REVISTA
SAN GREGORIO

http: 10.36097/rsan.v1i44.1607
Artículo Original

Method for Hdd Reliability Multiparametric Assessment

Método para la evaluación multiparamétrica de confiabilidad del disco duro

Authors

Iskandar Nailovich Nasyrov[1], Ildar Iskandarovich Nasyrov[2], Rustam Iskandarovich Nasyrov[3], Bulat Askarovich Khairullin[4]

[1]*Doctor of Economics, Professor, Department of Economics of Enterprises and Organizations, Naberezhnye Chelny Institute, Kazan Federal University, ecoseti@yandex.ru, +79172311451, Author ID 57216019750, ORCID 0000-0003-3293-6965*

[2]*Candidate of Engineering Sciences, Associate Professor, Department of Economics of Enterprises and Organizations, Naberezhnye Chelny Institute, Kazan Federal University, ildarec@mail.ru, Author ID 57213240407, ORCID 0000-0002-0186-2871*

[3]*Senior Lecturer, Department of Economics of Enterprises and Organizations, Naberezhnye Chelny Institute, Kazan Federal University, rinasyrov@gmail.com, Author ID 57204110317, ORCID 0000-0002-4923-4532*

[4]*Candidate of Economic Sciences, Associate Professor, Department of Economics of Enterprises and Organizations, Naberezhnye Chelny Institute, Kazan Federal University, presario87@yandex.ru, Author ID 57204114461, ORCID 0000-0003-4083-2760*

Fecha de recibido: 2020-11-30
Fecha de aceptado para publicación: 2021-02-01
Fecha de publicación: 2021-03-25

Abstract

A method of multiparametric assessment of information storage devices reliability based on the dependence on the operating time of the SMART parameter values characterizing the state of hard magnetic disks in computers is presented. The parameters, with an increase in the values of which the failure probability of disk storage devices increases, are considered. The scientific objective of the study is to establish the importance of these parameters in failed hard drives from the standpoint of developing a method for assessing the reliability of information storage devices of various manufacturers. In the course of the research, the drives of such brands as HGST, Hitachi, Samsung, ST, Toshiba, WDC, operated in the largest commercial data centre Backblaze, were analysed. The analysis revealed the following set of the most important parameters: 5, Reallocated sectors count (number of reallocated sectors), 196, Reallocation event count (number of successful and unsuccessful reallocation attempts), 197, Current pending sector count (number of sectors that are candidates for replacement), 1, Read error rate (the frequency of errors when reading data from the disk), 9, Power-on hours (the number of hours spent in the on state), 7, Seek error rate (the frequency of errors when positioning the magnetic head unit), 3, Spin-up time (the time to spin up the packet disks from rest to operating speed), 10, Spin-up retry count (the number of retries to spin up disks to operating speed if the first attempt was unsuccessful). It is taken into account that the very existence of the considered parameters values depends on the hard drives manufacturer.

Keywords: information, storage, hard disk, reliability, parameter, method.

Resumen
Se presenta un método de evaluación multiparamétrico de la confiabilidad de los dispositivos de almacenamiento de información basado en la dependencia del tiempo de operación de los valores de los parámetros SMART que caracterizan el estado de los discos magnéticos duros en las computadoras. Se consideran los parámetros, con un aumento en los valores de los cuales aumenta la probabilidad de falla de los dispositivos de almacenamiento en disco. El objetivo científico del estudio es establecer la importancia de estos parámetros en discos duros averiados desde el punto de vista de desarrollar un método para evaluar la fiabilidad de los dispositivos de almacenamiento de información de varios fabricantes. En el curso de la investigación, se analizaron las unidades de marcas como HGST, Hitachi, Samsung, ST, Toshiba, WDC, que operan en el centro de datos comercial más grande, Backblaze. El análisis reveló el siguiente conjunto de los parámetros más importantes: 5, Recuento de sectores reasignados (número de sectores reasignados), 196, Recuento de eventos de reasignación (número de intentos de reasignación exitosos y no exitosos), 197, Recuento de sectores pendientes actuales (número de sectores que son candidatos para reemplazo), 1, Tasa de error de lectura (la frecuencia de errores al leer los datos del disco), 9, Horas de encendido (la cantidad de horas pasadas en el estado de encendido), 7, Tasa de error de búsqueda (la frecuencia de errores al colocar la unidad de cabezal magnético), 3, tiempo de giro (el tiempo para girar los discos de paquetes desde el reposo a la velocidad de funcionamiento), 10, recuento de reintentos de giro (el número de reintentos para girar los discos a la velocidad si el primer intento no tuvo éxito). Se tiene en cuenta que la existencia misma de los valores de los parámetros considerados depende del fabricante de los discos duros.

Palabras clave: información, almacenamiento, disco duro, confiabilidad, parámetro, método.

Introduction

The paper considers the change in the operating time of SMART parameters values (self-monitoring, analysis and reporting technology) used for the internal assessment of computer hard disk state, as well as for predicting its possible failure. The selected parameters are critical in the sense that if their values increase, then failure probability of information storage devices on hard disks increases. The scientific task of the study is to establish the importance of these parameters in failed hard drives from the standpoint of developing a method for multi-parameter assessment of information storage devices reliability supplied by various manufacturers.

In the course of the research, the drives used in the largest commercial data centre Backblaze were analysed. These were hard drives of HGST (Hitachi Global Storage Technologies), Hitachi (later HGST), Samsung, ST (Seagate), Toshiba, WDC (Western Digital) brands. The analysis revealed the following set of parameters (in decreasing order of importance): 5, Reallocated sectors count (number of reallocated sectors), 196, Reallocation event count (number of successful and unsuccessful reallocation attempts), 197, Current pending sector count (number of sectors that are candidates for replacement), 1, Read error rate (error rate when reading data from a disk), 9, Power-on hours (the number of hours spent in the on state), 7, Seek error rate (error rate when positioning the magnetic head unit), 3, Spin-up time (time to spin up a disk package from rest to operating speed), 10, Spin-up retry count (number of retries to spin up disks to operating speed if the first attempt was unsuccessful). It is taken into account that the very existence of considered parameters values depends on the manufacturer of the hard drives.

