentally investigate the occurrence of rockburst. They also collected rock fragments after the experiments. The fragments were calculated based on their size frequency, mass frequency, length-to-thickness frequency and ratio frequency. They found that when they are observably accompanied by blocky characteristics, the crushing degree of impact rockburst fragments is higher.

During the progress of the research, GeoLab Zhang, Yu et.al. (2013) encountered some dilemmas that restricted the development of rockburst research technologies. The most important dilemma is data storage, meaning how to save and manage large amounts of experimental data efficiently. Let us take one rockburst experiment, numbered “yqsii7#,” as an example. “yqsii7#” evenly produced 41523 txt files within one hour. These 41523 txt files occupied 15GB of hard disk space. Until now, the amount of GeoLab’s rockburst experimental data had reached 800T, but they had analyzed only approximately 4 percent of the data. During the procedure of rockburst experiments, a huge amount of experiment data is generated. The amount level of the data is in TB and even in PB. Thus, our focus in this paper is how to effectively store and manage these data, a problem that we refer to as the data storage dilemma.

Zhou, Hui et. al.(2015) selected four representative cases and analyze rockbursts in the deep underground, and they classified rockbursts into three types after a preliminary analysis. Ma, T. H et. al.(2015) employed a microseismic monitor device in rockbursts and achieved the full process monitoring about rocks’ movements. By doing so, they get a large number of data. Further, they found the space-time relationship in the microseismic activity of rockbursts. They believe that this approach can enable observation of the rockburst and early forecasts of this phenomenon.. In addition, Feng, Guang-Liang et.al.(2015) also proposed a real-time method that can predict rockburst risk during tunnel excavations. They built a database in which the data are from a microseismic device and contain several typical rockburst events. They established some formulas to calculate the rockburst risk level from different intensity of the rock. Novak, A et.al.(2015) used different nonlinear indicators to quantify that there is high nonlinearity in metal’s elastic deformation. They used acoustic emissions and nonlinear relaxation work together, and found that rocks and metals relax their energy following a logarithmic function. Zhong L.et. al.(2015) researched rockburst prognostication using Acoustic Emission wavelet signals. They obtained the wave of the AE device during the mining procedure and analyzed these waves in real time in order to predict the rockburst in time. At first, they divided waves into 6 groups based on their frequency, which is from 0 KHz to 31.25 KHz, from 31.25 KHz to 62.5 KHz, from 62.5 KHz to 125 KHz, from 125 KHz to 250 KHz, from 250 KHz to 500 KHz, and from 500 KHz to 1000 KHz. After this study, they found that 3 of the 6 groups are key groups for rockburst, and they are the groups from 31.25 KHz to 62.5 KHz, from 62.5 KHz to 125 KHz, and from 125 KHz to 250 KHz. This work has a guiding significance for Acoustic Emission research in the rockburst domain. Kong Biao et. al.(2017) also predicted the rockburst risk using AE wave data. They performed some experiments using the one-axial pressure test under different temperatures and collected the wave data. Then, the frequency distribution of the wave is analyzed and the main frequency band is obtained. Finally, they concluded that temperature has a significant impact on the frequency, and the higher the temperature is, the larger the frequency is. Their work is very valuable, but there is still a great difference between real rockbursts and their work, because their work is one-axial.

In recent years, Zhu, Ting, et.al.(2015) has combined big data technology and wireless communication technology, and the data sources of the big data are from wireless sensor networks. Zhu J et.al.(2016) set a wireless sensor network as their data source and proposed a data collection protocol that has four wireless stages, which are cluster the network, plan the route, combine the routes and collect the data. This protocol can achieve more movement with a lower energy exhaust. Krishnaveni, S. et.al.(2015) proposed four steps for big data, which are how to generate the data, how to acquire the data, how to save the data, and how to analyze the data. Gao, Honghao, et.al.(2018) have researched the quality of data obtained on the WIFI networks, Bluetooth networks, 5G communications systems, and Wireless Sensor

Networks. When we mention big data technology, Hadoop will be mentioned first. Hadoop can process data in a reliable, efficient and scalable way. Dean, J. & Ghemawat, S. (2008) have proposed Hadoop is derived from the Google File System (GFS) and Wang, C. et.al.(2010) have proposed Hadoop is derived from the MapReduce (GMR). Liao, C et.al.(2016) have proposed HDFS (Hadoop Distributed File System) is good at saving huge amounts of data reliably, and it can provide these data to user destinations at a very high speed. Designed for big data, Dede, E. et. al.(2016) have proposed Hadoop can evaluate streaming data and realize online algorithms in real-time. Gao, Honghao, et.al.(2017) have proposed an approach to data consistency checking method. Rashid, M. et.al.(2017) , Li, X. et.al(2017) and Zhu, Ting. et. al.(2015) all have proposed big data technology combined with wireless sensor networks has made some progress.

In this paper, we focus on the dilemma of data storage, which means we try to provide a method to solve the problem of saving and managing rockburst experiment data efficiently. In this way, we want to promote the study of rockburst mechanisms and even forecast rockbursts. We use wireless sensor networks to collect the data, as illustrated in section 2.2 of this paper.

2 BIG DATA AND ROCKBURST EXPERIMENT

2.1 Big data of rockburst experiment

WE have mentioned that GeoLab's rockburst experimental data have reached 800T. Take the same example as in section 1. The "yqsii7#" rockburst experiment generated 41523 txt files within one hour. These 41523 txt files occupied 15GB of hard disk space. If we use the mainstream 2T~3T hard disk to save these files, it will take 133hours~200hours to fill the hard disk with data. We can also see similar phenomena in Table 2, which is the table titled Details of the three randomly selected rockburst experiments. It is obvious that the rockburst experimental data are a huge amount of data, or big data. Therefore, the biggest problem faced by GeoLab is how to efficiently store, manage and analyze these rockburst big data.

2.2 Acquisition method of rockburst big data

2.2.1 Indoor rockburst experiment system

GeoLab performed the indoor rockburst experiment using special equipment, which was named the three-axial rockburst experiment system. The system was developed by Manchao He, who is the director of GeoLab and an academician at the China Engineering Academy. Furthermore, the system was patented under the China National Invention Patents under grant No.ZL200710099297.1 early in 2007. Figure 1 is the three-axial rockburst experiment system.

