

## МОДЕЛЬ НАДЕЖНОСТИ ДОЛГОВРЕМЕННОГО ХРАНЕНИЯ ИНФОРМАЦИИ НА ОПТИЧЕСКИХ ДИСКАХ, ОБЪЕДИНЕННЫХ В ГИБРИДНЫЕ СТРУКТУРЫ

### Аннотация.

*Актуальность и цели.* Одной из задач проектирования долговременных электронных хранилищ является минимизация количества носителей информации при достижении максимально возможной надежности хранения информации. При использовании в качестве носителей оптических дисков эта задача может быть решена объединением носителей в структуры RAID. Однако при этом для архивов, не использующих дорогостоящие роботизированные библиотеки, возникает проблема оперативного доступа к хранимой информации. Проблема может быть решена объединением носителей в гибридные структуры, где наряду с массивами RAID информация хранится также и на одинарных носителях. Целью исследования является построение математической модели надежности хранения информации на оптических дисках, объединенных в гибридные структуры (названы в работе как G5, G6, G15, G16), и определение с помощью этой модели наиболее выгодной структуры.

*Материалы и методы.* Построение модели базируется на известных методах оценки надежности массивов RAID, разработанных для жестких дисков, которые были дополнены выражениями для учета надежности хранения информации на группах одиночных дисков. В качестве исходных данных для модели рассматриваются: общий объем архивного хранилища, емкость одного оптического диска, вероятность выхода из строя оптического диска при хранении между проверками целостности, условное количество дисков с информацией в массиве RAID, количество изготавливаемых запасных копий.

*Результаты.* Построена модель и проведено моделирование для различных значений емкости электронного архива. Для сравнения выполнено также моделирование для архивов на базе RAID наиболее распространенных типов (5, 6, 15, 16), а также для архива на базе одиночных дисков с запасными копиями. Подтверждено, что хотя гибридные структуры и проигрывают по надежности «чистым» RAID, но значительно превосходят архивы на одиночных дисках.

*Выводы.* По результатам исследования для создания архивов на оптических дисках, объединенных в гибридные структуры, можно однозначно рекомендовать структуру G16.

**Ключевые слова:** долговременное электронное хранилище информации, гибридные структуры оптических дисков, надежность хранения информации

А. В. Chernyshov

## MODEL OF RELIABILITY OF LONG-TERM STORAGE OF INFORMATION ON OPTICAL DISKS UNITED IN HYBRID STRUCTURES

© Чернышов А. В., 2020. Данная статья доступна по условиям всемирной лицензии Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), которая дает разрешение на неограниченное использование, копирование на любые носители при условии указания авторства, источника и ссылки на лицензию Creative Commons, а также изменений, если таковые имеют место.

**Abstract.**

*Background.* One of the tasks of designing long-term electronic storage is to minimize the number of data carriers while achieving the maximum possible reliability of information storage. When using optical disks as media, this problem can be solved by combining the media into RAID structures. However, for archives that do not use expensive robotic libraries, there is a problem of rapid access to stored information. The problem can be solved by combining media in hybrid structures, where along with RAID arrays; information is also stored on single media. The aim of the study is to build a mathematical model of the reliability of information storage on optical disks combined in hybrid structures (named in the paper as G5, G6, G15, G16), and determine the most profitable structure using this model.

*Materials and methods.* The model is based on well-known methods for evaluating the reliability of RAID arrays developed for hard disks, which were supplemented with expressions to account for the reliability of storing information on groups of single disks. The initial data for the model is considered: the total volume of archive storage, the capacity of a single optical disk, the probability of failure of the optical disk during storage between integrity checks, the conditional number of disks with information in the RAID array, and the number of spare copies being made.

*Results.* A model is constructed and simulations are performed for various values of the capacity of the electronic archive. For comparison, we also performed simulations for RAID-based archives of the most common types (5, 6, 15, 16), as well as for an archive based on single disks with spare copies. It is confirmed that although hybrid structures are less reliable than "pure" RAID, they are significantly superior to archives on single disks.

