

# АНАЛИЗ ПОКАЗАТЕЛЕЙ НАДЕЖНОСТИ ИЗБЫТОЧНЫХ ДИСКОВЫХ МАССИВОВ

П.А. Рахман

## Введение

В последние три десятилетия наблюдается бурное развитие информационных технологий и их внедрение в самые различные сферы деятельности человека, и информация, представленная в электронном виде, стала ключевой частью жизни и работы не только организаций, но и каждого отдельного человека. Более того, сохранность и доступность информации для ее пользователей, как правило, имеет критическую важность, а потеря данных нередко может приводить к катастрофическим последствиям. В связи с этим, на протяжении нескольких десятилетий учеными и специалистами разрабатываются самые разнообразные технические средства и организационные подходы по обеспечению надежного хранения данных и их высокой доступности для пользователей. В такой ситуации анализ показателей надежности дисковых массивов имеет достаточно высокую актуальность, особенно для предприятий среднего и крупного масштабов, поскольку такой анализ также позволяет косвенно оценивать риски потери данных и принимать соответствующие решения, и при необходимости внедрять дополнительные технические средства.

Современная теория надежности [1, 2], в основном базируется на математический аппарат теории вероятностей [3] и дополнительно на множество других разделов математики. Здесь следует отметить, что в отечественной литературе по теории надежности достаточно глубоко рассматриваются теоретические аспекты, однако, примеры и задачи, приводимые в них, как правило, очень далеки от технологий и средств, используемых в современных системах хранения данных. В такой ситуации возникает глубокий научный и академический «пробел» между теоретическими моделями и современными информационными технологиями, что в конечном итоге приводит к фактическому игнорированию теоретических аспектов надежности современными специалистами по проектированию и эксплуатации систем хранения данных, вынуждая их в основном полагаться на практический опыт и интуицию. За рубежом, в частности в США, ситуация обстоит лучше, и на протяжении последних двух десятилетий было проведено множество исследований и опубликовано ряд статей и диссертаций [4, 5, 6, 7], посвященных моделям надежности избыточных дисковых массивов.

В рамках данной статьи автор не ставит целью изложение и интерпретацию результатов западных исследований по рассматриваемому вопросу. Вместо этого автором сделана собственная попытка «увязывания» теории и практики касательно задачи анализа показателей надежности дисковых массивов, на основе собственного многолетнего изучения теоретических моделей и многолетней практики проектирования и эксплуатации дисковых систем. Следует отметить, что в рамках небольшого исследования невозможно было всесторонне охватить рассматриваемый вопрос, и автор сосредотачивался только на оценке наиболее важного показателя надежности – среднего времени наработки до разрушения (безвозвратной частичной или полной потери данных). В западной литературе такой показатель обозначается как: MTTDL – Mean Time To Data Loss (среднее время до потери данных). Также невозможно было рассмотреть всевозможные технологии и практические реализации отказоустойчивого хранения данных, и автор ограничивался рамками известных дисковых массивов RAID-0, RAID-5, RAID-6 и RAID-1.

Также следует отметить, что в современной практике систем хранения данных получили распространение двухуровневые каскадные RAID-массивы (RAID-01, RAID-10, RAID-50, RAID-60, RAID-51, RAID-61), позволяющие более гибко выбирать между избыточностью, отказоустойчивостью и быстродействием, но в рамках данной статьи они не рассматриваются.

# 1. Обзор дисковых массивов RAID

## 1.1. Описание дисковых массивов RAID-0, 1, 5 и 6

**RAID** (Redundant Array of Inexpensive Disks) – избыточный массив недорогих дисков. Технология RAID применяется для объединения нескольких дисков в единый массив большой емкости. Существует несколько разновидностей массивов RAID:

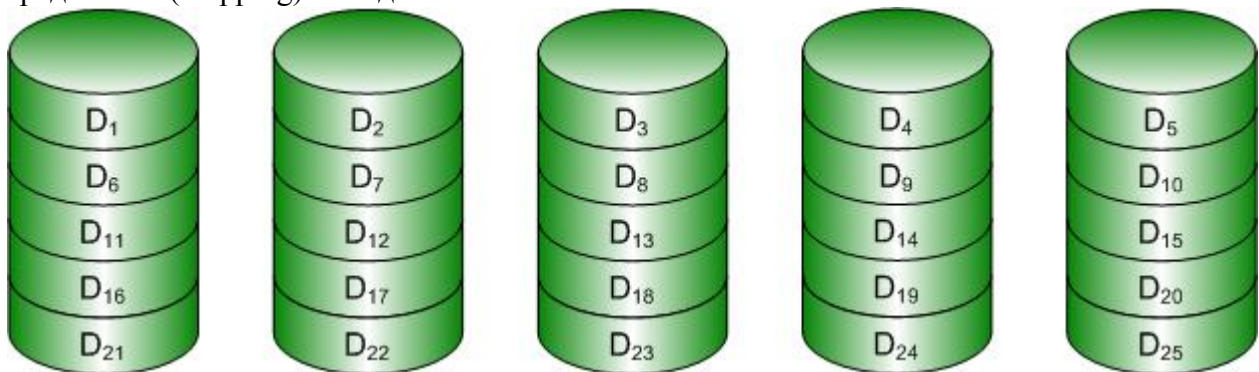
**Массив RAID-0.** Дисковый массив RAID-0 состоит из  $n \geq 2$  независимых дисков одинаковой емкости и является чисто последовательной системой с точки зрения надежности, поскольку сохраняет работоспособность только при условии работоспособности всех  $n$  дисков.

**Массив RAID-0 не обладает избыточностью и не обеспечивает никакой отказоустойчивости. При отказе любого из дисков все пользовательские данные всего массива безвозвратно теряются.**

Полезная емкость массива составляет суммарную емкость всех дисков массива.

Пользовательские данные разбиваются на блоки фиксированного размера и методом «чередования» распределяются по дискам – такой подход дает существенный выигрыш по производительности на операциях чтения данных по сравнению с отдельным диском, особенно при множестве одновременных запросов на чтение данных.

Ниже приводится упрощенная схема распределения блоков данных методом чередования (striping) по 5 дискам массива RAID-0.



**Массив RAID-1.** Дисковый массив RAID-1 состоит из  $n \geq 2$  независимых дисков одинаковой емкости и является число параллельной системой с точки зрения надежности, поскольку сохраняет работоспособность до тех пор, пока не откажут все  $n$  дисков.

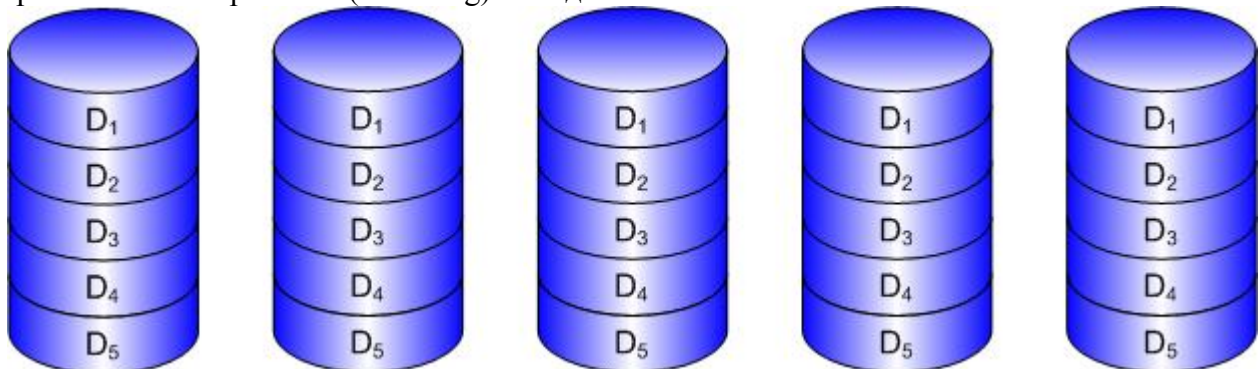
Массив RAID-1 обладает самой большой отказоустойчивостью и, соответственно, избыточностью, поскольку все диски массива хранят одни и те же пользовательские данные.

**Пользовательские данные теряются только при отказе всех дисков.**

Полезная емкость массива составляет  $1 / n$  часть от суммарной емкости всех дисков.

По этой причине на практике чаще всего используются RAID-1 массивы из  $n = 2$  дисков.

Ниже приводится упрощенная схема распределения блоков данных методом зеркального копирования (mirroring) по 5 дискам массива RAID-1.



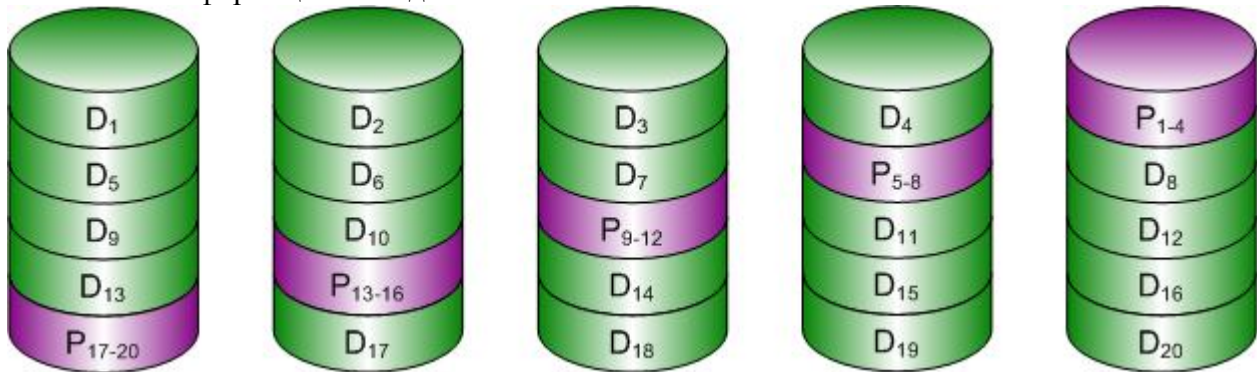
**Массив RAID-5.** Массив RAID-5 состоит из  $n \geq 3$  независимых дисков одинаковой емкости и сохраняет работоспособность при отказе не более одного (любого) диска.

Массив RAID-5 является компромиссом между отказоустойчивостью и избыточностью. На каждом из дисков  $1/n$  часть дискового пространства отводится для хранения избыточной (контрольной) информации, вычисляемой по пользовательским данным, хранящимся на других дисках. Это позволяет в случае отказа любого одного из дисков вычислять недостающую информацию по пользовательским данным и контрольной информации, хранящимся на оставшихся дисках.

**При отказе любых двух дисков, равно как и при отказе любого второго диска до того, как произойдет замена и полное восстановление ранее отказавшего другого диска, все пользовательские данные всего массива теряются.**

Полезная емкость массива составляет  $(n - 1) / n$  часть от суммарной емкости дисков.

Ниже приводится упрощенная схема распределения пользовательских данных и избыточной информации по 5 дискам массива RAID-5.



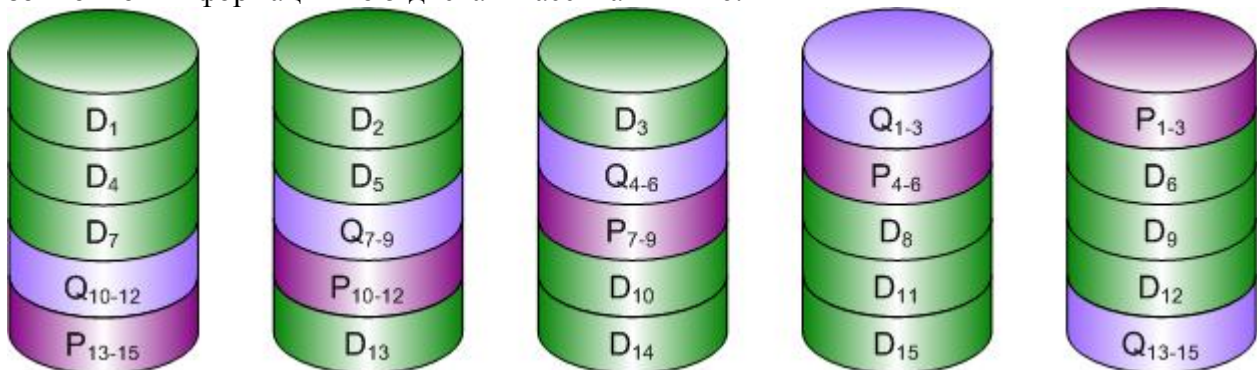
**Массив RAID-6.** Массив RAID-6 состоит из  $n \geq 4$  независимых дисков одинаковой емкости и сохраняет работоспособность при отказе не более двух (любых) дисков.

Массив RAID-6 является компромиссом между отказоустойчивостью и избыточностью. На каждом из дисков  $2/n$  часть дискового пространства отводится для хранения избыточной (контрольной) информации, вычисляемой по пользовательским данным, хранящимся на других дисках. Это позволяет в случае отказа любого одного или двух дисков вычислять недостающую информацию по пользовательским данным и контрольной информации, хранящимся на оставшихся дисках.

**При отказе любых трех дисков, равно как и при отказе любого третьего диска до того, как произойдет замена и полное восстановление одного из двух ранее отказавших других дисков, все пользовательские данные всего массива теряются.**

Полезная емкость массива составляет  $(n - 2) / n$  часть от суммарной емкости дисков.

Ниже приводится упрощенная схема распределения пользовательских данных и избыточной информации по 5 дискам массива RAID-6.