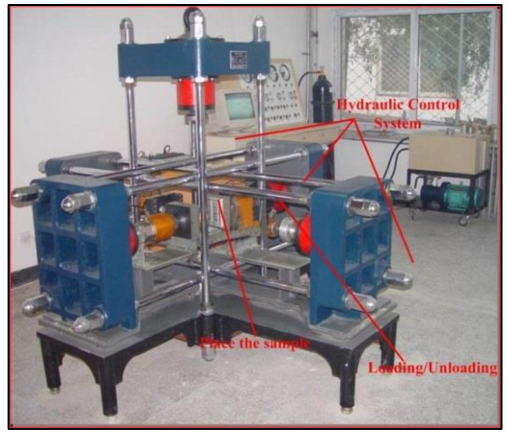

Figure 1. **Three-axial rockburst experiment system.**

As illustrated in figure 1, we can see that the experiment system is a three-axial machine that is able to achieve a one-sided sudden unloading. The size of the experiment system is 2240mm×1960mm×1800mm, and the experiment system weighs 2300kg. The maximum press of the system is 450kN. It can load independently in three directions and unload suddenly on one surface, which can simulate the formation of the free surface in field engineering due to excavation. In the three-axial rockburst experiment system process, when unloading on one side, the surface of the test piece was rapidly exposed. In this way, we can simulate the procedure of stress transformed during the procedure of excavation indoors.

The hydraulic control system consists of two pump stations and one major component console. The console weighs 100kg with a size of 950mm×840mm×1570mm. The rated power is 2.2kW and the rated Voltage is 380V. The working principle of the hydraulic control system is not our focus in this paper. We provide a simple sketch of its mechanism in figure 2.

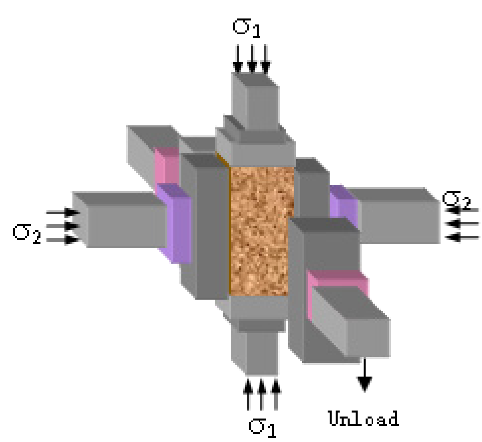

Figure 2. **Sketch of the three-axial rockburst experiment system mechanism.**

2.2.2 Acoustic emission data collection cloud platform

The acoustic emission data collection cloud platform is constituted by a DSG9803 dynamic strain amplifier and a USB8516 Portable Data Acquisition Instrument Composition. The acoustic emission data collection cloud platform is illustrated in Figure 3.

The acoustic emission data collection system is able to capture the fast nonlinear dynamic characteristics of rock stress and the displacement of rockburst under the combined effect of certain stress. For example, figure 4 is a typical vertical load curve of a granite rockburst.

From figure 4, we can see that the vertical load has a greater drop process in a very short period of time when a rockburst occurs. By this method, the vertical direction of the force can be observed in detail in figure 4. It is the foundation of analyzing the mechanical behavior during rockbursts.

The Acoustic Emission (AE) is generated by the fast unloading inside the material's local area where then elastic energy is released. The principle of acoustic emission testing technology is illustrated as follows. During the experiment, elastic waves are caused by a serious of injuries such as crack closures, crack expansions and crack penetrations. The elastic waves spread to the acoustic emission sensors. Then, we get the original waveform of the acoustic emission data after acoustic-electric conversion by sensors and strengthen it using the preamplifier. After the processing procedure of the original data, we can analyze the AE Laws in the process of the destruction of rocks. This procedure is the base of the mechanism of rock materials generating acoustic emissions.

Because a rockburst is one kind of rock damage, the rockburst process will produce a large number of acoustic emissions. It is very important to analyze the rockburst mechanism by analyzing the rockburst acoustic emissions data from the rockburst experiments. The acoustic emissions data can be analyzed using the acoustic emission waveform, acoustic emission energy, acoustic emission numbers, etc.

Acoustic emission testing needs to convert acoustic emission signals into electrical signals using sensors. Therefore, how to select sensors with high reliability and high quality in acoustic emission testing is our goal. Since different sensors can test different frequency areas, we need to select the sensor for the frequency that is suitable for rockburst or suitable for rock. Figure 5 is the acoustic emission frequency range and its corresponding acoustic emission technology applications.

Figure 5 gives the special frequencies of different rocks in the indoor experiments. From figure 5, we can see that the rockburst frequency is at the 101 Hz level.

To clearly and accurately obtain the character of rockburst acoustic emission data, we use two types of sensors from two different manufacturers. One is the WD from the Physical Acoustics Corporation (PAC) in America, and the other is the PXR 150 from Pengxiang in China. In Table 1, we give some key parameters of the two types of sensors. In addition, in figure 6, we can see the distribution of the sensors in the rockburst experiment system.

