An Analysis on Security and Reliability of Storage Devices

Ana-Maria DINCĂ, Gabriel PETRICĂ, PhD
Faculty of Electronics, Telecommunications and Information Technology,
National University of Science and Technology POLITEHNICA Bucharest, Romania
ana_maria.dinca@stud.etti.upb.ro, gabriel.petrica@upb.ro

Abstract

The secure storage of information is an essential objective for companies, especially if that information has a classification level that requires medium or maximum protection. This paper analyzes two components of dependability: ensuring the security of backup data must be complemented with the analysis of the reliability of storage media. For this, S.M.A.R.T. technology provides information about the wear of a storage unit (magnetic, optical or flash memory) and allows the prevention of data loss when the storage equipment is nearing the end of its useful life.

Index terms: backup security, data storage, information classification, reliability, S.M.A.R.T. technology

1. Classification of information

Classification of data into well-defined categories has long been a process left solely to the discretion of the user, but it can now be automated within organizations, establishing processes that allow users to categorize the documents they create, send or modify. Alternatively, organizations can classify their existing data using a process of scanning file structures and reporting the results. In October 2022, the ISO/IEC 27001:2022 [1] standard was published, whereby information is classified into following 4 categories, depending on the required level of protection (Figure 1):

- *Public*: the information is intended for the public and can be made public without implications for the company (no impact of disclosure / security breaches). Information integrity is important, but not vital.
- "*Internal Only*" use (medium sensitivity): access to information is limited only from within the organization and must be protected against external access. Unauthorized access could impact the operational effectiveness of the organization, cause significant financial loss, provide significant growth to a competitor, or cause a major decrease in customer confidence. Information integrity is vital.
- *Restricted* (high degree of sensitivity): information received from customers, in any form, for processing in production by the company. The information must not be changed in any way without the written permission of the customer. The highest levels of integrity, confidentiality and availability are necessary and vital.
- *Highly Confidential* (secret): the information collected and used by the organization in carrying out the activity: staff hiring, authentication and fulfillment of customer requirements, management of all aspects related to financial aspects, etc. Access to this information is highly restricted within the organization. The highest levels of integrity, confidentiality and availability are necessary and vital.

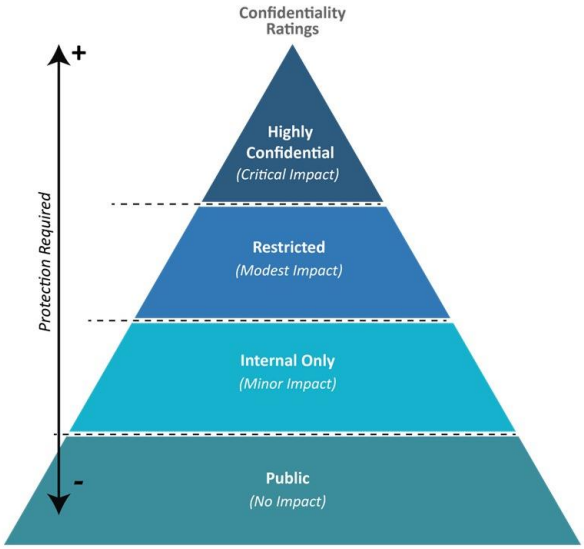

Fig. 1. Classification of information [2]

Every computer system (both servers and workstations) should be protected against the loss of information confidentiality (C), integrity (I) and availability (A), the three components (C-I-A Triad) that define information security according to ISO 27000.

Determining the protection level for a system is mainly based on the type of information stored and processed. Considering the potential impact of a security breach in a system, security levels can be divided into three categories: low, moderate and high [3]. A low level of security may be adopted if the loss of C-I-A of information will have limited consequences (no or minor impact) on the organization's or users' operations and assets. A moderate level of security is chosen if the loss of C-I-A of information will have significant consequences for the operations and assets of the organization or individuals (major impact). This category includes incidents as a result of which the organization can carry out its basic activities, but their efficiency is significantly reduced, there are large financial losses, there are major employee accidents (which do not involve loss of human life). Finally, a high level of security should be chosen if the loss of C-I-A of information will have catastrophic consequences (critical impact) on the operations and assets of the organization or individuals. Such very serious consequences can be the inability of the company to carry out its core activities for a limited period, very large financial losses, or a major employee accident occurring with possible loss of life.

