

ЗАВИСИМОСТЬ СКОРОСТИ ПРОЦЕССА РЕЗЕРВНОГО КОПИРОВАНИЯ ЦИФРОВЫХ АРХИВОВ ТИПОГРАФИЙ И ИЗДАТЕЛЬСТВ ОТ ПАРАМЕТРОВ ОПЕРАЦИОННОЙ СИСТЕМЫ

Н.Е. Проскуряков¹

д.т.н., профессор, e-mail: vippne@mail.ru

Б.С. Яковлев¹

к.т.н., доцент, e-mail: bor_yak@mail.ru

Н.Н. Архангельская¹

к.т.н., доцент, e-mail: arhangel_nataly@mail.ru

Е.В. Трапезников²

старший преподаватель, e-mail: evtrapeznikov@yandex.ru

¹Тульский государственный университет, Тула, Россия

²Омский государственный технический университет, Омск, Россия

Аннотация. Проведены исследования процесса резервного копирования цифровых архивов с использованием операционных систем Windows и Linux. Были выявлены основные параметры, влияющие на скорость процесса резервного копирования цифровых архивов. В результате экспериментальных и аналитических исследований установлено, что самыми существенными факторами являются тип операционной системы и форма интерфейса используемых служебных утилит. Обнаружено влияние направления копирования данных, особенно внутри жёстких дисков, где скорость передачи данных всегда ниже, чем в случае копирования файлов на другие носители. Выявлено, что самыми низкими показателями скорости обладают варианты копирования данных в условиях одного жёсткого диска. Даны рекомендации по оптимальному размеру файла для процесса резервного копирования при его использовании на одной ЭВМ и без использования сетевых технологий и внешних жёстких дисков.

Ключевые слова: резервное копирование, отказоустойчивая система, надёжность хранения данных, утилиты, gobocopy, cp, cwrscopy, gcupc, скорость процесса резервного копирования, Windows, Linux Server, C#, Python3.

Введение

В современных реалиях велика вероятность потери данных вследствие деятельности злоумышленников или конкурентов. Данная проблема наиболее актуальна для полиграфических предприятий по причине возможности безвозвратной потери данных, на которые предприятие потратило значительные

усилия и время. Поэтому целесообразно использовать резервное копирование данных.

Цифровые технологии активно применяются во всех отраслях человеческой деятельности, в том числе и в полиграфии. Они особенно важны на этапах подготовки макетов, общении с заказчиками, организации печати по требованию (через сайты) и т. п.

Данные технологии хорошо развиты, показывают свою эффективность. Без них сложно представить современный цикл производства книг и другой печатной продукции.

Наибольший интерес у исследователей в данный момент времени вызывает организация резервного копирования на основе аппаратной защиты данных, основанная на RAID-массивах. К подобным работам можно отнести [1–3].

Однако, на наш взгляд, эти технологии в корне отличаются от задач стандартного резервного копирования, т. к. в случае заражения компьютера вирусом заражённые данные, находящиеся на отказоустойчивых системах, сохраняют заражённые файлы, не будет возможности восстановить данные, потому что не с чего восстанавливать, все файлы находятся на месте, но они не годятся для работы, их версии невозможно восстановить.

Проблематично использовать RAID-массив и для защиты данных от банального человеческого фактора, например, от случайного стирания файлов, каталогов. Технология RAID сохранит изменения и, более того, станет причиной очень больших затруднений при попытках восстанавливать данные, т. к. по своей структуре файлы хранятся в разрозненном виде и обычно разделены между дисками. Собрать такую информацию воедино очень тяжело, а иногда невозможно.

В силу того, что программные реализации в основном сводятся к постоянному копированию файлов, полагаясь на возможности, службы самой операционной системы (ОС), а также по причине того, что под ОС Linux практически нет таких решений, вопрос по реализации, оптимизации программных решений в сфере создания резервных копий стоит очень остро до сих пор.

Поэтому целью данного исследования являлось определение параметров, влияющих на скорость процесса резервного копирования, для различных операционных систем, что является логическим обобщением и продолжением ранее проведённой работы [4].