Methods

We analysed the dependences of the parameter values on the operating time of failed data storage devices on hard magnetic disks listed on the Backblaze company website (Backblaze, 2020). 45 SMART parameters of 92,530 drives were considered for 93 models of 6 brands, namely, HGST, Hitachi, Samsung, ST, Toshiba, and WDC for the period from April 10, 2013 to December 31, 2016. It was found that at the end of the study

period 79.58% drives from their total continued to work normally, 14.74% were decommissioned ahead of schedule, and 5.68% were out of order.

In total, information about the semantic meaning of more than 80 SMART parameters out of 256 is available, but most of them are not used by manufacturers (Nasyrov et al., 2019). Therefore, Backblaze specialists recorded only 40 of them in 2013-2014, and since 2015 there were 45 of them with numbers 1-5, 7-13, 15, 22, 183, 184, 187-201, 220, 222-226, 240 -242, 250-252, 254, 255 (22, 220, 222, 224, 226 were added in 2015).

Results and Discussion

For displaying the surface state of hard drives, the best parameter is 5, Reallocated sectors count. It is always used for all drives and is interpreted in the same way by all manufacturers. The proof of the priority of the number of reallocated sectors when assessing the health of a hard disk is presented in (Pinheiro et al., 2007), which shows the results of 100,000 drives study in servers around the world, carried out by Google. It was found that its changes coincide with changes in parameters 1, 196, 197. An example of such a coincidence for a failed hard drive of the HGST brand is shown in Figure 1.

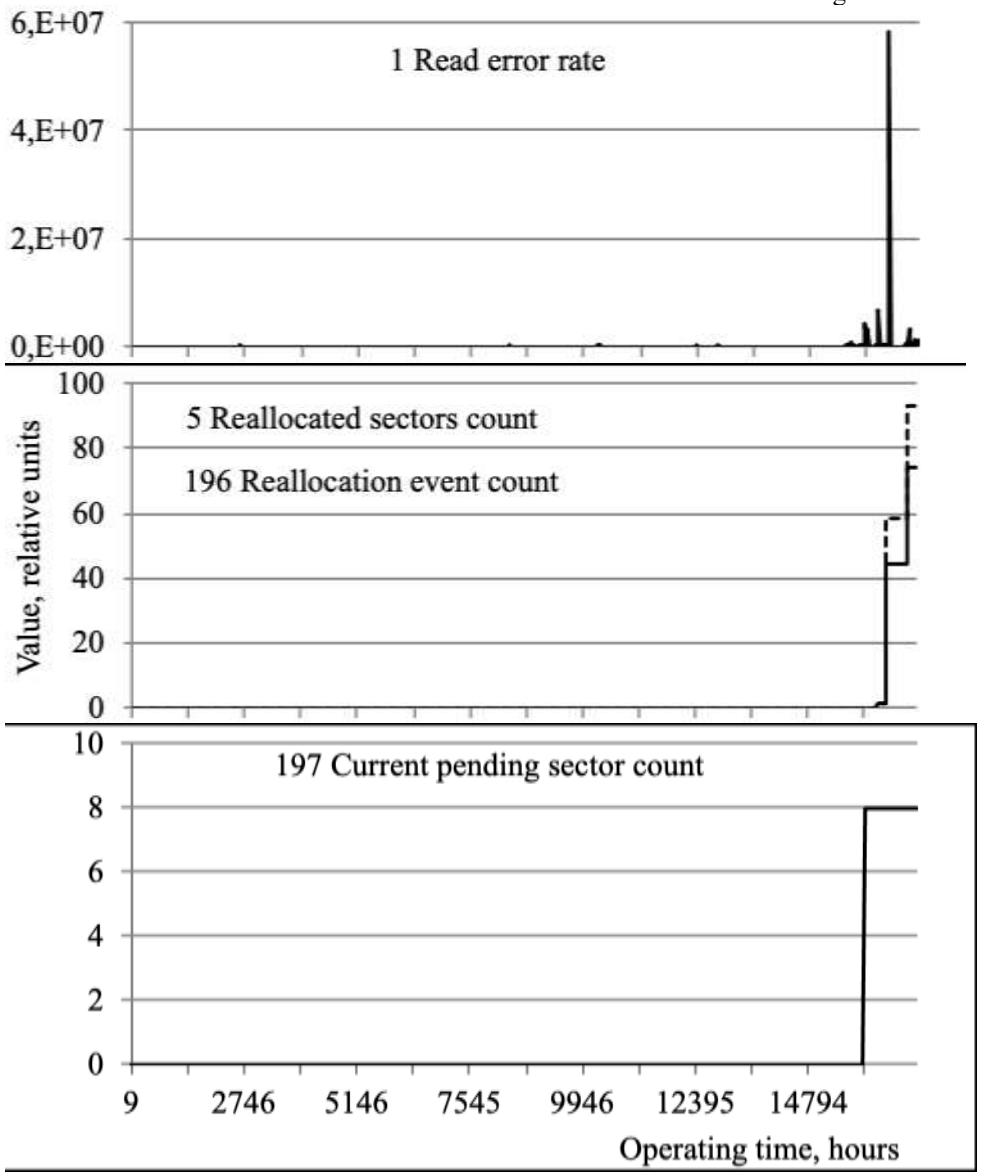

Figure 1. Parameter values 1, 5, 197 (solid line) and 196 (dashed line) depending on the operating time for a failed hard disk of the HGST HMS5C4040ALE640 model with the number PL2331LAGSTGLJ and a capacity of 4 TB

A similar coincidence is observed for failed disks manufactured by Hitachi (Figure 2), Toshiba (Figure 3), and WDC (Figure 4).