*Conclusions.* According to the results of the study, we can definitely recommend the G16 structure for creating archives on optical disks combined in hybrid structures.

**Keywords:** long-term storage of electronic information, hybrid structures of optical discs, reliability of information storage

## Введение

При проектировании архивных хранилищ электронной информации, призванных обеспечить сохранность информации в течение десятилетий, всегда встает вопрос об обеспечении надежности хранения информации с одновременным обеспечением доступности информации потенциальным потребителям. Одной из проблем здесь является естественная деградация с течением времени применяемых для хранения носителей информации. Существующие методические рекомендации, посвященные вопросам оцифровки «бумажных» документов и обеспечению надежности сохранения полученных электронных копий [1, 2], как правило, много внимания уделяют довольно сложному порядку обеспечения доступа потребителей к информации через использование дубликатов носителей, но при этом почти ничего не говорят о способах организации надежного хранения оригинальных носителей с архивной информацией, ограничиваясь общими рекомендациями о необходимости использовать надежные технические решения начиная от систем хранения данных (СХД) и заканчивая съемными носителями с необходимым количеством копий.

Между тем действующий стандарт [3] прямо указывает, что для хранения электронной информации в долговременных архивах должны использо-

ваться носители типа WORM, что на сегодня однозначно указывает на оптические носители однократной записи, т.е. носители типов: CD-R, DVD+/-R, BD-R и им подобные.

Как показывают исследования, для хранилищ большого объема типа СХД даже для информации, имеющей статус оперативной, большая часть информации запрашивается редко, а часть информации – вообще никогда [4]. Запросы на доступ к информации, находящейся в долговременном архиве, как правило, поступают редко. Поэтому хранение архивной информации на съемных носителях, к числу которых относятся и оптические диски однократной записи, является технически и экономически оправданным решением.

Как правило, информация записывается на одиночные архивные носители, с которых делаются дубликаты для предоставления информации пользователям. При этом архивные носители для обеспечения необходимой надежности хранения информации сразу записывают в виде нескольких запасных копий. Если обозначить через  $n$  количество необходимых запасных копий, то можно считать, что носители информации в архиве образуют структуры типа  $n$ -кратных RAID10 (хотя, в отличие от объединяемых в RAID10 жестких дисков в СХД, они продолжают использоваться только поодиночке). В дальнейшем будем обозначать эту структуру как R1.

В развитие этой идеи в работах [5, 6] было показано, что объединение архивных носителей информации в  $n$ -кратные структуры типа RAID5, RAID6, RAID15, RAID16 (обозначим их как R5, R6, R15, R16) способно при сохранении сопоставимого уровня надежности хранения информации уменьшить количество носителей в архиве по сравнению с  $n$ -кратными структурами R1. Такой подход, кроме экономии средств на приобретение носителей, должен обеспечить экономию пространства для хранения носителей, а заодно и экономию времени на регулярный контроль целостности информации, хранимой на носителях, что в свою очередь позволяет увеличить емкость архива.

Проблема, однако, заключается в том, что съемные носители, составляющие структуры типа R5, R6, R15, R16, при воспроизведении информации должны быть реально объединены в RAID. Это не составляет большой проблемы при использовании дорогостоящих роботизированных библиотек типа [7], но во многих других случаях, когда, прежде всего по экономическим соображениям, дорогостоящее оборудование применено быть не может, время доступа к архивной информации может существенно возрасти.

Например, несложно представить себе архив небольшой организации с одним сервером и одним пишущим/воспроизводящим приводом. Информация к записи на съемные носители накапливается на жестком диске сервера. При этом на жестком диске выделены файлы с образцами для записи съемных носителей, эти файлы логически объединены в структуру RAID, на которую и выполняется запись информации, помещаемой в архив. Как только очередная структура RAID будет заполнена информацией, каждый из файлов образцов последовательно (с применением одного пишущего привода) записывается на  $n$  съемных носителей, что и обеспечивает формирование в архиве соответствующих структур RAID из съемных носителей.

