

А. В. Чернышов, канд. техн. наук, доц. кафедры ИУ-6 МФ, sch@mgul.ac.ru
Мытищинский филиал МГТУ им. Н. Э. Баумана, Мытищи, Российская Федерация

Исследование свойств долговременных электронных архивных хранилищ информации на оптических дисках, организованных в структуры RAID-5

Рассмотрена проблема создания локальных архивных хранилищ электронной информации для небольших организаций. Наряду с надежностью хранения информации большое внимание уделено минимизации стоимости создания и владения архивным хранилищем. Для достижения цели предложено использовать оптические диски однократной записи, объединенные в массивы RAID-5. Показано, что в большинстве случаев наилучшие результаты достигаются при использовании структур RAID-5 с самостоятельными запасными копиями каждого диска в каждом массиве RAID-5. Но при минимальном размере архива (порядка 1 Тбайт) более выгодным может оказаться использование одиночных дисков с тремя копиями. Показано, что возможно увеличение числа дисков в массивах RAID-5 для сокращения удельной стоимости хранения информации при соответствующем уменьшении надежности хранения информации.

Ключевые слова: электронное архивное хранилище информации, длительное хранение, оптические носители информации, надежность хранения информации, минимизация затрат

Введение

В мире существует большое число небольших организаций, которым по тем или иным причинам необходимо локальное файловое хранилище. Например, в России одних только школьных музеев (число сотрудников часто менее 5) на 2010 г. существовало более 11 000 [1]. Объемы фондов таких музеев при оцифровке, как правило, не превышают 10 Тбайт, но в некоторых случаях могут достигать нескольких десятков терабайт. Обычно такие организации не могут использовать дорогостоящее профессиональное оборудование для создания локального файлового хранилища. Использование же для этих целей облачных сервисов нежелательно по целому ряду причин [2].

Между тем сегодня появилась возможность создания архивных файловых хранилищ довольно больших объемов из недорогих компьютерных комплектующих, которые можно приобрести в обычном компьютерном магазине.

Современный стандарт архивного хранения информации [3] требует использовать носители информации однократной записи, т. е. оптиче-

ские диски однократной записи (CD-R, DVD-R, BD-R). Для записи или воспроизведения каждого такого диска требуется специальный привод. На рынке представлены роботизированные библиотеки оптических дисков [4, 5]. Но цена этих устройств [6, 7] не позволяет небольшим организациям применять их для создания своих локальных файловых архивных хранилищ. Однако при работе с дисками "вручную" ничто не мешает созданию архивного файлового хранилища на оптических дисках с минимальными затратами средств. Необходимы лишь привод оптических дисков, жесткий диск большой емкости (обычно являются частью самого компьютера), сами оптические диски однократной записи и место для их хранения.

По разным причинам оптические диски время от времени все же выходят из строя [8, 9]. Причем предсказать априори вероятность отказа оптических дисков из конкретной партии практически невозможно. Поэтому при размещении в архивном файловоом хранилище очередного оптического диска сразу необходимо изготовление нескольких его копий, также помещаемых на архивное хранение.

Снизить стоимость хранения данных можно за счет объединения оптических дисков в отказоустойчивые массивы RAID, наиболее дешевым из которых оказывается RAID-5. Некоторые современные библиотеки оптических дисков предлагают такую возможность в качестве одного из режимов работы [4, 6].

С практической точки зрения представляет интерес определение оптимального числа дисков в группе, организованной в RAID-5, а также числа копий дисков, при которых заданный уровень надежности хранения информации заданного объема будет достигнут с минимальной стоимостью при априори неизвестной надежности применяемых оптических дисков.

Практически все опубликованные ранее исследования (в частности, [10–12]) рассматривают массивы RAID большого объема, построенные на жестких дисках (вероятность отказа одного диска, как правило, 10^{-15} и менее) и без учета стоимости решения.

В данной работе рассматривается вопрос надежности архивов относительно небольшого объема, построенных на дисках однократной записи (вероятность отказа одного диска в общем случае непредсказуема и может принимать значения вплоть до 10^{-1} , а в отдельных случаях еще хуже) с требованием минимизации стоимости хранения информации.

Постановка задачи

Пусть необходимо создать долговременное электронное хранилище информации (файловый архив) на оптических дисках однократной записи для небольшой организации. Необходимо минимизировать затраты на создание и сопровождение файлового архива, но при этом обеспечить заданное значение надежности хранения информации при априори неизвестной вероятности отказа оптических дисков, применяемых для хранения информации, путем создания нескольких копий каждого диска.

Прогнозируемый максимальный объем сохраняемой в архиве информации задается при проектировании структуры архива и может лежать в пределах от единиц до десятков терабайт.