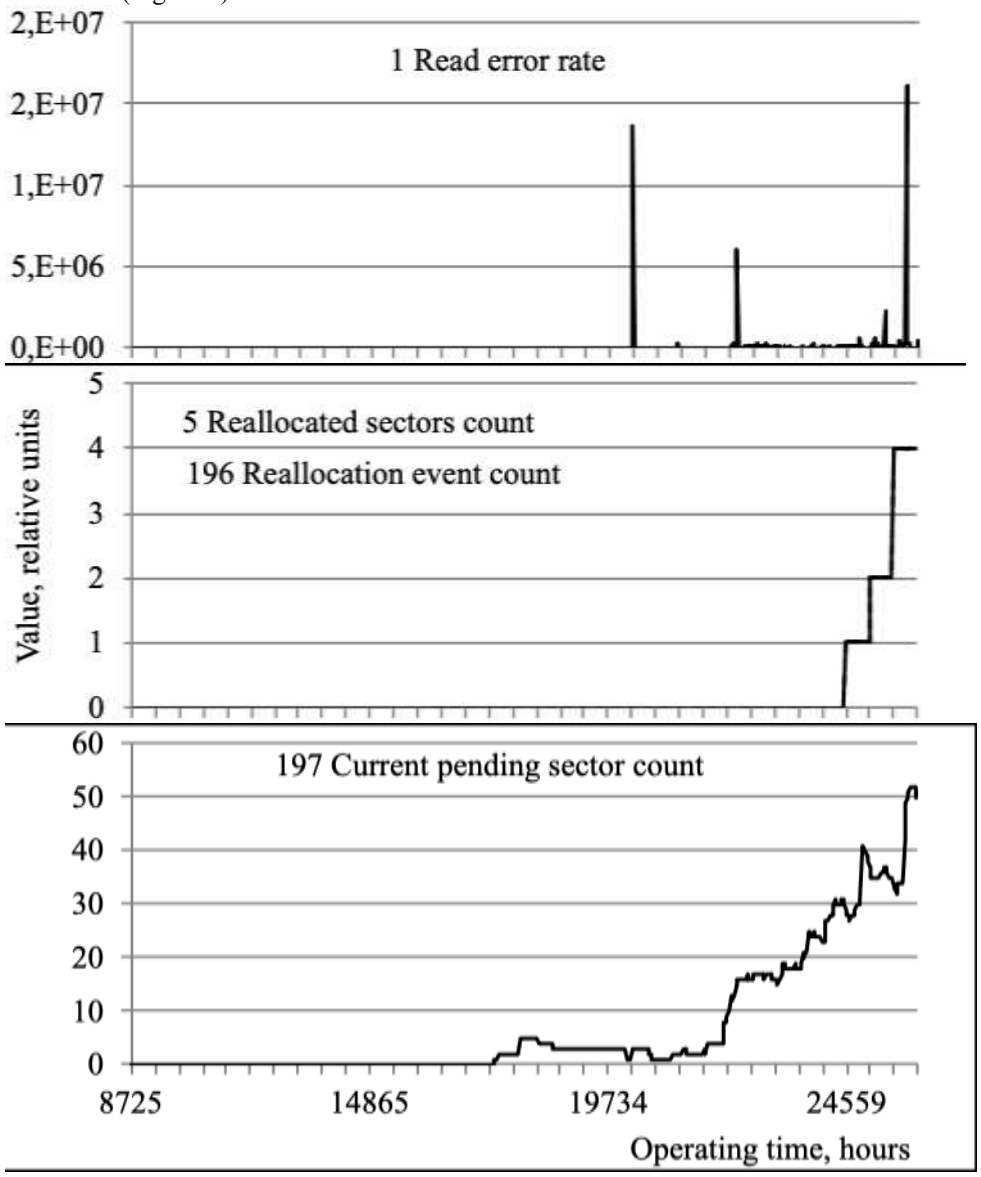

Figure 2. Parameter values 1, 5, 197 (solid line) and 196 (dashed line) depending on the operating time for the failed hard disk of the Hitachi HDS5C3030ALA630 model with the number MJ1311YNG72PXA and a capacity of 3 TB