1. Постановка задачи

Основными задачами, которые необходимо решить в ходе работы, являются: максимальное сокращение времени, затрачиваемого на операции резервного копирования; сравнение характеристик работы операционных систем Windows и Linux; выявление параметров, влияющих на скорость и качество процесса сохранения и архивации данных: разработка программного обеспечения и выработка рекомендаций для специалистов в области хранения и защиты данных.

2. Теория

Можно выделить два основных типа систем хранения данных: централизованная система и децентрализованная система [7–10].

В централизованных системах данные хранятся на самостоятельном, независимом от других носителя информации или облачном хранилище. Резервное копирование в таких системах обеспечивается программными способами.

В децентрализованных системах хранения эти данные разбиваются на блоки на уровне байт (или даже на уровне бит) и записываются одновременно на несколько носителей. Для этого, соответственно, требуется несколько жёстких дисков и использование технологии RAID-массивов. Этот тип системы хранения данных особенно хорош при малых объёмах данных, так как от этого в значительной мере зависит стоимость резервного копирования данных.

Мы склоняемся к тому, что значительно проще работать с данными при централизованном подходе хранения архивных данных на предприятиях. Для этого существует несколько причин.

1. В большинстве случаев для восстановления данных необходимы простые копии файлов. Этот вариант обеспечивает только централизованный вариант хранения данных.
2. Далеко не все программы восстановления данных способны работать с рейд массивами. Для этого необходимы значительно большие затраты в вычислениях и другие подходы, обычно подобные функции раскрываются в платных версиях подобных продуктов.
3. Централизованная система хранения данных обычно дублируется, поэтому в случае выхода из строя одного из носителей, почти на 100 % будет сохранен второй.
4. С учётом пункта 3 многократно возрастает стоимость децентрализованных систем и RAID, т. к. их стоимость значительно превосходит стоимость обычных жёстких дисков.

Также можно отметить, что согласно общепринятой теории, системы, не имеющие интерфейсов, работают быстрее «desktopных» вариантов, что относится также к ОС.

В более ранней работе проводился эксперимент на стандартной ОС Windows, полученные материалы многое позволили выявить и уточнить, но отсутствие результатов исследования, проведённых для ОС Linux, делает работу незаконченной. Например, влияют ли установленные ранее параметры для ОС Windows в той же степени на скорость процесса копирования файлов для ОС Linux.

Также здесь присутствует попытка проанализировать — имеет ли значение «направление» копирования. Под «направлением» подразумевается система источник — приёмник и их смена. Неясно, повлияет ли данное обстоятельство на ход процесса.

3. Результаты экспериментов

Исследования проводились на двух ОС и ЭВМ с применением разработанного ПО на базе языков C# и Python 3.

Характеристики ЭВМ с установленной ОС Windows 7 были описаны в работе [4].

Характеристики ЭВМ с установленной ОС Linux практически идентичны ЭВМ из эксперимента [4]:

1. Процессор: Intel(R) Xeon(R) CPU X5460 3.16GHz;
2. Оперативная память: 8 Гб;
3. Тип системы: 64-разрядная операционная серверная система Ubuntu 18.04.2 LTS (GNULinux 4.15.0-50-generic x86_64);
4. Жёсткий диск WD Red SATA 6 Тб, файловая система EXT4, скорость вращения — 5500 мин^{-1} (2 шт);
5. Жёсткий диск WD Blue SATA 6 Тб, файловая система EXT4, скорость вращения — 7200 мин^{-1} (1 шт);
6. Жёсткий диск WD Red SATA 3 Тб, файловая система EXT4, скорость вращения — 5500 мин^{-1} (2 шт).

В качестве инструментов проверки для ОС Linux использовались утилиты *ср*, *rcsync*. *Ср* — является встроенным инструментом в ОС, а *rcsync* — программа от разработчиков *сwRsync*. Причём *rcsync* является первичной разработкой, а *сwRsync* был экспортирован на Windows.