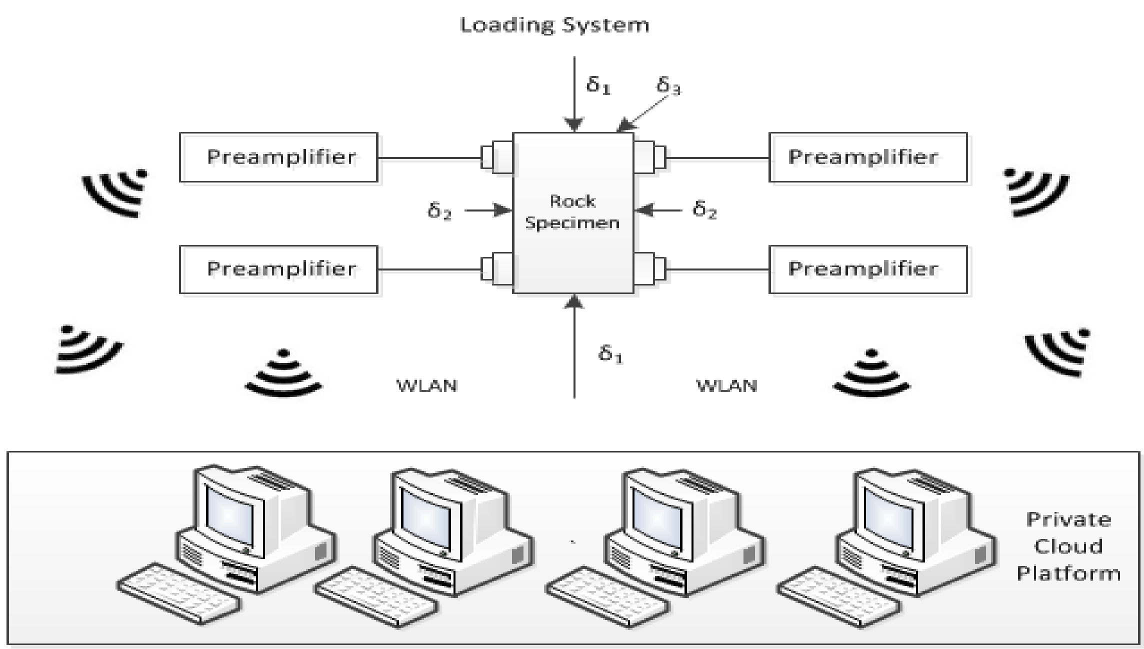

Figure 3. Acoustic emission data collection cloud platform.

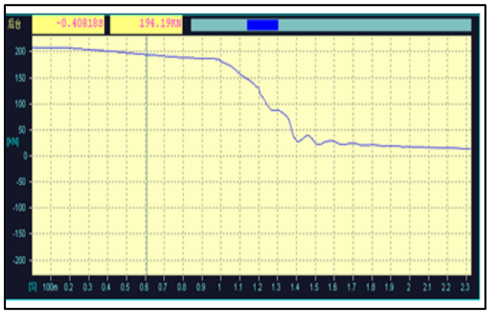

Figure 4. A typical vertical load curve of a granite rockburst.

2.2.3 Data storage dilemma

Take the same rockburst experiment example as in section 1, which is number "yqsii7#".

Table 1. Key parameters of the two types of sensors

	WD-PAC	PXR150-Pengxiang
Frequency range	100KHz~300KHz	100 KHz~1 MHz
Response Frequency	150 KHz	55 KHz
Front Amplify	40 dB	20 dB
Threshold Value	40 dB	50 dB

i) "yqsii7#" generated 41523 txt files in one hour, and these files occupied 15GB of hard disk storage space.

ii) Each txt file represents an acoustics emission waveform. We can draw four diagrams in our research from each txt file. The four diagrams are described as follows.

(a) Time domain-waveform diagram

(b) Frequency spectrum diagram

(c) Time-frequency diagram

(d) Three-dimensional diagram

These diagrams are not our focus in this paper. We focus on the size of them. These diagrams are in JPG format, and each of them occupies 300KB of hard disk space. Thus, we can use formula 1 to calculate the total disk space needed to store the data generated by the rockburst experiment "yqsii7#" in each hour.

$$\text{TotalODS}=\text{ODS}+N*300\text{KB/picture}*4\text{picture} \quad (1)$$

In formula (1), ODS represents the occupied disk space, which is occupied by the original txt files, and N is number of txt files. Since we can draw four diagrams in our research from each txt file, the TotalODS can be calculated by formula 1. The TotalODS means the total occupied disk space after the experiment.

As in formula (1), the rockburst experiment that is numbered "yqsii7#" will occupy 64.83G of disk space to store one hour of data as in formula (2).

$$15\text{GB}+41523*300\text{KB}*4=64.83\text{GB} \quad (2)$$

At this speed, since the normal hard disk capacity is 2TB~3TB, one disk will be full of data within 31hours~46hours. In other words, currently, for every 46 hours of rockburst experiments, GeoLab will need to buy a new hard disk to save all these data.

We randomly select three rockburst experiments as examples, which are illustrated in table 2. To protect privacy, we use the laboratory internal code as the experiment name instead of the specific name.

From table 2 and formula (1), we can calculate the TotalODS of the three randomly selected rockburst experiments as follows.

$$\text{TotalODSO-1\#}=1.50\text{GB}+29438*300\text{KB}*4=36.83\text{GB} \quad (3)$$

$$\text{TotalODSG-1\#}=2.11\text{GB}+41645*300\text{KB}*4=52.08\text{GB} \quad (4)$$

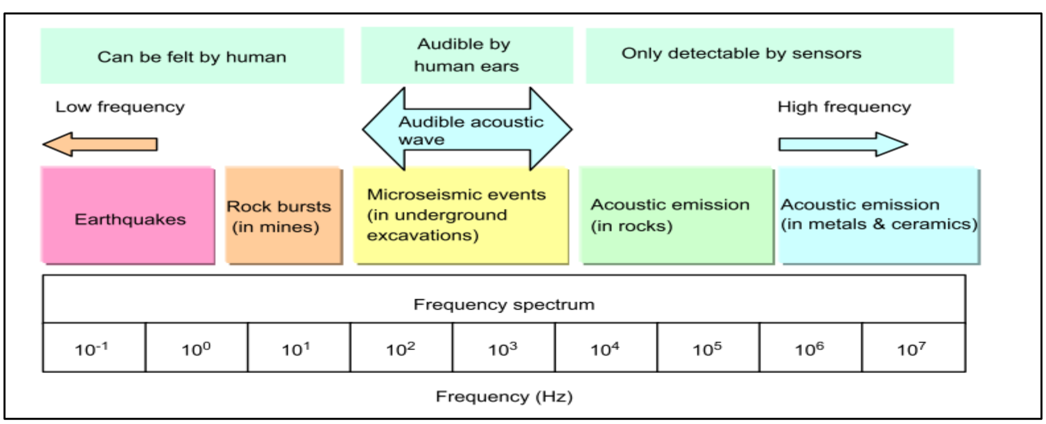

Figure 5. Acoustic emission frequency range and its applications.

