

МЕТОД ПОВЫШЕНИЯ НАДЕЖНОСТИ ХРАНЕНИЯ ИНФОРМАЦИИ В ДОЛГОВРЕМЕННЫХ ЭЛЕКТРОННЫХ ХРАНИЛИЩАХ НА ОПТИЧЕСКИХ ДИСКАХ, ОРГАНИЗОВАННЫХ В МАССИВЫ RAID-6 ЗА СЧЕТ СМЕШИВАНИЯ ДИСКОВ ЗАПАСНЫХ КОПИЙ

А.В. Чернышов

sch@mgul.ac.ru

Мытищинский филиал МГТУ им. Н.Э. Баумана, г. Мытищи,
Московская обл., Российская Федерация

Аннотация

Предложен и обоснован метод обеспечения высокой надежности хранения информации в долговременных электронных хранилищах на оптических дисках. Построена математическая модель надежности хранения информации. На примере расчетов для роботизированной библиотеки Panasonic LB-DH8 показано, что метод позволяет минимизировать число запасных копий информации, обеспечивающих заданные при проектировании хранилища информации показатели надежности. Продемонстрировано использование модели для расчета максимально допустимого числа оптических дисков в массиве RAID-6 для снижения избыточности дискового массива

Ключевые слова

Электронное архивное хранилище информации, длительное хранение, оптические носители информации, надежность хранения информации, массив RAID-6

Поступила в редакцию 29.11.2016
© МГТУ им. Н.Э. Баумана, 2017

Введение. Перевод документооборота учреждений и организаций на электронную основу приводит к необходимости создания и поддержания долговременных электронных архивных хранилищ, в которых должны собираться и храниться электронные документы. В соответствии со стандартами [1, 2], в качестве носителей информации в таких архивных хранилищах должны использоваться только носители однократной записи, что однозначно приводит к необходимости использования однократно записываемых оптических дисков [3]. В связи с этим существует проблема выбора метода размещения информации на носителях (оптических дисках) в целях минимизации вероятности потери информации из-за возможного выхода из строя отдельных носителей.

При записи архивной информации на отдельные носители обеспечить необходимую надежность хранения информации в электронном архиве возможно путем создания некоторого числа запасных копий каждого носителя. Математическая модель для расчета надежности хранения информации в зависимости от числа копий для этого метода построена и исследована в [4]. Отметим, что объединение носителей информации в структуры типа RAID-6 должно снизить вероятность потери информации в каждой копии, т. е. для достижения заданного показателя надежности хранения будет требоваться меньше запасных копий.

Математическая модель для расчета надежности хранения информации и необходимого числа запасных копий электронного архива, в котором носители объединены в массивы RAID-6 и эти массивы рассматриваются как неделимое целое, приведена на рис. 1.

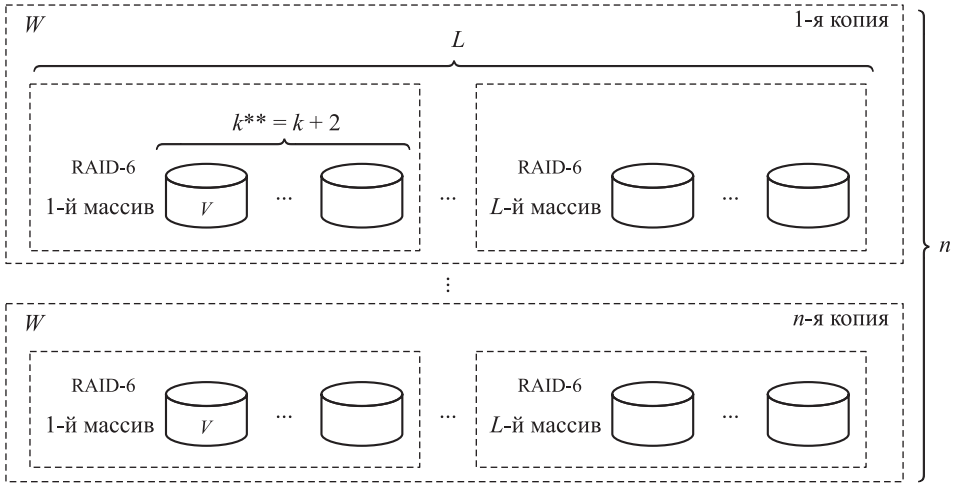


Рис. 1. Структура электронного файлового архива с дисками, организованными в массивы RAID-6, при создании запасных копий RAID-6