Напомним, что ранее в [4] в качестве исходных данных брались каталоги объёмом 30 Гб, состоящие из наборов файлов одинакового размера: 100 Кб; 1, 10, 100, 500 Мб и 1 Гб, которые были получены командой *fsutil.exe*. Каталоги, полученные таким образом, содержали в себе следующее количество подкаталогов и файлов (табл. 1).

Суть эксперимента заключается в том, что из каталогов, состоящих из одинаковых по размерам файлов (как 100 Кб... 500 Мб, так и 10, 20, 30 Гб) удалялся определённый объём с целью эмулировать «потерю» файлов в результате вирусной атаки или человеческого фактора. Результаты экспериментов представлены в таблицах и графических зависимостях в работе [4].

В данной работе были выполнены те же эксперименты, но для ОС Linux Ubuntu Server и далее проведены сопоставления результатов исследования из работы [4].

Отличительной особенностью практически всех серверных версий Linux является отсутствие визуального оформления. То есть в таких системах нет такого понятия, как рабочий стол, оформление ПО. Существует лишь текстовый интерфейс — терминал типа DOS, где для выполнения команд применяется текстовая команда.

Таблица 1. Состав тестовых каталогов для проверки гипотезы влияния среднего размера файлов на скорость процесса резервного копирования с учётом общего объёма 30 гб

Средний размер файла	Количество каталогов и подкаталогов	Количество файлов
1 Гб	3	30
500 Мб	40	60
100 Мб	100	300
10 Мб	394	3 000
1 Мб	3 394	30 000
100 Кб	33 400	300 000

Отсутствие графического интерфейса сильно снижает нагрузку на процессор, видеокарту и оперативную память, поэтому такие системы считаются более быстрыми, чем ОС Windows и даже Mac OS.

Также стоит отметить, что на Linux практически нет вирусов и крайне жёстко выполняется правило по доступу к файлам и папкам пользователей ОС. К примеру, нельзя получить доступ к файлам пользователя, даже находящимся в общем каталоге локальной сети, если они были переписаны пользователем, зарегистрированным на серверной ЭВМ. Это позволяет добиться крайне устойчивой к заражению среды, которая практически идеально подходит для хранения файлов пользователей.

На базе таких ЭВМ создают файловые серверы, мультимедийные станции, WEB- и FTP-серверы. Поэтому нам нужно было проверить, насколько быстро работает данное решение для поставленных задач.

Для этого были повторены некоторые ранее проведённые эксперименты [4], но теперь на базе Linux Ubuntu Server.

В данном эксперименте нас интересует, насколько быстрее работает решение в подобных ОС. Поэтому необходимо было проверить быстроту операции резервного копирования, которая проводится в реалиях одного жёсткого диска, копирование данных между двух жёстких дисков с одинаковой и разной скоростью вращения шпинделя.

В ОС Linux чаще всего для решения подобных задач используется *cwRsync* (тот же, что и под ОС Windows), а также утилита *cp*. На их основе и будет проводиться исследование. Для этого были проведены опыты, результаты которых показаны в табл. 2–6. В них присутствуют данные по утилите *cwRsync* из работы [4] (для общего сравнения). Также стоит отметить, что направление копирования было от жёсткого диска HDD WD Blue 6 Tb (частота вращения 7200 мин⁻¹ в HDD WD Red 6 Tb (частота вращения 5500 мин⁻¹).

Результаты проверки влияния направления копирования данных приведены в табл. 6. Надо учесть, что восстанавливался весь объём, т. е. 30 Гб, и вос-

Таблица 2. Общее время анализа и восстановления 1 Гб данных из материалов объемом 30 Гб для ОС Linux

Средний размер файла	Программы (утилиты)		
	<i>cp</i> (ОС Linux)	<i>rsync</i> (ОС Linux)	<i>cwRsync</i> (ОС Windows)
1 Гб	6,52	0,13	0,18
500 Мб	7,05	0,08	0,22
100 Мб	4,8	0,15	0,25
10 Мб	4,85	0,17	0,25
1 Мб	6,08	0,18	0,87
100 Кб	7,15	0,4	8,27