By classifying data, two objectives are achieved: data security is improved and compliance with the regulations in force is ensured. In order to ensure the security of critical data (within the organization, of customers or partners, etc.) it is first necessary to know and understand this data, and for this purpose the following aspects must be analyzed:

- type of sensitive data held - Intellectual Property (IP), medical records (Protected Health Information, PHI), personal data (Personally Identifiable Information, PII), financial or banking information, etc. According to NIST, Personally Identifiable Information is "any information about an individual maintained by an agency, including (1) any information that can be used to distinguish or trace an individual's identity, such as name, social security number, date and place of birth, mother's maiden name, or biometric records; and (2) any other information that is linked or linkable to an individual, such as medical, educational, financial, and employment information" [4].
- where this sensitive data is located;
- who can access / modify / delete this data;
- how the organization's activity will be affected if this data is disclosed, destroyed or modified inappropriately.

To comply with current regulations, organizations must protect specific data such as information about bank cardholders (according to PCI DSS - Payment Card Industry Data Security Standard, 2004), medical records (HIPAA - Health Insurance Portability and Accountability Act, 1996), financial data (SOX - Sarbanes-Oxley Act, 2002) or personal data of European Union residents (GDPR - General Data Protection Regulation, 2016). Identifying and classifying data will locate sensitive types of data, select necessary security controls, and comply with regulatory requirements regarding data search and monitoring. Thus, by complying with the regulations in force, an organization's chances of successfully passing various audits and controlling the flow of sensitive data increase.

2. Ensuring the security and availability of electronic data

The operation of data storage media and the long-term availability of stored data may be affected by factors such as hardware failures, storage device failures, cyber-attacks, natural disasters or human errors. Frequent use of backups reduces the risks associated with these aspects and ensures efficient and complete data recovery in the event of an unfortunate event. However, there are some issues that need to be considered to ensure the integrity and accessibility of these backups over time.

One of the main aspects is the selection of a suitable storage medium for backups. This involves evaluating the various storage technologies available, such as magnetic disks, optical media, flash memory and cloud storage. Each technology has advantages and disadvantages in terms of long-term data reliability and durability. High-quality storage media that are resistant to wear and corrosion provide better protection for stored data. In addition, environmental conditions such as temperature, humidity and exposure to external factors such as sunlight or electromagnetic sources can affect the reliability of storage media [5].

In addition to the physical quality of the storage media, the implementation of a backup system plays a crucial role in ensuring long-term data reliability. Backup systems must be able to perform regular data backups and provide effective recovery options in the event of data loss or corruption. It is also important to consider using multiple storage locations for backups to minimize the risk of total loss in extreme situations such as fire or natural disasters. Data security is another crucial aspect in the long-term reliability of storage media. Backups must be protected against unauthorized access and cyber-attacks. The use of strong encryption and authentication methods, as well as the implementation of strict security and monitoring policies, help protect stored data [6].

However, there are also challenges regarding the reliability of long-term data storage media. Technologies evolve rapidly, and some storage media may become obsolete or incompatible over time. Therefore, planning and constantly updating the storage infrastructure is essential to ensure long-term data compatibility and availability. One of the major challenges is the physical degradation of storage media over time. Storage devices such as hard drives, SSDs or CDs are susceptible to wear and tear as they are exposed to factors such as extreme temperatures, humidity, mechanical shocks or electromagnetic radiation. This can lead to reading errors or data corruption in the long run. Thus, it is important to consider storage conditions and regularly monitor the condition of storage media to prevent potential problems [7].

3. Reliability analysis of storage equipment using SMART technology

In this chapter we compared 4 units for data storage with 2 different technologies: 2 HDD (hard-disk drive, magnetic storage) and 2 SSD (solid-state drive, NAND flash memory). Their technical specifications are presented in Table 1.

