

## Модель планирования количества оптических дисков, необходимых для создания и поддержания долговременного электронного архива

А. В. Чернышов

Мытищинский филиал Московского государственного технического университета им. Н. Э. Баумана, Российская Федерация, 141005, Мытищи, Московская обл., 1-я Институтская ул., 1

**Для цитирования:** Чернышов А. В. Модель планирования количества оптических дисков, необходимых для создания и поддержания долговременного электронного архива // Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления. 2019. Т. 15. Вып. 4. С. 578–591. <https://doi.org/10.21638/11702/spbu10.2019.413>

В статье предложена аналитическая модель, позволяющая оценить количество оптических дисков, необходимых ежегодно на создание и поддержание в целостном состоянии электронного архива длительного хранения информации, а также связанных с этими процессами затрат времени. Модель разработана для архивов на базе одиночных дисков (диски не объединены в структуры RAID). Исходными данными служили следующие параметры: предполагаемая (проектная) емкость архива; емкость используемых оптических дисков; оценки времени записи, чтения и контроля целостности оптических дисков; время, выделяемое на технические работы в архиве (собственно запись новых данных, контроль ранее записанных дисков, восстановление отказавших); вероятность выхода из строя оптического диска при хранении; вероятность выхода из строя диска при записи (брак). Рассмотрены 4 варианта функционирования архива в зависимости от пропускной способности аппаратуры и описаны этапы функционирования каждого варианта. Для каждого варианта выведены аналитические выражения, дающие возможность оценить ежегодно необходимое количество оптических дисков (что во многом определяет ежегодные материальные затраты), а также моменты времени смены этапов. Дополнительно для некоторых вариантов получены выражения, позволяющие оценить возможность достижения проектной емкости архива.

*Ключевые слова:* долговременные архивы электронной информации, оптические диски, обеспечение целостности информации, ежегодные затраты.

**Введение.** При организации и эксплуатации долговременных электронных архивных хранилищ наряду с проблемой надежности сохранения информации [1–4] и ее доступности [5–7] существует и проблема затрат на поддержание архива. Согласно действующему стандарту [8], для создания долговременных электронных архивов должны использоваться только диски однократной записи, т. е., как правило, оптические диски типа CD-R, DVD+/-R, DB-R. Однако особенности применяемой аппаратуры, организации процесса размещения данных в архиве и контроля целостности хранимых дисков, необходимость восстанавливать диски, отказавшие в процессе хранения, а также интенсивность поступления самой информации в архив могут существенно влиять на параметры архива.

Вопросы построения и исследования моделей надежности хранения информации на оптических дисках, организованных в различные структуры, были рассмотрены в статьях [9–12]. Этим же вопросам, применительно к использованию жестких дисков, были посвящены многие публикации (например, [13–16]).

Вопросы планирования ежегодно необходимого количества оптических дисков

и затрат времени для создания и поддержания долговременных архивов к настоящему времени не исследованы.

В настоящей статье предложено построение модели ежегодно необходимого количества оптических дисков и затрат времени для архивов, в которых диски не объединены в структуры RAID (рассматриваются как самостоятельные). Эти архивы широко распространены на практике. Кроме того, именно на такие типы архивов в России разработаны методические рекомендации по их созданию и эксплуатации [17–19]. Они являются привлекательными для применения в небольших организациях, имеющих существенные ограничения в возможностях организации необходимого по мощности IT-подразделения.

**Общие соображения.** Обозначим максимальную (конструкционную) емкость архива  $W$ . Если необходимо сохранить больше информации, то должен быть создан еще один архив емкостью  $W_1$ , для которого должны быть проведены полностью аналогичные расчеты. Для сравнения также должны быть осуществлены расчеты для архива емкостью  $(W + W_1)$ . Архивы создаются на оптических дисках однократной записи емкостью  $V$  и стоимостью  $c$ . Вероятность выхода из строя диска с записью при хранении в течение года обозначим  $q$ ; вероятность отказа диска в момент записи —  $q_w$ .

Пусть  $t$  — время в годах существования архива;  $T$  — время в течение года, выделенное на технические работы в архиве (запись новых дисков и контроль целостности ранее записанных). В это время не входит время доступа к информации, предоставляемое пользователям архива.

Пусть изначальная стоимость необходимого аппаратного обеспечения  $C_{\text{П}}$ , а затраты на создание и поддержание архива в год с номером  $t$  составляют  $C(t)$ .

Дополнительно введем функцию  $f_C(t)$ , описывающую стоимость поддержания работоспособности этого оборудования в год с номером  $t$ . Хотя обычно принято считать, что данный параметр должен носить функциональный характер, но в действительности очень часто для конкретных типов аппаратуры вид такой функции неизвестен, а современные поставщики оборудования предпочитают указывать ежегодные эксплуатационные расходы на поставляемое оборудование в виде постоянной суммы (например, 1/10 от стоимости приобретения, как для роботизированной библиотеки оптических дисков Panasonic LB-DH8 [20]). Потому здесь вид функции уточнять не будем.

Для заданного уровня надежности хранения информации необходимо создать  $n$  копий дисков.

Обозначим  $l = \lceil W/V \rceil$  — количество дисков емкости  $V$ , которое обеспечивает запись информации объема  $W$ .

В действительности при создании и эксплуатации таких архивов возможны следующие варианты (рисунок):

1 — пропускная возможность аппаратуры не ограничена, вследствие чего предельная емкость архива  $W$  может оказаться достигнутой уже в первый год эксплуатации. В таком случае остальные годы затраты будут связаны исключительно с задачами хранения информации;

2 — интенсивность поступления информации в архив невысока и стабильна (фактически, постоянна) или во всяком случае не превышает некоторой заданной величины, что приводит к заполнению архива постепенно в течение нескольких лет. Данный вариант предполагает, что пропускная способность аппаратуры позволяет выполнить запись новой информации и контроль ранее записанной без ограничений;

3 — возможности применяемой аппаратуры (или организации самого процесса) ограничивают скорость записи информации на оптические диски (ограничена пропускная способность), что также приводит к тому, что архив будет заполняться в течение нескольких лет;

4 — возможен «комбинированный» случай, когда сначала скорость заполнения архива лимитируется интенсивностью поступления информации, но по мере заполнения архива начинают сказываться ограничения пропускной способности аппаратуры или применяемой организации процесса.