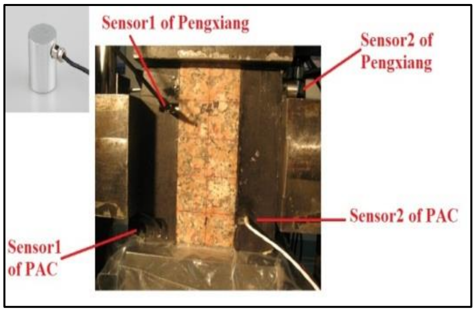

Figure 6. **Sensors and their parameters.**

Table 2. **Details of three randomly selected rockburst experiments**

Experiment Internal Code	Amount of TXT files	Values of all TXT files	ODS (GB)
O-1#	29438	120578048	1.50
G-1#	41645	170577920	2.11
GO-1#	71351	292253696	3.66

$$\text{TotalODSGO-1\#}=3.66\text{GB}+71351*300\text{KB}*4=89.28\text{GB} \quad (5)$$

$$\begin{aligned}\text{AverageRatio}&=(36.83/1.50+52.08/2.11+89.28/3.66)/3\\&=(24.53+24.64+24.37)/3\\&=24.51\end{aligned} \quad (6)$$

Up to now, GeoLab's rockburst experiment data has reached 800T. We must be aware that 800T is only the data directly obtained from the experiments. If we take the average AverageRatio 24.51, which is calculated according to formula (6), there will be approximately 19608TB data, or nearly 2PT data. It is truly big data. As the research continues, the number of experiments and the number of files generated by the experiments will increase exponentially.

From table 2, we can also see that each rockburst experiment produced a number of TXT files at the level of 105, and the produced values of all TXT files at the level of 109.

$$\begin{aligned}&(120578048/1.50+170577920/2.11+292253696/3.66)/3\\&=(80385365+80842616+79850736)/3\\&=80359572\end{aligned} \quad (7)$$

From formula (7), we can see that each GB of data has 80359572 values on average.

$$\begin{aligned}&(800\text{TB}/1\text{GB})*80359572\\&\approx 64\text{T}\end{aligned} \quad (8)$$

As we mentioned, GeoLab's rockburst experiment data have reached 800T. According to formula (8), there still are approximately 64T values waiting to be processed.

Because there is not enough hard disk space and existing data have yet to be analyzed, GeoLab had to decelerate the speed of performing rockburst experiments. That is, the dilemma of data storage seriously affected the speed of rockburst research. In this paper, we proposed a new rockburst experiment data compression storage algorithm based on big data technology and 5G wireless sensor networks to solve the data storage dilemma in rockburst experiments.

3 DATA COMPRESSION STORAGE ALGORITHMS BASED ON BIG DATA TECHNOLOGY

3.1 *System structure*

FIRST, we designed and implemented a BDSS system, which is short for a big data storage system. This system can efficiently store and manage data, provide these data to a user destination at very high speed, and analyze these data in real time. The system consists of four parts, which are the loading machine cluster, query machine cluster, metadata node machine cluster and storage node machine cluster. The structure of this system is illustrated in figure 7.

1）Loading machine cluster

The loading machine cluster is the data loading terminal of the system. With the process as the unit, multiple concurrent loading clients can simultaneously be established on multiple devices. The loading efficiency of the system can be improved by the concurrent load. In the BDSS, the loading machine cluster can simultaneously cache the data that have recently been stored. After a fixed period of time, the cached data are written to the data storage management device through the cloud platform.

2) Query machine cluster

The user sends a query instruction on the query machine to establish a query plan. The query machine distributes the query task to the storage node according to the metadata information stored in the metadata node cluster, summarizes the query results returned by multiple storage nodes and then submits them to the user.

3) Metadata node machine cluster

The metadata node machine cluster is used to coordinate the work of the entire cluster and save the metadata information required for the entire system. The metadata stored in metadata node cluster includes the following.

- Data node status,
- Specific storage location of index shard,
- Table-space metadata,
- Some assistant information in table-space,
- System log, and
- Other information.

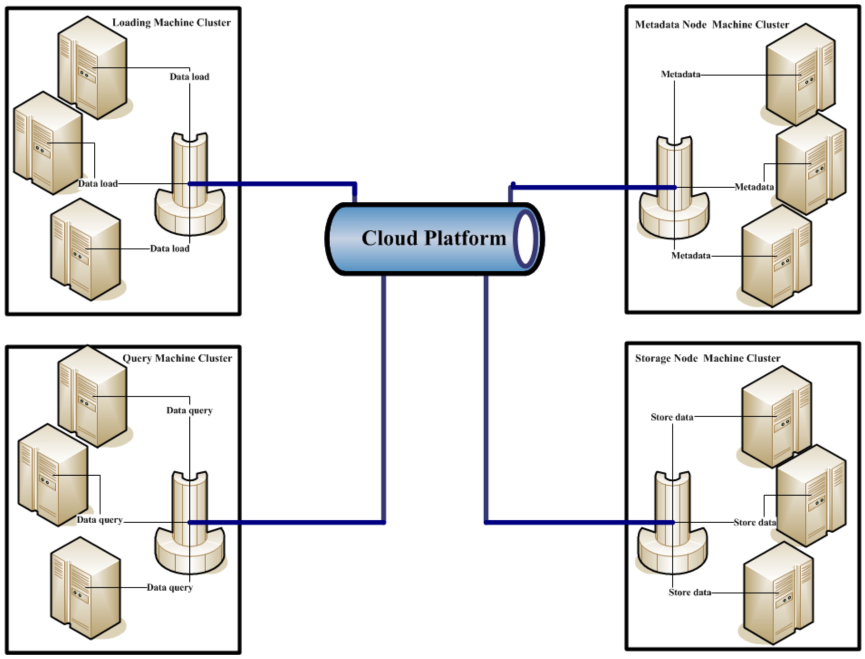

Figure 7. Structure of big data storage systems.