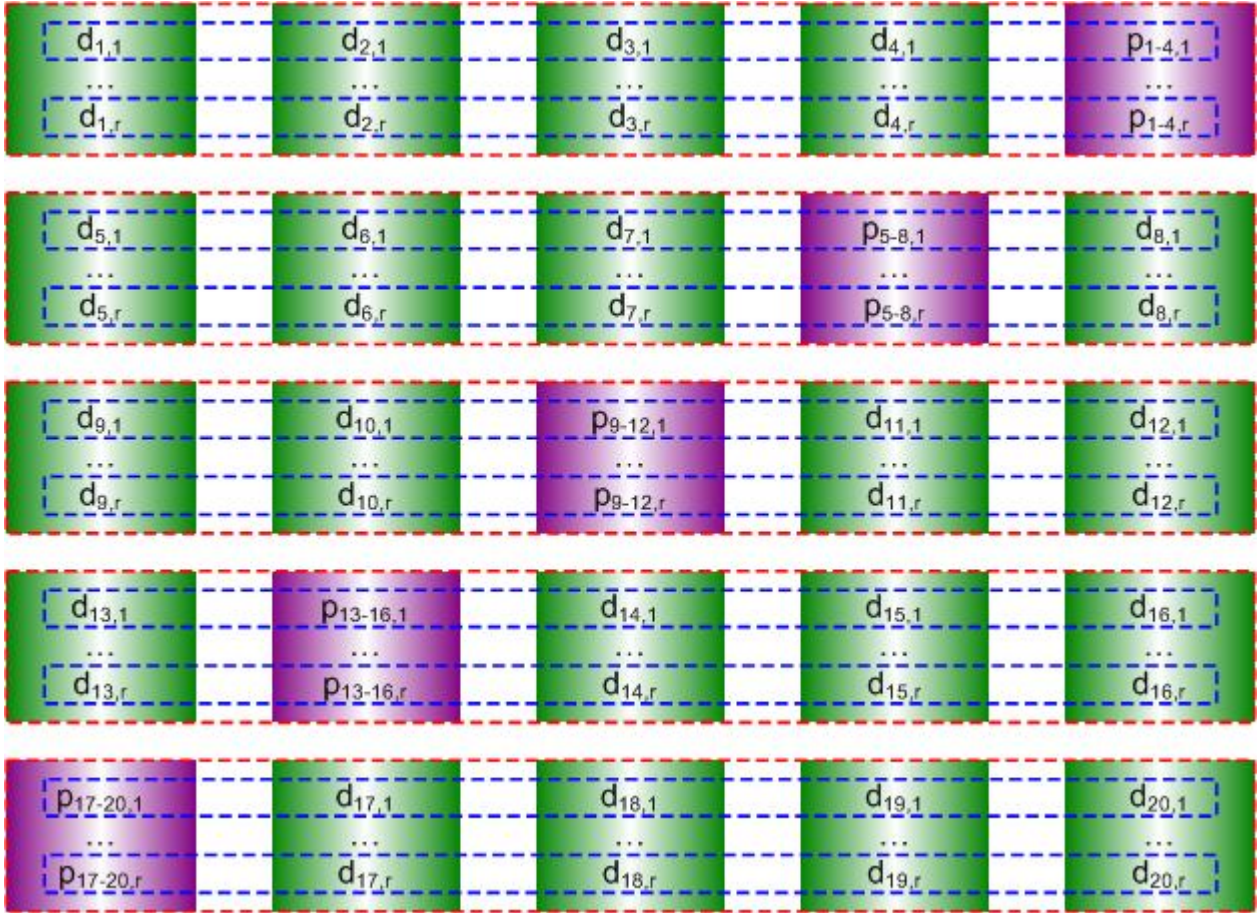




## 1.2. Особенности реализации технологий RAID-5 и RAID-6

**Технология RAID-5.** Технология RAID-5 разбивает данные на блоки и размещает их по «горизонтальным сечениям», охватывающим по одному блоку каждого из  $n$  дисков. В каждом «сечении»  $n-1$  информационных блоков  $\{D\}$  и один контрольный блок  $P$ . В свою очередь, блоки содержат массивы байтов, причем контрольные байты, расположенные в контрольном блоке  $P$ , независимо и параллельно вычисляются по соответствующим информационным байтам, расположенных в информационных блоках  $\{D\}$ .

Следует отметить, что количество сечений определяется емкостью диска  $V$  (в байтах) и размером блока  $r$  (в байтах) и составляет  $m = V / r$ . Заметим, что  $m$  вовсе не обязательно делиться нацело на число дисков  $n$  (на рисунке  $m = 5$  и  $n = 5$  лишь для наглядности).



Контрольная сумма для информационных байтов вычисляется, как их побитовая сумма по модулю 2 (побитовая операция XOR), которая по сути является арифметической операцией сложения элементов поля Галуа  $GF(2^8)$ , причем каждый информационный байт, равно как и контрольная сумма интерпретируются как элементы поля  $GF(2^8)$ . Описание поля Галуа  $GF(2^8)$  и его арифметики приводится в приложении.

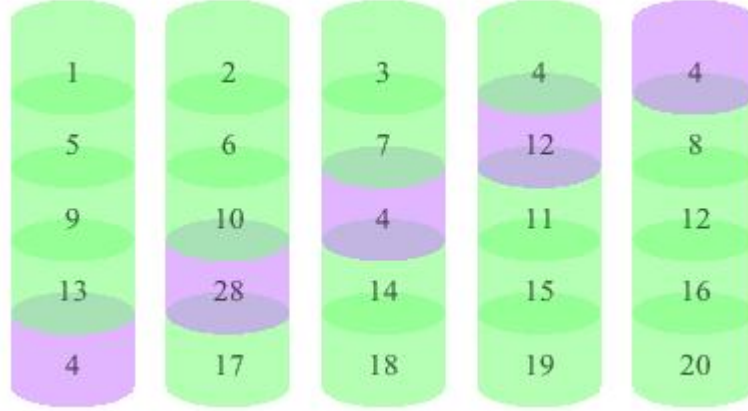
Иными словами для каждого  $k$ -го «горизонтального сечения» для каждой  $j$ -й группы из  $n-1$  информационных байтов, размещенных в соответствующих информационных блоках  $\{D\}$ , вычисляется контрольный байт  $p$ , размещаемый в контрольном блоке  $P$ :

$$p_{a-b,j} = \overbrace{d_{a,j} \oplus \dots \oplus d_{b,j}}^{GF(2^8)} = \sum_{i=a}^b d_{i,j} \quad (1.1.1)$$

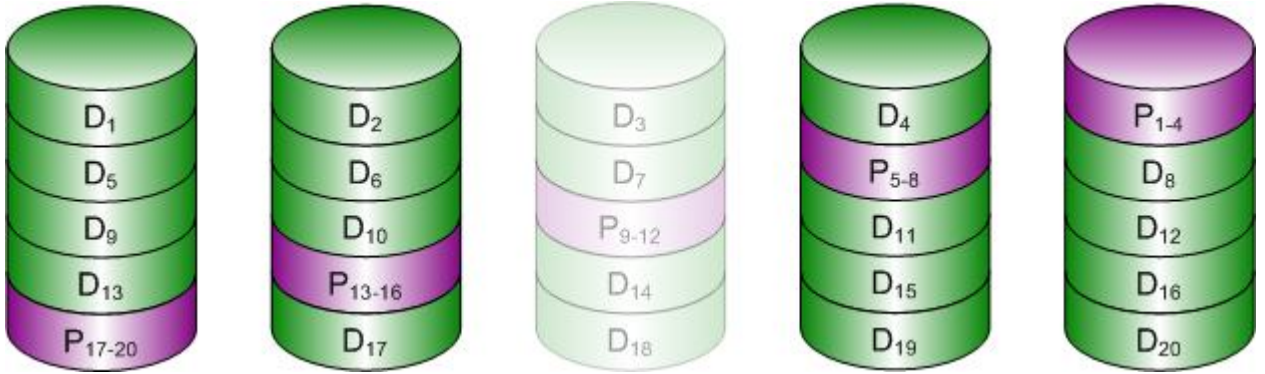
$$j = 0 \dots r-1. \quad a = (n-1) \cdot k + 1. \quad r = BlockSize$$

$$k = 0 \dots (V/r) - 1. \quad b = (n-1) \cdot (k+1). \quad V = DiskSize$$

Ниже приведен пример расчета контрольных блоков  $P$  по информационным блокам  $\{D\}$ . Для упрощения примера размер блока выбран 1 байт, иными словами, 1 блок содержит 1 байт. В качестве «пользовательских данных», хранящихся в информационных блоках взяты порядковые номера этих блоков. Соответственно, в контрольных блоках  $P$  отображены вычисленные значения.



**Массив RAID-5 при отказе одного диска.** При отказе любого одного диска оказываются недоступными все блоки, расположенные на этом диске, причем для некоторых «горизонтальных сечений» недоступным оказывается один из информационных блоков  $D$ , а для других – контрольный блок  $P$ :



В «горизонтальных сечениях», в которых стал недоступным контрольный блок  $P_{a-b}$ , этот блок легко вычисляется по  $n-1$  информационным блокам  $D_i : a \leq i \leq b$ :

$$p_{a-b,j} = \sum_{i=a}^b \overbrace{d_{i,j}}^{GF(2^8)} ; \quad j = 0 \dots r-1; \quad a = (n-1) \cdot k + 1; \quad (1.1.2)$$

$$k = 0 \dots (V/r) - 1; \quad b = (n-1) \cdot (k+1);$$

В «горизонтальных сечениях», в которых стал недоступным информационный блок  $D_u : a \leq u \leq b$ , этот блок вычисляется по контрольному блоку  $P_{a-b}$  и по оставшимся  $n-2$  информационным блокам  $D_i : a \leq i \leq b; i \neq u$ :

$$d_{u,j} = p_{a-b,j} \oplus \sum_{i=a}^b \overbrace{d_{i,j}}^{GF(2^8)} ; \quad j = 0 \dots r-1; \quad a = (n-1) \cdot k + 1; \quad (1.1.3)$$

$$k = 0 \dots (V/r) - 1; \quad b = (n-1) \cdot (k+1);$$

$$i \neq u$$

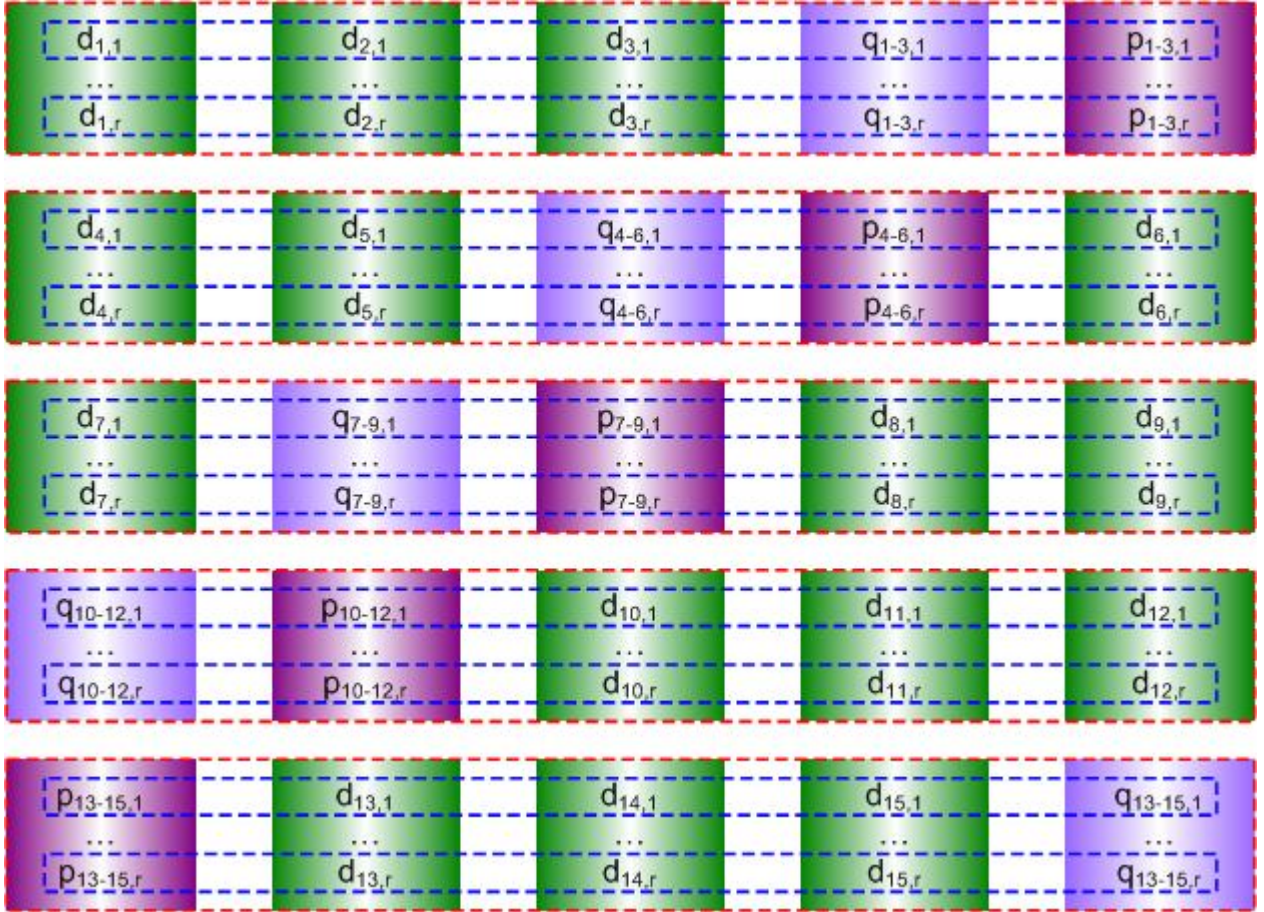
Таким образом, обеспечивается вычисление «на лету» отсутствующих блоков.

После замены вышедшего из строя диска, выполняется вычисление и запись отсутствующих блоков на новый диск.