Все четыре случая схематично показаны на рисунке, где каждый столбик изображает время в течение года, необходимое на технические работы, причем незаштрихованная часть — это время записи на диски новой информации, одинарная штриховка — время контроля целостности ранее записанных дисков, двойная штриховка — время восстановления вышедших из строя при хранении дисков.

На рисунке  $t_{full}$  — номер года, в который архив достиг проектной емкости  $W$ ,  $\tilde{t}$  — номер года, в который достигнута пропускная способность аппаратуры (вариант 2 переходит в вариант 3).

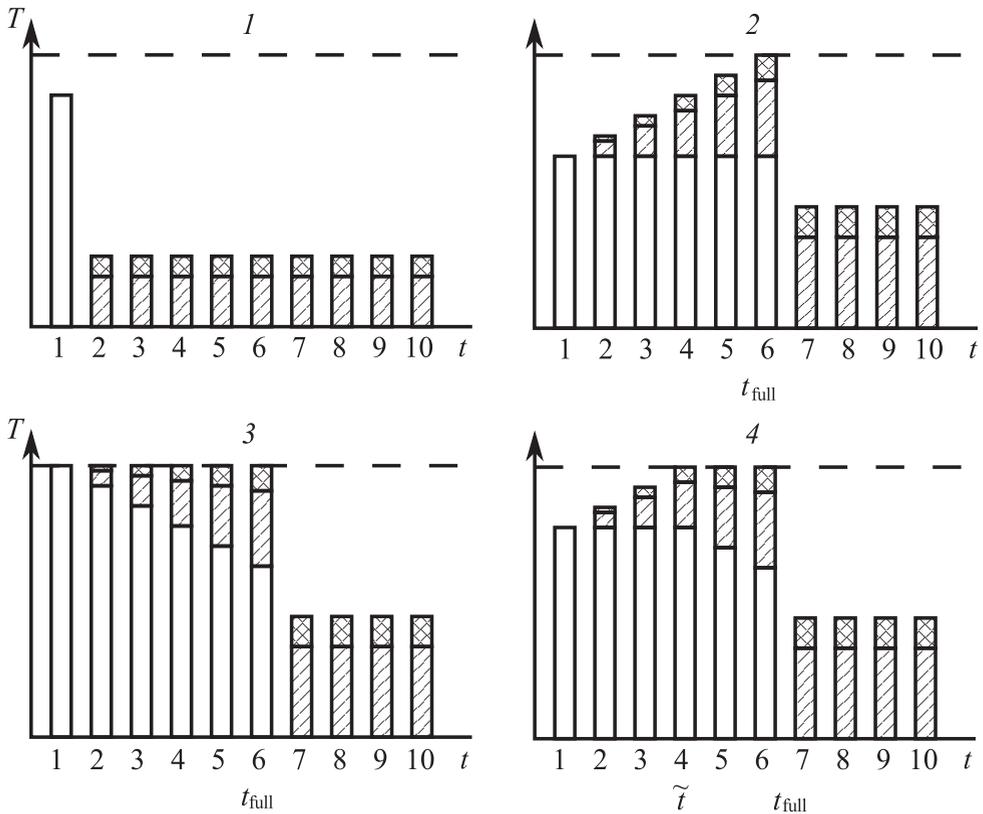


Рисунок. Четыре варианта организации и эксплуатации рассматриваемых архивов (объяснение в тексте)

Здесь и далее будем считать, что  $q = \text{const}$ , поскольку на практике для реальных расчетов обычно удается получить только точечную оценку ее математического ожидания.

Надо также заметить, что через 30 или 50 лет, когда закончится срок эксплуатации оптических дисков (обычно указываемый производителями), необходимо будет выполнить полную перезапись информации на новые носители. Однако к этому времени, скорее всего, будут доступны совершенно другие технологии длительного хранения электронной информации. Потому можно лишь предположить (весьма условно), что  $C(30)$  будет по «структуре цены» соответствовать  $C(1)$ ,  $C(31) - C(2)$  и т. д.

**Построение моделей. Модель варианта 1.** В первый год будет записано (и должно быть приобретено)

$$N(1) = \left[ \frac{nl}{1 - q_w} \right]$$

дисков с учетом оказавшихся бракованными. При этом в архиве будет размещено  $nl$  дисков. Запись  $(1 - q_w)$  позволяет учесть дополнительные расходы, связанные с некачественными экземплярами дисков, отказывающимися сразу после записи или даже в процессе записи.

Во второй и каждый последующий годы из-за отказавших при хранении дисков придется приобрести и перезаписать

$$N(t) = \left[ \frac{[qnl]}{1 - q_w} \right] \quad (1)$$

дисков, имея в виду, что при проведении ежегодного контроля целостности дисков необходимо будет восстанавливать из запасных копий  $[qnl]$  отказавших дисков.

Формула (1) будет справедлива для всех рассматриваемых вариантов при полностью заполненном архиве. Модель затрат на создание/содержание такого архива:

– в первый год  $C(1) = cN(1) + C_{\Pi}$ ;

– во второй и каждый последующий годы  $C(t) = cN(t) + f_C(t)$ ,  $t > 1$ .

Таким образом, возможно вычислить ежегодно необходимое количество оптических дисков, ежегодные затраты на работу с архивом, а также суммарные затраты, которые понесет за  $t$  лет владелец архива.

