Influence of parameters of the operating system on process speed backup of digital archives of printing industry

N E Proskuriakov¹, B S Yakovlev¹, N N Arkhangelskaia¹, E V Trapeznikov²
¹Tula State University, 92, Lenin Ave., Tula, 300012, Tula, Russia
²Omsk State Technical University, 11, Mira Ave., Omsk, 644050, Russia
E-mail: vippne@mail.ru

Abstract. Results of pilot studies of comparison of process of backup on the Windows and Linux operating systems are presented. Experimental and analytical means revealed the key parameters influencing backup process speed. As showed a research, the most significant parameters are the form of the interface of the software and type of the operating system. Also the limit at which the speed of copying of data remains rather balanced was set. The research showed that "direction" of copying of data is also very important. The software capable to make backup in the environment of operating systems Windows and Linux was developed.

1. Introduction
In modern realities data loss probability owing to activity of malefactors or competitors is high. This problem the most relevant for the printing industry for the reason of a possibility of irrevocable data loss for which the enterprise spent considerable efforts and time. Therefore, it is reasonable to use data backup.

Digital technologies are actively applied in all industries of human activity including printing industry. They are especially important at stages of preparation of models, communication with customers, the organization of printing on demand (via the websites), etc.

These technologies are well-developed show the efficiency. Without them, it is difficult to present a modern cycle of production of books and other printed materials.

The greatest interest of researchers of time is attracted at present by the organization of backup based on hardware data protection, using RAID arrays. It is possible to carry to similar articles [1-3]. However, in our opinion these technologies differ radically from problems of standard backup since in case of computer infection with a virus the infected data, which are on failsafe systems, will save the infected files. Therefore, it will not be possible to restore the data, because there is no place, where it can be restored. Although all files are available, they are not suitable for operation, and their versions cannot be restored.

It is problematic to use RAID an array and for data protection from a banal human factor, for example, from accidental erasing of files, directories. The RAID technology will save changes and will become the reason of very big difficulties in attempts to recover data since on the structure files are stored in a separate look, and are usually divided between disks. It is very hard to aggregate such information, and sometimes it is impossible.

Program realizations in the sphere of backup copies creation in printing industry generally comes down to constant copying of files, relying on opportunities and services of the most operating system
(OS). Taking into account that under OS Linux now, there are practically no ready decisions, the question on realization and optimization of backup copies creation still is not solved yet. Therefore, an objective of this research was determination of the parameters influencing backup process speed for different operating systems that is logical generalization and continuation of earlier carried out article [4].

2. Problem definition
The main problems, which need to be solved during this article, are:
– The maximum reduction of time spent for backup operations;
– Comparison of characteristics of work of the Windows and OS LINUX;
– Identification of the parameters influencing the speed and quality of process of saving and archiving of data;
– Software development and creating of recommendations for specialists in storage areas and data protection.

3. Methods
It is possible to select two main types of data storage systems: the centralized system and decentralized system [5-8].
Data are stored in the centralized systems on the information medium or cloudy storage, independent, independent of others. Backup in such systems it is provided in the program ways.
In decentralized systems of storage, these data break into blocks at the level of bytes (or even at the level of bits) and register at the same time in several carriers. For this purpose respectively, several hard drives and use of RAID technology of arrays are required. This type of a data storage system is especially good at the small volumes of these data as the cost of data backup considerably depends on it.
We incline that it is much simpler to work with data at the centralized approach of archived data storage at the enterprises. For this purpose there are several reasons:
1. In most cases, simple copies of files are necessary for data recovery. This option provides only the centralized data storage option;
2. Not all programs of data recovery are capable to work about raid with arrays. Big expenses in calculations and other approaches are for this purpose necessary considerably, usually similar functions reveal in paid versions of similar products;
3. The centralized data storage system usually is duplicated therefore in case of failure of one of carriers, almost for 100% the second will be saved;
4. Taking into account Paragraph 3 the cost of decentralized systems and RAID since their cost considerably exceeds the cost of normal hard drives repeatedly increases.
It is also possible to note that according to the standard theory, the systems, which do not have interfaces, work quicker than "desktop" options that belongs, also to OS.
In earlier article in the experiment on standard OS Windows the results were received, many questions were allowed to discover and specify, but the lack of the results of a research, which are carried out for OS Linux, does this article unfinished. For example, whether the parameters for OS Windows set earlier in the same degree influence the speed of process of copying of files for OS Linux. In addition, there is an attempt to analyze – whether "direction" of copying matters. "Direction" is meant as a system a source – the receiver and their change. Not clearly, whether this circumstance will affect the process course.