Table 1. The analyzed data storage units

Type	SSD	SSD	HDD	HDD
Model	SA400S37240G	860 EVO 1TB	HTS727550A9E364	WD10EADS-00L5B1
Manufacturer	Kingston	Samsung	Hitachi	Western Digital
Drive Capacity	240 GB	1000 GB	500 GB	1000 GB
Controller	Serial ATA 6Gb/s	Serial ATA 6Gb/s (USB)	Serial ATA 3Gb/s	Serial ATA 3Gb/s (USB)
Security Feature	Supported	Supported	Supported	Supported
Enhanced Security Erase	Supported	Supported	Supported	Supported
S.M.A.R.T. feature	Present, Active	Present, Active	Present, Active	Present, Active

Self-Monitoring, Analysis and Reporting Technology (S.M.A.R.T.) is an industry standard that can be used as a reliability prediction indicator for IDE/ATA and SCSI storage units. Proposed by IBM in 1992, S.M.A.R.T. refers to a method of signaling between the sensors in the disk drive and the host computer [8]. The technology monitors the computer's physical disk drives to detect and report various indicators. In this way, failures can be predicted, and users are warned of impending failure of the entire disk drive, allowing for early drive replacement to avoid data loss and/or unexpected service interruptions. S.M.A.R.T. can only warn of predictable errors, which result from slow processes (such as mechanical ones or wear) and can be anticipated by analyzing certain indicators. Unpredictable failures, such as a sudden mechanical failure resulting from an electrical surge, cannot be monitored and analyzed.

S.M.A.R.T. uses a multitude of operating parameters expressed as a *raw value*, which can only take values between certain manufacturer-dependent limits (e.g. 0-100, 0-200 or 0-253) or a *normalized value* calculated with the formula INT [x - (raw_value / max_raw_value) * x], where *max_raw_value* represents the maximum value that the parameter can take. Normalization is used to represent the performance of a device independent of the *max_raw_value* (which is manufacturer dependent). Normalized values are usually mapped so that higher values are better (with some exceptions).

In general, S.M.A.R.T. parameters start from a maximum value and decrease throughout the life of the storage unit. Other terms used by S.M.A.R.T. are "*worst*" - the worst normalized value recorded for a parameter and "*threshold*" - the threshold value that, once reached, triggers an alarm about the need for action (for example, replacing the drive). In Figure 2 we used Hard Disk Sentinel monitoring and analysis software [9] for our 4 data storage units analyzed in this paper.

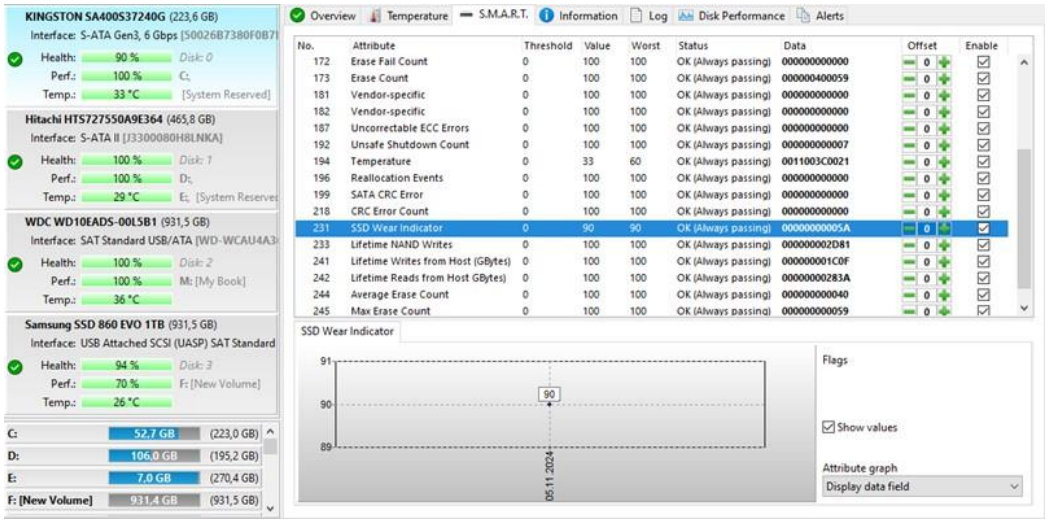

Fig. 2. Threshold, current and worst values for S.M.A.R.T. attributes displayed in Hard Disk Sentinel