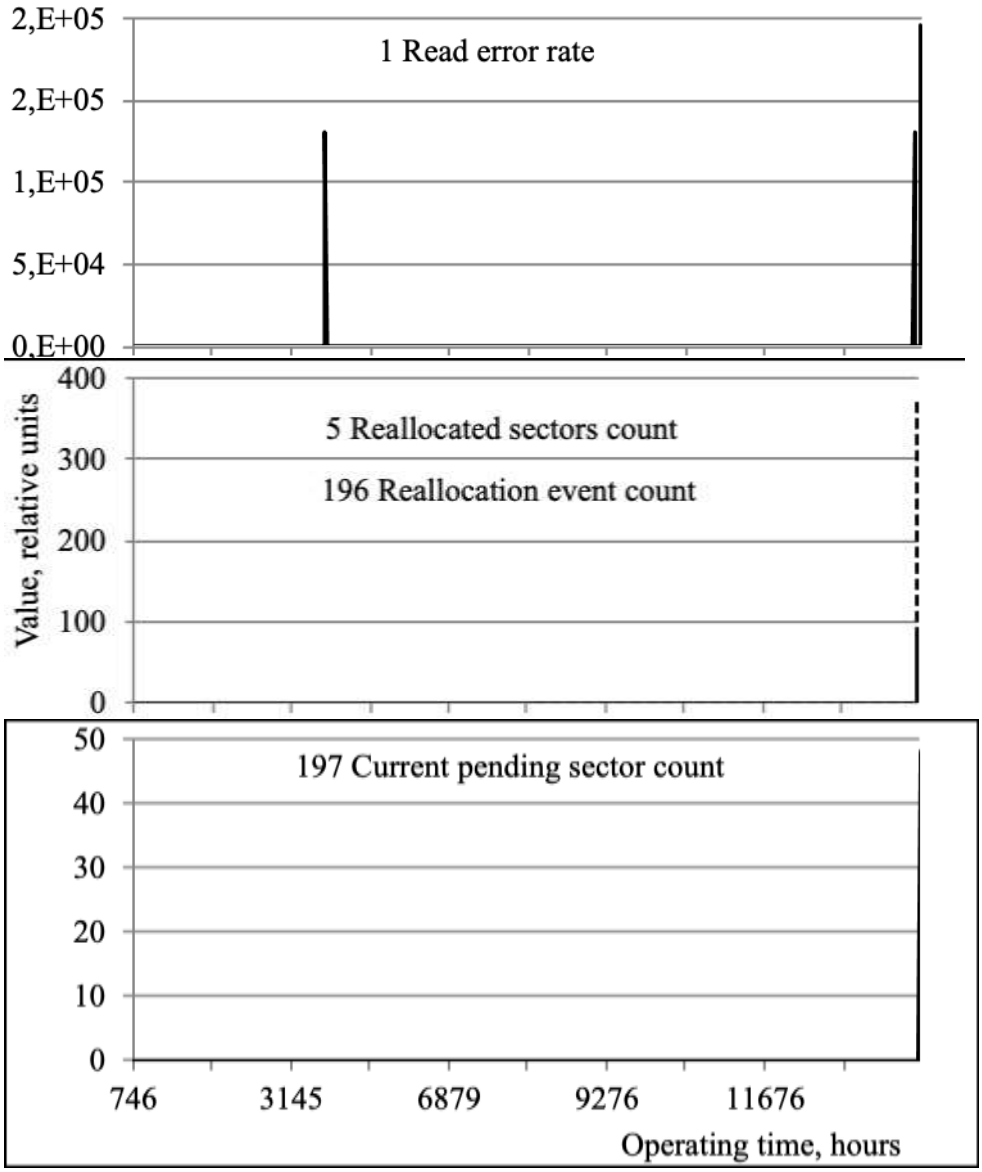

Figure 3. Parameter values 1, 5, 197 (solid line) and 196 (dashed line) depending on the operating time for the failed hard disk of the Toshiba DT01ACA300 model with the number Z252A34AS and a capacity of 3 TB