Таблица 3. Общее время анализа и восстановления 10 Гб данных из материалов объёмом 30 Гб для ОС Linux

Средний размер файла	Программы (утилиты)		
	<i>cp</i> (ОС Linux)	<i>rsync</i> (ОС Linux)	<i>cwRsync</i> (ОС Windows)
1 Гб	5,75	1,4	1,92
500 Мб	6,18	1,43	2,02
100 Мб	4,45	1,4	2,18
10 Мб	4,82	1,57	2,13
1 Мб	5,08	1,33	5,97
100 Кб	6,13	2,1	42,63

становливался только при помощи утилиты *rsync*, т. к. она показала лучшее время относительно утилиты *cp* и позволяет создать полное зеркало изначального каталога, что является значительно более важным, чем скорость процесса. На наш взгляд, при использовании восстановления 30 Гб данных даёт более правильное отражение хода процесса резервного копирования.

4. Обсуждение результатов

На основе результатов экспериментов из табл. 2...5 было выявлено, что независимо от количества восстанавливаемых данных существует точка, после прохождения которой показатели скорости резервного копирования начинают

Таблица 4. Общее время анализа и восстановления 20 Гб данных из материалов объёмом 30 Гб для ОС Linux

Средний размер файла	Программы (утилиты)		
	<i>cp</i> (ОС Linux)	<i>rsync</i> (ОС Linux)	<i>cwRsync</i> (ОС Windows)
1 Гб	4,95	2,75	3,92
500 Мб	5,1	2,77	4,05
100 Мб	4,37	2,82	4,40
10 Мб	4,65	3,27	4,17
1 Мб	4,37	2,68	11,28
100 Кб	5	3,72	86,77

Таблица 5. Общее время анализа и восстановления 30 Гб данных (полное уничтожение данных) для ОС Linux

Средний размер файла	Программы (утилиты)		
	<i>cp</i> (ОС Linux)	<i>rsync</i> (ОС Linux)	<i>cwRsync</i> (ОС Windows)
1 Гб	4,18	4,22	5,77
500 Мб	4,12	4,2	6,28
100 Мб	4,13	4,25	6,72
10 Мб	4,65	4,75	6,22
1 Мб	3,55	4	16,27
100 Кб	4,48	5,38	128,48

нормализоваться и постепенно увеличиваются. При этом не имеет значение какая выбрана операционная система — Windows или Linux. Данный эффект проиллюстрирован на рис. 1.

Это явно свидетельствует о том, что существует определённый размер файла, который можно считать минимально подходящим для копирования файлов. Он составляет около 10 Мб. Можно отметить, что показатели скорости до этого размера файла ведут себя по-разному, хаотично, и предсказать поведение скорости практически невозможно. Также рис. 1 показывает, что все операции службами ОС Linux проходят быстрее, со значительным отрывом. Скорее всего, это связано с тем, что в этом случае нет нужды обрабатывать формы приложений при реализации программ в консольном виде. С учётом данных

Таблица 6. Общее время анализа и восстановления 30 Гб данных (полное уничтожение данных) в зависимости от направления копирования для ОС Linux

Средний размер файла	Направление копирования				
	Внутри жёсткого диска с частотой вращения 5500 мин ⁻¹	Внутри жёсткого диска с частотой вращения 7200 мин ⁻¹	От жёст- кого диска с частотой вращения 5500 мин ⁻¹ к диску с частотой вращения 7200 мин ⁻¹	От жёст- кого диска с частотой вращения 7200 мин ⁻¹ к диску с частотой вращения 5500 мин ⁻¹	Между двух жёстких дисков с частотой вращения 5500 мин ⁻¹
1 Гб	9,90	8,65	4,8	4,22	4,65
500 Мб	9,83	8,63	4,7	4,2	4,48
100 Мб	9,82	8,67	4,63	4,25	4,45
10 Мб	10,15	9,3	5,03	4,75	4,85
1 Мб	9,1	7,75	4,38	4	4,17
100 Кб	10,03	9,42	5,53	5,33	5,3