Однако теперь при необходимости воспроизведения архивной информации необходимо выполнить обратный перенос на диск сервера полного

объема одного из массивов RAID в виде файлов образов и только после этого RAID может быть логически собран, обеспечив доступ к информации.

Описанная проблема может быть решена, если из  $n$  запасных копий в архиве только  $(n - 1)$  будут представлять собой структуры RAID, а одна копия будет записана на одиночных носителях. Назовем такие структуры гибридными. Такие структуры могут быть созданы на базе любой из структур типа R5, R6, R15, R16. Обозначим их, соответственно, как G5, G6, G15, G16.

Долговременные архивы, выполненные на съемных носителях, объединенных в гибридные структуры, ранее в литературе не рассматривались. Для проектирования и последующей эксплуатации таких архивов необходимо построить модель, позволяющую оценить надежность хранения информации в зависимости от ряда факторов, включая количество запасных копий  $n$ .

## 1. Материал и методика

Для построения модели зададимся следующими параметрами:  $W$  – общий объем создаваемого архивного хранилища;  $V$  – емкость одного съемного носителя информации.

Информация будет записываться в архив блоками, определяемыми размером RAID применяемого типа.

Отталкиваясь от моделей, применяемых для рассмотрения массивов RAID, построенных на жестких дисках [8–10], опишем их в общем виде. Пусть  $k$  – условное количество носителей с хранимой информацией ( $k > 1$ ),  $m$  – условное количество носителей с контрольной информацией, тогда

$$k^* = k + m, k > 1, m = 1 \text{ для R5 и R15, } m = 2 \text{ для R6 и R16.}$$

Обозначим  $n$  – количество запасных копий информации,  $q$  – вероятность выхода из строя одного носителя информации за период между последовательными контролями исправности (обычно 1 год),  $L$  – количество блоков, которые будет содержать архив ( $L = (l / k)$ , где  $l = (W / V)$ ).

Для структур типа G5, G6 имеем следующие соображения.

В блоке имеется  $k$  одиночных дисков с информацией, причем на каждом диске информация уникальна. Информация будет хотя бы частично потеряна, если будет испорчен хотя бы один диск. Следовательно, вероятность потери информации будет оценена как

$$Q_{\text{од}} = 1 - (1 - q)^k.$$

В блоке также имеется несколько массивов R5 или R6, дублирующих информацию одиночных дисков в  $(n - 1)$  копий. Вероятность потери информации на одном массиве запишется как

$$Q_R = 1 - \sum_{i=0}^m \left( q^i (1 - q)^{k-i} C_k^i \right) = \sum_{i=m+1}^k \left( q^i (1 - q)^{k-i} C_k^i \right),$$

где  $C_k^i = \frac{k!}{i!(k-i)!}$ .

Таким образом, вероятность потери информации в одном блоке можно оценить как  $Q_{од}Q_R^{n-1}$ , а в  $L$  блоках соответственно

$$Q = 1 - \left(1 - Q_{од}Q_R^{n-1}\right)^L.$$

Для структур типа G15, G16 для одного блока значение  $Q_{од}$  не изменится, а вероятность потери информации на массивах R15, R16 в одном блоке для  $(n - 1)$  копии составит

$$Q_{R1} = 1 - \sum_{i=0}^m \left( \left( q^{n-1} \right)^i \left( 1 - q^{n-1} \right)^{k-i} C_k^i \right).$$

Для всего архива вероятность потери информации составит

$$Q = 1 - \left(1 - Q_{од}Q_{R1}^{n-1}\right)^L.$$

Модель построена.

## 2. Результаты

Для построенных моделей было проведено численное моделирование для архивов различной емкости  $W$  в пределах от 1 до 90 Тбайт. Моделирование выполнялось для архивов на оптических дисках емкостью  $V = 25$  Гбайт (BD-R). Использовалось наиболее вероятное значение вероятности выхода диска из строя  $q = 0,1$  [11]. Расчеты были выполнены для наиболее надежных (и, соответственно, наиболее избыточных) массивов RAID при  $k = 2$ .