Модели надежности хранения информации в архиве из оптических дисков

Построение и исследование модели надежности хранения информации в архиве из неза-

висимых оптических дисков было выполнено в работе [13]. Обозначим для краткости этот тип хранилища как R1. Согласно полученной модели вероятность Q_n потерять хотя бы часть информации составит

$$Q_n = 1 - (1 - q^n)^l.$$

Стоимость такого решения составит

$$C_n = c \cdot l \cdot n + C_{\text{п}},$$

где $l = [W/V]$ — число оптических дисков, необходимых для сохранения информации объема W на дисках объемом V ; $[\cdot]$ — операция округления до большего целого; q — вероятность выхода из строя одного оптического диска в течение периода между контролями исправности; n — число запасных копий каждого диска; c — стоимость одного оптического диска; $C_{\text{п}}$ — стоимость одного привода оптических дисков.

Из выкладок [12] для структур RAID-5 вероятность потери хотя бы части информации при отсутствии запасных копий:

— для одного массива RAID-5

$$Q_{\text{RAID5}} = 1 - (1 - q)^k(1 + kq);$$

— для всего объема сохраняемой информации

$$Q_5 = 1 - ((1 - q)^k(1 + kq))^L,$$

где k — число дисков в одном массиве RAID-5 без учета избыточного диска ($k > 2$); L — число массивов RAID-5, необходимое для сохранения всей информации объема W .

Для увеличения надежности сохранения информации в архиве необходимо создание запасных (резервных) копий архивных дисков. Для структур RAID-5 возможно два варианта создания таких копий:

1) каждая группа RAID-5 из k^* дисков ($k^* = k + 1$) рассматривается и используется как неделимое целое; создаются n копий каждой группы RAID-5 (рис. 1); обозначим этот вариант как R5;

2) создаются n копий каждого отдельного диска в архиве (можно условно считать, что эти копии образуют структуру RAID-1, хотя в действительности диски записывают и используют поочередно); для считывания информации массив RAID-5 может быть создан из любых копий каждого из k^* дисков группы (условно можно рассматривать получившуюся структуру как RAID-15) (рис. 2); обозначим этот вариант как R15.

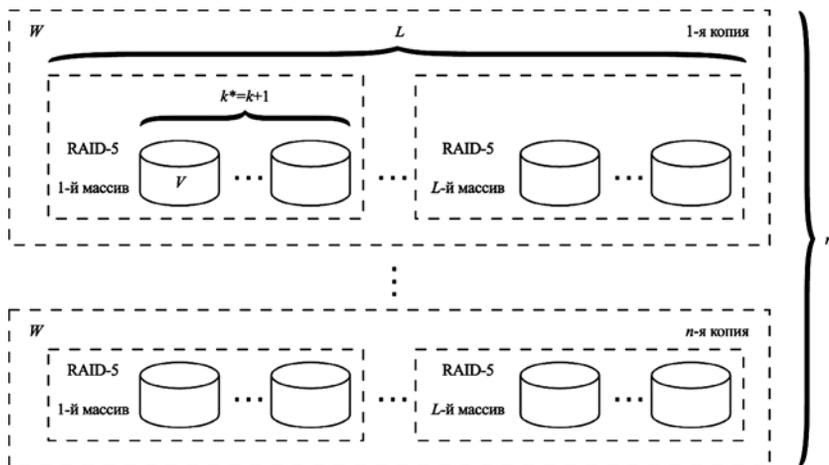


Рис. 1. Структура электронного файлового архива с дисками, организованными в структуры RAID-5, при создании запасных копий RAID-5 (структура R5): W — общий объем данных, сохраняемый в архиве; V — емкость одного оптического диска

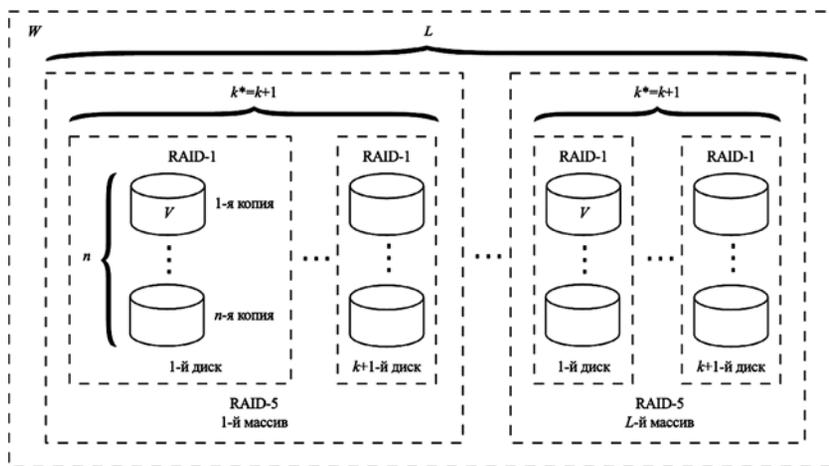


Рис. 2. Структура электронного файлового архива с дисками, организованными в структуры RAID-5, с запасными копиями каждого диска (структура R15)

В работе [12] исследована надежность этих структур дисковых массивов, но только для $n = 2$ и только для жестких дисков. Рассмотрим эти структуры с учетом специфики дисковых массивов, построенных из однократно записываемых оптических дисков.