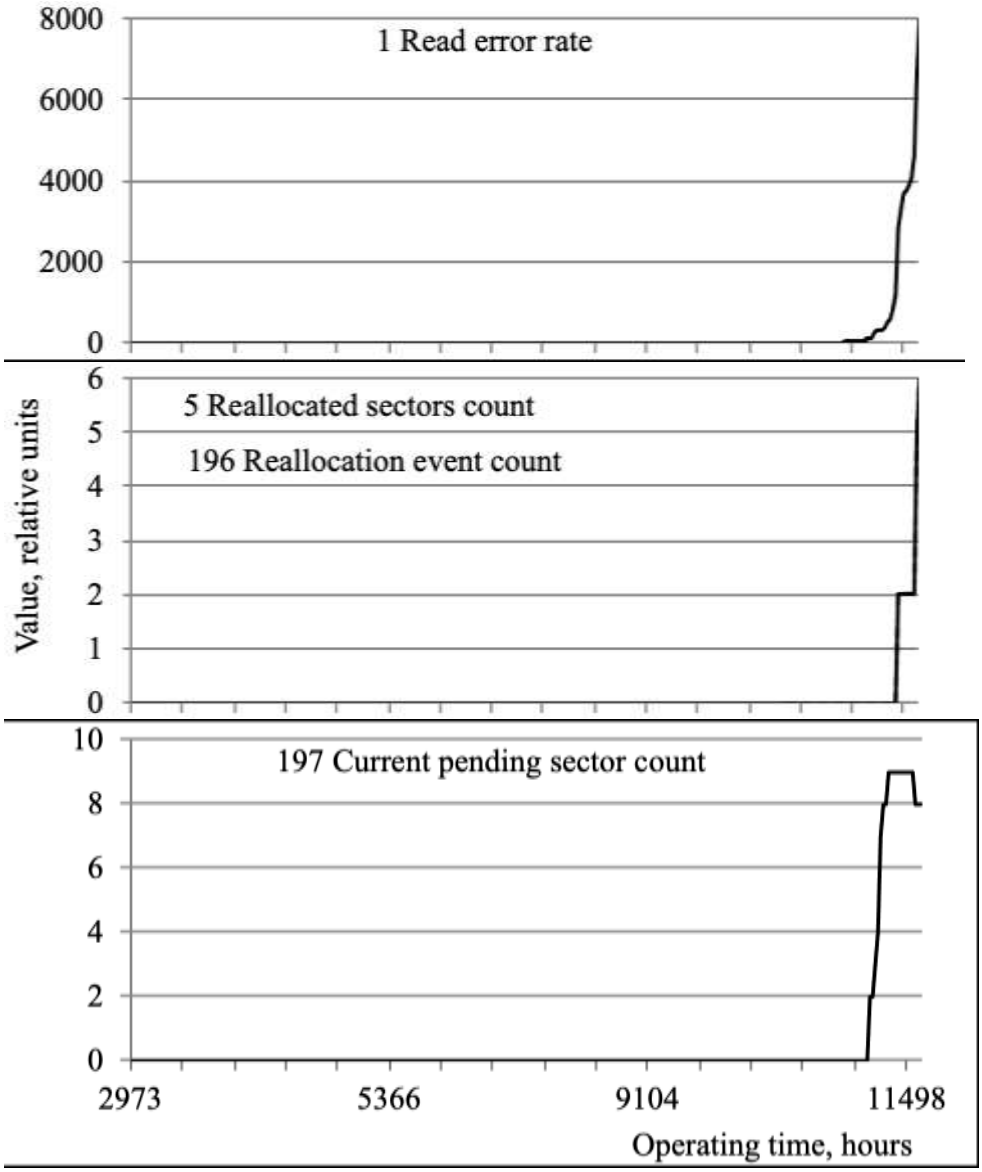

Figure 4. Parameter values 1, 5, 197 (solid line) and 196 (dashed line) depending on the operating time for a failed hard disk of the WDC WD30EFRX model with the number WD-WMC1T0811229 and a capacity of 3 TB

The parameters 5 and 196 were equal to zero that belonged only to the disk from such manufacturer as Samsung that lost its functionality. Nevertheless, the coincidence of changes in parameters 1 and 197 is still observed (Figure 5). Here, parameters 13 and 183 are additionally given in order to indicate another reason for the failure of the drive.

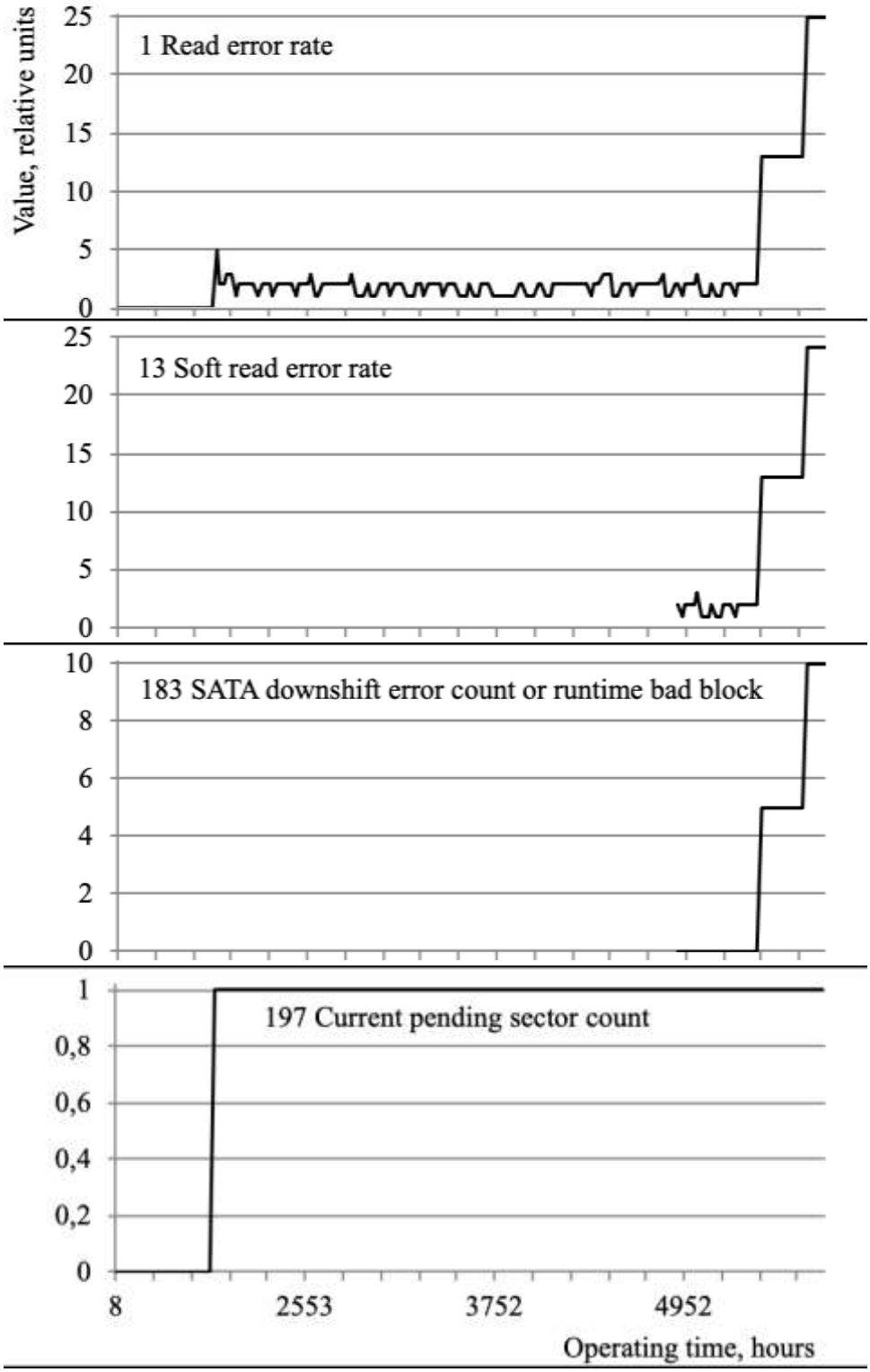

Figure 5. Parameter values 1, 13, 183, and 197 depending on the operating time for the failed hard disk of the Samsung HD154UI model with number S2CHJR0Z900286 and a capacity of 1.5 TB (Nasyrov et al., 2019)

There is a coincidence similar to the previous ones, but without parameter 196, which is absent in all drives from Seagate, and with a chaotic change in parameter 1, for failed drives of the ST brand (Figure 6).