4. Results of experiments
Researches were conducted on two OS and a computer, using the developed software on the basis of the C# and Python 3.
Characteristics of a computer from the installed Windows 7 OS were described in article [4].
Characteristics of a computer from the installed OS LINUX are almost identical to a computer from an experiment [4]:
1. Processor: Intel (R) of Xeon (R) CPU X5460 @ 3.16GHz;
2. Random access memory: 8 Gb;
3. System type: 64-bit operating Ubuntu 18.04.2 LTS server system (GNU/Linux 4.15.0-50-generic x86_64);
4. The hard drive WD Red SATA 6 Tb, EXT4 file system, rotational speed – 5500 min-1 (2 pieces);
5. The hard drive WD Blue SATA 6 Tb, EXT4 file system, rotational speed – 7200 min-1 (1 piece);
6. The hard drive WD Red SATA 3 Tb, EXT4 file system, rotational speed – 5500 min-1 (2 pieces).

As instruments of check for OS LINUX the utilities cp, rcync were used. Cp – is the built-in tool in OS, and rcync – the program from cwRsync developers. And rcync is primary development, and cwRsync was exported in Windows.

Let's remind that earlier in [4] as basic data the directories with a capacity of 30 Gb consisting of file sets of the identical size undertook: 100 Kb; 1, 10, 100, 500 Mb and 1 Gb which were received by the fsutil.exe command. The directories received thus comprised the following number of subdirectories and files (table 1).

| Average file size, Mb | Number of directories and subdirectories | Number of files |
|----------------------|------------------------------------------|----------------|
| 1.000                | 3                                        | 30             |
| 500                  | 40                                       | 60             |
| 100                  | 100                                      | 300            |
| 10                   | 394                                      | 3,000          |
| 1                    | 3,394                                    | 30,000         |
| 0.1                  | 33,400                                   | 300,000        |

The essence of an experiment is that from directories, which consisting files from identical by the sizes (both 100 Kb…500 Mb, and 10, 20, 30 Gb), a certain volume with the purpose to emulate "loss" of files as a result of the virus attack or a human factor was removed. Results of experiments are presented in tables and graphic dependences in article [4].

In this article, the same experiments were executed, but for Linux Ubuntu Server OS and further comparisons of results, researches from article are conducted [4].

Distinctive feature practically of all server versions of OS Linux is the lack of a visual design. I.e. in such systems, there is no such concept as a desktop, a design of software. There is only a text interface – the DOS terminal where the text command is applied to execution of commands.

The lack of the graphic interface strongly reduces load of the processor, the video card and random access memory therefore such systems are considered as faster, than OS Windows and even Mac OS. In addition, it should be noted that in OS Linux, there are practically no viruses and the rule on access to files and folders of users of OS is extremely rigidly carried out. For example, it is impossible to get access to files of the user, even being in the shared directory of a local area network if the user registered on a server computer rewrote them. It allows to achieve the environment, extremely steady against infection, which is practically ideal for storage of files of users.

Based on such computers create file servers, multimedia stations, WEB- and FTP-servers. Therefore, we needed to check this solution for objectives how fast works.

Some earlier made experiments [4], but now based in OS Linux Ubuntu Server were for this purpose repeated.

In this experiment interests us, the solution in similar OS works how quicker. Therefore, it was necessary to check speed of backup operation, which is carried out in realities of one hard drive, copying of data between two hard drives with the identical and different rotational speed.
In OS Linux most often for a solution of similar tasks `Rsync` (the same as under OS Windows) and the `cp`-utility is used. Based on these utilities, we conducted a research. Experiments were for this purpose made, and the results are shown in table 2-6. At them, there are data on the `cwRsync` utility from article [4] (for the general comparison). In addition, it should be noted, that the copying direction was from the hard drive HDD WD Blue 6 Tb (rotating speed of 7200 min^{-1}) to HDD WD Red 6 Tb (rotating speed of 5500 min^{-1}).

