

Disk Failure Prediction Model for Information Systems based on SMART Technology

Yin Yang^{1,a}, *Wei Liang^{2,b} and Wenyi Li^{1,c}

¹School of Accounting, Wuhan Textile University, Wuhan, China

²The 722 Institute of China Shipbuilding Industry Corporation, Wuhan, China

^acs_yangyin@hust.edu.cn, ^b38863835@qq.com, ^cnewsone260@126.com

Keywords: SMART; failure prediction; disk reliability; prediction model; reliability model

Abstract. This paper proposes disk failure prediction model using disk SMART technology for information systems, which is based on the actual data, threshold and attribute value of the SMART parameters to establish parameter reliability model, through theoretical discussion and practical test and analysis to choose the weight of important parameters, according to the parameter reliability model and the parameter weights to create disk reliability model. Final, disk reliability failure threshold is determined by the analysis the SMART parameter actual data and parameter weights case, combined with the disk reliability model and disk reliability failure threshold to classify the level of disk reliability. Experimental results show that this technology can satisfies the actual usage and achieves the prediction result.

Introduction

Though disk is a reliable data and information storage equipment in information systems, and its broken-down rate every year is merely 0.3%-3% [1]. Once it breaks down, catastrophes like data losses would probably happen whose cost has far surpassed the cost of replacing disk hardware. In order to improve the reliability of storage system, RAID technology emerges as the times require [2]. But the reconstruction method needs a longer time to recover data. When member disk number increases, disk array controller becomes the bottle neck; disk read-write speed can't reach the peak value, so data recovery time would further increase. Moreover there is probably a second disk failure, which will then surpass possibly tolerance ability of the RAID, and will thus cause the data loss [3].

Consequently, this paper proposes disk failure prediction model using SMART technology. When systems run normally, it can monitor the running conditions including disk, array and computer-case. We first create parameter reliability model according to the actual data (AD), threshold (TH) and attribute value (AV) of disk SMART parameters, and analyze characteristics and significance of SMART parameters from different vendors, select the important parameters and determine the weight of important parameters. In terms of parameter reliability model and the weight of important parameters, the disk reliability model can be created; then the value of parameter reliability model and disk reliability model can be got. Finally, by analyzing the number of parameters whose DT has changed and the weight of parameters whose DT has changed, the disk reliability failure threshold can be confirmed, we compare the value of disk reliability model with disk reliability failure threshold, thereby the potential predictable disk failures can be predicted.

Disk failure prediction model

Parameter reliability model. This section defines SMART parameter Reliability Model as SRM. Therefore, the computational formula is $SRM = 100 \times ((AD - TH) / (AV - TH))$. The TH is designated by the disk manufacturers and calculated by specific formulas. If one AV is under corresponding threshold, it means the disk will become unreliable. The AV is the disk's default maximum normal

value. The AD during the operation of the disk's every detection items. Many items are aggregate-value. To the formula of SRM, AV and TH values of any parameter does not change after the disk leaves factory, but AD has constantly been changing. Since AV and AD can be compared and analyzed with TH at the same time, therefore, the comparison between AD and TH is the numerator while the comparison between AV and TH is the denominator.

Disk reliability model. The establishment of disk reliability model needs to choose SMART parameter that every manufacturer generally uses and possesses actual threshold as shown in Table 1. Disk manufactures set threshold with actual data, making the actual data of the parameter could easily reach the threshold, indicating that these parameters possessing huge influence on the disk's reliability.

Table 1 SMART parameter with actual threshold.

ID	Parametric description	Attribute value	Threshold
01	Raw Read Error Rate (RER)	200	6
04	Start/Stop Count (SSC)	100	20
05	Reallocated Sectors Count (RSC)	100	36
07	Seek Error Rate (SER)	100	30
0A	Spin Retry Count (SRC)	100	97
0C	Power Cycle Count (PCC)	100	20
BE	Airflow Temperature (WDC)	100	45

Because of the different characteristics possessed of SMART parameters, analysis of the importance of parameters needs to be conducted so as to confirm the weight of these significant parameters. By the discussions of relevant researches and observations on the changing of AD in practical tests, we can be sure that RSC, SRC, RER, PCC and SER have a huge influence on disk reliability and we can compare actual data with threshold. We suppose that five important parameters respectively represent the abbreviations and their reliability value. According to the above SMART parameter weight analysis and test result analysis, we confirm that RSC weight is about 20%; SRC weight is about 20%; RER weight is about 40%; PCC weight is about 10%; SER weight is about 10%. Therefore, the computational formula of Disk Reliability Model is $DRM = 0.2 \times RSC + 0.2 \times SRC + 0.4 \times RER + 0.1 \times PCC + 0.1 \times SER$.