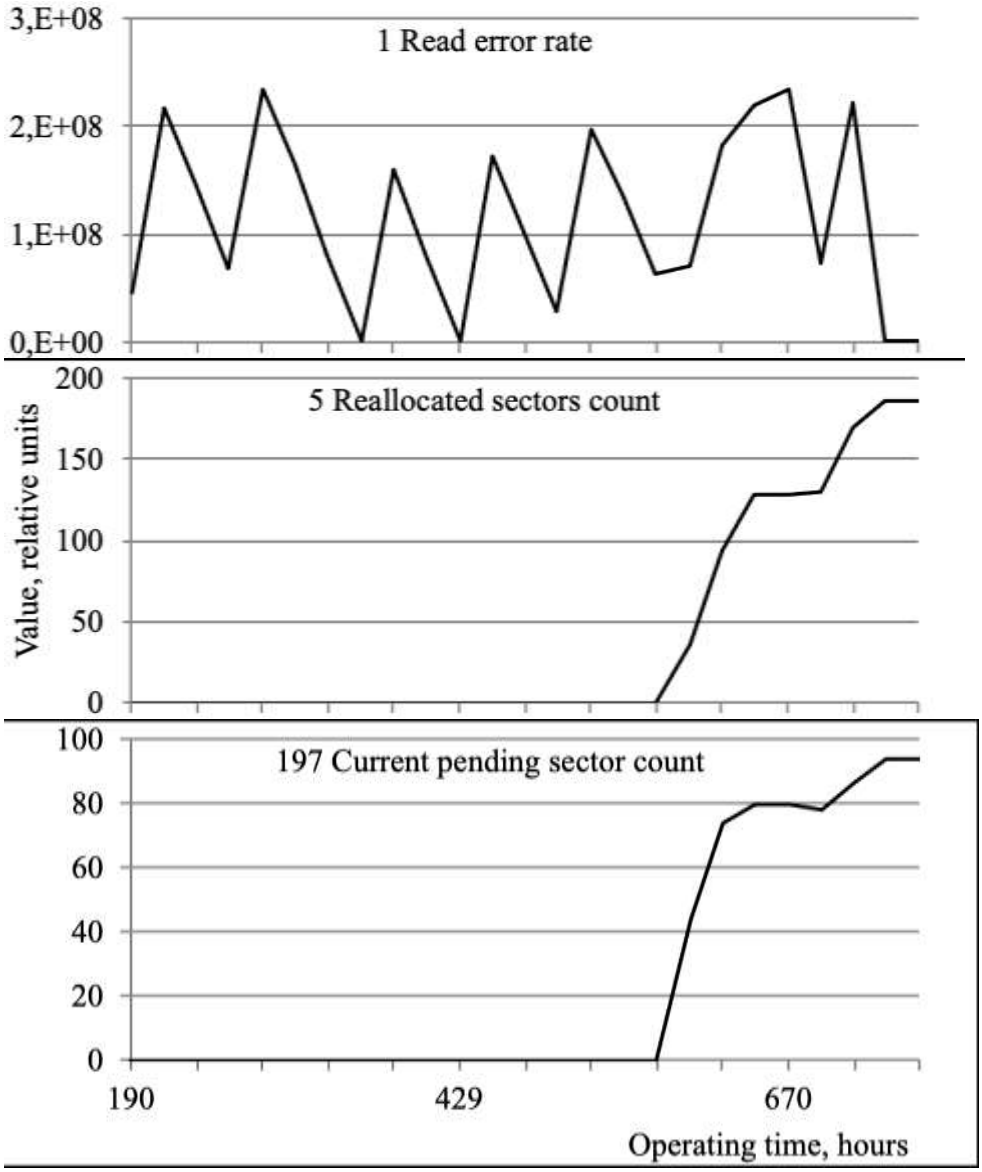

Figure 6. Parameter values 1, 5, 197 via the operating time for the failed hard disk of the ST4000DM000 model with number Z300WWF9 and a capacity of 4 TB

As you can see from Figures 1-6, parameter 1 usually changes earlier than all others. Here it characterizes the appearance of the very first write / read errors. Then parameter 197 is changed, which shows the number of sectors in which these write / read difficulties are observed. Then parameters 196 and 5 are changed with all or only successful reallocation attempts, respectively.

Parameter 1, Read error rate, characterizes the frequency of errors, when reading data from the disk, and the origin of which is caused by the disk hardware. This is the number of internal data corrections made before being issued to the interface. Huge numbers can be produced. Parameter 5, Reallocated sectors count, characterizes the number of reallocated sectors. When the disk detects a write / read error, it marks the sector as "reallocated" and transfers the data to a dedicated spare area. An increase in the value of this attribute may indicate deterioration in the surface condition of the disk platters. Parameter 196, Reallocation event count, shows the number of reallocation operations. The attribute characterizes the total number of attempts to transfer information from the reallocated sectors to the spare area. Both successful and unsuccessful attempts are counted. The parameter 197, Current pending sector count, shows the number of sectors that are candidates for replacement. They have not yet been identified as bad, but reading from them

is different from reading from a stable sector; these are the so-called suspicious or unstable sectors. In the event of a successful subsequent reading of the sector, it is excluded from the list of candidates. In case of repeated erroneous reads, the drive tries to restore it and performs a reallocation operation. An increase in the value of this attribute may indicate physical degradation of the hard disk.

In the analysis, it is proposed to take into account the priority for users of both reliability data and the type of hard disk failures. The highest priority is the group of parameters 5, 196, 197, 1, related to the state of the memory space. These parameters are always available for all drives of all types of almost all manufacturers (196 is not available for Samsung and Seagate (trademark ST) hard drives).
It is imperative to apply the standard definition of reliability as time to failure, which is second in its priority and corresponds to the always available parameter 9, Power-on hours (Nasyrov et al., 2018).

The third priority is the parameter 7, Seek error rate, which characterizes the frequency of errors during positioning of the magnetic head unit (Nasyrov et al., 2019).