**Table 2.** General time of the analysis and recovery of 1 Gb of data from the files with a capacity of 30 Gb.

| Average file size, Mb | Programs (utilities) |  
|----------------------|----------------------|  
|                      | `cp (OS Linux)`      | `rcync (OS Linux)` | `cwRsync (OS Windows)` |
| 1,000                | 6.52                 | 0.13               | 0.18                   |
| 500                  | 7.05                 | 0.08               | 0.22                   |
| 100                  | 4.8                  | 0.15               | 0.25                   |
| 10                   | 4.85                 | 0.17               | 0.25                   |
| 1                    | 6.08                 | 0.18               | 0.87                   |
| 0.1                  | 7.15                 | 0.4                | 8.27                   |

**Table 3.** General time of the analysis and recovery of 10 Gb of data from files volume 30 Gb.

| Average file size, Mb | Programs (utilities) |  
|----------------------|----------------------|  
|                      | `cp (OS Linux)`      | `rcync (OS Linux)` | `cwRsync (OS Windows)` |
| 1,000                | 5.75                 | 1.4                | 1.92                   |
| 500                  | 6.18                 | 1.43               | 2.02                   |
| 100                  | 4.45                 | 1.4                | 2.18                   |
| 10                   | 4.82                 | 1.57               | 2.13                   |
| 1                    | 5.08                 | 1.33               | 5.97                   |
| 0.1                  | 6.13                 | 2.1                | 42.63                  |

**Table 4.** General time of the analysis and recovery of 20 Gb of data from files volume 30 Gb.

| Average file size, Mb | Programs (utilities) |  
|----------------------|----------------------|  
|                      | `cp (OS Linux)`      | `rcync (OS Linux)` | `cwRsync (OS Windows)` |
| 1,000                | 4.95                 | 2.75               | 3.92                   |
| 500                  | 5.1                  | 2.77               | 4.05                   |
| 100                  | 4.37                 | 2.82               | 4.40                   |
| 10                   | 4.65                 | 3.27               | 4.17                   |
| 1                    | 4.37                 | 2.68               | 11:28                  |
| 0.1                  | 5.0                  | 3.72               | 86.77                  |

**Table 5.** General time of the analysis and recovery of 30 Gb of data (removal of data) for OS Linux.

| Average file size, Mb | Programs (utilities) |  
|----------------------|----------------------|  
|                      | `cp (OS Linux)`      | `rcync (OS Linux)` | `cwRsync (OS Windows)` |
| 1,000                | 4.18                 | 4.22               | 5.77                   |
Results of check of influence of the direction of copying of data are given in table 6. It is necessary to consider that all volume was recovered, i.e. 30 Gb and it was recovered only by means of the `rcync` utility since it showed the best time concerning the `cp` utility, and allows to create a full mirror of the initial directory that is much more important, than process speed. In our opinion, when using recovery of 30 Gb of data, the result is given more correct reflection of the course of backup process.

Table 6. General time of the analysis and recovery of 30 Gb of data (removal of data) depending on the copying direction.

| Average file size, Mb | In the hard drive with rotating speed 5500 min⁻¹ | In the hard drive with rotating speed 7200 min⁻¹ | From the hard drive with rotating speed 5500 min⁻¹ to a disk with rotating speed 7200 min⁻¹ | From the hard drive with rotating speed 7200 min⁻¹ to a disk with rotating speed 5500 min⁻¹ | Between two hard drives with rotating speed 5500 min⁻¹ |
|----------------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|
| 1.000                | 9.90                                         | 8.65                                         | 4.8                                          | 4.22                                         | 4.65                                         |
| 500                  | 9.83                                         | 8.63                                         | 4.7                                          | 4.2                                         | 4.48                                         |
| 100                  | 9.82                                         | 8.67                                         | 4.63                                         | 4.25                                         | 4.45                                         |
| 10                   | 10:15                                        | 9.3                                          | 5.03                                         | 4.75                                         | 4.85                                         |
| 1                    | 9.1                                          | 7.75                                         | 4.38                                         | 4                                           | 4.17                                         |
| 0.1                  | 10:03                                        | 9.42                                         | 5.53                                         | 5.33                                         | 5.3                                          |

5. Discussion of results

Based on results of experiments from table 2-5, it was revealed that irrespective of the number of the recovered data there is a point after which passing indicators of speed of backup begin to be normalized and gradually increase. At the same time does not matter what the OS - Windows or Linux is selected. This effect is illustrated in figure 1.