Вариант R5 [12] обеспечивает меньшую надежность сохранения информации. Однако для роботизированных библиотек, работающих с оптическими дисками, объединенными в картриджи [4], технически возможным оказывается только он. Вариант же R15 возможен либо при полностью "ручной" сборке RAID-5, либо на оптических библиотеках, не использующих картриджи, и только если в такой библиотеке имеется не менее трех приводов оптических дисков. Например, в работе [14] описана роботизированная библиотека, которая может иметь до 64 одновременно работающих приводов.

Для варианта R5: вероятность потери информации на одном массиве RAID-5 при n копиях равна $(Q_{\text{RAID5}})^n$, вероятность сохранения информации на n копиях массива RAID-5 равна $1 - (Q_{\text{RAID5}})^n$, вероятность сохранения полного объема информации W на L массивах RAID-5 с n копиями равна $(1 - (Q_{\text{RAID5}})^n)^L$, а вероятность потери хотя бы части информации в таком архиве равна

$$Q_{5n} = 1 - (1 - (Q_{\text{RAID5}})^n)^L = 1 - (1 - (1 - (1 - q)^k(1 + kq)))^n)^L.$$

Для варианта R15: можно считать, что каждый массив RAID-5 образован дисковыми массивами RAID-1 по n дисков в каждом, вероятность потери информации на каждом из которых равна q^n .

Используя выведенную ранее формулу вероятности потери хотя бы части информации для L массивов RAID-5 (параметр Q_5), получим формулу вероятности потери хотя бы части информации для L структур типа R15 (заменяв значение q вероятности отказа одного диска на значение q^n вероятности отказа массива RAID-1):

$$Q_{15} = 1 - ((1 - q^n)^k(1 + kq^n))^L.$$

Стоимость обоих вариантов архива будет одинакова и составит

$$C_{5n} = C_{15} = ncl^* + C_{\text{п}}k^*,$$

где $l^* = Lk^*$ — общее число дисков для сохранения одной копии информации объемом W . В формуле учтено, что для работы с массивами RAID-5 потребуется одновременное использование k^* приводов оптических дисков.

Исследование построенных моделей

Для файловых архивов небольших организаций имеет смысл рассмотреть построенные модели для значений W , равных 1, 10 и 90 Тбайт. Последнее значение близко к емкости промышленной библиотеки оптических дисков Panasonic LB-DH8 [4] в базовой комплектации, имея

Таблица 1

| Тип диска | Емкость диска V , Гбайт | Абсолютная цена c , руб. | | Удельная цена c/V , руб./Гбайт | |
|-----------|---------------------------|----------------------------|--------------|----------------------------------|--------------|
| | | Минимальная | Максимальная | Минимальная | Максимальная |
| CD-R | 0,7 | 20 | 67 | 28,57 | 95,71 |
| DVD-R | 4,7 | 21 | 106 | 4,47 | 22,55 |
| BD-R | 25 | 65 | 140 | 2,6 | 5,6 |
| BD-R DL | 50 | 160 | 510 | 3,2 | 10,2 |
| BD-R XL | 100 | 1100 | 3606 | 11 | 36,06 |

в виду, что для электронных архивов небольших организаций объем данных, сохраняемых в архиве, обычно меньше этого значения.

В качестве оптического диска в данной работе будет рассматриваться только носитель типа BD-R емкостью 25 Гбайт ($V = 25$). Этот выбор обусловлен необходимостью минимизации затрат на закупку дисков и отвечает реалиям рынка оптических носителей в России. Как следует из табл. 1, построенной по данным [15–17], диски BD-R и для минимальной, и для максимальной цены на рынке обладают наименьшей удельной стоимостью на единицу сохраняемой информации.

Поскольку при создании архива вероятность отказа используемых оптических дисков априори неизвестна, будем при исследовании моделей использовать "заведомо плохое" значение $q = 0,1$ (то есть отказывает каждый десятый диск). Это значение q следует, в частности, из работы [8]. Будем считать, что результаты данного исследования покажут "заведомо наихудший" случай; реальная картина сохраняемости информации в архиве должна быть лучше.

Построим графики зависимости $Q(q)$ для различных значений W и для разного числа копий информации n (рис. 3).

Графики для структур R5 и R15 построены для $k = 2$, обеспечивающем минимальное (наилучшее) значение параметра Q .

Условимся считать, что Q создаваемого архива должно быть по крайней мере не хуже q оптических дисков, которые его составляют, т. е. $Q < q$. Это условие выполняется для графиков, оказавшихся под штриховой линией, соответствующей функции $Q(q) = q$. Оно никогда не выполнится для архива R1 при $n = 1$ (без копий) [13], поэтому эти графики на рис. 3 не показаны и далее при анализе не рассматриваются.