Drives can report a S.M.A.R.T. status usually reported as one of two values, typically "*drive OK*" / "*drive fail*" or "*threshold not exceeded*" / "*threshold exceeded*". A "*drive fail*" or "*threshold exceeded*" value indicates that there is a high probability that the unit will fail soon. However, the failure may not be catastrophic, with the S.M.A.R.T. status indicating that the drive will not perform according to the manufacturer's stated specifications (e.g. the drive will run more slowly).

Manufacturers do not necessarily agree on the precise definitions of all attributes and reference values, so in general there are known attributes, supported by IDE and Serial ATA drives, but also non-standard attributes, specific to each manufacturer, used for various purposes (even commercial secrets) [10]. Also, some codes are specific to certain types of drives (fixed magnetic media, flash memory, etc.), and drives may use different codes for the same parameter [11]. Raw values with higher values may be better or worse depending on the attribute and manufacturer.

Using HWINFO [12] we extracted relevant S.M.A.R.T. parameters, specific for the four analyzed disk units (see Table 2):

Table 2. S.M.A.R.T. parameters

SSD Kingston SA400S37240G

Self-Monitoring, Analysis and Reporting Technology (S...	
[01] Raw Read Error Rate:	100/Always OK, Worst: 100
[09] Power-on Hours/Cycle Count:	100/Always OK, Worst: 100 (3914 hours / 163.1 days)
[0C] Power Cycle Count:	100/Always OK, Worst: 100 (Data = 1077, 0)
[94] Unknown	100/Always OK, Worst: 100
[95] Unknown	100/Always OK, Worst: 100
[A7] SSD Protect Mode:	100/Always OK, Worst: 100
[A8] SATA PHY Error Count:	100/Always OK, Worst: 100
[A9] Total Bad Block Count:	100/Always OK, Worst: 100 (Data = 9, 0)
[AA] Bad Block Count:	100/10, Worst: 100 (Data = 8, 0)
[AC] Erase Fail Count (Total):	100/Always OK, Worst: 100
[AD] Erase count:	100/Always OK, Worst: 100 (Data = 4194393, 0)
[B5] Program Fail Count (Total):	100/Always OK, Worst: 100
[B6] Erase Fail Count (Total):	100/Always OK, Worst: 100
[BB] Uncorrectable Errors:	100/Always OK, Worst: 100
[C0] Unsafe Shutdown Count:	100/Always OK, Worst: 100 (Data = 6, 0)
[C2] Temperature:	35/Always OK, Worst: 60 (35.0 °C)
[C4] Later Bad Block Count:	100/Always OK, Worst: 100
[C7] SATA CRC Error Count:	100/Always OK, Worst: 100
[DA] CRC Error Count:	100/Always OK, Worst: 100
[E7] SSD Life Left:	90/Always OK, Worst: 90 (Data = 90, 0)
[E9] Lifetime Writes to Flash:	100/Always OK, Worst: 100 (Data = 11618, 0)
[F1] Host Writes:	100/Always OK, Worst: 100 (Data = 7163, 0)
[F2] Host Reads:	100/Always OK, Worst: 100 (Data = 10259, 0)
[F4] Average Erase Count:	100/Always OK, Worst: 100 (Data = 64, 0)
[F5] Max Erase Count/Total Media Writes:	100/Always OK, Worst: 100 (Data = 89, 0)
[F6] Total Erase Count:	100/Always OK, Worst: 100 (Data = 226416, 0)
Drive Remaining Life	90%