Figure 1. Backup speed indicators at recovery of 30 Gb of data for OS Windows and Linux.
It obviously demonstrates that there is a certain file size which can be considered minimum suitable for copying of files. It is about 10 MB. It is possible to note that speed indicators to this file size behave differently, chaotically and it is almost impossible to predict behavior of speed. Also figure 1 shows that all operations by OS services of Linux take place quicker, with a considerable separation. Most likely it is connected with the fact that in this case there is no need to process forms of applications at implementation of programs in a console look. Taking into account data from table 6 it turns out that copying operation depends also on "direction" of copying. Data backup comes quicker from the hard drive with high speed to disks with smaller. These distinctions are small, but obviously exist that is shown in figure 2.

![Figure 2](image)

**Figure 2.** Speed indicators depending on the direction of copying of 30 Gb of data for OS Linux.

From figure 2 it is visible that the lowest indicators of backup speed options are in the conditions of one hard drive possess. The diagram represents results of experiments only for OS Linux, but probability is high that these effects will be shown also in terms of use of other OS.

6. **Conclusions**

The main conclusions by experimental results for OS Linux are given and their comparison with analogical results for OS Windows was carried out.

1. As show results of the made experiments, utilities in OS Linux work quicker, than in OS Windows.

2. Recovery of full data volume are expedited by utilities in OS Linux relatively and slightly differ from each other. However, at recovery of directories with already available files the `Rsync` utility proves to be much better, than the `cp` utility that is confirmed by results of experiments on time of the analysis and data recovery.

3. The conducted research on influence of the direction of copying of data revealed that in hard drives data transmission rate is always lower, than in case of file transfer from disks on other carriers.
4. The direction of copying also depends on that what rotational speed of a spindle the hard disk receiver possesses. So data transmission rate turns out lower if the receiver are hard disks with the rotational speed of 5500 min\(^{-1}\) and above if the hard disk with rotation 7200 min\(^{-1}\) is the receiver.

5. Irrespective of the direction of copying of a directory, consisting of files with an average size of 100 Kb and 10 Mb, showed the worst figure of copying speed and dropped out of an overall picture of results of an experiment. As these sizes of files showed these deviations in all cases of copying, in our opinion it is not an error of experience, but the revealed dependence.

6. Rational file size for process of backup at its use on one computer, without use of network technologies and external hard drives is set. Its average value makes 700 Mb.

7. If data volume increases, then the copying speed becomes stable at the sizes of files from 10 Mb. It is characteristic as of OS Windows, and OS Linux. Thus, it is possible to draw a conclusion that file size in 10 Mb is the minimum file size, which gives stability and stability to backup process, when using hard drives.

References

[1] Ruben G A 2011 How to automatically test and validate your database backup and recovery strategy J. of Phys.: Conf. Series 331 042031
[2] Rahman P A. and Novikova F S 2018 Reliability model of disk arrays RAID-5 with data striping J. of Phys., Conf. Series.: Materials Science and Engineering 327 022087
[3] Rahman P A 2017 Using a specialized Markov chain in the reliability model of disk arrays RAID-10 with data mirroring and striping J. of Phys., Conf. Series.: Materials Science and Engineering 177 012087.
[4] Proskuriakov N E and Yakovlev B S 2019 Determination of parameters of the Automated backup process of digital data for printing houses and publishing houses without use of external network technologies transformations J. of Phys., Conf. Series 1210 012116
[5] Kalita A, Ozhiganova M and Tishchenko E 2019 Basics of Adaptive Information Security Systems NBI technologies 13 11–15
[6] Mansour E 2017 A survey of digital information literacy (DIL) among academic library and information professionals Digital Library Perspectives 33 166–188
[7] Bakhacheva Yu 2017 Development of structure of a subsystem of information security of the automated system MATEC Web of Conf. 132(6) 04005 1–4
[8] Sorokin A D, Kazarin O V, Petrov S T and Tarasov A A 2014 Application of a method of the analysis of hierarchies in the field of saving of digital heritage Modern problems and tasks of information security support. Proc. of the Intern. Sc. and Pract. Conf. «SIB–2014» ed O A Makarova (Moscow: Moscow Financial and Legal University) pp 67–73