На рис. 3 можно видеть, что:

— для $W = 1$ Тбайт условие $Q < q$ выполняется при $n > 2$ для архива R1 и при $n > 1$ для архивов R5, R15 уже при $q = 0,1$, а с уменьшением q до $\sim 10^{-2}$ становится выполнимым для любых структур;

— для $W = 10$ Тбайт это условие при $q = 0,1$ выполняется при $n > 3$ для архива R1 и при $n > 2$ для архивов R5, R15 (для R15 оно выполняется и для $n = 2$); для любых структур оно начинает выполняться при уменьшении значения q до $\sim 10^{-3}$;

— для $W = 90$ Тбайт это условие при $q = 0,1$ выполняется при $n > 2$ для R5, R15 и при $n > 4$ для R1; согласно выполненным расчетам, для любых структур оно начинает выполняться при уменьшении значения q до $\sim 10^{-4}$;

— для любого из рассмотренных значений W можно говорить о сопоставимости значений Q и их характера изменения при изменении q :

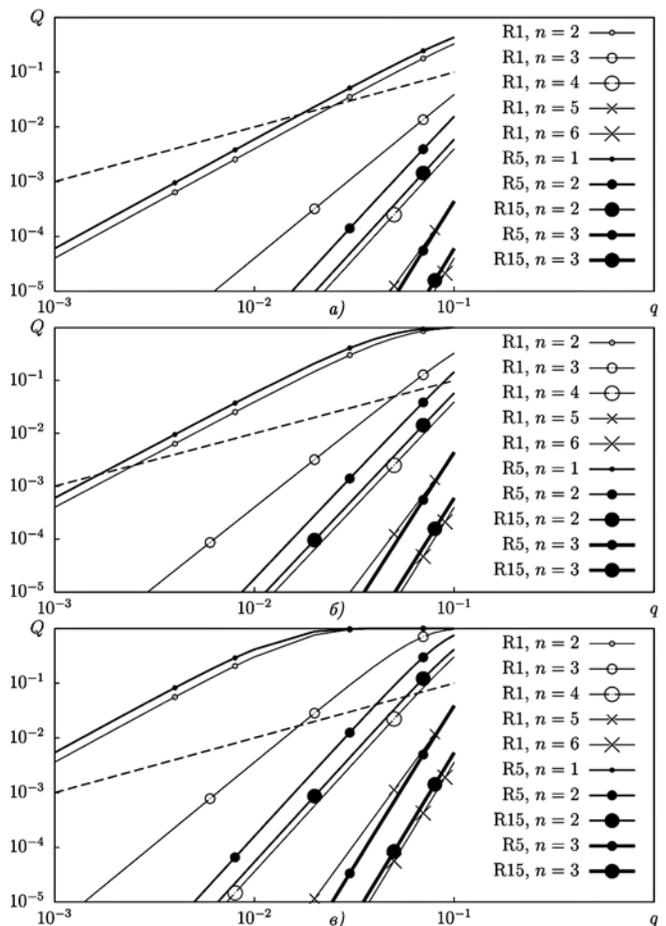


Рис. 3. Характер изменения вероятности потери информации Q при изменении вероятности отказа одного диска q в районе $q = 0,1$ при $k = 2$:

a — для $W = 1$ Тбайт; b — для $W = 10$ Тбайт; c — для $W = 90$ Тбайт

— для R1 с $n = 2$ и R5 с $n = 1$;
 — для R1 с $n = 4$ и R5, R15 с $n = 2$. В последнем случае структура R1 с точки зрения значения параметра Q имеет заметное на графике преимущество перед R5 и незначительное перед R15, но по числу оптических дисков R1 проигрывает R5 и R15, а структура R1 с $n = 3$ имеет одинаковое со структурами R5, R15 число дисков, но проигрывает им в значении параметра Q , что особенно заметно при уменьшении значения параметра q .

Анализируя графики на рис. 3, построим сводную рекомендательную таблицу выбора значений n для различных структур архива, организованного на оптических дисках DB-R ($V = 25$ Гбайт), чтобы выполнялся сформулированный выше минимальный критерий $Q < q$ для выбранного нами "наихудшего" значения $q = 0,1$ (табл. 2).

Анализируя табл. 2, необходимо помнить, что в каждой строке $Q(R1, n_{\text{табл}}) \gg Q(R15, n_{\text{табл}})$, но при этом $Q(R15, n_{\text{табл}})$ довольно близко к $Q(R1, n_{\text{табл}} + 1)$, хотя и не достигает его. В табл. 3 выполнен расчет числа оптических дисков, необходимых для создания архива в соответствии с числом копий, приведенным в табл. 2.

По табл. 2 и 3 можно сделать вывод, что если пренебречь стоимостью двух дополнительных оптических приводов (при $k = 2$), то с учетом более низкого (лучшего) значения Q уже при $W = 1$ Тбайт наилучшим (наиболее надежным и наименее затратным) решением будет создание архива на структурах R15 при $n = 2$.