**Технология RAID-6.** Технология RAID-6 разбивает данные на блоки и размещает их по «горизонтальным сечениям», охватывающим по одному блоку каждого из  $n$  дисков. В каждом «сечении»  $n-2$  информационных блоков  $\{D\}$  и два контрольных блока  $P$  и  $Q$ . В свою очередь, блоки содержат массивы байтов, причем контрольные байты, расположенные в контрольном блоке  $P$  и контрольные байты, расположенные в блоке  $Q$ , независимо и параллельно вычисляются по соответствующим информационным байтам, расположенным в информационных блоках.

Следует отметить, что количество «сечений» определяется емкостью диска  $V$  (в байтах) и размером блока  $r$  (в байтах) и составляет  $m = V / r$ . Заметим, что  $m$  вовсе не обязательно делиться нацело на число дисков  $n$  (на рисунке  $m = 5$  и  $n = 5$  лишь для наглядности).



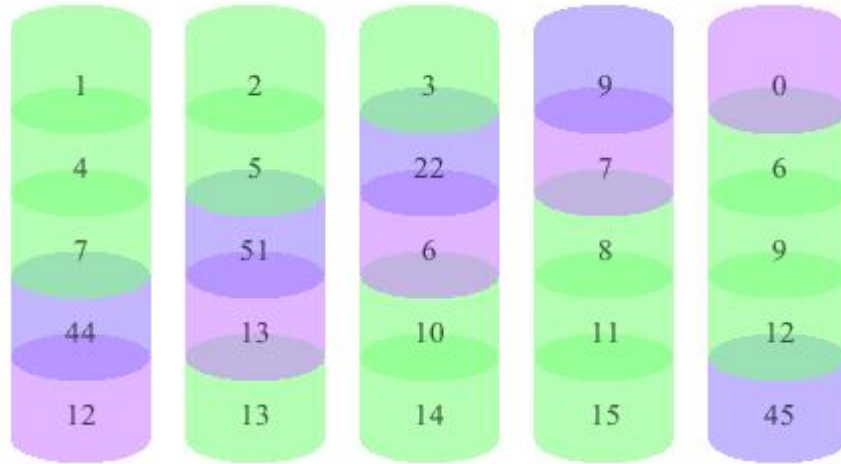
Для каждого  $k$ -го «горизонтального сечения» для каждой  $j$ -й группы из  $n-2$  информационных байтов, размещенных в соответствующих информационных блоках  $\{D\}$ , вычисляются контрольные байты  $p$  и  $q$ , размещаемые в соответствующих контрольных блоках  $P$  и  $Q$ , следующим образом:

$$\begin{aligned}
 p_{a-b,j} &= \overbrace{d_{a,j} \oplus d_{a+1,j} \oplus \dots \oplus d_{b,j}}^{GF(2^8)} = \sum_{i=a}^b d_{i,j} \\
 q_{a-b,j} &= \overbrace{d_{a,j} \cdot \alpha^0 \oplus d_{a+1,j} \cdot \alpha^1 \oplus \dots \oplus d_{b,j} \cdot \alpha^{n-3}}^{GF(2^8)} = \sum_{i=a}^b \left( d_{i,j} \cdot \alpha^{i-a} \right)
 \end{aligned} \tag{1.2.1}$$

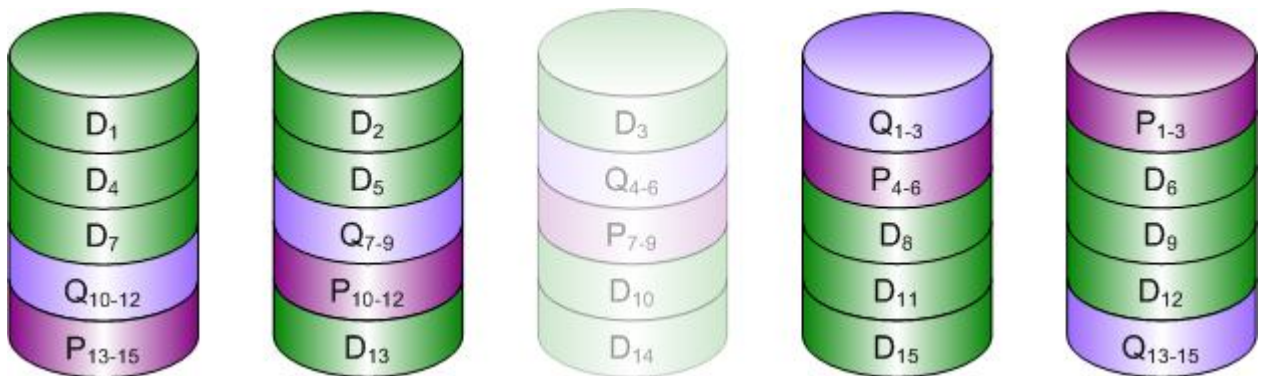
$$\begin{aligned}
 j &= 0 \dots r-1. & a &= (n-2) \cdot k + 1. & r &= \text{BlockSize} \\
 k &= 0 \dots (V/r) - 1. & b &= (n-2) \cdot (k+1). & V &= \text{DiskSize}
 \end{aligned}$$

Заметим, что контрольный байт  $p$  вычисляется точно так же, как и в технологии RAID-5, и является побитовой суммой по модулю 2 соответствующих информационных байтов. Что же касается контрольного байта  $q$ , то он вычисляется как аддитивно-взвешенная сумма, где в качестве весов выступают степени примитивного элемента  $\alpha$  поля Галуа  $GF(2^8)$ . Такой подход выбран не случайно, он обеспечивает линейную независимость двух уравнений, благодаря которой становится возможным однозначное решение системы двух уравнений с двумя неизвестными, и это позволяет вычислять недостающие информационные и контрольные байты при двух отказавших дисках.

Ниже приведен пример расчета контрольных блоков  $P$  и  $Q$  по информационным блокам  $\{D\}$ . Для упрощения примера размер блока выбран 1 байт, иными словами, 1 блок содержит 1 байт. В качестве «пользовательских данных», хранящихся в информационных блоках взяты порядковые номера этих блоков. Соответственно, в контрольных блоках  $P$  и  $Q$  отображены вычисленные значения.



**Массив RAID-6 при отказе одного диска.** При отказе любого одного диска оказываются недоступными все блоки, расположенные на этом диске, причем для некоторых «горизонтальных сечений» недоступным оказывается один из информационных блоков  $D_u$ , для других – контрольный блок  $P$ , для третьих – контрольный блок  $Q$ :



В «горизонтальных сечениях», в которых стал недоступным контрольный блок  $P_{a-b}$ , этот блок легко вычисляется по  $n-2$  информационным блокам  $D_i : a \leq i \leq b$ :

$$p_{a-b,j} = \sum_{i=a}^b \overbrace{d_{i,j}}^{GF(2^8)} ; \quad j = 0 \dots r-1; \quad a = (n-2) \cdot k + 1; \quad k = 0 \dots (V/r) - 1; \quad b = (n-2) \cdot (k+1); \quad (1.2.2)$$

В «горизонтальных сечениях», в которых стал недоступным контрольный блок  $Q_{a-b}$ , этот блок легко вычисляется по  $n-2$  информационным блокам  $D_i : a \leq i \leq b$ :

$$q_{a-b,j} = \sum_{i=a}^b \left( d_{i,j} \cdot \alpha^{i-a} \right) ; \quad j = 0 \dots r-1; \quad a = (n-2) \cdot k + 1; \quad k = 0 \dots (V/r)-1; \quad b = (n-2) \cdot (k+1); \quad (1.2.3)$$

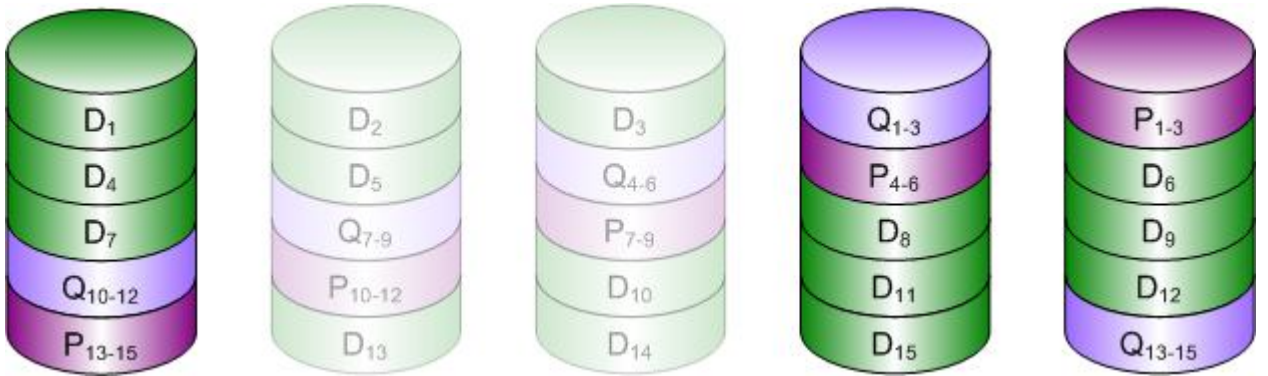
В «горизонтальных сечениях», в которых стал недоступным информационный блок  $D_u : a \leq u \leq b$ , этот блок вычисляется по контрольному блоку  $P_{a-b}$  и по оставшимся  $n-3$  информационным блокам  $D_i : a \leq i \leq b; i \neq u$ :

$$d_{u,j} = p_{a-b,j} \oplus \sum_{\substack{i=a \\ i \neq u}}^b d_{i,j} ; \quad j = 0 \dots r-1; \quad a = (n-2) \cdot k + 1; \quad k = 0 \dots (V/r)-1; \quad b = (n-2) \cdot (k+1); \quad (1.2.4)$$

Таким образом, обеспечивается вычисление «на лету» отсутствующих блоков.

После замены вышедшего из строя диска, выполняется вычисление и запись отсутствующих блоков на новый диск (процедура rebuild).

**Массив RAID-6 при отказе двух дисках.** При отказе любых двух различных дисков оказываются недоступными все блоки, расположенные на этих дисках, причем для некоторых «горизонтальных сечений» оказываются недоступными контрольные блоки  $P$  и  $Q$ , для других – контрольный блок  $P$  и информационный блок  $D_u$ , для третьих – контрольный блок  $Q$  и информационный блок  $D_u$ , наконец, для четвертых – информационные блоки  $D_u$  и  $D_v$ .



В «горизонтальных сечениях», в которых оказываются недоступными два контрольных блока  $P_{a-b}$  и  $Q_{a-b}$ , эти блоки вычисляются по имеющимся  $n-2$  информационным блокам  $D_i : a \leq i \leq b$ .

$$p_{a-b,j} = \sum_{i=a}^b d_{i,j} ; \quad q_{a-b,j} = \sum_{i=a}^b \left( d_{i,j} \cdot \alpha^{i-a} \right) \quad (1.2.5)$$

$$j = 0 \dots r-1. \quad a = (n-2) \cdot k + 1$$

$$k = 0 \dots (V/r)-1. \quad b = (n-2) \cdot (k+1)$$



В «горизонтальных сечениях», в которых оказываются недоступными контрольный блок  $Q_{a-b}$  и информационный блок  $D_u : a \leq u \leq b$ , сначала вычисляется неизвестный информационный блок  $D_u$ , а затем уже контрольный блок  $Q_{a-b}$ :

$$\begin{aligned}
 & \overbrace{GF(2^8)}^{GF(2^8)} \\
 1) \ d_{u,j} &= p_{a-b,j} \oplus \sum_{\substack{i=a \\ i \neq u}}^b d_{i,j} \quad ; \quad 2) \ q_{a-b,j} = \sum_{i=a}^b \left( d_{i,j} \cdot \alpha^{i-a} \right) \\
 & j = 0 \dots r-1. \quad a = (n-2) \cdot k + 1 \\
 & k = 0 \dots (V/r) - 1. \quad b = (n-2) \cdot (k+1)
 \end{aligned} \tag{1.2.6}$$

В «горизонтальных сечениях», в которых оказываются недоступными контрольный блок  $P_{a-b}$  и информационный блок  $D_u : a \leq u \leq b$ , сначала вычисляется неизвестный информационный блок  $D_u$ , а затем уже контрольный блок  $P_{a-b}$ :

$$\begin{aligned}
 & \overbrace{GF(2^8)}^{GF(2^8)} \\
 1) \ d_{u,j} &= \frac{1}{\alpha^{u-a}} \cdot \left( q_{a-b,j} \oplus \sum_{\substack{i=a \\ i \neq u}}^b \left( d_{i,j} \cdot \alpha^{i-a} \right) \right) \quad ; \quad 2) \ p_{a-b,j} = \sum_{i=a}^b d_{i,j} \\
 & j = 0 \dots r-1. \quad a = (n-2) \cdot k + 1 \\
 & k = 0 \dots (V/r) - 1. \quad b = (n-2) \cdot (k+1)
 \end{aligned} \tag{1.2.7}$$

Наиболее трудоемкий случай, когда в «горизонтальном сечении» оказываются недоступными два информационных блока  $D_u : a \leq u \leq b$  и  $D_v : a \leq v \leq b$ . В этом случае их можно найти из системы двух уравнений с двумя неизвестными, используя остальные имеющиеся  $n-4$  информационных блока, и два контрольных блока  $P_{a-b}$  и  $Q_{a-b}$ :

$$\left\{ \begin{aligned}
 & \overbrace{GF(2^8)}^{GF(2^8)} \\
 & p_{a-b,j} = d_{u,j} \oplus d_{v,j} \oplus \sum_{\substack{i=a \\ i \neq u \quad i \neq v}}^b d_{i,j} \\
 & \overbrace{GF(2^8)}^{GF(2^8)} \\
 & q_{a-b,j} = \left( d_{u,j} \cdot \alpha^{u-a} \right) \oplus \left( d_{v,j} \cdot \alpha^{v-a} \right) \oplus \sum_{\substack{i=a \\ i \neq u \quad i \neq v}}^b \left( d_{i,j} \cdot \alpha^{i-a} \right)
 \end{aligned} \right.$$

$$\begin{aligned}
 & j = 0 \dots r-1. \quad a = (n-2) \cdot k + 1 \\
 & k = 0 \dots (V/r) - 1. \quad b = (n-2) \cdot (k+1)
 \end{aligned}$$

Выразим из системы двух уравнений два неизвестных информационных блока:

$$\begin{aligned}
 d_{u,j} &= \frac{\overbrace{\left( \alpha^{v-a} \cdot \left( p_{a-b,j} \oplus \sum_{\substack{i=a \\ i \neq u, i \neq v}}^b d_{i,j} \right) \oplus \left( q_{a-b,j} \oplus \sum_{\substack{i=a \\ i \neq u, i \neq v}}^b d_{i,j} \cdot \alpha^{i-a} \right) \right)}^{GF(2^8)}}{\alpha^{u-a} \oplus \alpha^{v-a}} \\
 d_{v,j} &= \frac{\overbrace{\left( \alpha^{u-a} \cdot \left( p_{a-b,j} \oplus \sum_{\substack{i=a \\ i \neq u, i \neq v}}^b d_{i,j} \right) \oplus \left( q_{a-b,j} \oplus \sum_{\substack{i=a \\ i \neq u, i \neq v}}^b d_{i,j} \cdot \alpha^{i-a} \right) \right)}^{GF(2^8)}}{\alpha^{u-a} \oplus \alpha^{v-a}} \\
 j &= 0 \dots r-1. \quad a = (n-2) \cdot k + 1 \\
 k &= 0 \dots (V/r) - 1. \quad b = (n-2) \cdot (k+1)
 \end{aligned} \tag{1.2.8}$$

Таким образом, обеспечивается вычисление «на лету» отсутствующих блоков.