**Модель варианта 2.** Интенсивность поступления информации будет определять максимальный объем информации  $W_y$ , записываемый в год, количество записываемых дисков в год  $l_y$  и общее количество записываемых дисков  $nl_y$  с учетом запасных копий.

В данном случае важными являются два обстоятельства:

–  $l_y = \text{const}$ ,  $W_y = \text{const}$ ;

– значения  $l_y$  и  $W_y$  не могут быть выше, чем позволяет пропускная способность аппаратуры с учетом необходимости контроля на тех же самых приводах ранее записанных дисков в последний год записи. При этом все работы по записи новых дисков и контролю целостности, ранее записанных в течение года, должны быть выполнены за специально выделенное время  $T$ .

Пусть на запись одного диска требуется время  $t_w$ , включая верификацию информации сразу после записи; на плановый контроль одного диска — время  $t_{r0}$ ; для изготовления копии отказавшего диска из запасной копии — время  $t_r + t_w$  (сначала считать данные с запасной копии и выполнить верификацию считанных данных ( $t_r$ ), потом записать, включая верификацию). Будем считать, что чтение информации с запасной копии выполняется всегда без сбоев (вероятность сбоя ничтожно мала, поскольку диск перед этим был проверен). Тогда для произвольного года с номером  $t$  необходимо:

- проконтролировать  $(t - 1)nly$  дисков (были записаны в предыдущие  $(t - 1)$  годы), общее время контроля  $(t - 1)nlyt_{r0}$ ;
- записать новые диски с учетом бракованных:

$$\left\lceil \frac{nly}{1 - qw} \right\rceil;$$

- перезаписать отказавшие диски  $\lceil (t - 1)nlyq \rceil$  и на это потребуется приобрести дисков (также с учетом бракованных):

$$\left\lceil \frac{\lceil (t - 1)nlyq \rceil}{1 - qw} \right\rceil;$$

- чтобы общее время записи новых дисков составляло

$$\left\lceil \frac{nly}{1 - qw} \right\rceil t_w;$$

- чтобы время на перезапись отказавших дисков было равно

$$\lceil (t - 1)nlyq \rceil t_r + \left\lceil \frac{\lceil (t - 1)nlyq \rceil}{1 - qw} \right\rceil t_w.$$

Последняя запись предполагает, что в случае сбоя при записи перезаписываемого диска повторное считывание информации с запасного диска не требуется.

Тогда

$$T \geq \left\lceil \frac{nly}{1 - qw} \right\rceil t_w + (t - 1)nlyt_{r0} + \lceil (t - 1)nlyq \rceil t_r + \left\lceil \frac{\lceil (t - 1)nlyq \rceil}{1 - qw} \right\rceil t_w.$$

Из получившегося выражения следует вычислить максимально допустимое значение  $ly$ , при котором такой вариант работы архива будет иметь место. К сожалению, множество операций приведения к целому не позволяет сделать это однозначно. Потому поступим следующим образом — оценим граничные значения.

Для нижней границы просто уберем приведение к «верхнему» целому и получим, что

$$T \geq nly \left( \frac{t_w}{1 - qw} + (t - 1) \left( t_{r0} + qt_r + \frac{qt_w}{1 - qw} \right) \right).$$

Для верхней границы, убирая каждое приведение к целому, добавим 1:

$$T < \left( \frac{nly}{1 - qw} + 1 \right) + (t - 1)nlyt_{r0} + ((t - 1)nlyq + 1)t_r + \left( \frac{(t - 1)nlyq + 1}{1 - qw} + 1 \right) t_w.$$

После этого имеем

$$T \geq nly \left( \frac{t_w}{1 - qw} + (t - 1) \left( t_{r0} + qt_r + \frac{qt_w}{1 - qw} \right) \right) + 2t_w + t_r + \frac{t_w}{1 - qw}.$$

Обозначим

$$A = \frac{t_w}{1 - qw}, \quad B = t_{r0} + qt_r + \frac{qt_w}{1 - qw}, \quad t_{wr} = 2t_w + t_r + \frac{t_w}{1 - qw}.$$

Тогда после очевидных преобразований будем иметь

$$\frac{T - t_{wr}}{n(A + (t - 1)B)} < l_y \leq \frac{T}{n(A + (t - 1)B)}. \quad (2)$$

Расчеты, проведенные с реальными значениями параметров, показывают, что обычно разность левой и правой частей не превышает 1 диска. Поэтому, исходя из условий задачи, окончательно находим

$$l_y = \left\lfloor \frac{T}{n(A + (t - 1)B)} \right\rfloor.$$

Для использования формулы, кроме знания величин  $t_w$ ,  $t_r$ ,  $t_{r0}$ , которые характеризуют и пропускную способность аппаратуры, и применяемые диски, а также  $q$  и  $q_w$ , оценивающих такие диски, необходимо также определить время (номер года) заполнения архива  $t_{full}$ .

Выполним это следующим образом.

Хотя в последний год записи архива для достижения полной емкости архива  $W$  и может потребоваться записать меньше, чем  $l_y$  дисков, в предельном случае будем иметь соотношение (2). При этом  $l_y t_{full} \geq l = \lceil W/V \rceil$ , т. е.  $l_y \geq l/t_{full}$ , следовательно,

$$\frac{nl(A - B)}{T - nBl} \leq t_{full} < \frac{nl(A - B)}{T - t_{wr} - nBl},$$

что по условиям задачи ( $t$  — целое количество лет) дает

$$t_{full} = \left\lceil \frac{nl(A - B)}{T - nBl} \right\rceil.$$

Так как рассматриваются годовые циклы эксплуатации архива, то  $t_{full}$  должно быть округлено до большего целого, но в последний год, вероятнее всего, будет записано меньше, чем  $l_y$ , дисков.

В произвольный год с номером  $t$  ( $t < t_{full}$ ) необходимо будет приобрести для записи

$$N(t) = \left\lceil \frac{nl_y}{1 - q_w} \right\rceil + \left\lceil \frac{[(t - 1)nl_y q]}{1 - q_w} \right\rceil \quad (3)$$

дисков (эта формула справедлива для любого года, начиная с  $t = 1$ ).