4) Storage node machine cluster

In this article, we discuss only the storage node machine cluster that is combined with several hard disks and can be added anytime. It can provide high bandwidth and high concurrent access in order to ensure that the BDSS can achieve online loading data in real time. We proposed two data compression storage algorithms based on big data technology and the cloud platform as follows.

3.2 Data compression storage algorithm in BDSS

Second, we designed and realized a compression storage algorithm in the BDSS. We will discuss the BDSS algorithm in another paper in detail. Here, we introduce several important steps of the algorithm.

i) Table 2 shows that there are 105 number of TXT files in one rockburst experiment. First, we put them together into one txt file. It should be noted that the new TXT file is too large to be open by Windows applications. The new TXT file can be read and written by our algorithm designed for big data.

ii) The combined process is under the principle of {SN, RN1RN2RN3RN4, Data}.

SN represents that the data are located in a set txt file. Txt files are numbered 1, 2, 3, N, in the order in which they are produced.

RN1RN2RN3RN4 is a four-digit number. RN1RN2RN3RN4 represents the data location, which is the row number in a TXT file SN.

The data represent the wave value collected by an Acoustic Emission System.

Therefore, {SN, RN1RN2RN3RN4, Data} is the specific data in row RN1RN2RN3RN4 in the SNth TXT within the same experiment.

iii) The last but not the least principle is that if the data are equal to zero, they will not be recorded in the new TXT file. This means that a zero value will be compressed.

3.3 Data compression storage algorithm in BDSS

Third, based on the data compression storage algorithm in BDSS, we designed and implemented a new data compression storage algorithm to improve the former algorithm's performance. In the latter part of this paper, the data compression storage algorithm in the BDSS is abbreviated as the BDSS algorithm, and the new data compression storage algorithm in the BDSS is abbreviated as the NDCS algorithm. The NDCS is illustrated as follows.

i) Similar to the BDSS algorithm, the NDCS first combined all txt files that were produced in one rockburst experiment into a new single txt file.

ii) The combined process is under the principle of {DATA, NC}.

The data represent the wave values collected by the Acoustic Emission System.

NC represents the number of consecutive data.

Therefore, {DATA, NC} is a sequence of data. The sequence of the rows represents the data appearance sequence.

4 THEORETICAL ANALYSIS AND EXPERIMENTS

4.1 Distribution characteristics of rockburst experiment data

WE first analyze the characters of the three randomly selected rockburst experiments data, and draw figure 8 and figure 9. Because the lines of figure 8 are denser and the details are not clear, we draw figure 9 to show the amplification of the key part in figure 8.

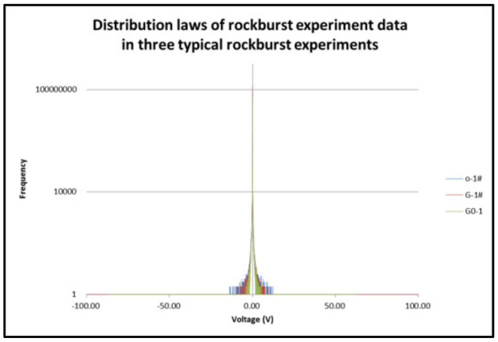

Figure 8. **Characters of the three randomly selected rockburst experiments' data.**

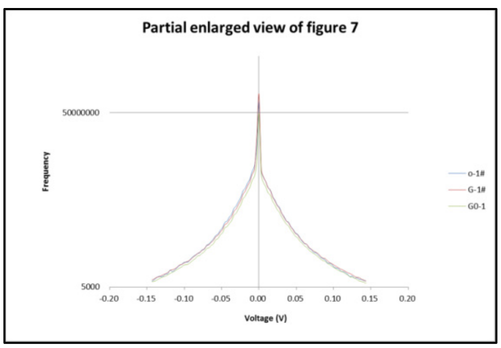

Figure 9. **Amplification of the key part in figure 8.**

From figures 8 and 9 we can see that the characters of the three randomly selected rockburst experiments' data are normally distributed. Furthermore, zero data is the majority. Therefore, the BDSS algorithm should have a good performance. Then, we conducted some further research on it.

4.2 Distribution characteristics of rockburst experiment data frequency

Second, we obtained the distribution characteristics of the rockburst experiments' data frequency in three randomly selected rockburst experiments, as shown in figures 10, 11 and 12 as follows. In figures 10, 11 and 12, "frequency value n" indicates that the value appears n times in all txt files, which is the same with the times that the value appears in the new TXT file. For example, "frequency value 1" represents the proportion of the values that only occur once in all experimental results. Similarly, "frequency value 2" represents the proportion of the values that occur twice in all experimental results. In Figs. 10, 11, and 12, we have calculated the statistics of these three sets of experimental data from "frequency value 1" to "frequency value 15".