После замены вышедших из строя дисков, выполняется вычисление и запись отсутствующих блоков на новые диски (процедура rebuild).

Примечание. Следует отметить, что в случае двух неизвестных информационных блоков для их вычисления нам приходится выполнять деление на выражение  $(\alpha^{u-a} \oplus \alpha^{v-a})$ , которое в силу свойств поля Галуа может обращаться в нуль даже с учетом того, что  $u$  и  $v$  заведомо разные номера информационных блоков.

Дело в том, что в поле Галуа  $GF(2^8)$ , во-первых, сумма любых двух одинаковых

элементов равна нулю  $\overbrace{a+a}^{GF(2^8)} = 0$ , а во-вторых,  $\alpha^0 = \alpha^{(2^8-1)} = 1$ , а в остальных случаях  $1 \leq w \leq 2^8 - 2 : \alpha^w \neq 1$ .

Тогда, учитывая что  $(\alpha^{u-a} \oplus \alpha^{v-a}) = \alpha^{v-a} \cdot (\alpha^{u-v} \oplus 1) = \alpha^{u-a} \cdot (\alpha^{v-u} \oplus 1)$ , мы видим, что если окажется  $|u-v| = 2^8 - 1 = 255$ , то выражение  $(\alpha^{u-a} \oplus \alpha^{v-a})$  обратится в нуль. Отсюда, следует важное ограничение на номера информационных блоков в пределах одного «горизонтального сечения»:  $\forall u, v \in [a, b] : |u-v| \leq 254$ . Поскольку, в крайнем случае,  $u = b$  и  $v = a$  (или же наоборот), то получаем ограничение:  $|b-a| \leq 254$ . Наконец, учитывая, что  $a = (n-2) \cdot k + 1$  и  $b = (n-2) \cdot (k+1)$ , получаем:  $n-3 \leq 254 \Rightarrow n \leq 257$ .

**Таким образом, количество дисков в массиве RAID-6 не должно превышать 257 в силу специфики используемого математического аппарата.**

## 2. Марковские модели надежности дисковых массивов

В основе модели лежит аппарат цепей Маркова и метод Колмогорова-Чепмена для составления системы дифференциальных уравнений и расчета вероятностей всех состояний системы. Каждое состояние отражает конкретное количество дисков, находящихся в ожидании замены или в процессе восстановления информации после замены. То есть состояние «0» означает, что все диски работают, «1» означает, что вышел из строя один из дисков, причем какой именно – не имеет значения, и так далее. Из всех состояний, кроме последнего состояния «s», возможно восстановление без какой-либо потери данных. Состояние «s» является финальным состоянием разрушения (безвозвратной частичной или полной потери данных), при котором дисковый массив теряет смысл своего существования. Само число  $s$  будем называть порогом разрушения, причем  $1 \leq s \leq n$ , где  $n$  – количество дисков в массиве.

Дисковая система содержит в себе множество из  $n$  независимых самостоятельно невосстанавливаемых дисков с точки зрения сохранности данных. То есть, под отказом отдельно взятого диска понимается такой отказ, который обязательно влечет частичную или полную потерю данных на нем, и потерянные данные могут быть восстановлены только благодаря избыточной информации, хранящейся на остальных дисках, при помощи соответствующего RAID-контроллера, после замены или ремонта отказавшего диска. Соответственно, в качестве среднего времени наработки до разрушения для отдельно взятого диска принимается средняя наработка до первого отказа самого диска, приводящего к частичной или полной потере данных на нем. Диски независимы по отказам. Время отказа рассматривается автором как случайная величина, распределенная по экспоненциальному закону. Интенсивность отказов обозначается как  $\lambda$ .

Однако, практический опыт автор показывает, что хотя интенсивности отказов самих дисков неизменны, в то же время наблюдается следующее явление. В массиве, состоящем из  $n$  дисков, интенсивность перехода от состояния «0» к состоянию «1», в полном согласии с теорией надежности, равна  $n \cdot \lambda$ . Что же касается интенсивностей переходов в последующие состояния «2», «3» и т.п., на практике они нередко оказываются значительно выше, чем теоретические оценки  $(n-1) \cdot \lambda$ ,  $(n-2) \cdot \lambda$  и т.п. То есть, отказ даже одного диска нередко приводит к некоторой деградации дисковой системы, и дальнейшая деградация вплоть до полного разрушения происходит значительно быстрее теоретических оценок. Связано это с повышением вычислительной и обменной нагрузки на RAID-контроллер в «деградированном режиме», когда ему «на лету» необходимо вычислять недостающие данные и, кроме того, параллельно восстанавливать информацию на замененный диск. Для учета этого явления, автором вводится, так называемая, добавочная интенсивность ошибок, обозначаемая  $\varepsilon$ .

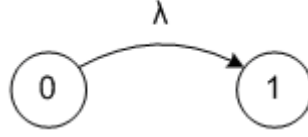
В целом дисковая система благодаря наличию специализированного RAID-контроллера, а также из допущения, что вышедшие из строя, заменяются вручную специалистом по эксплуатации или автоматически при помощи технологии «горячего резерва», считается восстанавливаемой системой. Под временем восстановления по отдельно взятому диску понимается совокупное время, требуемое на замену вышедшего из строя диска и последующего процесса пересчета информации (rebuild), выполняемого RAID-контроллером. Для простоты время восстановления также считается случайной величиной, распределенной по экспоненциальному закону. Интенсивность восстановления обозначается как  $\mu$ . С точки зрения восстановления диски имеют ограничение – одновременно могут восстанавливаться не более  $r$  дисков. При  $r = n$  диски оказываются независимыми с точки зрения восстановления.

Перейдем теперь к математическому описанию полученных автором марковских моделей надежности дисковых массивов и выведенных формул для расчета среднего времени наработки до разрушения с потерей данных.



## 2.1. Марковская модель надежности невосстанавливаемого диска

Рассмотрим марковскую модель надежности отдельного диска, не обладающего никакой отказоустойчивостью:



$$\left\{ \begin{array}{l} P_0(0) = 1; \quad P_1(0) = 0; \\ P_0(t) + P_1(t) = 1 \\ \frac{dP_0(t)}{dt} = -\lambda \cdot P_0(t) \\ \frac{dP_1(t)}{dt} = \lambda \cdot P_0(t) \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} P_0(t) = e^{-\lambda \cdot t} \\ P_1(t) = 1 - e^{-\lambda \cdot t} \end{array} \right.$$

Вероятность отказа **диска** равна вероятности состояния «1» модели:

$$Q_{Disk}(t) = P_1(t) = 1 - e^{-\lambda \cdot t} \quad (2.1.1)$$

Среднее время наработки до отказа невосстанавливаемого **диска**:

$$T_{Disk} = \int_0^{\infty} t \cdot \frac{dQ(t)}{dt} \cdot dt = \frac{1}{\lambda} \quad (2.1.2)$$

## 2.2. Модель надежности невосстанавливаемого дискового массива RAID-0

Рассмотрим марковскую модель надежности дискового массива RAID-0, состоящего из  $n \geq 2$  элементов, не обладающего никакой отказоустойчивостью:



$$\left\{ \begin{array}{l} P_0(0) = 1; \quad P_1(0) = 0; \\ P_0(t) + P_1(t) = 1 \\ \frac{dP_0(t)}{dt} = -n \cdot \lambda \cdot P_0(t) \\ \frac{dP_1(t)}{dt} = n \cdot \lambda \cdot P_0(t) \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} P_0(t) = e^{-n \cdot \lambda \cdot t} \\ P_1(t) = 1 - e^{-n \cdot \lambda \cdot t} \end{array} \right.$$

Вероятность отказа дискового массива **RAID-0** равна вероятности состояния «1» марковской модели:

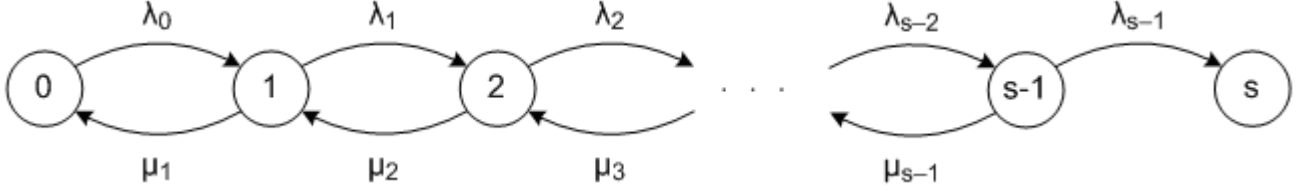
$$Q_{RAID0}(t) = P_1(t) = 1 - e^{-n \cdot \lambda \cdot t} \quad (2.2.1)$$

Среднее время наработки до отказа невосстанавливаемого дискового массива **RAID-0**:

$$T_{RAID0} = \int_0^{\infty} t \cdot \frac{dQ(t)}{dt} \cdot dt = \frac{1}{n \cdot \lambda} \quad (2.2.2)$$

### 2.3. Модель надежности системы из $n$ элементов с порогом разрушения $s \leq n$

Рассмотрим марковскую модель надежности обобщенной системы, состоящей из  $n \geq 2$  элементов с порогом разрушения  $2 \leq s \leq n$ :



Тогда, система дифференциальных уравнений Колмогорова-Чепмена для марковской модели надежности получается следующей:

$$\left\{ \begin{array}{l} P_0(0) = 1; \quad P_1(0) = 0 \quad \dots \quad P_s(0) = 0 \\ P_0(\infty) = 0 \quad \dots \quad P_{s-1}(\infty) = 0; \quad P_s(\infty) = 1 \\ P_0(t) + P_1(t) + \dots + P_s(t) = 1 \\ \frac{dP_0(t)}{dt} = -\lambda_0 \cdot P_0(t) + \mu_1 \cdot P_1(t) \\ \frac{dP_1(t)}{dt} = \lambda_0 \cdot P_0(t) - (\mu_1 + \lambda_1) \cdot P_1(t) + \mu_2 \cdot P_2(t) \\ \frac{dP_2(t)}{dt} = \lambda_1 \cdot P_1(t) - (\mu_2 + \lambda_2) \cdot P_2(t) + \mu_3 \cdot P_3(t) \\ \vdots \\ \frac{dP_{s-1}(t)}{dt} = \lambda_{s-2} \cdot P_{s-2}(t) - (\mu_{s-1} + \lambda_{s-1}) \cdot P_{s-1}(t) \\ \frac{dP_s(t)}{dt} = \lambda_{s-1} \cdot P_{s-1}(t) \end{array} \right.$$

$$Q_{sys}(t) = P_s(t); \quad 2 \leq s \leq n$$

**Среднее время наработки до разрушения** отказоустойчивой системы, состоящей из  $n \geq 2$  элементов, с порогом отказа  $s \leq n$  определяется по формуле, выведенной автором (вывод формулы приведен ниже в примечании):

$$T_{sys} = \int_0^{\infty} t \cdot \frac{dP_s(t)}{dt} \cdot dt = \sum_{k=0}^{s-1} \left( \sum_{q=0}^k \left( \frac{1}{\lambda_q} \cdot \prod_{j=1}^{s-1-k} \left( \frac{\mu_{q+j}}{\lambda_{q+j}} \right) \right) \right) \quad (2.3)$$

Частный случай. Невосстанавливаемая система  $\forall j = 1 \dots s-1: \mu_j = 0$ :

$$T_{sys} = \sum_{q=0}^{s-1} \left( \frac{1}{\lambda_q} \right)$$

Примечание. Найти аналитическое решение системы дифференциальных уравнений в общем случае не представляется возможным, и нахождение среднего времени наработки до разрушения также на первый взгляд затруднительно. Однако, среднее время наработки до разрушения можно также найти через изображение по Лапласу функции  $P_s(t)$ .

Среднее время наработки до разрушения:  $T_{sys} = \int_0^{\infty} t \cdot \frac{dP_s(t)}{dt} \cdot dt$ . Тогда интегрируя по

частям, имеем:  $T_{sys} = t \cdot P_s(t) \Big|_0^{\infty} - \int_0^{\infty} P_s(t) \cdot dt$ . Поскольку  $P_s(0) = 0$  и  $P_s(\infty) = 1$ , то первую

часть интеграла можно переписать следующим образом:  $t \cdot P_s(t) \Big|_0^{\infty} = t \Big|_0^{\infty} = \int_0^{\infty} dt$ . Тогда

имеем:  $T_{sys} = \int_0^{\infty} dt - \int_0^{\infty} P_s(t) \cdot dt = \int_0^{\infty} (1 - P_s(t)) \cdot dt$ . Умножим теперь обе части равенства на

выражение  $e^{-pt}$ , при этом устремив  $p \rightarrow 0$ , при этом также учтем, что  $\lim_{p \rightarrow 0} (e^{-pt}) = 1$ .