При  $t > t_{full}$  количество дисков, приобретаемых ежегодно на поддержание архива, будет полностью соответствовать записанному ранее выражению (1).

Важно подчеркнуть, что необходимо получить  $t_{full} < 30$ , если считать, что срок службы оптических дисков составляет 30 лет. В противном случае должны быть скорректированы исходные данные (например, либо уменьшен размер создаваемого архива  $W$ , либо обеспечена большая интенсивность поступления информации, если это позволяет пропускная способность аппаратуры).

**Модель варианта 3.** В этом случае  $l_y = l_f(t) \neq \text{const}$ .

В первый год ( $t = 1$ ) можно записать максимальное количество дисков. При этом

$$T \geq \left\lceil \frac{nl_f(1)}{1 - q_w} \right\rceil t_w,$$

откуда, используя выполненные выше обозначения,  $nl_f(1)A \leq T < nl_f(1)A + t_w$ , после преобразований получим выражение

$$\frac{T - t_w}{nA} < l_f(1) \leq \frac{T}{nA}.$$

С учетом физического смысла решаемой задачи можно записать, что

$$l_f(1) = \left\lfloor \frac{T}{nA} \right\rfloor.$$

В последующие годы количество записываемых дисков вынужденно уменьшится, так как часть пропускной способности системы уйдет на контроль целостности ранее записанных дисков и восстановление отказавших. Для любого последующего года получим неравенство

$$T \geq \left\lceil \frac{nl_f(t)}{1 - q_w} \right\rceil t_w + nt_{r0} \sum_{i=1}^{t-1} l_f(i) + \left\lceil nq \sum_{i=1}^{t-1} l_f(i) \right\rceil t_r + \left\lceil \frac{nq \sum_{i=1}^{t-1} l_f(i)}{1 - q_w} \right\rceil t_w.$$

Раскрывая ограничения целочисленности использованным выше способом, введя обозначение

$$t_{wr1} = t_r + \frac{t_w}{1 - q_w} + t_w$$

и применяя приведенные ранее обозначения, находим, что

$$\frac{T - t_w}{nA} - \frac{t_{wr1}}{nA} - \frac{B}{A} \sum_{i=1}^{t-1} l_f(i) < l_f(t) \leq \frac{T}{nA} - \frac{B}{A} \sum_{i=1}^{t-1} l_f(i).$$

В полученной формуле присутствуют рекурсивные суммы, от которых желательно избавиться.

Довольно просто показать (мы этого делать не будем), что справедливы такие выражения:

$$f(t) = x - y - z \sum_{i=1}^{t-1} f(i) = x(1 - z)^{t-1} - y(1 - z)^{t-2}, \quad t > 1, \quad f(1) = x, \quad (4)$$

$$f(t) = x - z \sum_{i=1}^{t-1} f(i) = x(1 - z)^{t-1}, \quad t \geq 1, \quad (5)$$

следовательно, можно записать неравенство

$$\frac{T - t_w}{nA} \left(1 - \frac{B}{A}\right)^{t-1} - \frac{t_{wr1}}{nA} \left(1 - \frac{B}{A}\right)^{t-2} < l_f(t) \leq \frac{T}{nA} \left(1 - \frac{B}{A}\right)^{t-1}.$$

Если сейчас выполнить вычисления с реальными значениями параметров для левой и правой частей этого неравенства для нескольких значений  $t$ , то можно увидеть, что разница между ними не превышает 1. Поскольку в конечном итоге необходимо оценить количество оптических дисков (целое число), мы не сделаем большой ошибки, если в дальнейшем везде будем использовать для оценок правую часть неравенства, что значительно упростит выкладки. Тогда будем иметь уравнение

$$l_f(t) = \left\lfloor \frac{T}{nA} \left(1 - \frac{B}{A}\right)^{t-1} \right\rfloor.$$

В этом случае с учетом восстанавливаемых дисков в год с номером  $t$  необходимо будет записывать

$$N(t) = \left\lceil \frac{nl_f(t)}{1 - q_w} \right\rceil + \left\lceil \frac{\left[ qn \sum_{i=1}^{t-1} l_f(i) \right]}{1 - q_w} \right\rceil.$$

Для расчета  $\sum_{i=1}^{t-1} l_f(i)$  воспользуемся установленным выше фактом (5). Выполнив несложные преобразования, получим, что

$$\sum_{i=1}^{t-1} l_f(i) = \frac{T}{nB} \left( 1 - \left( 1 - \frac{B}{A} \right)^{t-1} \right). \quad (6)$$

Тогда

$$N(t) = \left\lceil \frac{n}{1 - q_w} \left\lceil \frac{T}{nA} \left( 1 - \frac{B}{A} \right)^{t-1} \right\rceil \right\rceil + \left\lceil \frac{1}{1 - q_w} \left\lceil qn \left( \frac{T}{nB} \left( 1 - \left( 1 - \frac{B}{A} \right)^{t-1} \right) \right) \right\rceil \right\rceil.$$

Интерес представляет вычисление значения  $t_{\text{full}}$ . Определим его из соотношения

$$l \leq \sum_{i=1}^{t_{\text{full}}} l_f(i).$$

Неравенство здесь означает, что в последний год записи количество реально записанных дисков может оказаться меньше модельного значения  $l_f(t_{\text{full}})$ . Принимая во внимание выражение (6) и выполнив очевидные преобразования, находим, что

$$1 - \frac{nBl}{T} \geq \left( 1 - \frac{B}{A} \right)^{t_{\text{full}}}.$$

В результате получим неравенство

$$t_{\text{full}} \geq \log_{(1-B/A)} \left( 1 - \frac{nBl}{T} \right),$$

что по смыслу решаемой задачи дает

$$t_{\text{full}} = \left\lceil \log_{(1-B/A)} \left( 1 - \frac{nBl}{T} \right) \right\rceil.$$

Как и ранее, при  $t > t_{\text{full}}$  для поддержания архива необходимо будет приобретать  $N(t)$  (см. (1)) дисков. И точно также должно выполняться ограничение  $t_{\text{full}} < 30$ .