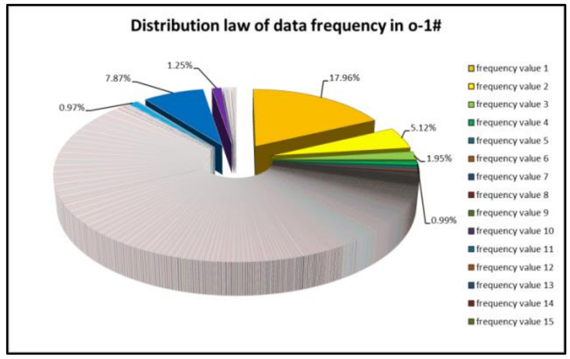

Figure 10. **O-1# data frequency.**

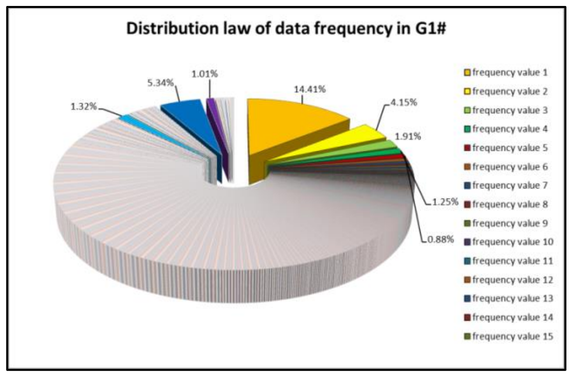

Figure 11. **G1# data frequency.**

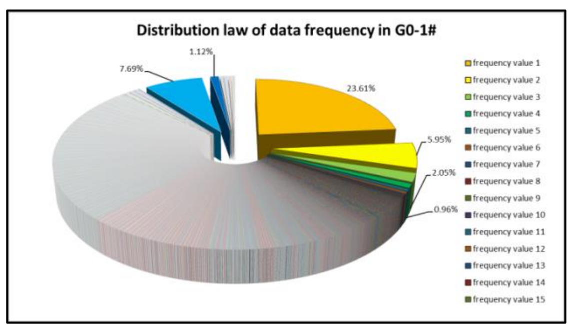

Figure 12. **GO-1# data frequency.**

From figure 10, we can see that frequency value 1 is 17.96%, which means that 82.04% of the values appear at least two times. Similarly, from figure 11, we can see that the frequency value 1 is 14.41%, which means that 85.59% of the values appear at least two times. From figure 12, we can see that frequency value 1 is 23.61%, which means that 76.39% of the values appear at least two times. In other words, the rockburst experiment data have high repeatability distribution characteristics, and so the NDCS should have a good compression ratio.

4.3 Distribution characteristics of rockburst experiment data frequency domain

Based on section 4.2, we continue to study the distribution characteristics of the rockburst experiment data frequency domain. The results are illustrated in figures 13, 14 and 15.

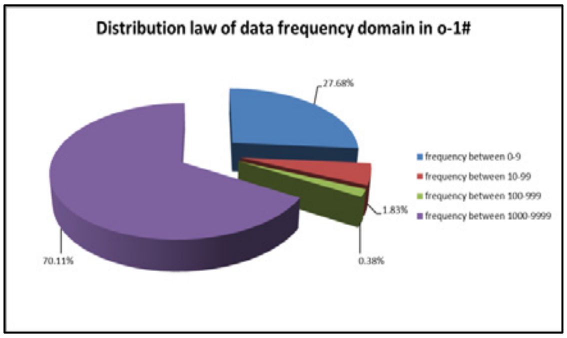

Figure 13. O-1# data frequency domain.

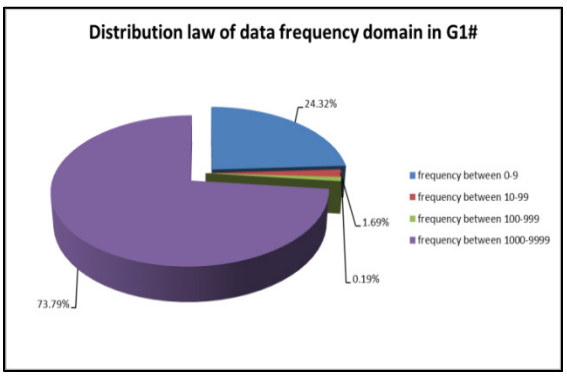

Figure 14. G1# data frequency domain.

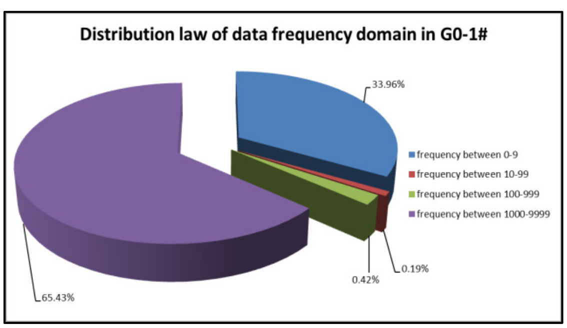

Figure 15. GO-1# data frequency domain.

In figures 13, 14 and 15, we divided the distribution characteristics of the rockburst experiment data frequency into four domains, including frequencies between 0-9, 10-99, 100-999 and 1000-9999. It can be seen from Figure 13 that for the experiment numbered O-1#, the probability of repeated occurrences in the range 0-9 is 27.68%, in the range 10-99 is 1.83%, in the range 100-999 is 0.38%, and in the range 1000-9999 is 70.11%. Similarly, it can be seen from Figure 14 that for the experiment numbered G1#, the probability of repeated occurrences in the range 0-9 is 24.32%, in the range 10-99 is 1.69%, in the range 100-999 is 0.19%, and in the range 1000-9999 is 73.79%. As shown in Figure 15, for the experiment numbered O-1#, the probability of repeated occurrences in the range 0-9 is 33.96%, in the range 10-99 is 0.19%, in the range 100-999 is 0.42%, and in the range 1000-9999 is 65.43%.

From figures 13, 14 and 15, we can see that the frequencies are mainly distributed in the 1000-9999 domains. That is, the values that appear from 1000 times to 9999 times are in the majority. Figure 13 is 70.11%, figure 14 is 73.79% and figure 15 is 65.43%. Therefore, the NDCS should have a good compression ratio and excellent performance.

4.4 Occupy disk space compares

WinRAR is one of the best compression software in the world. Therefore, we compare the BDSS and NDCS with WinRAR by their occupied disk space in figure 16. The blue column represents the original data's occupied disk space, the red one is the BDSS, the purple one is WinRAR and the green one is the NDCS. Please note that the values of the BDSS, WinRAR and the NDCS are after compressed by their own algorithm.

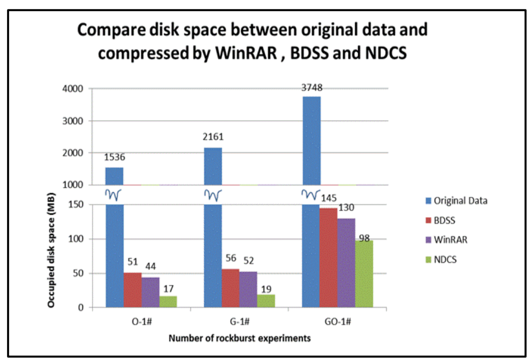

Figure 16. Occupy disk space compares.

