

## **СПЕЦИАЛЬНЫЕ МАРКОВСКИЕ МОДЕЛИ НАДЕЖНОСТИ КАСКАДНЫХ ДИСКОВЫХ МАССИВОВ RAID-01 и RAID-10**

### **ВВЕДЕНИЕ**

Современный мир практически невозможно представить без технических систем, ставших неотъемлемой частью как повседневной жизни и профессиональной деятельности человека, так и ключевыми элементами различных предприятий и отраслей экономики.

Помимо функциональных возможностей и основных технических характеристик, которые в первую очередь интересуют конечных потребителей и производителей, таких как: производительность, мощность, емкость и т.п., не менее важными являются показатели надежности, так как от них напрямую зависит эффективность и безопасность эксплуатации технических систем. Соответственно, разработка моделей и методов для расчета показателей надежности систем является актуальной и критически важной задачей.

На сегодняшний день имеется множество учебников и научных трудов, как отечественных [1, 2], так и зарубежных [3, 4, 5], посвященных основам теории надежности, общим и специализированным моделям и методам расчета показателей надежности систем. Одной из самых распространенных моделей восстанавливаемых технических систем является модель на базе цепей Маркова, позволяющих оценить такие показатели надежности системы как: коэффициент готовности, среднее время наработки на отказ, среднее время восстановления. В случае если система состоит из множества идентичных элементов, то в таком случае часто применяется известная марковская цепь гибели и размножения.

Однако следует особо отметить, что существуют системы, которые при достижении аварийного состояния не могут вернуться в предыдущее работоспособное состояние после замены минимально необходимого одного элемента, и требуют проведения ремонтных работ до исходного полностью исправного состояния. Примером таких систем являются дисковые системы RAID (избыточные массивы недорогих дисков), устойчивые к отказам некоторого количества дисков, и отказывающий вместе с потерей всех данных при отказе большего числа дисков, и требующий пересоздания массива «с нуля» и восстановления данных из резервной копии. Более того, «узким местом» надежности таких систем также является схема управления (контроллер) дискового массива, критические ошибки которого могут приводить систему в аварийное состояние из любого работоспособного состояния. Более того, в последние годы также стали широко распространяться каскадные RAID-массивы, которые объединяют массивы дисков одной RAID-технологии в более крупные массивы другой RAID-технологии, сочетая преимущества обеих технологий.

В последние годы автором был проведен ряд исследований в области анализа показателей надежности современных систем хранения, передачи и обработки информации, состоящих из нескольких идентичных элементов [6, 7, 8], в которых также использовалась традиционная марковская цепь гибели и размножения. В частности, она использовалась для оценки среднего времени наработки до потери данных традиционных дисковых массивов RAID-0, RAID-5, RAID-6 и RAID-1. Однако, при работе с каскадными RAID-массивами, в частности с массивами RAID-10 и RAID-01, традиционная марковская цепь гибели и размножения оказалась неподходящей для оценки показателей надежности, и возникла научная задача разработки специальных марковских моделей надежности.

В рамках статьи рассмотрены предложенные автором специальные марковские модели и выведенные формулы для расчета показателей надежности системы, состоящей из множества идентичных восстанавливаемых элементов. Также рассмотрены модели систем хранения данных на базе каскадных избыточных дисковых массивов RAID-10 и RAID-01 с резервной копией данных, а также примеры расчета показателей надежности.

Следует особо отметить, при разработке моделей был сделан упор на аналитическую разрешимость моделей и выведение расчетных формул, чтобы у специалистов была возможность быстрой оценки показателей надежности систем хранения данных, не прибегая к специальному математическому моделированию и программному обеспечению.

## 1. СПЕЦИАЛЬНАЯ МОДЕЛЬ ОТКАЗОУСТОЙЧИВОЙ СИСТЕМЫ (ТИП 1)

Пусть задана система, состоящая из множества идентичных элементов. Система имеет множество работоспособных состояний  $j = 0 \dots n$  и одно аварийное состояние F.