Такой вариант создания запасных копий электронного архива является единственно возможным для современных роботизированных библиотек оптических дисков, в которых единицей управления является картридж с группой дисков (например, Panasonic LB-DM9 или LB-DH8 [5–7]), поскольку массивы RAID в таких библиотеках могут быть организованы только в пределах каждого картриджа. Исследование этой модели показало, что вопреки ожиданиям для достижения приемлемого значения надежности необходимо создание не менее трех (а в ряде случаев, вероятно, потребуется и больше) копий информации, либо создание в каждом картридже не одного, а нескольких массивов RAID-6, что значительно снижает полезную емкость электронного архива. Требуется найти способ уменьшения числа запасных копий при сохранении заданных показателей надежности хранения информации.

Вариант решения задачи. Если отказаться от неделимости массивов RAID-6 в пределах картриджей, то надежность хранения информации в электронном архиве должна существенно возрасти. Снятие требования неделимости массивов в пределах картриджей фактически означает, что отдельные оптические диски из одного картриджа могут быть каким-либо способом переставлены в другой (точнее предполагается взаимобмен отдельными дисками между картриджами). Идею предлагаемого решения иллюстрирует рис. 2 (вышедшие из строя диски отмечены крестиком).

Из рис. 2 следует, что каждый из массивов RAID-6 вышел из строя (в каждом массиве вышло из строя более двух дисков) и не позволяет считать информацию.

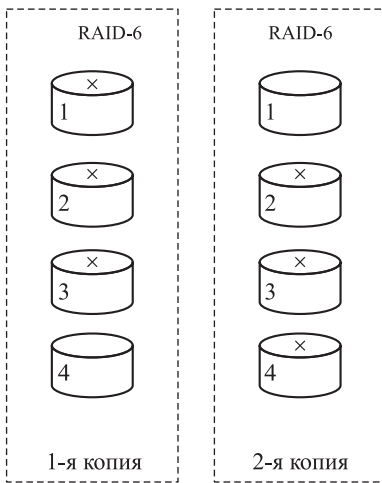


Рис. 2. Конфигурация отказов дисков, при которой возможно собрать RAID-6 без двух дисков

Однако, если объединить в массив RAID-6 оставшиеся исправными диски обоих массивов, то становится возможным получить доступ к информации (массив будет работать в критическом режиме, но чтение информации возможно) и, например, скопировать ее на новые носители, т. е. потери информации не произойдет.

Для современных роботизированных библиотек оптических дисков, оперирующих картриджами, такую работу можно выполнить лишь вручную, предварительно вынув картриджи с обмениваемыми дисками из библиотеки. Понятно, что такой подход будет эффективен лишь в том случае, если вероятность необходимости обмена дисков будет низка. Оценить вероятность этого события для конкретных параметров электронного архивного хранилища довольно легко с помощью математической модели (см. рис. 1). Она совпадает с вероятностью потери хотя бы части информации, хранящейся в архиве, Q_6 . Понятно, что чем больше информации хранится в архиве, тем эта вероятность больше. Поэтому, с точки зрения автора, для снижения этого показателя функция автоматического обмена отдельных дисков между картриджами в роботизированных библиотеках будущего была бы крайне полезна.

Выполним расчет надежности хранения информации в электронном архиве для предложенного метода.