The fourth in priority is the group of parameters related to the mechanics of hard disk drives: 10, Spin-up retry count, and 3, Spin-up time (Nasyrov et al., 2019).

The analysis of drive failures by relative (Nasyrov et al., 2018) or absolute (Nasyrov et al., 2018) values of these parameters confirms their importance for developing a method for assessing and predicting the failure of hard drives.
Therefore, we need to use sorting first by parameters 5, 196, 197 and 1, then by 9, then by 7, 3, and 10, when displaying drives by the degree of failure probability based on the meaning and availability of parameter values.

The data on the display are colour-coded for visualization clarity of the drive's state. For colour grading, the boundaries of the failure rate levels are used, taking into account the detected fact of reallocating sectors by whole tracks (divisible by 8).
The developed method provides that it is also possible to estimate the probability of a drive failure using a mathematical model, for example, based on neural networks. At present, this direction is actively studied by various groups of scientists in different countries of the world (Basak et al., 2019; Yiğit et al., 2018; Aussel et al., 2017).

The largest number of failed drives have low or zero parameter values depending on their operating time. Therefore, we propose, firstly, to consider not one parameter, but several to assess the reliability of hard drives; secondly, consider parameter 5 first, then 196, then 197 and 1, then the rest; thirdly, rank disks by parameter values in the order of the specified priority of parameters; fourthly, introduce colour rendering with grades of gradation according to the criteria of the danger of failures.
Hence the essence of the method for assessing the reliability of information storage devices, which includes the following sequence of tasks to be solved:
1) Selection of reliability parameters for information storage devices;
2) Determination of criteria for the danger levels of failures;
3) Formulation of a model for assessing the probability of failure;
4) Development of an algorithm for ranking drives according to the degree of reliability;
5) Drawing up a program for visualizing the danger of their failures.

Summary

It is shown that it is possible to develop a method for a multiparametric assessment of information storage devices reliability taking into account the probability of their failure, on the basis of the most important reliability parameters, the values of which are available for hard magnetic disks of all manufacturers.

The argument to justify the use of these parameters is that exactly the same parameters with practically the same meaning are used in solid-state information storage devices to characterize the state of the memory space (Kingston Technology Corporation. SMART Attribute Details; https://drive.google.com). Naturally, the parameters for positioning the read / write heads or disk rotation mechanisms in solid-state drives are not used, although the numbering remains the same. Therefore, in the future, no additional changes except for the exclusion of parameters 3, 7, and 10 will have to be introduced into the developed method,

with the transition of large data centres to storing information in drives of this type.

Conclusions

Similar studies on the same data with heterogeneous disk groups were carried out in (Rincón et al., 2017), where they searched for universal predictors of failures that could be applied to disks of all brands and models. The main problem was a significant number of SMART parameters, data on which were missing for most brands and models of drives. As a result, the authors had to discard parameters that were missing in at least 90% of disks, after which action only 21 parameters remained.

It was noted in (Mashhadi et al., 2018) that the same SMART parameters may not have exactly the same meaning for different manufacturers. A set of parameters was also defined there that correlates with the failure rate of hard drives, including parameters 1, 5, and 197.

The works (Gaber et al., 2017; Mashhadi & Behdad, 2017; Su & Huang, 2018) also used the SMART parameters of the specified data set from the data centre of the Backblaze company to determine the intensity and predict the failures of disk storage devices. In (Gopalakrishnan & Behdad, 2017; Botezatu et al., 2016), neural networks were used to predict disk failure. Machine learning algorithms were used in (Chaves et al., 2016; Qian et al., 2015).

Thus, the topic under consideration is relevant, and the method described in the paper has scientific novelty and allows the problem of a multiparametric assessment of the information storage device reliability to be solved using the parameters that are most important from this point of view and are identified as a result of the study.

Acknowledgements

The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University.

References

Backblaze. Hard Drive Data and Stats. URL: https://www.backblaze.com/b2/hard-drive-test-data.html (Accessed: 22.05.2020).

Nasyrov, I.N., Nasyrov, I.I., Nasyrov, R.I., Khairullin, B.A. (2019). Study of Failure Hazard Degree in Large Data Centers. *Helix*, *9*(5), 5345–5349. URL: http://helix.dnares.in/2019/10/31/loss-of-pressure-in-a-smooth-pipe-with-a-pulsating-turbulent-course/

Pinheiro, E., Weber, W. D., & Barroso, L. A. (2007). Failure trends in a large disk drive population. *Proceedings of the 5th USENIX Conference on File and Storage Technologies (FAST'07)*, San Jose, California, USA, 13-16 February, 17–28.

Nasyrov, I. N., Nasyrov, I. I., Nasyrov, R. I., & Khairullin, B. A. (2019). Reallocated Sectors Count Parameter for Analysing Hard Disk Drive Reliability. *Journal of Computational and Theoretical Nanoscience*, *16*(12), 5298-5302.

Nasyrov, I. N., Nasyrov, I. I., Nasyrov, R. I., & Khairullin, B. A. (2018). Dependence of reallocated sectors count on HDD power-on time. *International Journal of Engineering and Technology (UAE)*, *7*(4.7), 200-203.

Nasyrov, R., Nasyrov, I., & Khairullin, B. (2019). Positioning errors indication by seek error rate and other Hdd parameters. *Journal of Advanced Research in Dynamical and Control Systems*, *11*(8), 1797–1805.

Nasyrov, I. N., Nasyrov, I. I., Nasyrov, R. I., & Khairullin, B. A. (2019). Spin Retry Count Relation with Other hdd Parameters. *Journal of Computational and Theoretical Nanoscience*, *16*(12), 5303-5306.