Интенсивности перехода системы из состояния  $j = 0 \dots n$  в состояние  $j + 1$  (если  $j = n$ , то  $j + 1 = F$ ) будем обозначать  $\lambda_j$ . Интенсивности перехода системы из состояний  $j = 1 \dots n$  в состояние  $j - 1$  будем обозначать  $\mu_j$ .

Кроме того, пусть в системе возможны переходы, которые переводят систему из любого работоспособного состояния  $j = 0 \dots n$  напрямую в аварийное состояние F. Интенсивности перехода из состояний  $j = 0 \dots n$  в аварийное состояние будем обозначать  $\sigma_j$ .

Пусть после перехода в аварийное состояние F всегда требуется восстановление системы в полностью исправное состояние 0. Интенсивность восстановления системы из аварийного состояния будем обозначать  $\gamma$ .

Тогда специальная марковская модель отказоустойчивой системы (рис. 1):

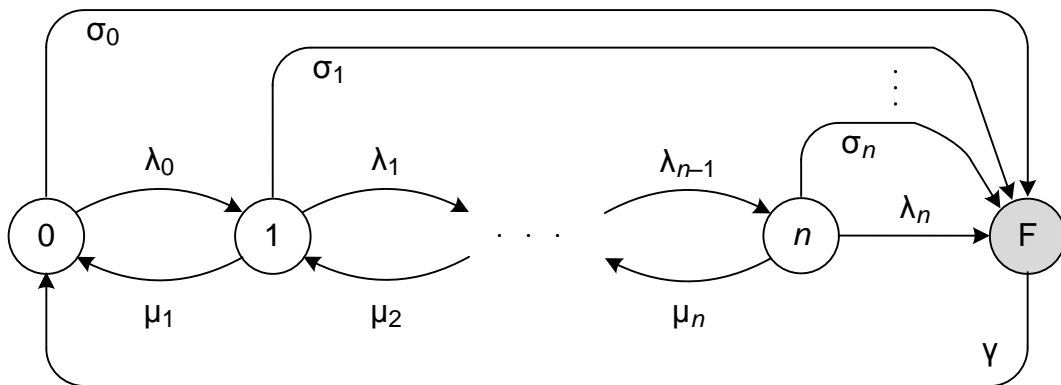


Рис. 1. Граф состояний отказоустойчивой системы (тип 1).

Соответственно, математическая модель (система уравнений Колмогорова-Чепмена) для расчета стационарных вероятностей состояний:

$$\left\{ \begin{array}{l} P_0 + P_1 + \dots + P_n + P_F = 1; \\ -(\lambda_0 + \sigma_0)P_0 + \mu_1 P_1 + \gamma P_F = 0; \\ \lambda_0 P_0 - (\mu_1 + \lambda_1 + \sigma_1)P_1 + \mu_2 P_2 = 0; \\ \vdots \\ \lambda_{n-2} P_{n-2} - (\mu_{n-1} + \lambda_{n-1} + \sigma_{n-1})P_{n-1} + \mu_n P_n = 0; \\ \lambda_{n-1} P_{n-1} - (\mu_n + \lambda_n + \sigma_n)P_n = 0; \\ \sigma_0 P_0 + \dots + \sigma_{n-1} P_{n-1} + (\lambda_n + \sigma_n)P_n - \gamma P_F = 0. \end{array} \right. \quad (1)$$

Тогда, стационарный коэффициент готовности системы:

$$K_{\Gamma} = \sum_{j=0}^n P_j.$$

Далее, учитывая, что система из аварийного состояния может перейти только в начальное состояние с интенсивностью  $\gamma$ , имеем среднее время восстановления системы:

$$T_B = 1/\gamma.$$

Наконец, среднее время наработки на отказ системы может быть определено из тождества  $K_{\Gamma} = T_0 / (T_0 + T_B)$ :

$$T_0 = K_{\Gamma} / (\gamma(1 - K_{\Gamma})).$$

Заметим, что решение системы для нахождения стационарных вероятностей является трудоемкой процедурой, обладающей кубической вычислительной сложностью  $\sim 2(n+2)^3$ .