Построение модели надежности хранения информации в архиве из оптических дисков, объединенных в массивы RAID-6, с возможностью смешивания дисков разных запасных копий. Будем использовать следующие обозначения (см. рис. 1, 2): W — общий объем данных, сохраняемый в электронном архиве; V — емкость одного оптического диска; l — число оптических дисков, необходимых для сохранения информации объема W ($l = [W / V]$, где $[.]$ — операция округления до большего целого); q — вероятность выхода из строя одного оптического диска в течение периода времени между контролями исправности; $1 - q$ — вероятность того, что диск в течение этого же периода времени останется исправным (упоминание периода времени для краткости далее будем опускать); k^{**} — число дисков в одном массиве RAID-6 ($k^{**} > 3$); поскольку избыточность массива RAID-6 соответствует емкости двух дисков массива, условно считаем, что информация хранится только на k дисках, т. е. $k^{**} = k + 2$. Тогда для сохранения всей информации объема W понадобится $L = [l / k]$ массивов RAID-6, а суммарное число оптических дисков составит $l^{**} = Lk^{**}$.

Согласно расчетам, вероятность потери хотя бы части информации из хранилища объемом W , в котором диски объединены в структуры RAID-6 без запасных копий, можно записать как

$$Q_6 = 1 - \left((1-q)^k (1+kq - q^2 (1+k - (k+2)!/(2k!))) \right)^L,$$

а вероятность потери хотя бы части информации из аналогичного хранилища с n запасными копиями каждого массива RAID-6 в виде

$$Q_{6n} = 1 - \left(1 - \left((1 - (1-q)^k (1+kq - q^2 (1+k - (k+2)!/(2k!))) \right)^n \right)^L.$$

Обозначим хранилище этого типа как R6.

Если допустить смешивание дисков разных копий одного и того же массива RAID-6, то можно считать, что создаются n копий каждого отдельного диска в архиве (фактически, эти копии образуют структуру RAID-1, но в отличие от «настоящих» RAID-1 диски с этими копиями обычно подключаются к компьютеру не одновременно); для считывания информации массив RAID-6 может быть создан из любых копий каждого k^{**} диска (можно условно рассматривать получившуюся структуру как RAID-16, хотя обычно аппаратура воспроизведения носителей в электронном архиве в каждый момент времени позволяет «собрать» только один массив RAID-6). Обозначим хранилище этого типа как R16.

При построении модели можно считать, что каждый массив RAID-6 образован не отдельными дисками, а дисковыми массивами RAID-1 по n дисков в каждом, вероятность потери информации на каждом равна q^n и всего для сохранения информации объема W имеется L таких структур (рис. 3).

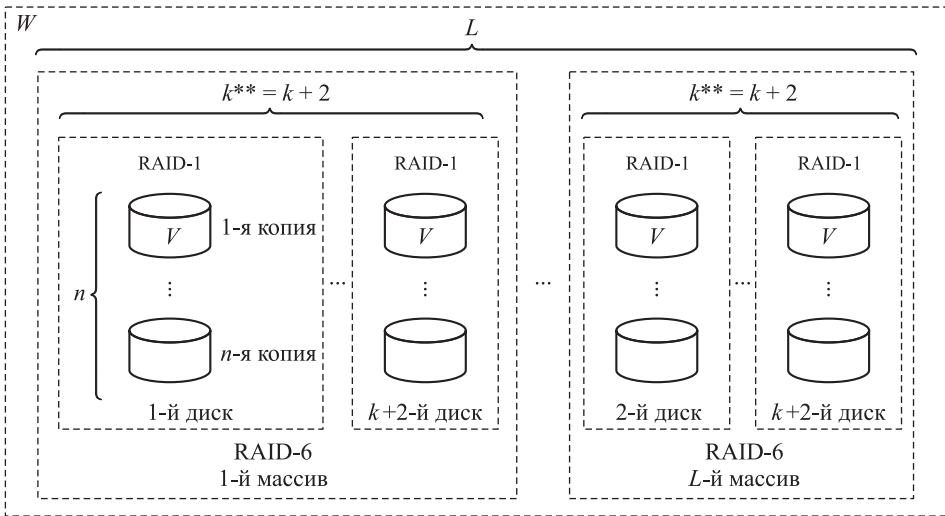


Рис. 3. Структура электронного файлового архива с дисками, организованными в массивы RAID-6 с запасными копиями каждого диска

