

Определение вероятности потери информации на частично записанном оптическом диске однократной записи при длительном хранении в электронном архиве

А.В. Чернышов
Мытищинский филиал МГТУ им. Н. Э. Баумана
Мытищи, Россия
sch-ru@yandex.ru

Аннотация—Рассмотрена задача определения вероятности потери информации на частично записанном оптическом диске однократной записи при длительном хранении в электронном архиве. Предложено два подхода к решению задачи. Приведены результаты численных расчётов, показавшие хорошее согласование искомого параметра, вычисленного обоими методами.

Ключевые слова— электронный архив длительного хранения информации, оптический диск однократной записи, вероятность потери информации.

1. ВВЕДЕНИЕ

Несмотря на развитие облачных технологий хранения больших объёмов информации, задача архивного (длительного) хранения электронных документов организаций должна по-прежнему решаться путём создания локальных электронных архивных хранилищ этих организаций. Международный опыт [1-4] склоняется к необходимости использования носителей однократной записи (WORM), к которым сегодня можно отнести только оптические диски типов CD-R, DVD+-R, BD-R.

При проектировании, создании и последующей эксплуатации такого архива естественным требованием является полная сохранность архивной информации. Для дисков однократной записи информация может быть потеряна частично или полностью из-за старения самого носителя — деградации диска [5, 6].

Проблема деградации оптических дисков с записью в процессе хранения мало изучена. Современные производители оптических дисков однократной записи предпочитают ограничиваться ускоренными тестами своих изделий [7], исходя из результатов которых обещают потребителям так называемый «срок службы» диска. Сегодня «срок службы» обычно называется в пределах 30 лет, иногда чуть больше.

Проблема заключается в том, что в течение указанного «срока службы» из-за деградации дисков всё равно существует опасность потери части хранимой информации из-за того, что какие-то диски из хранимого массива перестанут читаться частично или полностью. Она может быть решена путём создания архива с несколькими копиями хранимой информации и регулярным тестированием целостности информации на дисках [8, 9]. В случае обнаружения деградировавших дисков они могут быть заменены в архиве на новые с записью на них информации из запасных копий.

Однако для того, чтобы определить параметры эксплуатации архива (необходимое количество запасных копий, время, затрачиваемое на тестирование и замену дисков, необходимое количество ежегодно приобретаемых новых дисков для замены деградировавших) необходимо знать значение вероятности потери информации на одном оптическом диске q . А значение этого параметра никто из производителей дисков не сообщает.

На основании анализа доступных немногочисленных источников [6, 10, 11], а также по результатам проведённых собственных натуральных экспериментов [12, 13] автор может утверждать, что для дисков с заполнением, близким к их ёмкости, для практических расчётов надёжности архива можно принять вероятность потери информации на одном диске $q=0,1$.

Заполнение каждого диска близко к его максимальной ёмкости для его размещения в архиве является логичным и экономически оправданным.

Однако на практике встречаются специализированные архивы, при создании которых к заполнению дисков могут выдвигаться особые требования. Например, требование размещать на одном диске не больше одного файла или только файлы, относящиеся к какой-либо логически связанной группе.

В этом случае на диске будет оставаться незанятое пространство, деградация которого, не затрагивающая часть диска с записью, никак не скажется на воспроизведении с него информации. Следовательно, вероятность потери информации (обозначим её q^*) будет ниже.

На сегодняшний день отсутствуют данные, позволяющие определить вероятность потери информации в процессе хранения на не полностью заполненных дисках, которые были бы получены в результате натуральных экспериментов. Однако для практических расчётов таких архивов знание значения параметра необходимо. Попытаемся его оценить.

2. ОЦЕНКА ВЕРОЯТНОСТИ ПОТЕРИ ИНФОРМАЦИИ

А. Оценка геометрическим методом

Одним из вариантов оценки является геометрический подход. В этом случае будем просто считать, что вероятность потери информации линейно уменьшается с уменьшением занятой информацией площади на диске. Причём занятая площадь линейно зависит от объёма

записанной информации, выраженного в процентах ёмкости диска. То есть, например, для диска BD-R (полная ёмкость 25 гигабайт) при записи 5 гигабайт (20 % от полной ёмкости диска) получим $q^*=0,02$.

Б. Оценка вероятностным методом

Альтернативным подходом является попытка вычислить значение методами теории вероятностей.

Пусть архив создан из дисков ёмкостью V и q — вероятность потери хотя бы какой-то части информации на этом диске при хранении в течение года.

Пусть v — размер одного блока на диске и p — вероятность потери информации на нём. В данном случае не принципиально, о каком конкретно блоке в структуре записи информации на диске идёт речь. Но для корректности рассуждений будем считать, что запись информации на диск может выполняться только кратно размеру блока, то есть искомая величина соответствует вероятности потери информации в этом блоке.