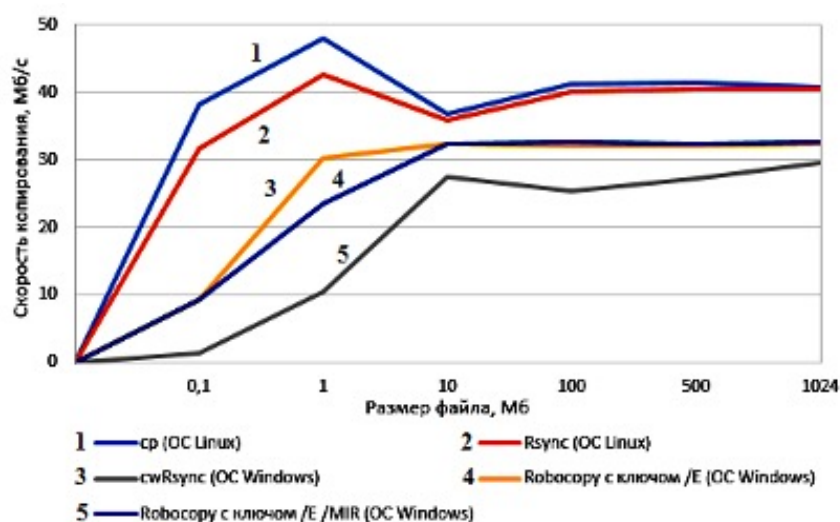


Рис. 1. Показатели скорости резервного копирования при восстановлении 30 Гб данных для ОС Windows и Linux

из табл. 6 получается, что операция копирования также зависит и от «направления» копирования. Резервное копирование данных происходит быстрее от жёсткого диска с большей скоростью к дискам с меньшей. Эти различия малы, но явно существуют, что показано на рис. 2.

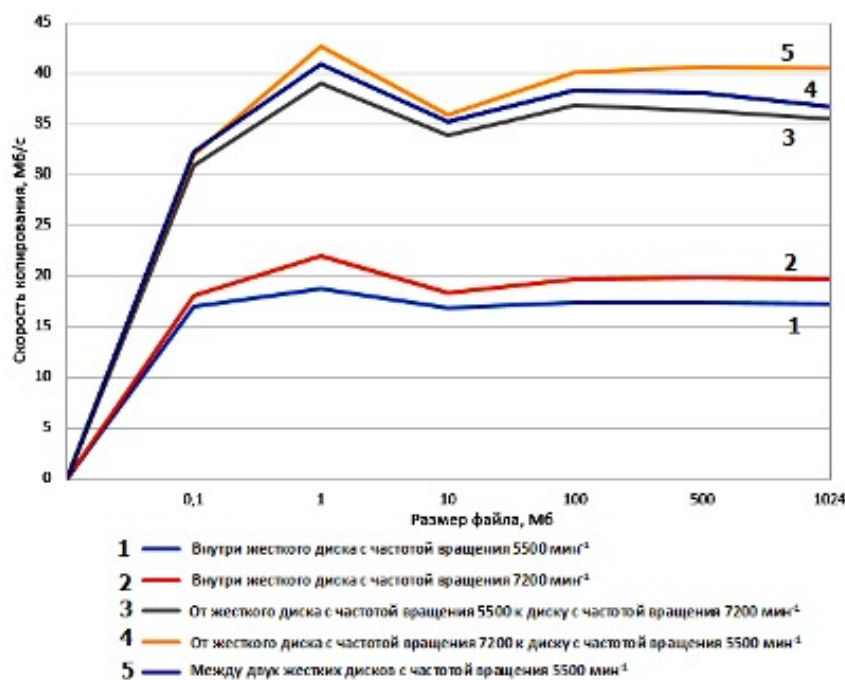


Рис. 2. Показатели скорости в зависимости от направления копирования 30 Гб данных для ОС Linux

Из рис. 2 видно, что самыми низкими показателями скорости обладают варианты копирования данных в условиях одного жёсткого диска. График представляет результаты экспериментов только для ОС Linux, но велика вероятность, что данные эффекты проявятся и в условиях использования других ОС.