Тогда:

$$\lim_{p \rightarrow 0} (e^{-pt} \cdot T_{sys}) = \lim_{p \rightarrow 0} \left( e^{-pt} \cdot \int_0^{\infty} (1 - P_s(t)) \cdot dt \right) = \lim_{p \rightarrow 0} \left( \int_0^{\infty} 1 \cdot e^{-pt} dt - \int_0^{\infty} P_s(t) \cdot e^{-pt} \cdot dt \right),$$

и в итоге получаем:  $T_{sys} = \lim_{p \rightarrow 0} \left( \frac{1}{p} - \int_0^{\infty} P_s(t) \cdot e^{-pt} \cdot dt \right)$ . Заметим, что интеграл в

полученном выражении – это не что иное, как преобразование Лапласа для функции  $P_s(t)$ .

Тогда, обозначив изображение функции  $P_s(t)$  по Лапласу как  $Y_s(p)$ , окончательно получаем:

$$T_{sys} = \lim_{p \rightarrow 0} \left( \frac{1}{p} - Y_s(p) \right)$$

Теперь для нахождения неизвестной  $Y_s(p)$  выполним преобразование по Лапласу

системы дифференциальных уравнений. Учитывая, что  $\frac{dP_i(t)}{dt} = p \cdot Y_i(p) - P_i(0)$ , а также учитывая  $P_0(0) = 1$ ;  $P_1(0) = 0 \dots P_s(0) = 0$ , получаем систему алгебраических уравнений:

$$\left\{ \begin{array}{l} p \cdot Y_0(p) - 1 = -\lambda_0 \cdot Y_0(p) + \mu_1 \cdot Y_1(p) \\ p \cdot Y_1(p) = \lambda_0 \cdot Y_0(p) - (\mu_1 + \lambda_1) \cdot Y_1(p) + \mu_2 \cdot Y_2(p) \\ p \cdot Y_2(p) = \lambda_1 \cdot Y_1(p) - (\mu_2 + \lambda_2) \cdot Y_2(p) + \mu_3 \cdot Y_3(p) \\ \vdots \\ p \cdot Y_{s-1}(p) = \lambda_{s-2} \cdot Y_{s-2}(p) - (\mu_{s-1} + \lambda_{s-1}) \cdot Y_{s-1}(p) \\ p \cdot Y_s(p) = \lambda_{s-1} \cdot Y_{s-1}(p) \end{array} \right.$$



Решая систему можно вывести:

$$Y_s(p) = \frac{\Omega_0}{p \cdot (p^{s-1} + \Omega_{s-2} \cdot p^{s-2} + \dots + \Omega_1 \cdot p + \Omega_0)}$$

Где,  $\Omega_0, \Omega_1, \dots, \Omega_{s-2}$  некоторые коэффициенты.

Тогда, подставляя  $Y_s(p)$  в формулу для среднего времени наработки до разрушения:

$$T_{sys} = \lim_{p \rightarrow 0} \left( \frac{1}{p} - Y_s(p) \right) = \lim_{p \rightarrow 0} \left( \frac{1}{p} - \frac{\Omega_0}{p \cdot (p^{s-1} + \Omega_{s-2} \cdot p^{s-2} + \dots + \Omega_1 \cdot p + \Omega_0)} \right) =$$

$$\lim_{p \rightarrow 0} \left( \frac{p^{s-2} + \Omega_{s-2} \cdot p^{s-3} + \dots + \Omega_1}{p^{s-1} + \Omega_{s-2} \cdot p^{s-2} + \dots + \Omega_1 \cdot p + \Omega_0} \right) = \frac{\Omega_1}{\Omega_0}.$$

Автором были выведены аналитические выражения для коэффициентов  $\Omega_0$  и  $\Omega_1$ :

$$\Omega_0 = \prod_{j=0}^{s-1} \lambda_j \text{ и } \Omega_1 = \sum_{k=0}^{s-1} \left( \sum_{q=0}^k \left( \frac{1}{\lambda_q} \cdot \left( \prod_{j=0}^q \lambda_j \right) \cdot \left( \prod_{j=q+1}^{s-1-k+q} \mu_j \right) \cdot \left( \prod_{j=s-k+q}^{s-1} \lambda_j \right) \right) \right).$$

На основе них была выведена формула для среднего времени наработки до разрушения:

$$T_{sys} = \frac{\Omega_1}{\Omega_0} = \frac{\sum_{k=0}^{s-1} \left( \sum_{q=0}^k \left( \frac{1}{\lambda_q} \cdot \left( \prod_{j=0}^q \lambda_j \right) \cdot \left( \prod_{j=q+1}^{s-1-k+q} \mu_j \right) \cdot \left( \prod_{j=s-k+q}^{s-1} \lambda_j \right) \right) \right)}{\prod_{j=0}^{s-1} \lambda_j} =$$

$$\sum_{k=0}^{s-1} \left( \sum_{q=0}^k \left( \frac{1}{\lambda_q} \cdot \frac{\left( \prod_{j=0}^q \lambda_j \right) \cdot \left( \prod_{j=s-k+q}^{s-1} \lambda_j \right)}{\prod_{j=0}^{s-1} \lambda_j} \cdot \left( \prod_{j=q+1}^{s-1-k+q} \mu_j \right) \right) \right) = \sum_{k=0}^{s-1} \left( \sum_{q=0}^k \left( \frac{1}{\lambda_q} \cdot \frac{\prod_{j=q+1}^{s-1-k+q} \mu_j}{\prod_{j=q+1}^{s-1-k+q} \lambda_j} \right) \right)$$

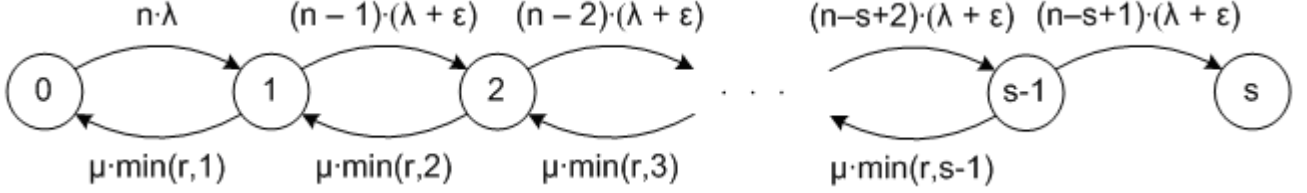
$$= \sum_{k=0}^{s-1} \left( \sum_{q=0}^k \left( \frac{1}{\lambda_q} \cdot \prod_{j=q+1}^{s-1-k+q} \left( \frac{\mu_j}{\lambda_j} \right) \right) \right) = \sum_{k=0}^{s-1} \left( \sum_{q=0}^k \left( \frac{1}{\lambda_q} \cdot \prod_{j=1}^{s-1-k} \left( \frac{\mu_{q+j}}{\lambda_{q+j}} \right) \right) \right).$$

Таким образом, среднее время наработки до разрушения определяется по формуле:

$$T_{sys} = \sum_{k=0}^{s-1} \left( \sum_{q=0}^k \left( \frac{1}{\lambda_q} \cdot \prod_{j=1}^{s-1-k} \left( \frac{\mu_{q+j}}{\lambda_{q+j}} \right) \right) \right).$$

## 2.4. Общая модель надежности массивов RAID из $n \geq 2$ дисков с порогом $s \leq n$

Рассмотрим общую марковскую модель надежности отказоустойчивого дискового массива (RAID) из  $n \geq 2$  дисков с порогом разрушения  $2 \leq s \leq n$ , возможностью одновременного восстановления  $r \geq 0$ , интенсивностью отказов дисков  $\lambda$ , интенсивностью восстановления  $\mu$ , и добавочной интенсивностью ошибок в процессе восстановления  $\varepsilon$ .



Тогда, система дифференциальных уравнений Колмогорова-Чепмена для марковской модели получается следующей:

$$\left\{ \begin{array}{l} P_0(0) = 1; \quad P_1(0) = 0 \quad \dots \quad P_s(0) = 0 \\ P_0(\infty) = 0 \quad \dots \quad P_{s-1}(\infty) = 0; \quad P_s(\infty) = 1 \\ P_0(t) + P_1(t) + \dots + P_s(t) = 1 \\ \frac{dP_0(t)}{dt} = -n \cdot \lambda \cdot P_0(t) + \mu \cdot \min(r, 1) \cdot P_1(t) \\ \frac{dP_1(t)}{dt} = n \cdot \lambda \cdot P_0(t) - (\mu \cdot \min(r, 1) + (n-1) \cdot (\lambda + \varepsilon)) \cdot P_1(t) + \mu \cdot \min(r, 2) \cdot P_2(t) \\ \frac{dP_2(t)}{dt} = (n-1) \cdot (\lambda + \varepsilon) \cdot P_1(t) - (\mu \cdot \min(r, 2) + (n-2) \cdot (\lambda + \varepsilon)) \cdot P_2(t) + \mu \cdot \min(r, 3) \cdot P_3(t) \\ \vdots \\ \frac{dP_{s-1}(t)}{dt} = (n-s+2) \cdot (\lambda + \varepsilon) \cdot P_{s-2}(t) - (\mu \cdot \min(r, s-1) + (n-s+1) \cdot (\lambda + \varepsilon)) \cdot P_{s-1}(t) \\ \frac{dP_s(t)}{dt} = (n-s+1) \cdot (\lambda + \varepsilon) \cdot P_{s-1}(t) \end{array} \right.$$

$$Q_{RAID}(t) = P_s(t); \quad 2 \leq s \leq n; \quad r \geq 0$$

**Среднее время наработки до разрушения** отказоустойчивого дискового массива RAID вычисляется по формуле, выведенной как частный случай общей формулы при  $\lambda_0 = n \cdot \lambda$ ,  $\lambda_i = (n-i) \cdot (\lambda + \varepsilon)$ ;  $i = 1 \dots s-1$  и  $\mu_i = \mu \cdot \min(r, i)$ ;  $i = 1 \dots s-1$ :

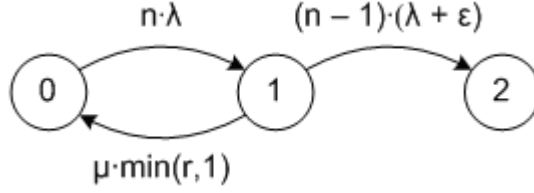
$$T_{RAID} = \sum_{k=0}^{s-1} \left( \frac{\mu^{s-1-k}}{(\lambda + \varepsilon)^{s-1-k}} \cdot \sum_{q=0}^k \left( \frac{1}{(n-q) \cdot (\lambda + \varepsilon \cdot \min(1, q))} \prod_{j=1}^{s-1-k} \left( \frac{\min(r, q+j)}{n-q-j} \right) \right) \right) \quad (2.4)$$

Частный случай. Невосстанавливаемая система  $\mu = 0$  и  $\varepsilon = 0$ :

$$T_{RAID} = \frac{1}{\lambda} \cdot \sum_{q=0}^{s-1} \frac{1}{n-q}$$

## 2.5. Модель надежности массива RAID-5 из $n \geq 3$ дисков с порогом $s = 2$

Рассмотрим модель отказоустойчивого дискового массива RAID-5 ( $n \geq 3$ ) с порогом разрушения  $s = 2$ , возможностью одновременного восстановления  $r \geq 0$ , интенсивностью отказов дисков  $\lambda$ , интенсивностью восстановления  $\mu$ , и добавочной интенсивностью ошибок в процессе восстановления  $\varepsilon$ .



$$\left\{ \begin{array}{l} P_0(0) = 1; \quad P_1(0) = 0; \quad P_2(0) = 0 \\ P_0(t) + P_1(t) + P_2(t) = 1 \\ \frac{dP_0(t)}{dt} = -n \cdot \lambda \cdot P_0(t) + \mu \cdot \min(r,1) \cdot P_1(t) \\ \frac{dP_1(t)}{dt} = n \cdot \lambda \cdot P_0(t) - (\mu \cdot \min(r,1) + (n-1) \cdot (\lambda + \varepsilon)) \cdot P_1(t) \\ \frac{dP_2(t)}{dt} = (n-1) \cdot (\lambda + \varepsilon) \cdot P_1(t) \end{array} \right.$$

$$Q_{RAID5}(t) = P_2(t); \quad n \geq 3; \quad r \geq 0$$

Для отказоустойчивого массива **RAID-5** можно вывести аналитическую формулу для расчета вероятности разрушения (вывод приводится ниже в примечании):

$$Q_{RAID5}(t) = P_2(t) = 1 - e^{-\alpha \cdot t} \cdot \left( ch\left(\sqrt{\alpha^2 - \beta^2} \cdot t\right) + \frac{\alpha}{\sqrt{\alpha^2 - \beta^2}} \cdot sh\left(\sqrt{\alpha^2 - \beta^2} \cdot t\right) \right)$$

$$\alpha = \frac{1}{2} \cdot (\mu \cdot \min(r,1) + (2 \cdot n - 1) \cdot \lambda + (n-1) \cdot \varepsilon); \quad \beta = \sqrt{\lambda \cdot (\lambda + \varepsilon) \cdot n \cdot (n-1)}; \quad \alpha \geq \beta$$

**Среднее время наработки до разрушения** отказоустойчивого массива **RAID-5** определяется по формуле, получаемой из общей формулы для RAID-массивов при  $s = 2$ :

$$T_{RAID5} = \frac{\mu \cdot \min(r,1) + (2 \cdot n - 1) \cdot \lambda + (n-1) \cdot \varepsilon}{\lambda \cdot (\lambda + \varepsilon) \cdot n \cdot (n-1)} \quad (2.5.1)$$

Частный случай 1. Однодисковое и безошибочное восстановление  $r = 1$  и  $\varepsilon = 0$ :

$$T_{RAID5} = \frac{\mu + (2 \cdot n - 1) \cdot \lambda}{\lambda^2 \cdot n \cdot (n-1)}$$

Частный случай 2. Невосстанавливаемая система  $\mu = 0$  и  $\varepsilon = 0$ :

$$T_{RAID5} = \frac{2 \cdot n - 1}{\lambda \cdot n \cdot (n-1)}$$



Примечание. Вывод формулы для вероятности разрушения осуществляется путем решения системы дифференциальных уравнений с использованием преобразования Лапласа.