В действительности размер создаваемого архива может ограничиваться также из-за того, что, начиная с некоторого года записи информации  $t_f$ , выделенного времени  $T$  будет хватать только на контроль целостности и восстановление ранее записанных дисков, что не позволит продолжить заполнение архива и плановый объем  $W$  никогда не сможет быть достигнут. Оценим значение  $t_f$  из соотношения  $l_f(t_f) < 1$ . Для упрощения дальнейших рассуждений выполним только верхнюю оценку. Тогда

$$\frac{T}{nA} \left( 1 - \frac{B}{A} \right)^{t_f-1} < 1,$$

т. е.

$$t_f > \log_{(1-B/A)} \left( \frac{nA}{T} \right) + 1,$$

что по смыслу решаемой задачи дает

$$t_f = \left\lceil \log_{(1-B/A)} \left( \frac{nA}{T} \right) \right\rceil + 1.$$

Для того чтобы объем архива  $W$  в принципе мог быть достигнут, должно выполняться условие  $t_{\text{full}} < t_f$ .

**Модель варианта 4.** В данном случае процесс функционирования архива распадается на следующие этапы:

– этап 1 условно соответствует ситуации варианта 2, когда  $l_y = \text{const} = \lceil W_y/V \rceil$  вплоть до момента, когда оно окажется ограничено пропускной способностью аппаратуры с учетом необходимости контроля ранее записанных дисков;

– этап 2 условно соответствует варианту 3, когда количество информации, записываемой в год (и соответственно количество записанных дисков), будет постепенно уменьшаться из-за необходимости тратить пропускную способность аппаратуры на контроль ранее записанных дисков, количество которых будет ежегодно возрастать;

– этап 3 соответствует состоянию «только хранение» — наступает после того, как будет достигнут плановый объем архива  $W$ , и полностью отвечает аналогичному состоянию в рассмотренных выше моделях.

Необходимо оценить, когда этап 1 закончится и начнется этап 2.

Обозначим  $\tilde{t}$  количество лет, в течение которых  $l_y = \text{const}$  (т. е. длится этап 1). В этом случае

$$T \geq \left\lceil \frac{nl_y}{1 - q_w} \right\rceil t_w + (\tilde{t} - 1)nl_y t_{r0} + \lceil (\tilde{t} - 1)nl_y q \rceil t_r + \left\lceil \frac{\lceil (\tilde{t} - 1)nl_y q \rceil}{1 - q_w} \right\rceil t_w.$$

И, как и ранее,

$$\frac{T - t_{wr} - nAl_y}{nBl_y} + 1 < \tilde{t} \leq \frac{T - nAl_y}{nBl_y} + 1,$$

т. е.

$$\tilde{t} = \left\lceil \frac{T - nAl_y}{nBl_y} \right\rceil + 1.$$

В течение этого периода времени в год потребуется записывать всего  $N(t)$  (см. (3)) дисков. Аналогично варианту 2 рассчитываются и ежегодные затраты на создание и поддержание архива. При этом за этап 1 в архиве окажется всего  $n\tilde{t}l_y$  дисков.

Теперь рассмотрим процесс дальнейшего заполнения архива на этапе 2. Он начнется с года  $t = \tilde{t} + 1$ . Количество дисков, записываемых в год, теперь представляет собой функцию  $l_f(t)$ . Для любого такого года имеем неравенство

$$T \geq \left\lceil \frac{nl_f(t)}{1 - q_w} \right\rceil t_w + t_{r0}n \left( \tilde{t}l_y + \sum_{i=1}^{t-\tilde{t}-1} l_f(\tilde{t} + i) \right) + \left\lceil qn \left( \tilde{t}l_y + \sum_{i=1}^{t-\tilde{t}-1} l_f(\tilde{t} + i) \right) \right\rceil t_r + \\ + \left\lceil \frac{qn \left( \tilde{t}l_y + \sum_{i=1}^{t-\tilde{t}-1} l_f(\tilde{t} + i) \right)}{1 - q_w} \right\rceil t_w.$$

Действуя по применявшейся ранее схеме, для года  $t > \tilde{t}$  получим выражение

$$\frac{T - t_{wr}}{nA} - \tilde{t}l_y \frac{B}{A} - \frac{B}{A} \sum_{i=1}^{t-\tilde{t}-1} l_f(\tilde{t} + i) < l_f(t) \leq \frac{T}{nA} - \tilde{t}l_y \frac{B}{A} - \frac{B}{A} \sum_{i=1}^{t-\tilde{t}-1} l_f(\tilde{t} + i).$$

Приняв  $ot = t - \tilde{t}$ , запишем неравенство

$$l_f(ot) \leq \frac{T}{nA} - \tilde{t}l_y \frac{B}{A} - \frac{B}{A} \sum_{i=1}^{ot-1} l_f(i).$$

В отличие от случая в (4), здесь член  $\tilde{t}l_y(B/A)$  будет присутствовать во всех итерациях (для всех  $ot$ ). Поэтому можно обозначить, что

$$D = \frac{T}{nA} - \tilde{t}l_y \frac{B}{A}.$$

Тогда, воспользовавшись (5), имеем, что

$$l_f(ot) \leq D \left(1 - \frac{B}{A}\right)^{ot-1},$$

т. е.

$$l_f(t) = \left[ \left( \frac{T}{nA} - \tilde{t}l_y \frac{B}{A} \right) \left(1 - \frac{B}{A}\right)^{t-\tilde{t}-1} \right], \quad t > \tilde{t}.$$