These parameters weight determined by two factors: technical report [4, 5] and research paper [3, 6] analyze and discuss the importance of SMART parameters; the statistical results of the AD of disk SMART parameter have changed in the actual test. If technical report and research paper analyze and discuss the SMART parameter is very important, and the amplitude of variation of the AD is large, this SMART parameter has a relatively greater weight (for example RER). If technical report and research paper analyze and discuss the SMART parameter is not very important, and the amplitude of variation of the AD is relatively small, these SMART parameters have a relatively smaller weight (for example PCC and SER). And, if technical report and research paper analyze and discuss the SMART parameter is relatively important, and the amplitude of variation of the AD is relatively large, these SMART parameters have a relatively big weight (for example RSC and SRC).

Disk reliability failure threshold

The value's formula mode is defined as DRFT (Disk Reliability Failure Threshold). The confirmation of the DRFT value is determined by two elements:

- (1) Among the five SMART parameters, which parameter's AD has changed

As we known that changes of SMART parameter actual data indicates that the parameter has negative influence on the disk reliability. If several parameter actual data change at the same time, more active negative influence would be made on the disk reliability.

We count the changes of parameter actual data with the method in Table 2 mainly in consideration that during the early use of the disk, SMART parameter reliability is very high, so some slight changes would not cause changes on the disk reliability. When every parameter actual data is close to

the threshold, the slight change of the value would cause the parametric reliability and disk reliability close to the DRFT.

Table 2 Five SMART parameter actual data changing processes and parametric reliability model.

AD changing serial process		1	2	3	4	5	6	7	8	9	10
RSC	AD	100	90	80	70	60	55	50	45	40	36
	TH	36	36	36	36	36	36	36	36	36	36
	AV	100	100	100	100	100	100	100	100	100	100
	SRM	100	84	69	53	38	30	22	14	6	0
SRC	AD	100	100	100	99	99	99	98	98	98	97
	TH	97	97	97	97	97	97	97	97	97	97
	AV	100	100	100	100	100	100	100	100	100	100
	SRM	100	100	100	67	67	67	50	50	50	0
RER	AD	200	180	150	110	80	50	30	20	10	6
	TH	6	6	6	6	6	6	6	6	6	6
	AV	200	200	200	200	200	200	200	200	200	200
	SRM	100	90	74	54	38	23	12	7	2	0
PCC	AD	100	90	80	70	60	50	40	30	25	20
	TH	20	20	20	20	20	20	20	20	20	20
	AV	100	100	100	100	100	100	100	100	100	100
	SRM	100	88	75	63	50	38	25	13	6	0
SER	AD	100	90	80	70	60	50	45	40	35	30
	TH	30	30	30	30	30	30	30	30	30	30
	AV	100	100	100	100	100	100	100	100	100	100
	SRM	100	86	71	57	43	29	21	14	7	0

(2) SMART parametric weight conditions with AD changes

The changes happened on SMART parameter actual data have some impact on the disk reliability; meanwhile, different values of SMART parameter weight of actual data can also have different impact on the disk reliability.

$$\begin{aligned}
 DRM &= 0.2 * 100 + 0.2 * 100 + 0.4 * 100 + 0.1 * 100 + 0.1 * 100 = 100 \\
 DRM &= 0.2 * 84 + 0.2 * 100 + 0.4 * 100 + 0.1 * 100 + 0.1 * 100 = 97 \\
 DRM &= 0.2 * 69 + 0.2 * 100 + 0.4 * 100 + 0.1 * 100 + 0.1 * 100 = 94 \\
 DRM &= 0.2 * 53 + 0.2 * 100 + 0.4 * 100 + 0.1 * 100 + 0.1 * 100 = 91 \\
 DRM &= 0.2 * 38 + 0.2 * 100 + 0.4 * 100 + 0.1 * 100 + 0.1 * 100 = 88 \\
 DRM &= 0.2 * 30 + 0.2 * 100 + 0.4 * 100 + 0.1 * 100 + 0.1 * 100 = 86 \\
 DRM &= 0.2 * 22 + 0.2 * 100 + 0.4 * 100 + 0.1 * 100 + 0.1 * 100 = 84 \\
 DRM &= 0.2 * 14 + 0.2 * 100 + 0.4 * 100 + 0.1 * 100 + 0.1 * 100 = 83 \\
 DRM &= 0.2 * 6 + 0.2 * 100 + 0.4 * 100 + 0.1 * 100 + 0.1 * 100 = 81 \\
 DRM &= 0.2 * 0 + 0.2 * 100 + 0.4 * 100 + 0.1 * 100 + 0.1 * 100 = 80
 \end{aligned}$$

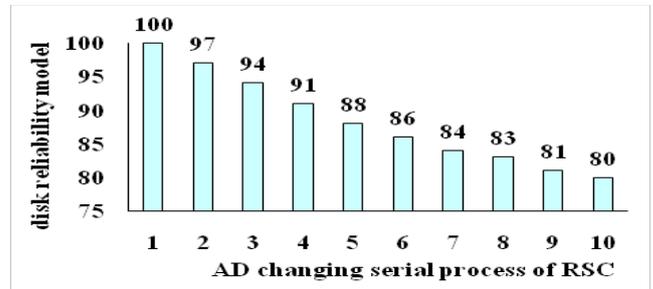


Fig. 1. The value of DRM changing with the AD of parameter RSC.