Используя выведенную ранее формулу вероятности потери хотя бы части информации для L массивов RAID-6 (параметр Q_6), получаем формулу вероятности потери хотя бы части информации для L структур типа RAID-16 (заменяв

значение q вероятности отказа одного диска на значение q^n вероятности отказа массива RAID-1):

$$Q_{16} = 1 - \left((1 - q^n)^k \left(1 + kq^n - (q^n)^2 (1 + k - (k + 2)! / (2k!)) \right) \right)^L.$$

Любопытно, что при равном количестве n запасных копий массивов RAID-6 стоимость хранения информации в электронном архиве, определяемая суммарной стоимостью носителей информации, и для хранилища типа R6, и для хранилища типа R16 одна и та же, поскольку суммарное число дисков L^* не изменяется.

Исследование построенной модели. Для удобства сопоставления результатов исследование модели проведем для электронного архива на базе роботизированной библиотеки Panasonic LB-DH8 [6], не забывая, что структура архива R16 в реальности без ручного вмешательства на этом устройстве не реализуется.

Библиотека в базовой комплектации оснащена 76 картриджами с оптическими дисками. Каждый картридж содержит 12 оптических дисков. Емкость каждого оптического диска $V = 100$ Гб. Таким образом, если не использовать объединение дисков в массивы RAID-6, емкость одного картриджа составляет 1,2 Тб, а всего библиотека может хранить 91,2 Тб информации. (Производитель особо подчеркивает, что в описании библиотеки 1 Тб = $1 \cdot 10^{12}$ байт.)

При объединении оптических дисков в массивы RAID-6 суммарная емкость библиотеки снижается. Поскольку максимальное число дисков в одном массиве RAID-6 по техническим причинам ограничено емкостью одного картриджа, возможно всего несколько вариантов объединения оптических дисков в массивы, позволяющие использовать емкость картриджа полностью (таблица). При этом сам производитель в распространяемых материалах говорит только о массиве RAID-6, организуемом на картридже в целом, т. е. на 12 дисках.

Варианты объединения оптических дисков в массивы

Дисков в массиве RAID-6 (k^*)	Массивов RAID-6 в картридже	Массивов RAID-6 в библиотеке (L)	Избыточность, %	Максимальная емкость библиотеки (W), Тб
4	3	228	50,0	45,6
6	2	152	33,3	60,8
12	1	76	16,7	76,0

Выполним вычисления для значения $q = 0,1$ и будем использовать следующий критерий: вероятность потери хотя бы части информации в архиве объема W должна быть не выше вероятности потери информации на одном носителе, т. е. $Q_{16} < q$.

Понятно, что для случая $n = 1$ (запасные копии отсутствуют) будет иметь место равенство $Q_6 = Q_{16}$. Результаты расчетов для значений n от 1 до 3 показаны на рис. 4. Штриховой линией показано значение параметра q (функция $Q_{16}(k^*) = q$).

Как можно видеть, уже при $n = 2$ критерий $Q_{16} < q$ выполняется для всех значений k^{**} из таблицы. Причем для $k^{**} = 12$ значение $Q_{16} < 0,016$, т. е. почти на порядок меньше значения q . Если же проектировщик архива посчитает это значение Q_{16} недостаточным, то при $n = 3$ для $k^{**} = 12$ получаем $Q_{16} < 1,7 \cdot 10^{-5}$.

Расчет допустимого числа дисков в массиве RAID-6. Для проектировщиков электронных архивов на оптических дисках представляет интерес вопрос о максимально возможном количестве носителей в одном массиве RAID-6, при котором останется приемлемым показатель надежности хранения информации в архиве. Поскольку в массиве дисков RAID-6 $k^{**} = k + 2$, то с увеличением k^{**} будет уменьшаться отношение $(k^{**} - k) / k^{**} = 2 / k^{**}$, т. е. будет уменьшаться избыточность и возрастет эффективность использования дискового пространства, что в свою очередь приведет к уменьшению удельной стоимости хранения единицы информации. Но в то же время с увеличением k^{**} будет увеличиваться вероятность выхода из строя массива RAID-6 и в конечном итоге вероятность потери информации в архиве Q вне зависимости от способа создания запасных копий.