На рис. 1 показаны графики зависимости вероятности потери информации в архиве  $Q$  от объема архива  $W$  для количества копий информации  $n$ , при котором  $Q < 0,0001$  (разработчик архива может рассмотреть и другие значения  $Q$ ). Видно, что вне зависимости от объема архива соотношение значений  $Q$  для разных структур архива остается практически постоянным. Следовательно, для определения наиболее выгодной структуры с точки зрения количества запасных копий  $n$  в большинстве случаев достаточно рассмотреть архив любой фиксированной емкости.

В табл. 1 приведены расчетные данные (количество носителей) / (количество копий) для  $W = 10$  Тбайт.

## 3. Обсуждение

Прежде всего необходимо заметить, что, хотя выкладки для структур G5, G6 и были сделаны, на практике нет большого смысла в их применении. Гибридные структуры являются компромиссом между скоростью доступа к информации и надежностью хранения, применяемым при невозможности использования дорогостоящих роботизированных библиотек. В этом случае все манипуляции с архивными носителями выполняются вручную, что позволяет с одинаковым успехом реализовать как структуры G5, G6, так и структуры G15, G16, но последние имеют значительно более высокую надежность хранения.

По значениям в табл. 1 можно видеть, что для достижения сопоставимой надежности хранения информации гибридным структурам требуется не-

много больше запасных копий, чем «чистым» RAID. Однако это количество копий все равно существенно меньше, чем для структуры R1.