Архив на базе структур R5 по надежности проигрывает R15, а при $W = 10$ Тбайт он

Таблица 2

| W, Тбайт | Значение n для структур архива: | | |
|----------|---------------------------------|------------|-------------|
| | R1 | R5 (k = 2) | R15 (k = 2) |
| 1 | 3 | 2 | 2 |
| 10 | 4 | 3 | 2 |
| 90 | 5 | 3 | 3 |

Таблица 3

| W, Тбайт | Суммарное число оптических дисков | | | |
|----------|-----------------------------------|---------------------------------|----------------------------------|------------------------------|
| | R1 ($n_{\text{табл}}$) | R5 ($k = 2, n_{\text{табл}}$) | R15 ($k = 2, n_{\text{табл}}$) | R1 ($n_{\text{табл}} + 1$) |
| 1 | 120 | 120 | 120 | 160 |
| 10 | 1600 | 1800 | 1200 | 2000 |
| 90 | 18 000 | 16 200 | 16 200 | 21 600 |

проигрывает даже структуре R1 (при выполнении условия $Q < q$). При $W = 90$ Тбайт R5 уже выигрывает у R1 по числу дисков, но, как ни странно, проигрывает ему по надежности хранения информации (по графикам это определить трудно — факт установлен анализом массива расчетных данных). Поэтому и здесь архив на структурах R15 ($n = 3$) оказывается предпочтительным.

В случае же, если стоит жесткое требование минимизации затрат на создание архива и при этом требуется обеспечить минимальное время доступа к информации (ограниченное временем ручного перемещения оптического диска из архива в оптический привод), то для небольшого архива с $W = 1$ Тбайт наилучшим решением оказывается структура R1 (приобретение дополнительных оптических приводов исключается) с трехкратным дублированием каждого диска ($n = 3$).

Для архива с объемом $W = 10$ Тбайт выбор должен быть сделан только между структурами R1 ($n = 4$) и R15 ($n = 2$). Определив, используя [15], стоимость одного оптического привода $C_{\text{п}} = 11\,890$ руб. (была выбрана модель привода с максимальной стоимостью), можно легко рассчитать, что приобретение двух дополнительных оптических приводов для реализации структуры R15 обойдется дешевле, чем приобретение 400 дополнительных оптических дисков для структуры R1 даже при минимальной стоимости каждого оптического диска (23 780 руб. против 26 000 руб.), а поскольку использование дешевых оптических дисков для записи архивной информации категорически нежелательно, то разница в затратах будет существенно больше.

Для архива с объемом $W = 90$ Тбайт очевидным выбором является структура R15 при $n = 3$.

Выбор оптимального числа оптических дисков в каждом массиве RAID

Рассмотрим теперь вопрос выбора оптимального значения параметра k .

Очевидно, что с увеличением k будет уменьшаться удельная стоимость хранения единицы информации и одновременно будет возрастать вероятность отказа отдельно взятого массива. Кроме того, для обеспечения оперативного доступа к архивной информации (сборки массива RAID-5) необходимо будет иметь $k + 1$ оптических приводов.

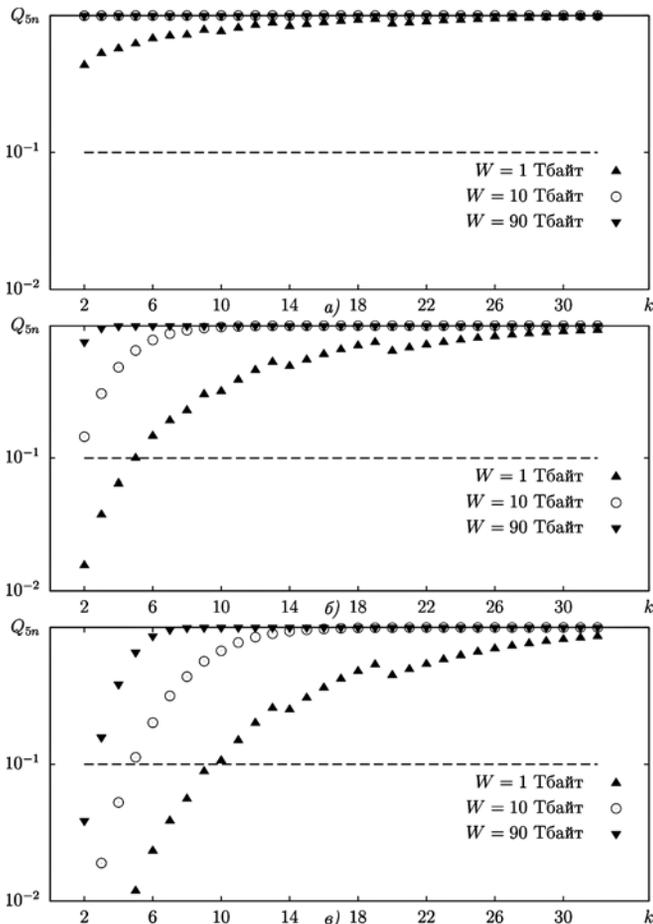


Рис. 4. Зависимость вероятности потери информации Q от числа дисков k в массиве RAID-5. Архив структуры R5: а — $n = 1$; б — $n = 1$; в — $n = 3$