Понятно, что ответ на этот вопрос зависит от конкретных параметров электронного архива, в частности от значений параметров W и V . В настоящей статье выполнен расчет для электронного архива, равного по максимальной емкости рассмотренной библиотеке Panasonic LB-DH8 ($W = 91,2$ ТБ) [6], и для носителей того же типа ($V = 100$ ГБ), но без учета аппаратной структуры библиотеки и считая, что k^{**} может принимать любые значения ($k^{**} > 3$) и при этом $W = \text{const}$. Интерес представляют только расчеты для структуры R16 при $n > 1$. Результаты расчетов для $3 < k^{**} < 33$ ($1 < k < 31$) показаны на рис. 5. Для сравнения на том же рисунке приведены результаты аналогичных расчетов для структуры R6, а также график изменения величины $(k^{**} - k) / k^{**}$ (избыточности массива RAID-6), выраженный в процентах.

На рис. 5 видно, что, как и следовало ожидать, для структуры R6 приемлемое значение параметра Q может быть получено только для небольших значений k^{**} и нескольких запасных копий, приводящих к высокой избыточности при хранении информации (даже для $n = 3$ $Q > q$ уже при $k^{**} = 12$ избыточность 16,7 %). Конечно, приемлемость значения параметра Q определяет проектировщик архива, исходя из конкретных соображений. Однако, как и ранее, будем считать, что должно выполняться условие $Q < q$ (т. е. в данном случае необходимо $k^{**} < 12$ даже при $n = 3$).

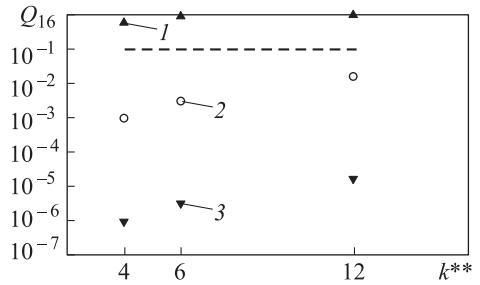


Рис. 4. Зависимость вероятности потери информации Q_{16} от числа дисков k^{**} в массиве RAID-6

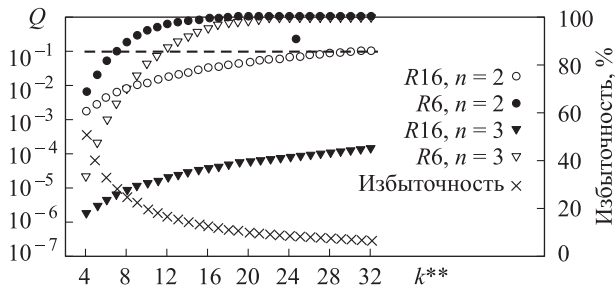


Рис. 5. Зависимость вероятности потери информации Q от числа дисков k^{**} в массиве RAID-6 для структур R6 и R16

Для структуры R16 уже при $n = 2$ значение k^{**} может достигать 29 ($Q < 0,098$, избыточность 6,9 %), а при $n = 3$ даже для $k^{**} = 32$ величина Q приближается к $1,5 \cdot 10^{-4}$ с достижением избыточности 6,25 %, т. е. возможно и дальнейшее увеличение параметра k^{**} .

Отметим, что на рынке присутствуют роботизированные библиотеки оптических дисков некоторых производителей (неунифицированные между собой по типам оптических дисков), допускающие в некоторых конфигурациях установку от 8 до 64 приводов оптических дисков [8, 9]. Однако в информационных материалах фирм-производителей этих библиотек отсутствует информация о возможности реализации на них режимов работы, допускающих создание массивов RAID (любого типа) из оптических дисков. Рассмотренная роботизированная библиотека Panasonic LB-DH8 в максимальной конфигурации может быть оснащена 7 группами приводов оптических дисков по 12 приводов в каждой группе, что дает 84 привода. Но производитель рассматривает эту возможность лишь как способ обеспечить параллельную работу на библиотеке нескольких пользователей и не предлагает создавать массивы RAID-6 из нескольких картриджей.