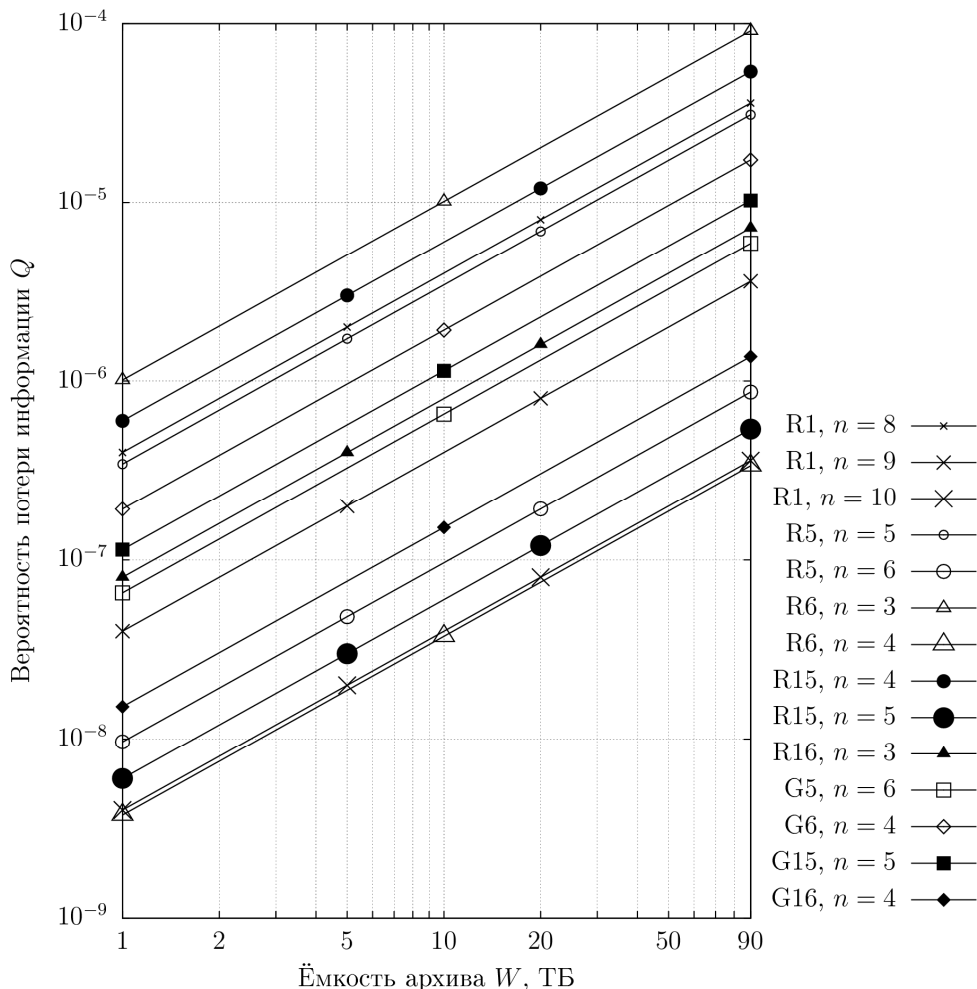


Рис. 1. Зависимость вероятности потери информации от емкости архива для различных структур, в которые организованы носители информации

Среди всех гибридных структур оптимальной с точки зрения максимальной надежности хранения информации при минимальном количестве носителей является структура G16.

### Заключение

В статье построена модель надежности хранения информации в долговременных электронных архивах, реализованных на оптических дисках, объединенных в гибридные структуры.

Показано, что с практической точки зрения представляют интерес только гибридные структуры типа G15, G16.

Проведено численное моделирование и на основе полученных результатов сделан вывод, что наиболее предпочтительной для применения в электронных архивах является структура типа G16.

Таблица 1

Тип структуры	Количество дисков / количество копий информации											
	хуже ← Q → лучше											
R1			3200/8						3600/9			
R5										3600/6		
R15		2400/4			3000/5						3000/5	
R6	2400/3											3200/4
R16							2400/3					
G5								3400/6				
G15							2800/5					
G6					2800/4							
G16										2800/4		

**Библиографический список**

1. **Юмашева, Ю. Ю.** Методические рекомендации по электронному копированию архивных документов и управлению полученным информационным массивом / Ю. Ю. Юмашева. – Москва : Росархив, ВНИИДАД, 2012. – 125 с.
2. Рекомендации по комплектованию, учету и организации хранения электронных архивных документов в государственных и муниципальных архивах. – Москва : Федеральное архивное агентство. ВНИИДАД, 2013. – 49 с.
3. ГОСТ Р 54989–2012 / ISO TR 18492:2005. Обеспечение долговременной сохранности электронных документов. – Москва, 2013.
4. **Корепанов, И.** Как сохранить архив на десятилетия? / И. Корепанов // Журнал сетевых решений/LAN. – 2008. – № 03. – URL: <http://www.osp.ru/lan/2008/03/4899898/> (дата обращения: 04.11.2015).
5. **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 / А. В. Чернышов // Вестник Московского государственного технического университета им. Н. Э. Баумана. Серия Приборостроение. – 2017. – № 4. – С. 88–97. – DOI 10.18698/0236-3933-2017-4-88-97.
6. **Чернышов, А. В.** Исследование свойств долговременных электронных архивных хранилищ информации на оптических дисках, организованных в структуры RAID-5 / А. В. Чернышов // Информационные технологии. – 2018. – Т. 24, № 9. – С. 586–593. – DOI 10.17587/it.24.586-593.
7. Data Archiver LB-DH8 series. – URL: <http://panasonic.net/avc/archiver/lb-dh8/> (дата обращения: 01.02.2016).
8. **Thomasian, A.** Hierarchical RAID: Design, performance, reliability, and recovery / A. Thomasian, Y. Tang, Y. Hu // J. Parallel Distrib. Comput. – 2012. – Vol. 72. – P. 1753–1769. – DOI 10.1016/j.jpdc.2012.07.002
9. **Thomasian, A.** Shortcut method for reliability comparisons in RAID / A. Thomasian // Journal of Systems and Software. – 2006. – Vol. 79. – P. 1599–1605.
10. **Yuan, D.** A Randomly Expandable Method For Data Layout Of Raid Storage Systems / D. Yuan, X. Peng, T. Liu, Z. Cui // International Journal of Innovative Computing, Information and Control. – 2018. – Vol. 14, № 3. – P. 1079–1094.
11. **Чернышов, А. В.** Определение значения вероятности выхода из строя однократно записываемого оптического диска с записью в процессе хранения в долговременном архиве / А. В. Чернышов // Лесной вестник / Forestry Bulletin. – 2019. – Т. 23, № 4. – С. 82–87. – DOI 10.18698/2542-1468-2019-4-82-87