Учитывая, что  $\frac{dP_i(t)}{dt} = p \cdot Y_i(p) - P_i(0)$ , а также  $P_0(0) = 1$ ;  $P_1(0) = 0$ ;  $P_2(0) = 0$ , из системы дифференциальных уравнений получаем систему алгебраических уравнений:

$$\begin{cases} p \cdot Y_0(p) - 1 = -n \cdot \lambda \cdot Y_0(p) + \mu \cdot \min(r, 1) \cdot Y_1(p) \\ p \cdot Y_1(p) = n \cdot \lambda \cdot Y_0(p) - (\mu \cdot \min(r, 1) + (n-1) \cdot (\lambda + \varepsilon)) \cdot Y_1(p) \\ p \cdot Y_2(p) = (n-1) \cdot (\lambda + \varepsilon) \cdot Y_1(p) \end{cases}$$

Решая систему, находим:

$$Y_2(p) = \frac{n \cdot (n-1) \cdot \lambda \cdot (\lambda + \varepsilon)}{p \cdot (p^2 + p \cdot (\mu \cdot \min(r, 1) + (2 \cdot n - 1) \cdot \lambda + (n-1) \cdot \varepsilon) + n \cdot (n-1) \cdot \lambda \cdot (\lambda + \varepsilon))}$$

Введем теперь следующие обозначения:  $\alpha = \frac{1}{2} \cdot (\mu \cdot \min(r, 1) + (2 \cdot n - 1) \cdot \lambda + (n-1) \cdot \varepsilon)$  и  $\beta = \sqrt{\lambda \cdot (\lambda + \varepsilon) \cdot n \cdot (n-1)}$ , и тогда получаем:

$$Y_2(p) = \frac{\beta^2}{p \cdot (p^2 + 2 \cdot \alpha \cdot p + \beta^2)}.$$

Выполним ряд преобразований полученной формулы:

$$\begin{aligned} Y_2(p) &= \frac{\beta^2}{p \cdot (p^2 + 2 \cdot \alpha \cdot p + \beta^2)} = \\ &= \frac{1}{p} - \frac{p + 2 \cdot \alpha}{p^2 + 2 \cdot \alpha \cdot p + \beta^2} = \frac{1}{p} - \frac{p + 2 \cdot \alpha}{p^2 + 2 \cdot \alpha \cdot p + \alpha^2 - \alpha^2 + \beta^2} = \frac{1}{p} - \frac{p + \alpha + \alpha}{(p + \alpha)^2 - (\alpha^2 - \beta^2)} = \\ &= \frac{1}{p} - \left( \frac{p + \alpha}{(p + \alpha)^2 - (\alpha^2 - \beta^2)} + \frac{\alpha}{\sqrt{\alpha^2 - \beta^2}} \cdot \frac{\sqrt{\alpha^2 - \beta^2}}{(p + \alpha)^2 - (\alpha^2 - \beta^2)} \right). \end{aligned}$$

Теперь, учтем что согласно преобразованию Лапласа:

$$e^{-at} \cdot ch(wt) \Leftrightarrow \frac{p + a}{(p + a)^2 - w^2}, \quad e^{-at} \cdot sh(wt) \Leftrightarrow \frac{w}{(p + a)^2 - w^2} \quad \text{и} \quad 1 \Leftrightarrow \frac{1}{p}.$$

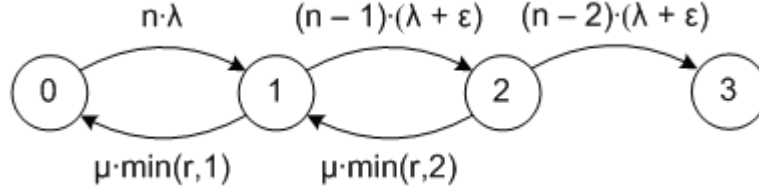
Тогда, мы можем перейти обратно от изображений по Лапласу к функциям времени, и записать окончательную формулу  $P_2(t)$ :

$$P_2(t) = 1 - e^{-\alpha \cdot t} \cdot \left( ch\left(\sqrt{\alpha^2 - \beta^2} \cdot t\right) + \frac{\alpha}{\sqrt{\alpha^2 - \beta^2}} \cdot sh\left(\sqrt{\alpha^2 - \beta^2} \cdot t\right) \right) \quad (2.5.2)$$

$$\alpha = \frac{1}{2} \cdot (\mu \cdot \min(r, 1) + (2 \cdot n - 1) \cdot \lambda + (n-1) \cdot \varepsilon); \quad \beta = \sqrt{\lambda \cdot (\lambda + \varepsilon) \cdot n \cdot (n-1)}; \quad \alpha \geq \beta$$

## 2.6. Модель надежности массива RAID-6 из $n \geq 4$ дисков с порогом $s = 3$

Рассмотрим модель отказоустойчивого дискового массива RAID-6 ( $n \geq 4$ ) с порогом разрушения  $s = 3$ , возможностью одновременного восстановления  $r \geq 0$ , интенсивностью отказов дисков  $\lambda$ , интенсивностью восстановления  $\mu$ , и добавочной интенсивностью ошибок в процессе восстановления  $\varepsilon$ .



$$\left\{ \begin{array}{l} P_0(0) = 1; \quad P_1(0) = 0; \quad P_2(0) = 0; \quad P_3(0) = 0 \\ P_0(t) + P_1(t) + P_2(t) + P_3(t) = 1 \\ \frac{dP_0(t)}{dt} = -n \cdot \lambda \cdot P_0(t) + \mu \cdot \min(r,1) \cdot P_1(t) \\ \frac{dP_1(t)}{dt} = n \cdot \lambda \cdot P_0(t) - (\mu \cdot \min(r,1) + (n-1) \cdot (\lambda + \varepsilon)) \cdot P_1(t) + \mu \cdot \min(r,2) \cdot P_2(t) \\ \frac{dP_2(t)}{dt} = (n-1) \cdot (\lambda + \varepsilon) \cdot P_1(t) - (\mu \cdot \min(r,2) + (n-2) \cdot (\lambda + \varepsilon)) \cdot P_2(t) \\ \frac{dP_3(t)}{dt} = (n-2) \cdot (\lambda + \varepsilon) \cdot P_2(t) \end{array} \right.$$

$$Q_{sys}(t) = P_3(t); \quad n \geq 4; \quad r \geq 0$$

**Среднее время наработки до разрушения** отказоустойчивого массива **RAID-6** определяется по формуле, получаемой из общей формулы для RAID-массивов при  $s = 3$ :

$$T_{RAID6} = \frac{1}{\lambda \cdot (\lambda + \varepsilon)^2 \cdot n \cdot (n-1) \cdot (n-2) + \mu \cdot ((\lambda + \varepsilon) \cdot (n-2) \cdot \min(r,1) + \lambda \cdot n \cdot \min(r,2)) + (\lambda + \varepsilon) \cdot ((\lambda + \varepsilon) \cdot (n^2 - 3 \cdot n + 2) + \lambda \cdot (2 \cdot n^2 - 3 \cdot n))} \cdot (\mu^2 \cdot \min(r,1) \cdot \min(r,2) + \dots) \quad (2.6)$$

Частный случай 1. Однодисковое и безошибочное восстановление  $r = 1$  и  $\varepsilon = 0$ :

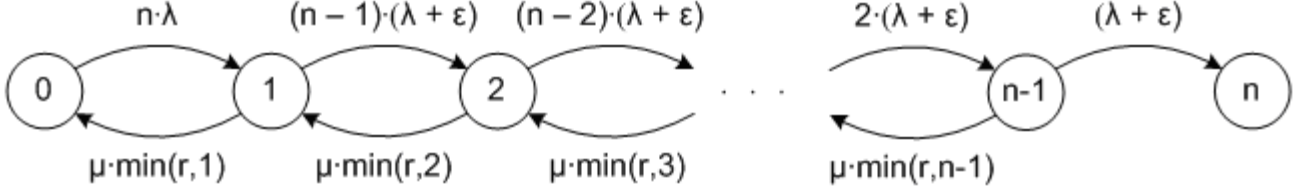
$$T_{RAID6} = \frac{\mu^2 + (2 \cdot n - 2) \cdot \mu \cdot \lambda + (3 \cdot n^2 - 6 \cdot n + 2) \cdot \lambda^2}{\lambda^3 \cdot n \cdot (n-1) \cdot (n-2)}$$

Частный случай 2. Невосстанавливаемая система  $\mu = 0$  и  $\varepsilon = 0$ :

$$T_{RAID6} = \frac{3 \cdot n^2 - 6 \cdot n + 2}{\lambda \cdot n \cdot (n-1) \cdot (n-2)}$$

## 2.7. Модель надежности массива RAID-1 из $n \geq 2$ дисков с порогом $s = n$

Рассмотрим модель отказоустойчивого дискового массива RAID-1 ( $n \geq 2$ ) с порогом разрушения  $s = n$ , возможностью одновременного восстановления  $r \geq 0$ , интенсивностью отказов дисков  $\lambda$ , интенсивностью восстановления  $\mu$ , и добавочной интенсивностью ошибок в процессе восстановления  $\varepsilon$ .



$$\left\{ \begin{array}{l} P_0(0) = 1; \quad P_1(0) = 0 \quad \dots \quad P_s(0) = 0 \\ P_0(\infty) = 0 \quad \dots \quad P_{s-1}(\infty) = 0; \quad P_s(\infty) = 1 \\ P_0(t) + P_1(t) + \dots + P_s(t) = 1 \\ \frac{dP_0(t)}{dt} = -n \cdot \lambda \cdot P_0(t) + \mu \cdot \min(r,1) \cdot P_1(t) \\ \frac{dP_1(t)}{dt} = n \cdot \lambda \cdot P_0(t) - (\mu \cdot \min(r,1) + (n-1) \cdot (\lambda + \varepsilon)) \cdot P_1(t) + \mu \cdot \min(r,2) \cdot P_2(t) \\ \frac{dP_2(t)}{dt} = (n-1) \cdot (\lambda + \varepsilon) \cdot P_1(t) - (\mu \cdot \min(r,2) + (n-2) \cdot (\lambda + \varepsilon)) \cdot P_2(t) + \mu \cdot \min(r,3) \cdot P_3(t) \\ \vdots \\ \frac{dP_{n-1}(t)}{dt} = 2 \cdot (\lambda + \varepsilon) \cdot P_{n-2}(t) - (\mu \cdot \min(r, s-1) + (\lambda + \varepsilon)) \cdot P_{n-1}(t) \\ \frac{dP_n(t)}{dt} = (\lambda + \varepsilon) \cdot P_{n-1}(t) \\ Q_{RAID1}(t) = P_n(t); \quad n \geq 2; \quad r \geq 0; \end{array} \right.$$

**Среднее время наработки до разрушения** отказоустойчивого массива RAID-1 определяется по формуле, получаемой из общей формулы для RAID-массивов при  $s = n$ :

$$T_{RAID1} = \sum_{k=0}^{n-1} \left( \frac{\mu^{n-1-k}}{(\lambda + \varepsilon)^{n-1-k}} \cdot \sum_{q=0}^k \left( \frac{1}{(n-q) \cdot (\lambda + \varepsilon \cdot \min(1, q))} \prod_{j=1}^{n-1-k} \left( \frac{\min(r, q+j)}{n-q-j} \right) \right) \right) \quad (2.7)$$

Частный случай 1. Однодисковое и безошибочное восстановление  $r = 1$  и  $\varepsilon = 0$ :

$$T_{RAID1} = \sum_{k=0}^{n-1} \left( \frac{\mu^{n-1-k}}{\lambda^{n-k}} \cdot \sum_{q=0}^k \left( \frac{1}{n-q} \cdot \prod_{j=1}^{n-1-k} \left( \frac{1}{n-q-j} \right) \right) \right)$$

Частный случай 2. Невосстанавливаемая система  $\mu = 0$  и  $\varepsilon = 0$ :

$$T_{RAID1} = \frac{1}{\lambda} \cdot \sum_{q=0}^{n-1} \frac{1}{n-q}$$

### 3. Примеры расчета среднего времени наработки до разрушения

В данном разделе в табличном виде приводятся результаты расчетов среднего времени (в часах) наработки до разрушения с потерей данных по выведенной автором формуле на примере дисковых массивов RAID-0, 5, 6 и 1 и при количестве дисков  $n = 2 \dots 6$ .

Интенсивность отказов диска для всех таблиц:  $\lambda = 1/120000$ . Соответственно, среднее время наработки до отказа диска составляет 120000 часов.

Интенсивность восстановления массива в первой таблице не учитывается  $\mu = 0$ , а в остальных таблицах  $\mu = 1/24$ , т.е. в среднем 24 часа на восстановление данных на 1 диск.

Максимальное количество одновременно восстанавливаемых дисков в первой таблице не учитывается  $r = 0$ , а в остальных либо  $r = 1$  (строго однодисковое восстановление), либо  $r = n$  (неограниченные возможности по одновременному восстановлению).

Добавочная интенсивность ошибок в первой таблице не учитывается  $\varepsilon = 0$ , в остальных таблицах либо больше, либо равна, либо меньше интенсивности восстановления.