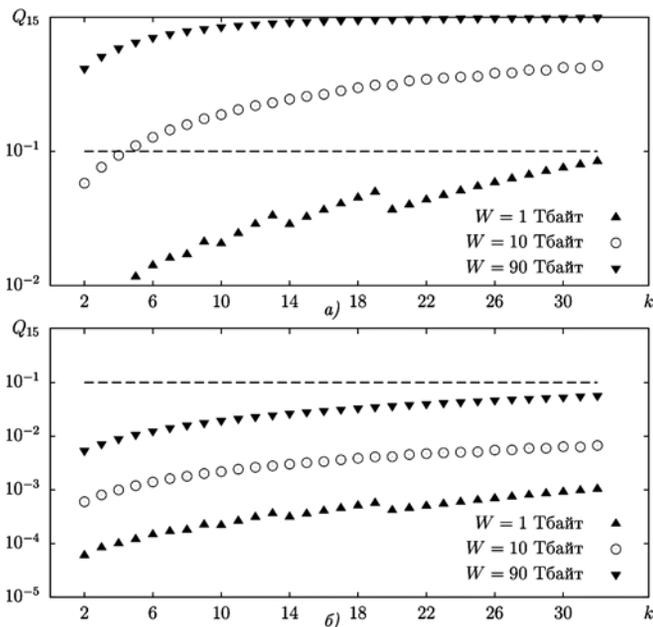


Рис. 5. Зависимость вероятности потери информации Q от числа дисков k в массиве RAID-5. Архив структуры R15: а — $n = 2$; б — $n = 3$

В связи с этим для небольших архивов (типа школьных музеев) вряд ли имеет смысл говорить о значении параметра k больше 2.

Однако изучение этого вопроса представляет интерес для понимания проблемы в общем. Например, для устройства, описанного в [4], производителем декларируется поддержка структур RAID на уровне картриджей оптических дисков, где один картридж содержит 12 оптических дисков.

На рис. 4 построены зависимости вероятности потери хотя бы части информации Q от числа дисков k в массиве RAID-5 для структуры R5 для разных значений n . На рис. 5 построены аналогичные зависимости для структуры R15. Для удобства сопоставления все зависимости, кроме приведенных на рис. 5, б, построены в одном масштабе по оси Y . В связи с этим некоторые точки для малых значений k на рисунках не показаны. Штриховая линия соответствует зависимости $Q(k) = q$.

По графикам на рис. 4, 5 можно сделать следующие практические выводы.

1. Подтвердился результат, что при $n = 1$ условие $Q < q$ недостижимо.

2. Для систем с оптическими дисками, упакованными в картриджи (структура R5):

— при $n = 2$ условие $Q < q$ выполнимо лишь для небольших значений k и лишь для объема архива $W < 10$ Тбайт;

— при $n = 3$ условие $Q < q$ может быть выполнено даже для $W = 90$ Тбайт, но лишь при $k = 2$; для $W = 1$ Тбайт оно выполняется для значений k от 2 до 9 включительно.

3. Для систем, в которых оптические диски в копиях массивов RAID-5 могут свободно обмениваться (структура R15):

— при $n = 2$ условие $Q < q$ выполняется при малых значениях k для $W = 10$ Тбайт, а для $W = 1$ Тбайт это условие будет выполняться для значений k по крайней мере до 32 (это на практике, вероятно, бессмысленно, но означает, что для практических целей может быть выбрано любое разумное значение параметра k);

— при $n = 3$ условие $Q < q$ выполняется для любого значения W вплоть до 90 Тбайт для значений k по крайней мере до 32.

Заключение

В работе исследованы вопросы проектирования электронных архивных хранилищ ин-

формации для небольших организаций, когда необходимо обеспечить заданный уровень надежности хранения информации в архиве при минимизации стоимости архива. Причем современный стандарт создания электронных архивных хранилищ фактически требует использовать в качестве носителей информации в таких хранилищах однократно записываемые оптические диски. А реалии российского рынка позволяют однозначно сделать выбор в пользу дисков BD-R емкостью 25 Гбайт.

Основной проблемой использования оптических дисков в реальных (существующих сегодня) архивных хранилищах является невозможность предсказания вероятности выхода их из строя.

Предложено при проектировании задавать с вероятностью потери информации в архиве меньше, чем вероятность отказа одного оптического диска. Очевидно, что достижение этой цели возможно лишь при создании копий каждого архивного оптического диска.

Рассмотрены различные структуры, которые могут образовывать оптические диски в архиве:

— независимые диски с несколькими копиями (структура R1);

— массивы RAID-5 с несколькими копиями с жестко фиксированным набором дисков в каждом массиве (структура R5);

— массивы RAID-5 с несколькими копиями с возможностью обмена дисков между копиями (структура R15).

Показано, что:

— для архивов с объемом до 1 Тбайт привлекательной является структура R1 с созданием трех копий каждого оптического диска;

— для архивов с объемом более 1 Тбайт в большинстве случаев наиболее подходящей является структура R15, причем в зависимости от объема архива (в пределах до 90 Тбайт) для обеспечения заданной надежности хранения информации может потребоваться две или три копии каждого оптического диска (но не более трех);

— для существующих сегодня на рынке библиотек оптических дисков, использующих картриджи с оптическими дисками, нельзя рекомендовать создание массивов RAID-5 на базе всего объема картриджей, поскольку получающаяся в этом случае структура R5 даже для трех копий не обеспечивает заданного уровня надежности хранения информации в архиве.