As shown in figure 16, for the experiment numbered as O-1#, the original data size is 1536MB. After being compressed by the BDSS, it takes up 51MB of hard disk space; after being compressed by WinRAR, it takes up 44MB of hard disk space; and after being compressed by the NDCS, it takes up only 17MB of hard disk space. Similarly, for the experiment numbered G-1#, the original data size is 2161MB. The hard disk spaces occupied by the original data are respectively 56MB, 52MB and 19MB after being compressed by the BDSS, WinRAR, and the NDCS. For the experiment numbered GO-1#, the original data size is 3748MB. The hard disk spaces occupied by the original data are respectively 145MB, 130MB and 98MB after being compressed by the BDSS, WinRAR, and the NDCS. Therefore, as shown in figure 16, although the BDSS compression effect is worse than WinRAR, the NDCS has a slight advantage ahead of WinRAR. Most important of all, the NDCS occupied less disk space than WinRAR. NDCS has the best performance among them.

4.5 Compression ratio comparison

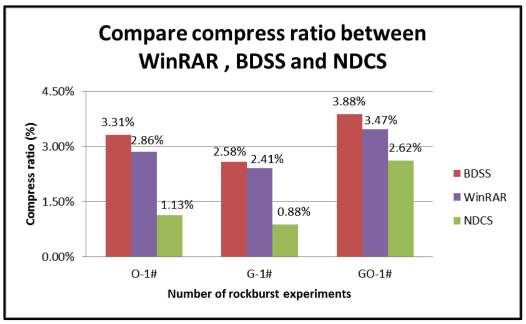

Figure 17. Compression ratio comparison.

We define the compression ratio as formula (9).

$$\text{Compression ratio} = \frac{\text{occupied disk space after being compressed}}{\text{occupied disk space of original data}} \tag{9}$$

From figure 17, it can be seen that for the experiment numbered O-1#, the compression ratio that can be achieved through using the BDSS is 3.31%, the compression ratio that can be achieved through using WinRAR is 2.86%, and the compression ratio that can be achieved through using the NDCS is 1.13%. Similarly, for the experiment numbered G-1#, the compression rates that can be achieved through using the BDSS, WinRAR and the NDCS compressions are respectively 2.58%, 2.41% and 0.88%. For the experiment numbered GO-1#, the compression rates that can be achieved through using the BDSS, WinRAR and the NDCS compressions are respectively 3.88%, 3.47%, and 2.62%. Therefore, from figure 17, we can see that the NDCS has a better compression ratio than WinRAR, and WinRAR has a better compression ratio than the BDSS. Once again, the NDCS is the best among them.

4.6 Compression time comparisons

Importantly, WinRAR needs time to rar and unrar the file. Therefore, when we want to use the data each time, we need time to rar or unrar the data. Because rockburst experiments need the online loading and analyzing of the data, WinRAR is not suitable for rockburst experiments.

Although before storing data, WinRAR, the BDSS and the NDCS all need time to compress data, they are different when using the data. On the one hand, when we need to use WinRAR compressed data, we need a decompression process called unrar. On the other hand, when we need to use the BDSS compressed data or the NDCS compressed data, we do not need a decompression process, which means that we can directly use the compressed data. It is obvious that the decompression process needs time to perform such a process. Therefore, in the view of time, the BDSS and NDCS are more efficient than WinRAR. Figure 18 shows the time consumption comparison among them.

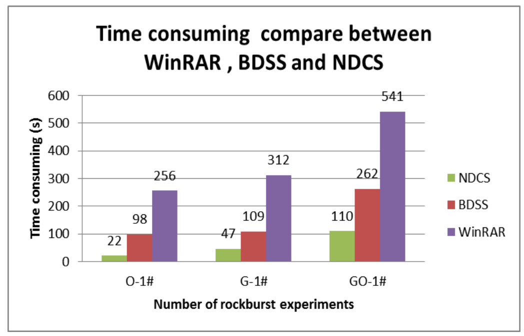

Figure 18. Time consumption comparison.

From figure 18, we find that the experiment numbered O-1# take 256 seconds when compressed by WinRAR, 98 seconds by the BDSS, and only 22 seconds by the NDCS. Similarly, the experiment numbered G-1# takes 312 seconds when compressed by WinRAR, 109 seconds by the BDSS, and only 47 seconds by the NDCS. The experiment numbered GO-1# take 541 seconds when compressed by WinRAR, 262 seconds by the BDSS, and only 110 seconds by the NDCS. Therefore, the NDCS is the fastest of the three, the BDSS is in the middle, and WinRAR is the worst. Because rockburst experiments need online loading and analyzing of the data, the NDCS is the most suitable algorithm for rockburst experiment data.

5 CONCLUSION

FROM all of the above, we can draw the conclusion that the NDCS algorithm has best efficiency in the rockburst domain, such as occupying disk space, its compression ratio, and its time consumption. Furthermore, we successfully solved the data storage dilemma in rockbursts by using the NDCS algorithm in this paper. In the future, we may try to solve the other two dilemmas in rockbursts and solidify the foundation of rockburst research.

6 ACKNOWLEDGMENT

THIS work was supported by National Key R&D Program of China (No.2016YFC0600901), and also The Fundamental Research Funds for Beijing Universities of Civil Engineering and Architecture (Responsed by ZhangYu & X18018 & 30860919020), and also Excellent Teachers Development Foundation of BUCEA (Responsed by

ZhangYu), and also Foundation of State Key Laboratory for GeoMechanics and Deep Underground Engineering (No.SKLGDUEK1523), and also Foundation of Beijing University of Civil Engineering and Architecture (No.00331615029), and also Institute of Energy and Environmental Sciences of Beijing Coal Mine Innovation Alliance (00362016055). This work was partly supported by Beijing Municipal Education Commission (No. KM201510016009).

7 DISCLOSURE STATEMENT

NO potential conflict of interest was reported by the authors.