Таблица 1.1. Невосстанавливаемая система:  $r = 0$ ,  $\mu = 0$ ,  $\varepsilon = 0$ :

	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$
RAID-0 ( $s = 1$ )	60000	40000	30000	24000	20000
RAID-5 ( $s = 2$ )	-	100000	70000	54000	44000
RAID-6 ( $s = 3$ )	-	-	130000	94000	74000
RAID-1 ( $s = n$ )	180000	220000	250000	274000	294000

Таблица 1.2. Однодисковое безошибочное восстановление (без учета добавочной интенсивности ошибок при восстановлении):  $r = 1$ ,  $\mu = 1/24$ ,  $\varepsilon = 0$ :

	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$
RAID-0 ( $s = 1$ )	60000	40000	30000	24000	20000
RAID-5 ( $s = 2$ )	-	$1,001 \cdot 10^8$	$5,007 \cdot 10^7$	$3,0054 \cdot 10^7$	$2,0044 \cdot 10^7$
RAID-6 ( $s = 3$ )	-	-	$1,2515 \cdot 10^{11}$	$5,008 \cdot 10^{10}$	$2,505 \cdot 10^{10}$
RAID-1 ( $s = n$ )	$3,0018 \cdot 10^8$	$5,004 \cdot 10^{11}$	$6,25625 \cdot 10^{14}$	$6,25751 \cdot 10^{17}$	$5,21563 \cdot 10^{20}$

Таблица 1.3. Многодисковое безошибочное восстановление (без учета добавочной интенсивности ошибок при восстановлении):  $r = n$ ,  $\mu = 1/24$ ,  $\varepsilon = 0$ :

	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$
RAID-0 ( $s = 1$ )	60000	40000	30000	24000	20000
RAID-5 ( $s = 2$ )	-	$1,001 \cdot 10^8$	$5,007 \cdot 10^7$	$3,0054 \cdot 10^7$	$2,0044 \cdot 10^7$
RAID-6 ( $s = 3$ )	-	-	$2,5025 \cdot 10^{11}$	$1,0013 \cdot 10^{11}$	$5,008 \cdot 10^{10}$
RAID-1 ( $s = n$ )	$3,0018 \cdot 10^8$	$1,0007 \cdot 10^{12}$	$3,75325 \cdot 10^{15}$	$1,50157 \cdot 10^{19}$	$6,25775 \cdot 10^{22}$



Таблица 1.4. Однодисковое восстановление с добавочной интенсивностью ошибок большей, чем интенсивность восстановления:  $r = 1$ ,  $\mu = 1/24$ ,  $\varepsilon = 9/24$ :

	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$
RAID-0 ( $s = 1$ )	60000	40000	30000	24000	20000
RAID-5 ( $s = 2$ )	-	42223	31111	24667	20444
RAID-6 ( $s = 3$ )	-	-	31175	24692	20457
RAID-1 ( $s = n$ )	66669	42473	31184	24698	20463

Таблица 1.5. Многодисковое восстановление с добавочной интенсивностью ошибок большей, чем интенсивность восстановления:  $r = n$ ,  $\mu = 1/24$ ,  $\varepsilon = 9/24$ :

	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$
RAID-0 ( $s = 1$ )	60000	40000	30000	24000	20000
RAID-5 ( $s = 2$ )	-	42223	31111	24667	20444
RAID-6 ( $s = 3$ )	-	-	31236	24717	20470
RAID-1 ( $s = n$ )	66669	42720	31281	24734	20480

Таблица 1.6. Однодисковое восстановление с добавочной интенсивностью ошибок равной интенсивности восстановления:  $r = 1$ ,  $\mu = 1/24$ ,  $\varepsilon = 1/24$ :

	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$
RAID-0 ( $s = 1$ )	60000	40000	30000	24000	20000
RAID-5 ( $s = 2$ )	-	60007	40005	30004	24003
RAID-6 ( $s = 3$ )	-	-	45019	32013	25010
RAID-1 ( $s = n$ )	120011	80035	50056	34070	25745

Таблица 1.7. Многодисковое восстановление с добавочной интенсивностью ошибок равной интенсивности восстановления:  $r = n$ ,  $\mu = 1/24$ ,  $\varepsilon = 1/24$ :

	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$
RAID-0 ( $s = 1$ )	60000	40000	30000	24000	20000
RAID-5 ( $s = 2$ )	-	60007	40005	30004	24003
RAID-6 ( $s = 3$ )	-	-	50021	34015	26011
RAID-1 ( $s = n$ )	120011	100039	80087	64166	52296

Таблица 1.8. Однодисковое восстановление с добавочной интенсивностью ошибок меньшей, чем интенсивность восстановления:  $r = 1$ ,  $\mu = 1/24$ ,  $\varepsilon = 1/216$ :

	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$
RAID-0 ( $s = 1$ )	60000	40000	30000	24000	20000
RAID-5 ( $s = 2$ )	-	219784	119910	77956	55978
RAID-6 ( $s = 3$ )	-	-	523886	239608	136838
RAID-1 ( $s = n$ )	599245	1835152	4153358	7490412	11242880

Таблица 1.9. Многодисковое восстановление с добавочной интенсивностью ошибок меньшей, чем интенсивность восстановления:  $r = n$ ,  $\mu = 1/24$ ,  $\varepsilon = 1/216$ :

	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$
RAID-0 ( $s = 1$ )	60000	40000	30000	24000	20000
RAID-5 ( $s = 2$ )	-	219784	119910	77956	55978
RAID-6 ( $s = 3$ )	-	-	927755	401188	217644
RAID-1 ( $s = n$ )	599245	3450304	22700604	161287651	1200034082

Из всех приведенных таблиц практике автора наиболее соответствуют теоретические оценки среднего времени наработки дискового массива до разрушения с потерей данных, приведенных в **таблице 1.8**. Проанализируем данные этой таблицы:

- **RAID-0** по определению не может рассматриваться в качестве метода повышения отказоустойчивости, и даже наоборот, для любого количества дисков, наработка массива в разы уступает наработке отдельно взятого диска. Такой массив годится только для хранения временных файлов или кэш-хранилищ (Proху cache), где главное – быстродействие, а потеря файлов не имеет никакого значения.
- **RAID-5** является самым дешевым компромиссом между отказоустойчивостью и избыточностью и обладает наработкой большей наработки отдельного диска только при  $n = 3$  дисках, при  $n > 3$  наработка массива уступает наработке отдельного диска. Такой массив может использоваться в системах резервного копирования.
- **RAID-6** является относительно неплохим компромиссом между отказоустойчивостью и избыточностью и обладает наработкой большей наработки отдельного диска при количестве дисков вплоть до  $n = 6$  (при  $n \geq 7$  уступает отдельному диску). Такой массив может использоваться в системах хранения основных данных, для которых периодически выполняется резервное копирование.
- **RAID-1** является самым дорогим компромиссом между отказоустойчивостью и избыточностью и обладает наработкой существенно большей наработки отдельного диска для любого количества дисков, более того, с ростом числа дисков наработка также значительно увеличивается. В силу самой высокой избыточности на практике чаще всего встречаются двухдисковые массивы. Такой массив может использоваться для хранения системных файлов операционной системы, баз данных систем управления ресурсами предприятия и другой информации, к сохранности и доступности которых предъявляются повышенные требования. В то же время, такой массив вовсе не отменяет необходимость выполнения резервного копирования.

## Приложение 1. Конечное поле Галуа $GF(2^8)$ и его арифметика.

Расширенное конечное поле Галуа  $GF(2^8)$  является частным случаем расширенных конечных полей  $GF(2^m)$  характеристики 2 и имеет широкое применение в технологиях помехоустойчивой передачи и хранения информации благодаря тому, что основной единицей информации в вычислительной технике является байт. Байт состоит 8 битов и с помощью него можно представить 256 различных символов, и поле  $GF(2^8)$  также содержит 256 элементов, которые можно представить в виде 8-разрядных двоичных чисел.

Поле Галуа  $GF(2^8)$  содержит 256 элементов, которые могут быть представлены как в виде многочленов, так и в виде соответствующих им двоичных числовых эквивалентов, которые, в свою очередь, всегда могут быть переведены в десятичные числа:

$a(x) :$	0	1	$x$	...	$x^7 + x^6 + x^5 + x^4 + x^3 + x^2 + x + 1$
$GF(2^8) : (a)_2 :$	00000000	00000001	00000010	...	11111111
$(a)_{10} :$	0	1	2	...	255

Поле Галуа  $GF(2^8)$ , по определению являющееся полем многочленов вида  $a(x) = a_7 \cdot x^7 + \dots + a_1 \cdot x + a_0$ , образуется на базе простого поля Галуа  $GF(2)$  и примитивного неприводимого многочлена 8-й степени. В технологиях отказоустойчивого хранения данных (RAID-5, 6) используется примитивный неприводимый многочлен следующего вида:

$$p(x) = x^8 + x^4 + x^3 + x^2 + 1$$

В двоичном представлении примитивный неприводимый многочлен выглядит как:  $(p)_2 = 100011101$ , а в десятичном представлении:  $(p)_{10} = 285$ . Примитивным элементом поля  $GF(2^8)$  является элемент  $\alpha(x) = x$ , и при помощи него можно получить все ненулевые элементы поля. В двоичном представлении примитивный элемент поля выглядит как  $(\alpha)_2 = 10$ , а в десятичном представлении, соответственно, как  $(\alpha)_{10} = 2$ .

Для формирования таблицы степеней примитивного элемента  $(\alpha)_{10} = 2$  используется процедура, являющаяся упрощением процедуры для поля  $GF(2^m)$  при  $m = 8$ . Таблицу логарифмов можно формировать параллельно с формированием таблицы степеней, используя двоичное или десятичное представление степеней примитивного элемента в качестве индексов таблицы логарифмов:

$$\left\{ \begin{array}{l} k = 1 \dots 2^8 - 2; \quad \alpha^0 = 1; \quad \log_{\alpha}(\alpha^0) = 0 \\ \alpha^k = \left\{ \begin{array}{l} \alpha^{k-1} \ll 1, \quad \alpha_7^{(k-1)} = 0 \\ (\alpha^{k-1} \ll 1) \oplus (100011101)_2, \quad \alpha_7^{(k-1)} = 1 \end{array} \right. \\ \log_{\alpha}(\alpha^k) = k \end{array} \right.$$

Под выражением  $\alpha^{k-1} \ll 1$  понимается сдвиг двоичного числа влево на один разряд. Под выражением  $(\alpha^{k-1} \ll 1) \oplus (100011101)_2$  понимается сдвиг двоичного числа влево на один разряд с последующей операцией «побитового» XOR результата сдвига с двоичным эквивалентом примитивного неприводимого многочлена.

Примечание. Поскольку в десятичном виде примитивный элемент поля  $GF(2^8)$  выглядит как  $(\alpha)_{10} = 2$ , то будем также обозначать  $k$ -ую степень примитивного элемента как  $2^k$ , а логарифм от элемента  $a$  по основанию примитивного элемента как  $\log_2 a$ .

Ниже в таблице 1.1 (в виде матрицы 16 x 16) приведены степени примитивного элемента  $(\alpha)_{10} = 2$  поля  $GF(2^8)$ , образованного при помощи примитивного неприводимого многочлена  $p(x) = x^8 + x^4 + x^3 + x^2 + 1$ . Степени примитивного элемента для компактности приведены в десятичном представлении, и расположены построчно (по 16 в строке), начиная с 0-й степени, и заканчивая 255-й. Заметим, что 255-я степень эквивалентна 0-й степени и равна 1 в силу свойств конечных полей Галуа  $GF(p^m)$ :  $\alpha^{(p^m-1)} = \alpha^0 \Rightarrow \alpha^{(2^8-1)} = \alpha^0$ .

Примечание 1. Для того, чтобы выбрать в таблице требуемую степень  $2^k$ , необходимо выбрать строку и столбец таким образом, чтобы сумма индексов строки и столбца (индексы строк приведены в левом заголовочном столбце серого цвета, индексы столбцов – в верхней заголовочной строке серого цвета) была равна показателю степени  $k$ .

**Таблица 1.1. Таблица степеней  $2^k$  для поля Галуа  $GF(2^8)$ .**

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	1	2	4	8	16	32	64	128	29	58	116	232	205	135	19	38
16	76	152	45	90	180	117	234	201	143	3	6	12	24	48	96	192
32	157	39	78	156	37	74	148	53	106	212	181	119	238	193	159	35
48	70	140	5	10	20	40	80	160	93	186	105	210	185	111	222	161
64	95	190	97	194	153	47	94	188	101	202	137	15	30	60	120	240
80	253	231	211	187	107	214	177	127	254	225	223	163	91	182	113	226
96	217	175	67	134	17	34	68	136	13	26	52	104	208	189	103	206
112	129	31	62	124	248	237	199	147	59	118	236	197	151	51	102	204
128	133	23	46	92	184	109	218	169	79	158	33	66	132	21	42	84
144	168	77	154	41	82	164	85	170	73	146	57	114	228	213	183	115
160	230	209	191	99	198	145	63	126	252	229	215	179	123	246	241	255
176	227	219	171	75	150	49	98	196	149	55	110	220	165	87	174	65
192	130	25	50	100	200	141	7	14	28	56	112	224	221	167	83	166
208	81	162	89	178	121	242	249	239	195	155	43	86	172	69	138	9
224	18	36	72	144	61	122	244	245	247	243	251	235	203	139	11	22
240	44	88	176	125	250	233	207	131	27	54	108	216	173	71	142	1

Также ниже в таблице 1.2 (в виде матрицы 16 x 16) приведены логарифмы по основанию примитивного элемента  $(\alpha)_{10} = 2$  поля  $GF(2^8)$ . Логарифмы расположены построчно (по 16 в строке) для всех элементов поля, начиная с  $(a)_{10} = 0$ , заканчивая  $(a)_{10} = 255$ . Заметим, что логарифм от 0 не существует (соответствующая ячейка «N/A»).