1. **Школьный музей:** из века XX — в XXI // Музей. 2010. N. 07. С. 70—73. URL: http://turcentrnf.ru/d/358156/d/001_persin-1.pdf (дата обращения: 2015.12.26).

2. **Каташин В. В., Чернышов А. В.** Проблема надежного хранения файлов в архивных электронных хранилищах небольших организаций (на примере музеев) // Перспективы развития информационных технологий: сборник материалов XIX Международной научно-практической конференции. Новосибирск: Изд-во ЦРНС, 2014. С. 12—23.

3. **ГОСТ Р 54989—2012 / ISO TR 18492:2005** Обеспечение долговременной сохранности электронных документов. (Вступил в силу 01.05.2013).

4. **Data Archiver LB-DH8 series.** URL: <http://panasonic.net/avc/archiver/lb-dh8/> (дата обращения: 2016.02.01).

5. **ODS-L10 Optical Disc Archive 10-Slot library.** URL: <http://www.pro.sony.eu/pro/lang/en/eu/product/archiving-storage-oda-small-library/ods-l10/overview/> (дата обращения: 2016.05.17).

6. **Робот-библиотекарь** поселился в новом кампусе СПбГУ "Михайловская дача" // Санкт-Петербургский государственный университет. URL: <http://spbu.ru/smi/o-nas-pishut/24726-robot-bibliotekar-poselilsya-v-novom-kampuse-spbgu-mikhajlovskaya-dacha.html> (дата обращения: 2015.11.13).

7. **ODS-L10 Optical Disc Archive 10-Slot library. List Price.** URL: <https://pro.sony.com/bbsc/ssr/cat-datastorage/cat-opticaldiscarchive/product-ODSL10/> — Дата обращения: 2016.05.17.

8. **Устинов В.** Хранение данных на CD- и DVD-дисках: на наш век хватит? // BROADCASTING: Телевидение и радиовещание. 2006. N 4. URL: http://www.broadcasting.ru/articles2/Oborandteh/hranenie_dannyh_na_CD_DVD_diskah (дата обращения: 10.12.2015).

9. **Zheng J., Slattery O. T.** NIST/Library of Congress Optical Disc Longevity Study: Final Report. 32 p. URL: https://www.loc.gov/preservation/resources/rt/NIST_LC_OpticalDiscLongevity.pdf (Accessed 2016.05.11).

10. **Weatherspoon H., Kubiawicz J.** Erasure Coding vs. Replication: A Quantitative Comparison // Peer-to-Peer Systems, Lecture Notes in Computer Science. 2002. Vol. 2429.

11. **Thomasian A., Tang Y., Hu Y.** Hierarchical RAID: Design, performance, reliability, and recovery // J. Parallel Distrib. Comput. 2012. Vol. 72. P. 1753—1769. Doi: 10.1016/j.jpdc.2012.07.002.

12. **Thomasian A.** Shortcut method for reliability comparisons in RAID // Journal of Systems and Software. 2006. Vol. 79. P. 1599—1605.

13. **Чернышов А. В.** К вопросу о применении оптических дисков для создания долговременных электронных архивных хранилищ информации небольших организаций // Информационные технологии. 2016. Т. 22. № 8. С. 635—640.

14. **Optical Jukeboxes and Libraries: KINTRONIX: IP Security Solutions.** URL: <https://kintronics.com/solutions/optical-jukeboxes-and-libraries/> (дата обращения: 2016.05.14).

15. **Жесткий диск HDD SATA** // Wikimart. URL: <http://computers.wikimart.ru/components/hdd/tag/hdd-sata/> (дата обращения: 17.12.2015).

16. **Blu-ray диски** для записи [Электронный ресурс] // Store52.ru: Магазин носителей информации. URL: http://www.store52.ru/catalog/bluray_diski_dlya_zapisi (дата обращения: 17.12.2015).

17. **BD-R DL** // Ресурс ПРО. URL: <http://resurs-pro.ru/catalog/bd-r-dl/> (дата обращения: 17.12.2015).

The Study of the Properties of Long-Term Electronic Archival Information Storage on Optical Disks Organized into a Structure of RAID-5

The problem is investigated of creating a local electronic archive information store for a small organization. Minimization of the cost of creating and maintaining archive is an important factor along with the reliability of information storage. Optical write-once disks organized in RAID-5 are proposed to be used for the purpose. A mathematical model of two alternatives such storages is constructed. It is proved that the best results in most cases are achieved when using the RAID-5 structure with backups of each disc. But with minimum size of storage (about 1 Tb) more profitable may be use single disks with three copy. It is shown that it is possible to increase the number of drives in the RAID-5 to reduce the unit cost of storing information with an acceptable decrease in reliability of storage.