8 REFERENCES

J. Dean & Ghemawat, S. (2008). MapReduce: simplified data processing on large clusters. Communications of the ACM, 51(1), 107-113.

E. Dede, et al. (2016). Processing cassandra datasets with Hadoop-streaming based approaches. IEEE Transactions on Services Computing, 9(1), 46-58.

G.-L. Feng, et al. (2015)."A microseismic method for dynamic warning of rockburst development processes in tunnels." Rock Mechanics and Rock Engineering 48.5, 2061-2076.

H. Gao, et al. (2018). "Applying improved particle swarm optimization for dynamic service composition focusing on quality of service evaluations under hybrid networks." International Journal of Distributed Sensor Networks 14.2: 1550147718761583.

H. Gao, et al. (2017). "An approach to data consistency checking for the dynamic replacement of service process." IEEE Access 5: 11700-11711.

W. Gong, et al. (2015)."Fracture Angle Analysis of Rockburst Faulting Planes Based on True-Triaxial Experiment." Rock Mechanics and Rock Engineering 48.3, 1017-1039.

M. He, et al. (2007). Experimental study on rockburst processes of granite specimen at great depth. Yanshilixue Yu Gongcheng Xuebao/Chinese Journal of Rock Mechanics and Engineering 26.5, 865-876.

M. He, et al. (2015). Rockburst laboratory tests database—application of data mining techniques. Engineering Geology 185, 116-130.

M. He, (2014). Latest progress of soft rock mechanics and engineering in China. Journal of Rock Mechanics and Geotechnical Engineering 6.3, 165-179.

B. Kong, et al. (2017). Acoustic emission signals frequency-amplitude characteristics of sandstone after thermal treated under uniaxial compression. Journal of Applied Geophysics 136: 190-197.

S. Krishnaveni, Satheesh, A., & Kannan, E. (2015). Review of Big Data on Student Information for Finding the Uncertainty in Higher Education Enrollment. International Journal of Grid and High Performance Computing (IJGHPC), 7(4), 21-32.

D. Li, Zhao, F., & Zheng, M. (2014). Fractal characteristics of cracks and fragments generated in unloading rockburst tests. International Journal of Mining Science & Technology, 24(6), 819-823.

X. Li, et al. (2017). A novel HBase data storage in wireless sensor networks. EURASIP Journal on Wireless Communications and Networking, 2017(1), 45.

C. Liao, Squicciarini, A., & Lin, D. (2017). LAST-HDFS: Location-Aware Storage Technique for Hadoop Distributed File System. IEEE, International Conference on Cloud Computing (pp.662-669). IEEE.

T. H. Ma, et al. (2015). Rockburst characteristics and microseismic monitoring of deep-buried tunnels for Jinping II Hydropower Station.Tunnelling and Underground Space Technology 49, 345-368.

A. Novak, et al. (2015). Nonlinear acoustics and acoustic emission methods to monitor damage in mesoscopic elastic materials. Emerging Technologies in Non-Destructive Testing VI.

M. M. Rashid, Gondal, I., & Kamruzzaman, J. (2017). Dependable large scale behavioral patterns mining from sensor data using Hadoop platform. Information Sciences, 379, 128-145.

X. Sun, et al. (2017) Experimental investigation of the occurrence of rockburst in a rock specimen through infrared thermography and acoustic emission[J]. International Journal of Rock Mechanics and Mining Sciences, 93: 250-259.

C. Wang, et al. (2010). MapDupReducer:detecting near duplicates over massive datasets. ACM SIGMOD International Conference on Management of Data, SIGMOD 2010, Indianapolis, Indiana, USA, June (pp.1119-1122). DBLP.

Y. Zhang, et al. (2013) .A big data based data storage systems for rockburst experiment. International Journal of Wireless and Mobile Computing 6.5, 463-472.

L. Zhong, et al. (2015).Wavelet energy features of acoustic emission signals under centrifugal pump cavitation conditions. Transactions of the Chinese Society of Agricultural Engineering 31.8 .

H. Zhou, et al. (2015). Analysis of rockburst mechanisms induced by structural planes in deep tunnels. Bulletin of Engineering Geology and the Environment 74.4, 1435-1451.

J. Zhu, et al. (2016). Mobility-assisted big data collecting in wireless sensor networks. International Journal of Distributed Sensor Networks, 12(8).

T. Zhu, et al. (2015).Big Data in Future Sensing. International Journal of Distributed Sensor Networks 2015, 1-2.

T. Zhu, et al. (2015). Emergent Technologies in Big Data Sensing: A Survey. International Journal of Distributed Sensor Networks, 2015, 1-13.

9 NOTES ON CONTRIBUTORS

Yu Zhang received his Doctor's degree of Computer Science and Engineering at Beijing Institute of Technology. Currently, he is Associate Professor and Master Supervisor at School of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture. His research interests include big data, cloud computing, distributed network, artificial intelligence and rockburst.

Yan-Ge Wang is a graduate student in Beijing University of Civil Engineering and Architecture, major in Control Science and Engineering and Computer Science. She graduated from the University in 2017, majoring in electrical and intelligent building, nowadays her research interests include big data, multi-source data fusion and data management. Her tutor researches big data, distributed network, artificial intelligence and rockburst.

Yan-Ping Bai received her doctor's degree of Economics in July 2007 at University of International Business and Economics. She is currently a teacher at college of management, Capital Normal University. Her research interests include Environmental Economics and Public Goods.

Yong-Zhen Li received his Master's degree in Computer Science and Engineering from Beijing Institute of Technology. Currently, he is studying for his Doctor's degree in Beijing University of Technology. His research interests include big data, virtualisation and code obfuscation.

Zhao-Yong Lv received his master's degree of control theory & engineering in March 2011 at Beijing University of Civil Engineering and Architecture. He is currently an experiment teacher at Computing Center. His research interests include Big Data and Control algorithm.

Hong-Wei Ding received her Master's degree of Beijing University of Civil Engineering and Architecture, major in Control Science and Engineering and Computer Science. Her tutor researches big data, distributed network, artificial intelligence and rockburst. Currently, she is an engineer in China IPPR International Engineering CO., LTO.