Nasyrov, I.N., Nasyrov, I.I., Nasyrov, R.I., Khairullin, B.A. (2018). Data mining for information storage reliability assessment by relative values. *International Journal of Engineering and Technology (UAE)*, *7*(7), 204–208. DOI: 10.14419/ijet.v7i4.7.20545. URL: https://www.sciencepubco.com/index.php/ijet/article/view/20545

Nasyrov, I. N., Nasyrov, I. I., Nasyrov, R. I., & Khairullin, B. A. (2018). Parameters selection for information storage reliability assessment and prediction by absolute values. *Journal of Advanced Research in Dynamical and*

Control Systems, *10*(2 Special Issue), 2248-2254.

Basak, S., Sengupta, S., & Dubey, A. (2019, June). Mechanisms for integrated feature normalization and remaining useful life estimation using lstms applied to hard-disks. In *2019 IEEE International Conference on Smart Computing (SMARTCOMP)* (pp. 208-216). IEEE.

Yigit, I. O., Arslan, Ş. Ş., & Zeydan, E. (2018, May). A visualization platfom for disk failure analysis. In *2018 26th Signal Processing and Communications Applications Conference (SIU)* (pp. 1-4). IEEE.

Aussel, N., Jaulin, S., Gandon, G., Petetin, Y., Fazli, E., & Chabridon, S. (2017, December). Predictive models of hard drive failures based on operational data. In *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)* (pp. 619-625). IEEE.

Kingston Technology Corporation. SMART Attribute Details. URL: https://drive.google.com/file/d/0B2RTg5K2_LNEZWpERlBjQ3BaM00/view (Accessed: 22.05.2020).

Micron Technology, Inc. Technical note: Client SATA SSD SMART Attribute Reference. URL: https://drive.google.com/file/d/0B2RTg5K2_LNETEF5aGhIVDgtNkU/view (Accessed: 22.05.2020).

Rincón, C. A., Pâris, J. F., Vilalta, R., Cheng, A. M., & Long, D. D. (2017, July). Disk failure prediction in heterogeneous environments. In *2017 International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS)* (pp. 1-7). IEEE.

Mashhadi, A. R., Cade, W., & Behdad, S. (2018). Moving towards real-time data-driven quality monitoring: a case study of hard disk drives. *Procedia Manufacturing*, *26*, 1107-1115.

Gaber, S., Ben-Harush, O., & Savir, A. (2017, May). Predicting hdd failures from compound smart attributes. In *Proceedings of the 10th ACM International Systems and Storage Conference* (pp. 1-1).

Mashhadi, A. R., & Behdad, S. (2017). Optimal sorting policies in remanufacturing systems: Application of product life-cycle data in quality grading and end-of-use recovery. *Journal of Manufacturing Systems*, *43*, 15-24.

Su, C. J., & Huang, S. F. (2018). Real-time big data analytics for hard disk drive predictive maintenance. *Computers & Electrical Engineering*, *71*, 93-101.

Kumare Gopalakrishnan, P., & Behdad, S. (2017, August). Usage of product lifecycle data to detect hard disk drives failure factors. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* (Vol. 58165, p. V004T05A031). American Society of Mechanical Engineers.

Botezatu, M. M., Giurgiu, I., Bogojeska, J., & Wiesmann, D. (2016, August). Predicting disk replacement towards reliable data centers. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 39-48).

Chaves, I. C., de Paula, M. R. P., Leite, L. G., Queiroz, L. P., Gomes, J. P. P., & Machado, J. C. (2016, October). Banhfap: A bayesian network based failure prediction approach for hard disk drives. In *2016 5th Brazilian Conference on Intelligent Systems (BRACIS)* (pp. 427-432). IEEE.

Qian, J., Skelton, S., Moore, J., & Jiang, H. (2015, August). P3: Priority based proactive prediction for soon-to-fail disks. In *2015 IEEE international conference on networking, architecture and storage (NAS)* (pp. 81-86). IEEE.

Iskandar Nailovich Nasyrov

He received his first higher education at Novosibirsk State University with a degree in Physics. He defended his thesis of a candidate of physical and mathematical sciences at the Institute of Nuclear Physics at the Academy of Sciences of the Uzbek SSR. He received his second higher education at the Kama Polytechnic Institute with a degree in Accounting and Audit. He defended his doctoral thesis in economics at the Moscow Academy of State and Municipal Administration. Now he is a professor at the

Naberezhnye Chelny Institute (branch) of the Kazan (Volga Region) Federal University.

Ildar Iskandarovich Nasyrov
He graduated from the St. Petersburg State Forestry Academy named after S.M. Kirov with a degree in Forest Engineering. He graduated from the postgraduate study of the Kama State Engineering and Economic Academy with a degree in Automation and Control of Technological Processes and Production (in Mechanical Engineering). He defended his thesis of candidate of engineering sciences at the Izhevsk State Technical University named after M.T. Kalashnikov. He has his work experience as the head of the department of technical systems and telecommunications in a bank. Now he is an assistant professor at the Naberezhnye Chelny Institute (branch) of the Kazan (Volga Region) Federal University.

Rustam Iskandarovich Nasyrov
He graduated from the Kama State Engineering and Economic Academy with a degree in Automated Information Processing and Management Systems. He also graduated from a postgraduate courses in the specialty "System analysis, management and information processing (engineering sciences)". He has experience as a project manager at an enterprise. He works now as a senior lecturer at the Naberezhnye Chelny Institute (branch) of the Kazan (Volga Region) Federal University.

Bulat Askarovich Khairullin
He graduated from the Kama State Engineering and Economic Academy with a degree in Finance and Credit. He also graduated from the postgraduate courses in the specialty "Economics and management of the national economy (business economics)". He defended his candidate degree thesis in Economics at the Kazan (Volga Region) Federal University. Now he works as an assistant professor at the Naberezhnye Chelny Institute (branch) of the Kazan (Volga Region) Federal University.