Выводы и заключение

В данном разделе помещены основные выводы по экспериментам с ОС Linux и сравнение результатов с ОС Windows, т. к. ранее в работе [4] были выполнены эксперименты для ОС Windows.

1. Как показывают таблицы 1–5, утилиты ОС Linux работают быстрее, чем на ОС Windows.
2. Из таблицы 5 видно, что восстановление полного объёма данных утилитами ОС Linux выполняется относительно быстро и мало отличается друг от друга. Однако когда нужно восстановить каталоги с уже имеющимися файлами, то утилита *cwRsync* показывает себя значительно лучше, чем *ср*. Это подтверждают результаты в табл. 2 – 5.

3. Исследование влияния направления копирования данных (таблица 6 и рисунок 2) выявило, что внутри жёстких дисков скорость передачи данных всегда ниже, чем в случае передачи файлов на другие носители.
4. Также стоит отметить, что не зависимо от направления копирования каталоги, состоящие из файлов со средним размером 100 Кб и 10 Мб продемонстрировали самый худший показатель скорости и выбиваются из результатов эксперимента. На наш взгляд, это не является следствием ошибочных замеров, т. к. данные средние размеры показали эти отклонения во всех направлениях копирования.
5. Был определён оптимальный размер файла для процесса резервного копирования при его использовании на одном ЭВМ и без использования сетевых технологий и внешних жёстких дисков. Его среднее значение составляет 700 Мб.
6. Чем больше объём данных, тем стабильнее скорость копирования при размерах файлов от 10 Мб. Это касается как ОС Windows, так и ОС Linux. Таким образом, можно сделать вывод, что размер файла в 10 Мб является минимальным размером файла, который даёт устойчивость и стабильность процессу резервного копирования на жёстких дисках HDD.

ЛИТЕРАТУРА

1. Ruben G.A. How to automatically test and validate your database backup and recovery strategy // Journal of Physics: Conference Series. 2011. V. 331.
2. Rahman P.A., Novikova F.S. Reliability model of disk arrays RAID-5 with data striping // IOP Conference Series: Materials Science and Engineering. 2018. V. 327.
3. Rahman P.A. Using a specialized Markov chain in the reliability model of disk arrays RAID-10 with data mirroring and striping // IOP Conference Series: Materials Science and Engineering. 2017. V. 177.
4. Proskuriakov N.E., Yakovlev B.S. Determination of parameters of the Automated backup process of digital data for printing houses and publishing houses without use of external network technologies transformations // Journal of Physics: Conference Series. 2019. V. 1210.
5. Михаил gotch Где мои файлы, ROBOCOPY? // Электронный ресурс Habr, 2015. URL: <https://habr.com/post/261359/> (дата обращения: 20.10.2018).
6. Malikov V. Используем Robocopy. Примеры. URL: <https://vladimirmalikov.com/\\используем-robocopy-примеры/> (дата обращения: 20.10.2018).
7. Гришина Н.В. Комплексная система защиты информации на предприятии. М. : Форум, 2015. 240 с.
8. Остроух А.В. Ввод и обработка цифровой информации: Учебник для нач. проф. образования. М. : ИЦ Академия, 2013. 288 с.
9. Малюк А.А., Пазизин С.В., Погожин Н.С. Введение в защиту информации в автоматизированных системах. М. : Горячая Линия — Телеком, 2011. 146 с.

10. Сорокин А.Д., Казарин О.В., Петров С.Т., Тарасов А.А. Применение метода анализа иерархий в области сохранения цифрового наследия // Современные проблемы и задачи обеспечения информационной безопасности. Труды Международной научно-практической конференции «СИБ - 2014». М. : МФЮА, 2014. С. 67-73.