According to the two elements, we can come to disk reliability failure threshold. We take only one parameter changing situation as example to introduce the calculation of the DRFT. For example, only when RSC parameter actual data changes, the reliability of other SMART parameters are 100, while reliability of RSC decreases from 100 to 36 with actual data (Table 2). According to the formula of DRM, when RSC actual data decreases from 100 to 36, and other SMART parameter reliabilities are 100, the changing process of disk reliability is shown as RSC in Fig. 1.

Evaluation methodology

SMART parameters tested are RSC, SRC, RER, PCC and SER. According to the formulas of SRM and DRM, we give the reliability results of the ten disks, the test results are listed in Table 3. From the analysis in this thesis, situation in this chart belongs to the changing situations of RER and SER (changes of parameter PCC is only 1, it does not belong to parameters in changing). Therefore, we can confirm that the DRFT in this situation is 50. Combining disk reliability classification principles and the comparison between disk reliability model and disk reliability failure threshold, we can come to the following conclusion: the DRFT and 10% of its value (This percentage can be increased or

decreased according to the situation) is $50+50*10\%=55$. If the scope of the disk reliability is 50-55, then the disk is judged as low reliable disk (disk 10); the DRFT plus its 30% , the warning threshold is $50+50*30\%=65$. If the scope of disk reliability is 55-65, then the disk is judged as medium reliable disk (disk 1 and 9); lastly, if the scope of disk reliability is 65-100, then the disk is judged as high reliable disk (disk 2 and 8).

Table 3 The test results of disk failure prediction model

Disks number	RSC			SRC			RER			PCC			SER			DRM
	AD	TH	AV													
D ₁	100	36	100	100	97	100	40	6	200	99	20	100	47	30	100	59
D ₂	100	36	100	100	97	100	117	6	200	99	20	100	51	30	100	75
D ₃	100	36	100	100	97	100	117	6	200	99	20	100	85	30	100	80
D ₄	100	36	100	100	97	100	114	6	200	100	20	100	84	30	100	79
D ₅	100	36	100	100	97	100	105	6	200	99	20	100	58	30	100	74
D ₆	100	36	100	100	97	100	118	6	200	99	20	100	90	30	100	81
D ₇	100	36	100	100	97	100	113	6	200	99	20	100	78	30	100	78
D ₈	100	36	100	100	97	100	117	6	200	99	20	100	67	30	100	77
D ₉	100	36	100	100	97	100	53	6	200	99	20	100	65	30	100	64
D ₁₀	100	36	100	100	97	100	23	6	200	99	20	100	37	30	100	54

Because disk reliability failure threshold (50) implies that the disk reaches the state of data lost, if the DHD of one disk is 50, then this disk can be considered to be a failure disk, in this condition the data in this disk cannot be recovered by active data migration technology. So it should make system send early-warning message before the value of DHD get to but close to the DRFT, and using data migration technology to protect the data in the disk. The value which is close to the DRFT is calculated from the DRFT plus a percentage, the percentage in this paper is 10%, but this percentage can be increased or decreased according to the situation. If storage system requires a higher reliability, this percentage could be increased so that the disk DHD could reach the DRFT more easily, and the early-warning message can be sent more easily.

Summary

This paper proposes a new disk failure prediction model based on SMART technology. This model create parameter reliability model according to the actual data (AD), threshold (TH) and attribute value (AV), then in terms of parameter reliability model and the weight of important parameters, the disk reliability model can be created. According the realization of disk reliability failure threshold, we can judge whether disk is failure or not. The model can effectively predict system fault in advance. Numerical results using many disks show that disk failure prediction model can effectively improve information systems reliability.

Acknowledgements

This work was financially supported by the Natural Science Foundation of Hubei Province (2015CFB235).

References

- [1] W. Jiang, C. Hu, Y. Zhou and A. Kanevsky: ACM Transactions on Storage Vol.4 (2008), p. 1-25
- [2] D. A. Patterson, G. Gibson and R. H. Katz: ACM SIGMOD Record Vol. 17 (1988), p. 109-116
- [3] Q. Xin, S. Cruz, E. L. Miller, T. Schwarz and D. E. E. Long, in: Proceeding 20th IEEE/11th NASA Goddard Conf. on Mass Storage Systems and Technologies (2003).
- [4] Information on <http://en.wikipedia.org/wiki/S.M.A.R.T>

- [5] H. D. S. Hungary, Technical Committee T13. (2007)
- [6] E. Pinheiro, W. D. Weber and L. A. Barroso, in: Proceeding 5th USENIX Conference on File and Storage Technologies (2007).