Keywords: *electronic archival storage of information, long-term storage, optical storage media, reliability of information storage, minimization of cost*

DOI: 10.17587/it.24.586-593

References

1. **Shkol'nyj muzej: iz veka XX — v XXI.** [School Museum: from the XX century — to XXI.] (in Russian). *Muzej* [Museum], 2010, vol. 07, pp. 70–73. URL: http://turcentrfr.ru/d/358156/d/001_persin-1.pdf. Access date: 2015.12.26.
2. **Katashin V. V., Chernyshov A. V. Problema nadjozhnogo hranenija fajlov v arhivnyh jelektronnyh hranilishhah nebol'shih organizacij (na primere muzeev).** [The problem of reliable storage of files in archival electronic storage of small organizations (on the example of museums)] (in Russian). *Perspektivy razvitiya informacionnyh tehnologij: sbornik materialov XIX Mezhdunarodnoj nauchno-prakticheskoj konferencii.* [Prospects of development of information technologies: proceedings of the XIX International scientific and practical conference]. Novosibirsk: CRNS, 2014, pp. 12–23.
3. **GOST R 54989—2012 / ISO TR 18492:2005 Obespechenie dolgovremennoj sohrannosti jelektronnyh dokumentov.** [Ensuring long-term preservation of electronic records.] (access date: 2013.01.05).
4. **Data Archiver LB-DH8 series.** URL: <http://panasonic.net/avc/archiver/lb-dh8/> (access date: 2016.02.01).
5. **ODS-L10 Optical Disc Archive 10-Slot library.** URL: <http://www.pro.sony.eu/pro/lang/en/eu/product/archiving-storage-oda-small-library/ods-l10/overview/> (access date: 2016.05.17).
6. **Robot-bibliotekar' poselilsja v novom kampuse SPbGU "Mikhajlovskaja dacha"** [Robot librarian settled in the new campus of SPSU "Mikhailovskaya dacha"] (in Russian). *Sankt-Peterburgskij gosudarstvennyj universitet* [St. Petersburg state University]. URL: <http://spbu.ru/smi/o-nas-pishut/24726-robot-bibliotekar-poselilsya-v-novom-kampuse-spbgu-mikhajlovskaja-dacha.html>. Publ. date: October, 2015 (access date: 13.11.2015).
7. **ODS-L10 Optical Disc Archive 10-Slot library. List Price.** [Electronic resource]. URL: <https://pro.sony.com/bbsc/ssr/cat-datastorage/cat-opticaldiscarchive/product-ODSL10/> (access date: 2016.05.17).
8. **Ustinov V. Hranenie dannyh na CD- i DVD-diskah: na nash vek hvatit?** [Data storage on CD and DVD: on our century will suffice?] [Electronic resource] (in Russian). *BROADCASTING: Televidenie i radioveshhanie* [Television and broadcasting]. URL: http://www.broadcasting.ru/articles2/Oborandteh/hranenie_dannyh_na_CD_DVD_diskah (access date: 2015.10.12).
9. **Zheng J., Slattery O. T. NIST/Library of Congress Optical Disc Longevity Study: Final Report.** 32 p. URL: https://www.loc.gov/preservation/resources/rt/NIST_LC_OpticalDiscLongevity.pdf (access date: 2016.05.11).
10. **Weatherspoon H., Kubiawicz J. Erasure Coding vs. Replication: A Quantitative Comparison, Peer-to-Peer Systems, Lecture Notes in Computer Science, 2002, vol. 2429.**
11. **Thomasian A., Tang Y., Hu Y. Hierarchical RAID: Design, performance, reliability, and recovery, J. Parallel Distrib. Comput., 2012, vol. 72, pp. 1753–1769. doi: 10.1016/j.jpdc.2012.07.002**
12. **Thomasian A. Shortcut method for reliability comparisons in RAID, Journal of Systems and Software, 2006, vol. 79, pp. 1599–1605.**
13. **Chernyshov A. V. K voprosu o primenenii opticheskikh diskov dlja sozdaniya dolgovremennyh jelektronnyh arhivnyh hranilishh informacii nebol'shih organizacij** [To the question of the optical discs application for long term digital archive storage of small organizations] (in Russian). *Informacionnye tehnologii.* [Information technologies]. 2016, vol. 22, no. 8, pp. 635–640.
14. **Optical Jukeboxes and Libraries: KINTRONIX: IP Security Solutions.** URL: <https://kintronics.com/solutions/optical-jukeboxes-and-libraries/> (access date: 2016.05.14).
15. **Hard disk with SATA interface** (in Russian), wikimart. URL: <http://computers.wikimart.ru/components/hdd/tag/hdd-sata/> (access date: 2015.12.17).
16. **Blu-ray diski dlja zapisi** [Blu-ray disks for writing.] (in Russian), *Store52.ru: Magazin nositelej informacii.* [Store52.ru: Shop of information media]. URL: http://www.store52.ru/catalog/bluray_diski_dlya_zapisi (access date: 2015.12.17).
17. **BD-R DL** (in Russian). URL: <http://resurs-pro.ru/catalog/bd-r-dl/> (access date: 2015.12.17).