При этом в год  $t$  ( $t > \tilde{t}$ ) для поддержания архива надо будет записать дисков (с учетом брака)

$$N(t) = \left[ \frac{nl_f(t)}{1 - q_w} \right] + \left[ \frac{qn \left( \tilde{t}l_y + \sum_{i=1}^{t-\tilde{t}-1} l_f(\tilde{t} + i) \right)}{1 - q_w} \right].$$

Приняв во внимание, что

$$l_f(t) \leq D - \frac{B}{A} \sum_{i=1}^{t-\tilde{t}-1} l_f(\tilde{t} + i) = D \left(1 - \frac{B}{A}\right)^{t-\tilde{t}-1},$$

получим равенство

$$\sum_{i=1}^{t-\tilde{t}-1} l_f(\tilde{t} + i) = \frac{DA}{B} \left(1 - \left(1 - \frac{B}{A}\right)^{t-\tilde{t}-1}\right). \quad (7)$$

Тогда

$$N(t) = \left[ \frac{n}{1 - q_w} \left[ \left( \frac{T}{nA} - \tilde{t}l_y \frac{B}{A} \right) \left(1 - \frac{B}{A}\right)^{t-\tilde{t}-1} \right] \right] + \left[ \frac{1}{1 - q_w} \left[ qn \left( \tilde{t}l_y + \frac{DA}{B} \left(1 - \left(1 - \frac{B}{A}\right)^{t-\tilde{t}-1}\right) \right) \right] \right].$$

Архив будет окончательно заполнен не позднее года с номером  $t_{\text{full}}$ . Определим его из соотношения

$$l \leq \tilde{t}l_y + \sum_{i=1}^{t_{\text{full}}-\tilde{t}} l_f(\tilde{t} + i).$$

Принимая во внимание (7), имеем

$$l \leq \tilde{t}l_y + \frac{DA}{B} \left( 1 - \left( 1 - \frac{B}{A} \right)^{t_{\text{full}}-\tilde{t}} \right).$$

После преобразования получим неравенство

$$t_{\text{full}} \geq \log_{(1-B/A)} \left( 1 + \frac{B}{DA} (\tilde{t}l_y - l) \right) + \tilde{t},$$

следовательно, исходя из физического смысла задачи,

$$t_{\text{full}} = \left\lceil \log_{(1-B/A)} \left( 1 + \frac{B}{DA} (\tilde{t}l_y - l) \right) \right\rceil + \tilde{t}.$$

Как и ранее, при  $t > t_{\text{full}}$  для поддержания архива необходимо будет приобретать  $N(t)$  (см. (1)) дисков. И также должно выполняться  $t_{\text{full}} < 30$ . Оценим, аналогично варианту 3, принципиальную возможность заполнения всего объема  $W$  архива, считав номер года  $t_f$ , после которого запись дисков с новой информацией окажется невозможной из-за необходимости расходовать все время  $T$  на поддержание архива. Изначально имеем  $l_f(t_f) < 1$ , т. е.

$$\left( \frac{T}{nA} - \tilde{t}l_y \frac{B}{A} \right) \left( 1 - \frac{B}{A} \right)^{t_f-\tilde{t}-1} < 1.$$

В конечном итоге находим, что

$$t_f = \left\lceil \log_{(1-B/A)} \left( \frac{1}{\frac{T}{nA} - \tilde{t}l_y \frac{B}{A}} \right) \right\rceil + \tilde{t} + 1.$$

**Заключение.** В статье рассмотрена задача построения модели ежегодно необходимого количества оптических дисков для архивов, в которых диски не объединены в структуры RAID. Рассмотрены возможные варианты функционирования таких архивов и определены ключевые параметры этапов их функционирования (время полного заполнения и др.), для оценки которых также построены аналитические модели.

Эти модели позволяют оценить материальные затраты на создание и обслуживание электронного архива (количество ежегодно приобретаемых дисков), а также необходимые затраты времени и предельные значения параметров архива (время заполнения, максимально достигаемый объем).

## Литература

1. *Verbatim Launches 100GB MDISC Blu-ray Storage Media* // CDRinfo. URL: <https://cdrinfo.com/d7/content/verbatim-launches-100gb-mdisc-blu-ray-storage-media> (дата создания: 18.05.2016 г.; дата обращения: 04.01.2019 г.).