SSD Samsung 860 EVO 1TB

Self-Monitoring, Analysis and Reporting Technology (S...	
[05] Reallocated Sector Count:	100/10, Worst: 100
[09] Power-on Hours/Cycle Count:	94/Always OK, Worst: 94 (29539 hours / 3.37 years)
[0C] Power Cycle Count:	99/Always OK, Worst: 99 (Data = 206, 0)
[B1] Wear Leveling Count:	94/Always OK, Worst: 94 (Data = 99, 0)
[B3] Used Reserved Block Count (Total):	100/10, Worst: 100
[B5] Program Fail Count (Total):	100/10, Worst: 100
[B6] Erase Fail Count (Total):	100/10, Worst: 100
[B7] Runtime Bad Block (Total):	100/10, Worst: 100
[BB] Uncorrectable Error Count:	100/Always OK, Worst: 100
[BE] Airflow Temperature:	73/Always OK, Worst: 53 (27.0 °C)
[C3] ECC Error Rate:	200/Always OK, Worst: 200
[C7] SATA CRC Error Count:	100/Always OK, Worst: 100
[EB] POR Recovery Count:	99/Always OK, Worst: 99 (Data = 101, 0)
[F1] Total Host Writes:	99/Always OK, Worst: 99 (Data = 419226767, 18)
Drive Remaining Life	94%

HDD Hitachi HTS727550A9E364

Self-Monitoring, Analysis and Reporting Technology (S...	
[01] Raw Read Error Rate:	100/62, Worst: 100
[02] Throughput Performance:	100/40, Worst: 100
[03] Spin Up Time:	180/33, Worst: 100 (Data = 2, 22)
[04] Start/Stop Count:	98/Always OK, Worst: 98 (Data = 4652, 0)
[05] Reallocated Sector Count:	100/5, Worst: 100
[07] Seek Error Rate:	100/67, Worst: 100
[08] Seek Time Performance:	100/40, Worst: 100
[09] Power-on Hours/Cycle Count:	80/Always OK, Worst: 80 (8856 hours / 1.01 years)
[0A] Spin Retry Count:	100/60, Worst: 100
[0C] Power Cycle Count:	99/Always OK, Worst: 99 (Data = 2390, 0)
[B7] SATA Interface Downshift / Runtime Bad Block:	100/Always OK, Worst: 100
[B8] End to End Error Detection Count:	100/97, Worst: 100
[BB] Reported Uncorrectable Errors:	100/Always OK, Worst: 100
[BC] Command Timeout Count:	100/Always OK, Worst: 99 (Data = 1, 0)
[BE] Airflow Temperature / Exceed Count:	70/45, Worst: 49 (30.0 °C)
[BF] G-Sense Error Rate:	90/Always OK, Worst: 90 (Data = 2561, 0)
[C0] Power-Off Retract Count:	100/Always OK, Worst: 100 (Data = 1245203, 0)
[C1] Load/Unload Cycle Count:	58/Always OK, Worst: 58 (Data = 428293, 0)
[C4] Reallocation Event Count:	100/Always OK, Worst: 100
[C5] Current Pending Sector Count:	100/Always OK, Worst: 100
[C6] Off-Line Uncorrectable Sector Count:	100/Always OK, Worst: 100
[C7] UltraDMA/SATA CRC Error Rate:	100/Always OK, Worst: 100
[DF] Load/Unload Retry Count:	100/Always OK, Worst: 100

HDD Western Digital WD10EADS-00L5B1

Self-Monitoring, Analysis and Reporting Technology (S...	
[01] Raw Read Error Rate:	200/51, Worst: 200
[03] Spin Up Time:	179/21, Worst: 154 (Data = 6033, 0)
[04] Start/Stop Count:	99/Always OK, Worst: 99 (Data = 1915, 0)
[05] Reallocated Sector Count:	200/140, Worst: 200
[07] Seek Error Rate:	100/Always OK, Worst: 253
[09] Power-on Hours/Cycle Count:	96/Always OK, Worst: 96 (3529 hours / 147.0 days)
[0A] Spin Retry Count:	100/Always OK, Worst: 100
[0B] Calibration Retry Count:	100/Always OK, Worst: 100
[0C] Power Cycle Count:	100/Always OK, Worst: 100 (Data = 465, 0)
[C0] Power-Off Retract Count:	200/Always OK, Worst: 200 (Data = 11, 0)
[C1] Load/Unload Cycle Count:	200/Always OK, Worst: 200 (Data = 1914, 0)
[C2] Temperature:	119/Always OK, Worst: 91 (31.0 °C)
[C4] Reallocation Event Count:	200/Always OK, Worst: 200
[C5] Current Pending Sector Count:	200/Always OK, Worst: 200
[C6] Off-Line Uncorrectable Sector Count:	200/Always OK, Worst: 200
[C7] UltraDMA/SATA CRC Error Rate:	200/Always OK, Worst: 200
[C8] Write/Multi-Zone Error Rate:	200/Always OK, Worst: 200