Тогда всего на диске блоков $n=V/v$. И можно записать

$$q = \sum_{i=1, n} p^i (1-p)^{n-i} C_n^i \quad (1)$$

Обратим внимание, что выражение

$$\sum_{i=0, n} p^i (1-p)^{n-i} C_n^i = (p+(1-p))^n \quad (2)$$

представляет собой бином Ньютона. Следовательно

$$q = (p+(1-p))^n - C_n^0 (1-p)^n p^0 \quad (3)$$

В результате получим

$$p = 1 - (1-q)^{1/n} \quad (4)$$

Если теперь задаться средней долей заполненности каждого диска a ($a < 1$), то можно записать, что вероятность потери информации на диске q^* можно вычислить как

$$q^* = 1 - (1-p)^{an} \quad (5)$$

3. СРАВНЕНИЕ РЕЗУЛЬТАТОВ

Сравним результаты обоих методов расчёта. Для этого построим несложную таблицу в LibreOffice Calc для значений от 0,1 до 1,0 с шагом 0,1. Будем выполнять вычисления для дисков типа BD-R ($V=25$ гигабайт). Примем размер одного блока $v=2048$ байт. Результаты расчётов приведены в таблице 1.

В приведённой таблице числа сознательно даны именно в той форме, как показывает их LibreOffice Calc. Можно видеть, что при небольших значениях вероятностный метод даёт значения q^* , незначительно превышающие геометрический метод. Выполнив округление до сотых стандартным методом, принятым в математике, получим совпадение значений в обоих методах.

4. ЗАКЛЮЧЕНИЕ

Учитывать ли получившиеся расхождения при проектных расчётах архива, должен решить сам проектировщик архива. К тому же для полностью заполненного диска результат вычисления для вероятностного метода явно содержит погрешности вычислений, что заставляет подумать о том же и для других результатов.

В целом можно констатировать, что оба метода вычисления значений q^* не противоречат друг другу, а скорее подтверждают результаты друг друга.

ТАБЛИЦА 1. СРАВНЕНИЕ РЕЗУЛЬТАТОВ ВЫЧИСЛЕНИЙ

a	q^*	
	Геометрический метод	Вероятностный метод
0,1	0,01	0,0104807417
0,2	0,02	0,0208516375
0,3	0,03	0,0311138387
0,4	0,04	0,0412684843
0,5	0,05	0,0513167017
0,6	0,06	0,0612596064
0,7	0,07	0,071098302
0,8	0,08	0,0808338808
0,9	0,09	0,0904674235
1,0	0,10	0,0999999995

ЛИТЕРАТУРА

- [1] ISO/TR 18492:2005 Long-term preservation of electronic document-based information / Last reviewed and confirmed in 2013. – URL: http://www.iso.org/iso/catalogue_detail.htm?csnumber=38716 (26.07.2021).
- [2] Zheng, J. NIST/Library of Congress Optical Disc Longevity Study: Final Report / J. Zheng, O.T. Slattery. – 2007. – 32 p. URL: https://www.loc.gov/preservation/resources/rt/NIST_LC_OpticalDiscLongevity.pdf (11.05.2016).
- [3] Haus, G. Preserving cultural heritage: A new approach to increase the life expectancy of optical discs / G. Haus, C. Polizzi, A. Visconti // Journal of Cultural Heritage. – 2018. – Vol. 29. – P. 67-74. DOI: 10.1016/j.culher.2017.08.004.
- [4] Bradley, K. Risks Associated With the Use of Recordable CDs and DVDs as Reliable Storage Media in Archival Collections / K. Bradley // Strategies and Alternatives. – UNESCO, Paris, 2006. – 31 p.
- [5] Smith, E. When Discs Die / E. Smith // Tedium // URL: <https://tedium.co/2017/02/02/disc-rot-phenomenon/> (18.11.2017).
- [6] Пилипчук, М.И. Рекомендации по обеспечению сохранности информации, записанной на оптических дисках (Тестирование выборочного массива документов федеральных архивов) / М.И. Пилипчук, А.Н. Балакирев, Л.В. Дмитриева, Г.З. Залаев. – М.: РГАНТД, 2011. – 52 с.
- [7] ISO/IEC 10995:2011(E) Information technology - Digitally recorded media for information interchange and storage // Test method for the estimation of the archival lifetime of optical media. – 2015. – 28 p.
- [8] Zalaev, G.Z. Some Issues of Long-Term Storage of Electronic Documents / G.Z. Zalaev, N.E. Kalenov, V.A. Tsvetkova // Sci. Tech. Inf. Proc. – 2016. – Vol. 43(4). – P. 268-274. DOI: 10.3103/S0147688216040110.
- [9] Corrado, E.M. Digital Preservation for Libraries, Archives, and Museums / E.M. Corrado, H.L. Rowman. – Moulaison & Littlefield Publishers, 2014. – 294 p.
- [10] van Ooijen, P.M.A. Accessibility of Data Backup on CD-R after 8 to 11 years / P.M.A. van Ooijen, A.R. Viddeleer, F. Meijer, M. Oudkerk // J Digit Imaging. – 2010. – Vol. 23. – P. 95. DOI: 10.1007/s10278-008-9161-9.
- [11] Устинов, В. Хранение данных на CD- и DVD-дисках: на наш век хватит? / В. Устинов // BROADCASTING: Телевидение и радиовещание. – 2006. – № 4. URL: http://www.broadcasting.ru/articles2/Oborandteh/hranenie_dannyh_na_CD_DVD_diskah (10.12.2018).
- [12] Чернышов, А.В. Оценка вероятности выхода из строя оптического диска DVD-R с записью по результатам натурального эксперимента / А.В. Чернышов // Современные технологии в задачах управления, автоматизации и обработки информации: Сборник трудов XXVII Международной научно-технической конференции. – Тамбов: ТГТУ, 2018. – С. 272.
- [13] 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. – 2019. – Т. 23, № 4. – С. 82-87. DOI: 10.18698/2542-1468-2019-4-82-87.