2. *Optical media longevity* // The X Lab.  
URL: <https://www.thexlab.com/faqs/opticalmedialongevity.html> (дата обращения: 29.09.2016 г.).
3. *Акимова Г. П., Пашикин М. А., Пашикина Е. В., Соловьев А. В.* Электронные архивы: возможные решения проблем долгосрочного хранения данных // Труды Ин-та системного анализа РАН. 2013. Т. 63. № 4. С. 39–49.
4. *Smith E.* When discs die. Tedium. <https://tedium.co/2017/02/02/disc-rot-phenomenon/> (дата создания: 02.02.2017 г.; дата обращения: 18.11.2017 г.).
5. *Маличенко Д. А.* Эвристический алгоритм расчета размеров памяти в многоуровневой системе хранения // Информационно-управляющие системы. 2015. № 5. С. 100–105.  
<https://doi.org/10.15217/issn1684-8853.2015.5.100>
6. *Zalaev G. Z., Kalenov N. E., Tsvetkova V. A.* Some issues of long-term storage of electronic documents // Scientific and Technical Information Processing. 2016. Vol. 43. N 4. P. 268–274.  
<https://doi.org/10.3103/S0147688216040110>
7. *Лобанов А. К.* Методы построения систем хранения данных // Jet Info Online. 2003. N 7.  
URL: <https://citforum.ru/hardware/data/db/> (дата обращения: 05.05.2016 г.).
8. *ГОСТ Р 54989–2012 / ISO TR 18492:2005.* Обеспечение долговременной сохранности электронных документов (Вступил в силу 01.05.2013 г.). М.: Стандартинформ, 2013.
9. *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.
10. *Чернышов А. В.* Модель надежности хранения информации на современных библиотеках оптических дисков, объединенных в массивы RAID–6 // Вестн. Моск. гос. технич. ун-та им. Н. Э. Баумана. Сер. Приборостроение. 2017. № 3. С. 65–75.  
<https://doi.org/10.18698/0236-3933-2017-3-65-75>
11. *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 // Вестн. Моск. гос. технич. ун-та им. Н. Э. Баумана. Сер. Приборостроение. 2017. № 4. С. 88–97. <https://doi.org/10.18698/0236-3933-2017-4-88-97>
12. *Чернышов А. В.* Исследование свойств долговременных электронных архивных хранилищ информации на оптических дисках, организованных в структуры RAID–5 // Информационные технологии. 2018. Т. 24. № 9. С. 586–593. <https://doi.org/10.17587/it.24.586-593>
13. *Weatherspoon H., Kubiatowicz J.* Erasure Coding vs. Replication: A quantitative comparison // Peer-to-Peer Systems, Lecture Notes in Computer Science. 2002. Vol. 2429. P. 328–337.
14. *Thomasian A., Tang Y., Hu Y.* Hierarchical RAID: Design, performance, reliability, and recovery // J. Parallel Distrib. Comput. 2012. Vol. 72. P. 1753–1769. <https://doi.org/10.1016/j.jpdc.2012.07.002>
15. *Thomasian A.* Shortcut method for reliability comparisons in RAID // Journal of Systems and Software. 2006. Vol. 79. P. 1599–1605.
16. *Yuan D., Peng X., Liu T., Cui Z.* A randomly expandable method for data layout of Raid Storage Systems // International Journal of Innovative Computing, Information and Control. 2018. Vol. 14. N 3. P. 1079–1094.
17. *Пилитчук М. И., Балакирев А. Н., Дмитриева Л. В., Залаев Г. З.* Рекомендации по обеспечению сохранности информации, записанной на оптических дисках (Тестирование выборочного массива документов федеральных архивов). М.: РГАНТД, 2011. 52 с.
18. *Юмашева Ю. Ю.* Методические рекомендации по электронному копированию архивных документов и управлению полученным информационным массивом. М.: Росархив, ВНИИДАД, 2012. 125 с.
19. *Рекомендации по комплектованию, учету и организации хранения электронных архивных документов в государственных и муниципальных архивах.* М.: Федеральное архивное агентство, ВНИИДАД, 2013. 49 с.
20. *Российские организации начинают строить электронные архивы на оптических дисках* // Ассоциация электронных торговых площадок. URL: <https://www.aetp.ru/market-news/item/400867> (дата создания: 06.11.2015 г.; дата обращения: 13.11.2015 г.).

Статья поступила в редакцию 6 февраля 2019 г.

Статья принята к печати 7 ноября 2019 г.

Контактная информация:

*Чернышов Александр Викторович* — канд. техн. наук, доц.; [sch-ru@yandex.ru](mailto:sch-ru@yandex.ru)

# Model for planning the number of optical disks needed to create and maintain a long-term electronic archive

A. V. Chernyshov

Mytischki Branch of N. E. Bauman Moscow State Technical University, 1, 1st Institutskaya ul., Mytischki, Moscow region, 141005, Russian Federation

**For citation:** Chernyshov A. V. Model for planning the number of optical disks needed to create and maintain a long-term electronic archive. *Vestnik of Saint Petersburg University. Applied Mathematics. Computer Science. Control Processes*, 2019, vol. 15, iss. 4, pp. 578–591. <https://doi.org/10.21638/11702/spbu10.2019.413> (In Russian)

Development of an analytical model that allows to estimate the number of optical disks needed annually to create and maintain a complete electronic archive of long-term storage of information, as well as the time costs associated with these processes. An analytical model is developed for archives based on single disks (disks are not combined into RAID structures). The model uses the following parameters as input data: the estimated (design) capacity of the archive; the capacity of the optical disks used; estimates of the time to write, read and control the integrity of optical disks; the time allocated for technical work in the archive (the actual recording of new data, the control of previously recorded disks, the recovery of failed); the probability of failure of the optical disk during storage; the probability of failure of the disk during recording (marriage). Considered 4 options for the functioning of the archive depending on the bandwidth of the equipment and describes the stages of each option. For each option, analytical expressions are obtained that allow to estimate annually the required number of optical disks (which largely determines the annual material costs), as well as the time of changing stages. In addition, for some variants, expressions are obtained to evaluate the possibility of achieving the design capacity of the archive.

*Keywords:* long-term archives of electronic information, optical discs, ensuring the integrity of information, annual costs.