**References**

1. Yumasheva Yu. Yu. *Metodicheskie rekomendatsii po elektronnomu kopirovaniyu arkhivnykh dokumentov i upravleniyu poluchennym informatsionnym massivom* [Methodical recommendations for electronic copying of archival documents and management of the received information array]. Moscow: Rosarkhiv, VNIIDAD, 2012, 125 p. [In Russian]
2. *Rekomendatsii po komplektovaniyu, uchetu i organizatsii khraneniya elektronnykh arkhivnykh dokumentov v gosudarstvennykh i munitsipal'nykh arkhivakh* [Recommendations for the acquisition, accounting and storage of electronic archival documents in state and municipal archives]. Moscow: Federal'noe arkhivnoe agentstvo. VNIIDAD, 2013, 49 p. [In Russian]
3. GOST R 54989–2012 / ISO TR 18492:2005. *Obespechenie dolgovremennoy sokhranosti elektronnykh dokumentov* [Ensuring long-term preservation of electronic documents]. Moscow, 2013. [In Russian]



4. Korepanov I. *Zhurnal setevykh resheniy/LAN* [Network Solutions Journal/LAN]. 2008, no. 03. Available at: <http://www.osp.ru/lan/2008/03/4899898/> (accessed 04.11.2015). [In Russian]
5. Chernyshov A. V. *Vestnik Moskovskogo gosudarstvennogo tekhnicheskogo universiteta im. N. E. Baumana. Seriya Priborostroenie* [Bulletin of Bauman Moscow State Technical University. Series: Instrument engineering]. 2017, no. 4, pp. 88–97. DOI 10.18698/0236-3933-2017-4-88-97. [In Russian]
6. Chernyshov A. V. *Informatsionnye tekhnologii* [Information Technology]. 2018, vol. 24, no. 9, pp. 586–593. DOI: 10.17587/it.24.586-593. [In Russian]
7. *Data Archiver LB-DH8 series*. Available at: <http://panasonic.net/avc/archiver/lb-dh8/> (accessed 01.02.2016).
8. Thomasian A., Tang Y., Hu Y. *J. Parallel Distrib. Comput.* 2012, vol. 72, pp. 1753–1769. DOI 10.1016/j.jpdc.2012.07.002
9. Thomasian A. *Journal of Systems and Software*. 2006, vol. 79, pp. 1599–1605.
10. Yuan D. A., Peng X., Liu T., Cui Z. *International Journal of Innovative Computing, Information and Control*. 2018, vol. 14, no. 3, pp. 1079–1094.
11. Chernyshov A. V. *Lesnoy vestnik* [Forestry Bulletin]. 2019, vol. 23, no. 4, pp. 82–87. DOI 10.18698/2542-1468-2019-4-82-87 [In Russian]

---

**Чернышов Александр Викторович**

кандидат технических наук, доцент,  
кафедра прикладной математики,  
информатики и вычислительной техники,  
Мытищинский филиал Московского  
государственного технического  
университета имени Н. Э. Баумана  
(национального исследовательского  
университета) (Россия,  
Московская область, Мытищи-5,  
ул. 1-я Институтская, 1)

E-mail: sch-ru@yandex.ru

**Chernyshov Aleksandr Viktorovich**

Candidate of engineering sciences, associate  
professor, sub-department of applied  
mathematics, computer science  
and engineering, Mytischki Branch  
of Bauman Moscow State Technical  
University (National Research University)  
(1 1st Institutskaya street, Mytischki-5,  
Moscow region, Russia)

---

**Образец цитирования:**

Чернышов, А. В. Модель надежности долговременного хранения информации на оптических дисках, объединенных в гибридные структуры / А. В. Чернышов // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2020. – № 2 (54). – С. 25–33. – DOI 10.21685/2072-3059-2020-2-3.