Заключение. Рассмотрен метод повышения надежности хранения информации в долговременных электронных хранилищах на оптических дисках, организованных в массивы RAID-6 за счет смешивания дисков запасных копий. Построена математическая модель надежности хранения информации в таком электронном архиве и на примере конкретной роботизированной библиотеки выполнен расчет, подтвердивший корректность метода.

Дополнительно для исходных данных той же роботизированной библиотеки выполнен расчет теоретически допустимого числа оптических дисков в массиве RAID-6. Показано, что при заданных условиях и критерии $Q_{16} < q$ массив RAID-6 может достигать 29 дисков при двух копиях информации ($n = 2$) и может существенно превышать 32 диска при $n = 3$.

ЛИТЕРАТУРА

1. ГОСТ Р 54989–2012. Обеспечение долговременной сохранности электронных документов. М.: Стандартинформ, 2013. 23 с.
2. ISO/TR 18492:2005. Long-term preservation of electronic document-based information. URL: http://www.iso.org/iso/catalogue_detail.htm?csnumber=38716 (дата обращения: 17.05.2016)
3. Российские организации начинают строить электронные архивы на оптических дисках // Ассоциация электронных торговых площадок. URL: <http://www.aetp.ru/market-news/item/400867> (дата обращения: 13.11.2015)
4. Чернышов А.В. К вопросу о применении оптических дисков для создания долго-временных электронных архивных хранилищ информации небольших организаций // Информационные технологии. 2016. Т. 22. № 8. С. 635–640.
5. Data archiver LB-DM9 series // Panasonic: веб-сайт компании. URL: <http://panasonic.net/avc/archiver/lb-dm9/LB-DM9series-PE.pdf> (дата обращения: 27.09.2016)
6. Data archiver LB-DH8 series // Panasonic: веб-сайт компании. URL: <http://panasonic.net/avc/archiver/lb-dh8> (дата обращения: 01.02.2016)
7. Робот-библиотекарь поселился в новом кампусе СПбГУ «Михайловская дача» // СПбГУ: веб-сайт университета. URL: <http://spbu.ru/smi/o-nas-pishut/24726-robot-bibliotekar-poselilsya-v-novom-kampuse-spbgu-mikhajlovskaya-dacha.html> (дата обращения: 13.11.2015)
8. Optical jukeboxes and libraries // Kintronics: веб-сайт компании. URL: <https://kintronics.com/solutions/optical-jukeboxes-and-libraries> (дата обращения: 14.05.2016)
9. New Sony Everspan library system delivers reliable optical disc archiving for data centers// CDRinfo: веб-сайт компании. URL: <http://www.cdrinfo.com/Sections/News/Details.aspx?NewsId=45763> (дата обращения: 14.05.2016)

Чернышов Александр Викторович — канд. техн. наук, доцент кафедры «Компьютерные системы и сети» Мытищинского филиала МГТУ им. Н.Э. Баумана (Российская Федерация, 141005, Московская обл., г. Мытищи 5, 1-я Институтская ул., д. 1).

Просьба ссылаться на эту статью следующим образом:

Чернышов А.В. Метод повышения надежности хранения информации в долговременных электронных хранилищах на оптических дисках, организованных в массивы RAID-6 за счет смешивания дисков запасных копий // Вестник МГТУ им. Н.Э. Баумана. Сер. Приборостроение. 2017. № 4. С. 88–97. DOI: 10.18698/0236-3933-2017-4-88-97

METHOD FOR IMPROVING INFORMATION STORAGE RELIABILITY IN LONG-TERM ELECTRONIC STORAGE ON OPTICAL DISKS ARRAYED IN RAID-6 BY MIXING BACKUP COPY DISKS

A.V. Chernyshov

sch@mgul.ac.ru

Mytischki Branch of Bauman Moscow State Technical University, Mytischki,
Russian Federation