**Таблица 1.2. Таблица логарифмов  $\log_2 a$  для поля Галуа  $GF(2^8)$ .**

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	N/A	0	1	25	2	50	26	198	3	223	51	238	27	104	199	75
16	4	100	224	14	52	141	239	129	28	193	105	248	200	8	76	113
32	5	138	101	47	225	36	15	33	53	147	142	218	240	18	130	69
48	29	181	194	125	106	39	249	185	201	154	9	120	77	228	114	166
64	6	191	139	98	102	221	48	253	226	152	37	179	16	145	34	136
80	54	208	148	206	143	150	219	189	241	210	19	92	131	56	70	64
96	30	66	182	163	195	72	126	110	107	58	40	84	250	133	186	61
112	202	94	155	159	10	21	121	43	78	212	229	172	115	243	167	87
128	7	112	192	247	140	128	99	13	103	74	222	237	49	197	254	24
144	227	165	153	119	38	184	180	124	17	68	146	217	35	32	137	46
160	55	63	209	91	149	188	207	205	144	135	151	178	220	252	190	97
176	242	86	211	171	20	42	93	158	132	60	57	83	71	109	65	162
192	31	45	67	216	183	123	164	118	196	23	73	236	127	12	111	246
208	108	161	59	82	41	157	85	170	251	96	134	177	187	204	62	90
224	203	89	95	176	156	169	160	81	11	245	22	235	122	117	44	215
240	79	174	213	233	230	231	173	232	116	214	244	234	168	80	88	175



Примечание 2. Для того, чтобы выбрать в таблице логарифм  $\log_2 a$  заданного элемента  $a$  поля  $GF(2^8)$ , необходимо выбрать строку и столбец таким образом, чтобы сумма индексов строки и столбца (индексы строк приведены в левом заголовочном столбце серого цвета, индексы столбцов – в верхней заголовочной строке серого цвета) была равна десятичному представлению (эквиваленту) элемента  $a$ .

Тогда с учетом всего вышесказанного имеем арифметику поля Галуа  $GF(2^8)$ , образованного с помощью неприводимого многочлена  $p(x) = x^8 + x^4 + x^3 + x^2 + 1$ , при котором примитивным элементом поля является элемент  $(\alpha)_{10} = 2$ :

$$\begin{aligned}
 & \bullet \quad \overbrace{a \pm b}^{GF(2^8)} = \left( \left( a_7 \oplus b_7 \right) \dots \left( a_0 \oplus b_0 \right) \right)_2 = a \oplus b \\
 & \bullet \quad \overbrace{a \cdot b}^{GF(2^8)} = \begin{cases} \overbrace{(\log_2 a + \log_2 b) \bmod (2^8 - 1)}^{< R, \{+, \cdot\} >} & a \neq 0 \ \& \ b \neq 0 \\ 0 & a = 0 \vee b = 0 \end{cases} \\
 & \bullet \quad \overbrace{a / b}^{GF(2^8)} = \begin{cases} \overbrace{(\log_2 a + ((2^8 - 1) - \log_2 b)) \bmod (2^8 - 1)}^{< R, \{+, \cdot\} >} & a \neq 0 \ \& \ b \neq 0 \\ 0 & a = 0 \ \& \ b \neq 0 \\ \text{Ошибка} & b = 0 \end{cases}
 \end{aligned}$$

Кроме того, если необходимо найти обратный элемент по умножению, то можно воспользоваться упрощенным вариантом формулы деления элементов:

$$\bullet \quad \overbrace{a^{-1}}^{GF(2^8)} = \begin{cases} \overbrace{((2^8 - 1) - \log_2 a) \bmod (2^8 - 1)}^{< R, \{+, \cdot\} >} & a \neq 0 \\ \text{Ошибка} & a = 0 \end{cases}$$

Наконец, для возведения в степень  $v$  заданного элемента  $a$  поля  $GF(2^8)$  можно использовать формулу, используя операцию умножения для кольца логарифмов  $\forall u, v \in LR(p^m - 1) \Rightarrow \frac{u \cdot v}{LR(p^m - 1)} = \frac{(u \cdot v) \bmod (p^m - 1)}{< R, \{+, \cdot\} >}$ , и полагая  $u = \log_2 a$ :

$$\bullet \quad \overbrace{a^v}^{GF(2^8)} = \begin{cases} \overbrace{(v \cdot \log_2 a) \bmod (2^8 - 1)}^{< R, \{+, \cdot\} >} & a \neq 0 \ \& \ v \geq 0 \\ 0 & a = 0 \ \& \ v > 0 \\ 1 & a = 0 \ \& \ v = 0 \end{cases}$$

Примечание 3. Отметим, что возведение в степень  $v$  заданного элемента  $a$  поля  $GF(2^8)$  по вышеприведенной формуле с аппаратной точки зрения требует сложного в реализации узла умножения логарифмов (умножения целых неотрицательных чисел по модулю  $2^8 - 1$ ). В действительности, на практике операцию возведения в степень удастся избежать при кодировании / декодировании информации за счет использования специальных подходов, например, использование схемы Горнера для вычисления значений полиномов при подстановке в них в качестве аргумента элемента поля  $GF(2^8)$ .

Примечание 4. Особо отметим, что при использовании десятичного представления элементов поля  $GF(2^8)$ , и элементы и логарифмы выглядят как десятичные числа, и следует соблюдать особую осторожность, и всегда знать, где логарифмы, а где элементы поля.

**Пример 1.** Найдем сумму элементов расширенного поля  $GF(2^8)$ , представленных в виде соответствующих чисел «123» и «231» в десятичной системе счисления. Имеем,

$$\underbrace{(123)_{10} + (231)_{10}}_{GF(2^8)} = \underbrace{(01111011)_2 + (11100111)_2}_{GF(2^8)} = \left\{ \begin{array}{c|c|c|c|c|c|c|c} \oplus 0 & \oplus 1 & \oplus 1 & \oplus 1 & \oplus 1 & \oplus 0 & \oplus 1 & \oplus 1 \\ \hline 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \end{array} \right\} =$$

$$= (10011100)_2 = (156)_{10}. \text{ Таким образом, } \underbrace{(123)_{10} + (231)_{10}}_{GF(2^8)} = (156)_{10}.$$

**Пример 2.** Найдем произведение элементов поля  $GF(2^8)$ , представленных в виде соответствующих чисел «20» и «11» в десятичной системе счисления. По «таблице логарифмов» для поля  $GF(2^8)$  имеем логарифмы элементов:  $\log_2(20)_{10} = 52$  и

$$\log_2(11)_{10} = 238. \text{ Тогда получаем: } \underbrace{(20)_{10} \cdot (11)_{10}}_{GF(2^8)} = 2^{\overbrace{(52 + 238) \bmod(255)}^{< R, \{+, \cdot\} >}} = 2^{35}. \text{ По «таблице}$$

степеней» для поля  $GF(2^8)$  имеем  $2^{35} = (156)_{10}$ . Таким образом:  $\underbrace{(20)_{10} \cdot (11)_{10}}_{GF(2^8)} = (156)_{10}$ .

**Пример 3.** Найдем отношение элементов поля  $GF(2^8)$ , представленных в виде соответствующих чисел «220» и «127» в десятичной системе счисления. По «таблице логарифмов» для поля  $GF(2^8)$  имеем логарифмы элементов:  $\log_2(220)_{10} = 187$  и

$$\log_2(127)_{10} = 87. \text{ Тогда получаем: } \underbrace{(220)_{10} / (127)_{10}}_{GF(2^8)} = 2^{\overbrace{(187 + (255 - 87)) \bmod(255)}^{< R, \{+, \cdot\} >}} = 2^{100}. \text{ По}$$

«таблице степеней» имеем  $2^{100} = (17)_{10}$ . Таким образом:  $\underbrace{(220)_{10} / (127)_{10}}_{GF(2^8)} = (17)_{10}$ .

**Пример 4.** Найдем обратный элемент по умножению для элемента «111», представленного в десятичном виде. По «таблице логарифмов» имеем логарифм элемента:

$$\log_2(111)_{10} = 61. \text{ Тогда обратный элемент: } \underbrace{((111)_{10})^{-1}}_{GF(2^8)} = 2^{\overbrace{(255 - 61) \bmod(255)}^{< R, \{+, \cdot\} >}} = 2^{194}. \text{ По}$$

«таблице степеней» имеем  $2^{194} = (50)_{10}$ . Таким образом:  $\underbrace{((111)_{10})^{-1}}_{GF(2^8)} = (50)_{10}$ .

**Пример 5.** Возведем элемент «13», представленный в десятичном виде, в заданную степень  $\nu = 17$ . По «таблице логарифмов» для поля  $GF(2^8)$  имеем логарифм элемента:

$$\log_2(13)_{10} = 104. \text{ Тогда по формуле получаем: } \underbrace{((13)_{10})^{17}}_{GF(2^8)} = 2^{\overbrace{(17 \cdot 104) \bmod(255)}^{< R, \{+, \cdot\} >}} = 2^{238}. \text{ По}$$

«таблице степеней» имеем  $2^{238} = (11)_{10}$ . Таким образом,  $\underbrace{((13)_{10})^{17}}_{GF(2^8)} = (11)_{10}$ .

## Приложение 2. Пример расчета в математическом пакете Maple 15.

```

restart;
with(student) :

# Mean Time To Data Loss for Fault-Tolerant Disk Systems with  $n \geq 2$  disks and critical threshold  $s \leq n$ ,
# rebuild capability  $r \leq n$ , mean time before failure of disk  $\frac{1}{\lambda}$ , mean time to disk replace and rebuild  $\frac{1}{\mu}$ ,
# and mean time to additional system errors during rebuild process  $\frac{1}{\varepsilon}$ , derived by Pavel A. Rahman, 2012.

MTTDL := (n, s, r, λ, μ, ε) → add( ( μs-1-k / (λ + ε)s-1-k · ( add( 1 / ((n-q) · (λ + ε · min(1, q))) · mul( min(r, q+j) / (n-q-j),
    j = 1..s-1-k ), q = 0..k ) ), k = 0..s-1 ) :

# Example of MTTDL calculations
λ := 1/120000 : μ := 1/24 : ε := 1/216 :

# Single calculation
n := 6 : s := 3 : r := 1 : # RAID-6
trunc( evalf( MTTDL(n, s, r, λ, μ, ε) ) );

136838

# Multiple calculations
nmax := 6 : r := 1 :
T := Matrix(5, nmax) :
T1,1 := 'N':
T2,1 := 'RAID-0': T3,1 := 'RAID-5':
T4,1 := 'RAID-6': T5,1 := 'RAID-1':
for i from 2 to nmax do T1,i := i : end:
for i from 2 to nmax do T2,i := trunc( evalf( MTTDL(i, 1, r, λ, μ, ε) ) ) : end:
for i from 3 to nmax do T3,i := trunc( evalf( MTTDL(i, 2, r, λ, μ, ε) ) ) : end:
for i from 4 to nmax do T4,i := trunc( evalf( MTTDL(i, 3, r, λ, μ, ε) ) ) : end:
for i from 2 to nmax do T5,i := trunc( evalf( MTTDL(i, i, r, λ, μ, ε) ) ) : end:
print(T);


```

N	2	3	4	5	6
RAID-0	60000	40000	30000	24000	20000
RAID-5	0	219784	119910	77956	55978
RAID-6	0	0	523886	239608	136838
RAID-1	599245	1835152	4153358	7490412	11242880

## Список литературы

1. **Черкесов Г.Н.** Надежность аппаратно-программных комплексов. – СПб.: Питер, 2005.
2. **Половко А.М., Гуров С.В.** Основы теории надежности. – 2-е изд., переработанное и дополненное. – СПб.: БХВ-Петербург, 2006.
3. **Гнеденко Б.В.** Курс теории вероятностей. – М.: Едиториал УРСС, 2005.
4. **G.A. Gibson.** Redundant Disk Arrays: Reliable, Parallel Secondary Storage. Ph.D. dissertation, Dept. Computer Science, UC Berkeley, 1991.
5. **Jon G. Elerath.** Reliability model and assessment of redundant arrays of inexpensive disks (RAID) incorporating latent defects and non-homogeneous Poisson process events. Ph.D. dissertation, University of Maryland, 2007.
6. **Jon G. Elerath, Michael Pecht.** A highly accurate method for assessing reliability of redundant arrays of inexpensive disks. IEEE Transactions on Computers, Vol. 58, No. 3, March 2009.
7. **Sam Siewert, Greg Scott.** Next Generation Scalable and Efficient Data Protection. Intel Developer Forum, 2011.

## Содержание

Введение.....	1
1. Обзор дисковых массивов RAID .....	2
1.1. Описание дисковых массивов RAID-0, 1, 5 и 6.....	2
1.2. Особенности реализации технологий RAID-5 и RAID-6.....	4
2. Марковские модели надежности дисковых массивов .....	11
2.1. Марковская модель надежности невосстанавливаемого диска.....	12
2.2. Модель надежности невосстанавливаемого дискового массива RAID-0 .....	12
2.3. Модель надежности системы из $n$ элементов с порогом разрушения $s \leq n$ .....	13
2.4. Общая модель надежности массивов RAID из $n \geq 2$ дисков с порогом $s \leq n$ .....	16
2.5. Модель надежности массива RAID-5 из $n \geq 3$ дисков с порогом $s = 2$ .....	17
2.6. Модель надежности массива RAID-6 из $n \geq 4$ дисков с порогом $s = 3$ .....	19
2.7. Модель надежности массива RAID-1 из $n \geq 2$ дисков с порогом $s = n$ .....	20
3. Примеры расчета среднего времени наработки до разрушения .....	21
Приложение 1. Конечное поле Галуа $GF(2^8)$ и его арифметика. ....	24
Приложение 2. Пример расчета в математическом пакете Maple 15.....	28
Список литературы .....	29
Содержание .....	30