In Table 2, SSD Life Left attribute for Kingston SA400S37240G (same as SSD Wear Indicator in Figure 2) indicates a wear level of 10% for the analyzed SSD drive. "SSD life left is based on actual usage and takes into account PE cycle consumption (life curve status) and Flash block retirement" [13]. For hard-disk drives, the attributes Power On Time Count (Hard Disk Sentinel) and Power-on Hours (Hwinfo) had values of 8,869 (Hitachi) and 3,532 (WDC), respectively, indicating a health level of 100 % ("the total expected lifetime of a hard disk in perfect condition is defined as 5 years... 43,800 hours)" [14].

4. Conclusions

Information is an asset that, like other important assets of a business or an individual, has a certain value and therefore must be properly protected. The limited resources that organizations invest in data protection lead to the need to develop a taxonomy that allows organizations to identify priorities and develop a plan that optimizes costs and effectively protects sensitive data. Data classification provides a solid foundation for a security strategy that correctly identifies areas of risk both within the network and in the cloud, enables more effective data protection and compliant use.

Therefore, data backup strategies (especially for confidential data) must consider the reliability of the equipment that stores this data. For the data to be safe and available, both scenarios regarding the limitation of access to them and the use of reliable storage media must be analyzed, an aspect to which S.M.A.R.T. technology can make its contribution.

References

[1]. ISO - International Organization for Standardization, ISO/IEC 27001:2022 Information security, cybersecurity and privacy protection — Information security management systems — Requirements, https://www.iso.org/standard/27001
[2]. SecureLink, Information Classification, https://www.securelinkme.net/information-classification
[3]. G. Petrică, S.D. Axinte, I.C. Bacivarov, Dependabilitatea sistemelor informatice, Matrix Rom, 2019, ISBN 978-606-25-0529-5.
[4]. NIST Special Publication 800-122. Guide to Protecting the Confidentiality of Personally Identifiable Information (PII), 2010, https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-122.pdf
[5]. S. Yarrapothu, "Effectiveness of Backup and Disaster Recovery in Cloud - A Comparative study on Tape and Cloud based Backup and Disaster Recovery", pp. 5-40.
[6]. D. Kaeli, "ACM Transactions on Architecture and Code Optimization", 2022, Volume19, Number 3, pp. 123-150, 179-201.
[7]. C. Yan, "Cloud Storage Services"- Thesis, Centria University of Applied Sciences, June 2017, pp. 4-18.
[8]. Samsung, S.M.A.R.T. - Self-Monitoring, Analysis and Reporting Technology, 2014, https://download.semiconductor.samsung.com/resources/others/SSD_Application_Note_SMART_final.pdf
[9]. Hard Disk Sentinel - HDD health and temperature monitoring, https://www.hdsentinel.com/
[10]. S.M.A.R.T. attribute list (ATA), https://www.hdsentinel.com/smart/smartattr.php
[11]. Hetman Software, SMART Parameters and Early Signs of a Failing Hard Disk, 2019, https://medium.com/hetman-software/smart-parameters-and-early-signs-of-a-failing-hard-disk-23dfec568808
[12]. HWINFO, Professional System Information and Diagnostics, https://www.hwinfo.com/.
[13]. Kingston, SMART Attribute Details, 2015, https://media.kingston.com/support/downloads/MKP_306_SMART_attribute.pdf
[14]. Hard Disk Sentinel Help - Power on time, https://www.hdsentinel.com/help/en/54_pot.html