DEPENDENCE OF THE SPEED OF THE PROCESS OF BACKING UP DIGITAL ARCHIVES OF PRINTING HOUSES AND PUBLISHERS ON THE PARAMETERS OF THE OPERATING SYSTEM

N.E. Proskuryakov¹

Dr.Sc. Professor, Department of Technological systems of food, printing and packaging industries, e-mail: vippne@mail.ru

B.S. Yakovlev¹

Ph.D. Associate Professor, Department of Technological systems of food, printing and packaging industries, e-mail: bor_yak@mail.ru

N.N. Arkhangel'skaya¹

Ph.D. Associate Professor, Department of Descriptive geometry, engineering and computer graphics, e-mail: arhangel_nataly@mail.ru

E.V. Trapeznikov²

Assistant Professor, Department of Complex Information Security, e-mail: evtrapeznikov@yandex.ru

¹Tula State University, Tula, Russia

²Omsk State Technical University, Omsk, Russia

Abstract. Research has been carried out on the process of backing up digital archives using Windows and Linux operating systems. The main parameters were identified that affect the speed of the process of backing up digital archives. As a result of experimental and analytical studies, it has been established that the most significant factors are the type of operating system and the form of the interface of the service utilities used. The influence of the direction of data copying was found, especially inside hard drives, where the data transfer rate is always lower than in the case of copying files to other media. It has been revealed that the lowest speed indicators have the options for copying data in a single hard disk. Recommendations on the optimal file size for the backup process when using it on one computer and without using network technologies and external hard drives are given.

Keywords: backup, fault tolerant system, data storage reliability, utilities, robocopy, cp, cwrscopy, rcync, backup process speed, Windows, Linux Server, C#, Python3..

REFERENCES

1. Ruben G.A. How to automatically test and validate your database backup and recovery strategy. Journal of Physics: Conference Series, 2011, vol. 331.
2. Rahman P.A. and Novikova F.S. Reliability model of disk arrays RAID-5 with data striping. IOP Conference Series: Materials Science and Engineering, 2018, vol. 327.

3. Rahman P.A. Using a specialized Markov chain in the reliability model of disk arrays RAID-10 with data mirroring and striping. IOP Conference Series: Materials Science and Engineering, 2017, vol. 177.
4. Proskuriakov N.E. and Yakovlev B.S. Determination of parameters of the Automated backup process of digital data for printing houses and publishing houses without use of external network technologies transformations. Journal of Physics: Conference Series, 2019, vol. 1210.
5. Mikhail gotch Gde moi faily, ROBOCOPY? Elektronnyi resurs Habr, 2015. URL: <https://habr.com/post/261359/> (20.10.2018). (in Russian)
6. Malikov V. Ispol'zuem Robocopy. Primery. URL: <https://vladimirmalikov.com/\\ispol'zuem-robocopy-primery/> (20.10.2018). (in Russian)
7. Grishina N.V. Kompleksnaya sistema zashchity informatsii na predpriyatii. Moscow, Forum Publ., 2015, 240 p. (in Russian)
8. Ostroukh A.V. Vvod i obrabotka tsifrovoy informatsii: Uchebnik dlya nach. prof. obrazovaniya. Moscow, ITs Akademiya Publ., 2013. 288 p. (in Russian)
9. Malyuk A.A., Pazizin S.V., and Pogozhin N.S. Vvedenie v zashchitu informatsii v avtomatizirovannykh sistemakh. Moscow, Goryachaya Liniya — Telekom Publ., 2011, 146 p. (in Russian)
10. Sorokin A.D., Kazarin O.V., Petrov S.T., and Tarasov A.A. Primenenie metoda analiza ierarkhii v oblasti sokhraneniya tsifrovogo naslediya. Sovremennye problemy i zadachi obespecheniya informatsionnoi bezopasnosti, Trudy Mezhdunarodnoi nauchno-prakticheskoi konferentsii «SIB - 2014», Moscow, MFYuA Publ., 2014, pp. 67-73. (in Russian)

Дата поступления в редакцию: 07.02.2021