## References

1. Verbatim Launches 100GB MDISC Blu-ray Storage Media. *CDRinfo*. Available at: <https://cdrinfo.com/d7/content/verbatim-launches-100gb-mdisc-blu-ray-storage-media> (created: May 18, 2016; accessed: January 04, 2019).
2. Optical media longevity. *The X Lab*. Available at: <https://www.thexlab.com/faqs/opticalmedialongevity.html> (accessed: September 29, 2016).
3. Akimova G. P., Pashkin M. A., Pashkina E. V., Solovyev A. V. Elektronnye arkhivy: vozmozhnye resheniya problem dolgosrochnogo khraneniya dannykh [Electronic archives: possible solutions to problems of long-term data storage]. *Trudy Instituta sistemnogo analiza Rossiyskoy Akademii nauk [Proceedings of the Institute of system analysis of the Russian Academy of Sciences]*, 2013, vol. 63, no. 4, pp. 39–49. (In Russian)
4. Smith E. *When discs die. Tedium*. Available at: <https://tedium.co/2017/02/02/disc-rot-phenomenon/> (created: February 02, 2017; accessed: November 18, 2017).
5. Malichenko D. A. Evristicheskiy algoritm rascheta razmerov pamyati v mnogourovnevnoy sisteme khraneniya [Heuristic algorithm for calculating the size of tiers in a hierarchical storage system]. *Informacionno-upravlyayushchie sistemy [Information management systems]*, 2015, no. 5, pp. 100–105. <https://doi.org/10.15217/issn1684-8853.2015.5.100> (In Russian)
6. Zalaev G. Z., Kalenov N. E., Tsvetkova V. A. Some issues of long-term storage of electronic documents. *Scientific and Technical Information Processing*, 2016, vol. 43, no. 4, pp. 268–274. <https://doi.org/10.3103/S0147688216040110>
7. Lobanov A. K. Metody postroeniya sistem khraneniya dannykh [A methods of build of data storage systems]. *Jet Info Online*, 2003, no. 7. Available at: <https://citforum.ru/hardware/data/db/> (accessed: May 05, 2016). (In Russian)
8. *GOST R 54989–2012 / ISO TR 18492:2005. Obespechenie dolgovremennoy sokhrannosti elektronnykh dokumentov [Ensuring long-term preservation of electronic records]*. (Entered to force May 01, 2013). Moscow, Standartinform Publ., 2013. (In Russian)

9. Chernyshov A. V. K voprosu o primeneni opticheskikh diskov dlya sozdaniya dolgovremennykh elektronnykh arkhivnykh khranilishch informatsii nebol'shikh organizatsiy [To the question of the optical discs application for long term digital archive storage of small organizations]. *Informacionnye tehnologii [Information technologies]*, 2016, vol. 22, no. 8, pp. 635–640. (In Russian)
10. Chernyshov A. V. Model' nadezhnosti khraneniya informatsii na sovremennykh bibliotekakh opticheskikh diskov, ob'edinennykh v massivy RAID–6 [Reliability model for data storage on modern libraries of optical disks united in RAID–6]. *Vestnik of Moscow State Technical University im. N. E. Bauman. Series Instrumentation*, 2017, no. 3, pp. 65–75. <https://doi.org/10.18698/0236-3933-2017-3-65-75> (In Russian)
11. Chernyshov A. V. Metod povysheniya nadezhnosti khraneniya informatsii v dolgovremennykh elektronnykh khranilishchakh na opticheskikh diskakh, organizovannykh v massivy RAID–6, za schet smeshivaniya diskov zapasnykh kopiy [Method for improving information storage reliability in long-term electronic storage on optical disks arrayed in RAID–6 by mixing backup copy disks]. *Vestnik of Moscow State Technical University im. N. E. Bauman. Series Instrumentation*, 2017, no. 4, pp. 88–97. <https://doi.org/10.18698/0236-3933-2017-4-88-97> (In Russian)
12. Chernyshov A. V. Issledovanie svoystv dolgovremennykh elektronnykh arkhivnykh khranilishch informatsii na opticheskikh diskakh, organizovannykh v struktury RAID–5 [The study of the properties of long-term electronic archival information storage on optical disks organized into a structure of RAID–5]. *Informacionnye tehnologii [Information technologies]*, 2018, vol. 24, no. 9, pp. 586–593. <https://doi.org/10.17587/it.24.586-593> (In Russian)
13. Weatherspoon H., Kubiawicz J. Erasure Coding vs. Replication: A quantitative comparison. *Peer-to-Peer Systems, Lecture Notes in Computer Science*, 2002, vol. 2429, pp. 328–337.
14. Thomasian A., Tang Y., Hu Y. Hierarchical RAID: Design, performance, reliability, and recovery. *J. Parallel Distrib. Comput.*, 2012, vol. 72, pp. 1753–1769. <https://doi.org/10.1016/j.jpdc.2012.07.002>
15. Thomasian A. Shortcut method for reliability comparisons in RAID. *Journal of Systems and Software*, 2006, vol. 79, pp. 1599–1605.
16. Yuan D., Peng X., Liu T., Cui Z. A randomly expandable method for data layout of Raid Storage Systems. *International Journal of Innovative Computing, Information and Control*, 2018, vol. 14, no. 3, pp. 1079–1094.
17. Pilipchuk M. I., Balakirev A. N., Dmitrieva L. V., Zalaev G. Z. *Rekomendacii po obespecheniyu sohranosti informacii, zapisannoj na opticheskikh diskah (Testirovanie vyborochnogo massiva dokumentov federal'nyh arhivov)* [Recommendations to ensure the safety of information recorded on optical discs (Testing of a sample array of Federal archives documents)]. Moscow, RGANDT Publ., 2011, 52 p. (In Russian)
18. Jumasheva Ju. Ju. *Metodicheskie rekomendacii po ehlektronnomu kopirovaniyu arhivnykh dokumentov i upravleniyu poluchennym informacionnym massivom [Guidelines for electronic copying of archival documents and management of the resulting information array]*. Moscow, Rosarhiv, VNIIDAD Publ., 2012, 125 p. (In Russian)
19. *Rekomendacii po komplektovaniyu, uchyotu i organizacii hraneniya ehlektronnykh arhivnykh dokumentov v gosudarstvennykh i municipal'nyh arhivah [Recommendations on acquisition, accounting and organization of storage of electronic archival documents in state and municipal archives]*. Moscow, Federalnoe arhives agency, VNIIDAD Publ., 2013, 49 p. (In Russian)
20. *Rossiyskie organizatsii nachinayut stroit' ehlektronnye arkhivy na opticheskikh diskakh [Russian organizations are beginning to build electronic archives on optical discs]*. Associaciya jelektronnykh trgovykh ploshhadok. Available at: <https://www.aetp.ru/market-news/item/400867> (created: November 06, 2015, accessed: November 13, 2015). (In Russian)

Received: February 06, 2019.

Accepted: November 07, 2019.

Author's information:

Alexandr V. Chernyshov — PhD in Technics, Associate Professor; sch-ru@yandex.ru