Abstract

The purpose of this study was to propose a method for ensuring high reliability of information storage in long-term electronic storage on optical disks arrayed in RAID-6 by mixing backup copy disks. We built a mathematical model of information storage reliability. The sample calculations for robotic library Panasonic LB-DH8 show that the method makes it possible to minimize the number of information backup copies ensuring the reliability targets specified when designing the information storage. The LB-DH8 library allows for combining in RAID-6 only the discs within one cartridge and we can obtain a spare copy of the cartridge on another cartridge. However, the exchange of individual disks between the cartridges in automatic mode is impossible. Therefore, for practical implementation of the method on the library LB-DH8 operator intervention is required to manually exchange the respective discs between the cartridges. We demonstrated the use of the model for calculating the maximum number of optical disks in RAID-6 to reduce the redundancy of the disks array

Keywords

Electronic archiving information storage, long-term storage, optical storage media, information storage reliability, RAID-6 array

REFERENCES

- [1] GOST R 54989–2012. Obespechenie dolgovremennoy sokhrannosti elektronnykh dokumentov [State standard R 54989–2012. Long-term preservation of electronic document-based information]. Moscow, Standartinform Publ., 2013. 23 p.
- [2] ISO/TR 18492:2005. Long-term preservation of electronic document-based information. Available at: http://www.iso.org/iso/catalogue_detail.htm?csnumber=38716 (accessed 17.05.2016)
- [3] Rossiyskie organizatsii nachinayut stroit' elektronnye arkhivy na opticheskikh diskakh [Russian organizations are beginning to build electronic archives on optical discs]. Assotsiatsiya lektronnykh torgovykh ploshchadok: web-site. Available at: <http://www.aetp.ru/market-news/item/400867> (accessed 13.11.2015)
- [4] Chernyshov A.V. To the question of the optical discs application for long term digital archive storage of small organizations. *Informatsionnye tekhnologii* [Information Technologies], 2016, vol. 22, no. 8, pp. 635–640 (in Russ.).
- [5] Data archiver LB-DM9 series. Panasonic: company web-site. Available at: <http://panasonic.net/avc/archiver/lb-dm9/LB-DM9series-PE.pdf> (accessed 27.09.2016)

[6] Data archiver LB-DH8 series. Panasonic: company web-site.

Available at: <http://panasonic.net/avc/archiver/lb-dh8> (accessed 01.02.2016)

[7] Robot-bibliotekar' poselilsya v novomkampuse SPbGU «Mikhaylovskaya dacha» [Robot-librarian moved in new "Mikhaylovskaya dacha" campus of SPbU]. SPbGU: university web-site.

Available at: <http://spbu.ru/smi/o-nas-pishut/>

24726-robot-bibliotekar-poselilsya-v-novom-kampuse-spbgu-mikhajlovskaya-dacha.html

(accessed 13.11.2015)

[8] Optical jukeboxes and libraries. Kintronics: company web-site.

Available at: <https://kintronics.com/solutions/optical-jukeboxes-and-libraries> (accessed 14.05.2016)

[9] New Sony Everspan library system delivers reliable optical disc archiving for data centers.

CDRinfo: company web-site. <http://www.cdrinfo.com/Sections/News/Details.aspx?NewsId=45763>

(accessed 14.05.2016)

Chernyshov A.V. — Cand. Sc. (Eng.), Assoc. Professor of Computer Systems and Networks Department, Mytischki Branch of Bauman Moscow State Technical University (1-ya Institut-skaya ul. 1, Mytischki 5, Moscow Region, 141005 Russian Federation).

Please cite this article in English as:

Chernyshov A.V. Method for Improving Information Storage Reliability in Long-Term Electronic Storage on Optical Disks Arrayed in RAID-6 by Mixing Backup Copy Disks. *Vestn. Mosk. Gos. Tekh. Univ. im. N.E. Baumana, Priborostr.* [Herald of the Bauman Moscow State Tech. Univ., Instrum. Eng.], 2017, no. 4, pp. 88–97. DOI: 10.18698/0236-3933-2017-4-88-97