В результате исследований автору удалось вывести аналитическое решение системы для общего случая в виде матричной формулы. Формула включает в себя произведение  $n + 1$  квадратных матриц размерности 3, содержащих все параметры надежности. Матричная формула имеет линейную вычислительную сложность  $\sim 138(n+1)$ . В итоге получается матрица, содержащая пять коэффициента  $U, V, W, M$  и  $D$ , и два из них,  $M$  и  $D$ , используются для вычисления коэффициента готовности системы и среднего времени наработки на отказ:

$$\Psi = \prod_{j=1}^n \begin{bmatrix} \lambda_{n+1-j} & 0 & 0 \\ 1 & \mu_{n+1-j} & \sigma_{n+1-j} \\ 1 & \mu_{n+1-j} & \lambda_{n+1-j} + \sigma_{n+1-j} \end{bmatrix};$$

$$\begin{bmatrix} U & 0 & 0 \\ V & 0 & W \\ M & 0 & D \end{bmatrix} = \Psi \times \begin{bmatrix} \lambda_0 & 0 & 0 \\ 1 & 0 & \sigma_0 \\ 1 & 0 & \lambda_0 + \sigma_0 \end{bmatrix}; \quad (2)$$

$$K_{\Gamma} = \frac{\gamma M}{\gamma M + D}; \quad T_0 = \frac{M}{D}; \quad T_B = \frac{1}{\gamma}.$$

Теперь заметим, что перемножение матриц можно значительно оптимизировать, учитывая то, что часть ячеек содержат 0 или 1, а также то, что операция умножения матриц обладает свойством ассоциативности  $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$ . Автором был выведена рекуррентная схема вычисления коэффициентов  $M$  и  $D$ , также обладающая линейной вычислительной сложностью  $\sim 36n$ , но вычисляющая значительно быстрее матричной формулы:

$$\left\{ \begin{array}{l} U^{(1)} = \lambda_0; \quad V^{(1)} = 1; \quad M^{(1)} = 1; \\ W^{(1)} = \sigma_0; \quad D^{(1)} = \lambda_0 + \sigma_0; \\ r = 1 \dots n; \\ \left\{ \begin{array}{l} U^{(r+1)} = \lambda_r U^{(r)}; \\ V^{(r+1)} = \sigma_r M^{(r)} + \mu_r V^{(r)} + U^{(r)}; \\ M^{(r+1)} = \lambda_r M^{(r)} + V^{(r+1)}; \\ W^{(r+1)} = \sigma_r D^{(r)} + \mu_r W^{(r)}; \\ D^{(r+1)} = \lambda_r D^{(r)} + W^{(r+1)}; \\ M = M^{(n+1)}; \quad D = D^{(n+1)}; \end{array} \right. \end{array} \right. \quad (3)$$

$$K_{\Gamma} = \frac{\gamma M}{\gamma M + D}; \quad T_0 = \frac{M}{D}; \quad T_B = \frac{1}{\gamma}.$$

## 2. СПЕЦИАЛЬНАЯ МОДЕЛЬ ОТКАЗОУСТОЙЧИВОЙ СИСТЕМЫ (ТИП 2)

Пусть задана система, состоящая из множества идентичных элементов. Система имеет множество работоспособных состояний  $j = 0 \dots n$  и одно аварийное состояние  $F$ .

Интенсивности перехода системы из состояния  $j = 0 \dots n$  в состояние  $j + 1$  (если  $j = n$ , то  $j + 1 = F$ ) будем обозначать  $\lambda_j$ .

Пусть из работоспособных состояний  $j = 1 \dots n$  система может восстанавливаться только в полностью исправное состояние 0. Интенсивности перехода системы из состояния  $j = 1 \dots n$  в состояние 0 будем обозначать  $\mu_j$ .

Кроме того, пусть в системе возможны переходы, которые переводят систему из любого работоспособного состояния  $j = 0 \dots n$  напрямую в аварийное состояние F. Интенсивности перехода из состояний  $j = 0 \dots n$  в аварийное состояние будем обозначать  $\sigma_j$ .

Пусть после перехода в аварийное состояние F всегда требуется восстановление системы в полностью исправное состояние 0. Интенсивность восстановления системы из аварийного состояния будем обозначать  $\gamma$ .

Тогда специальная марковская модель отказоустойчивой системы (рис. 2):

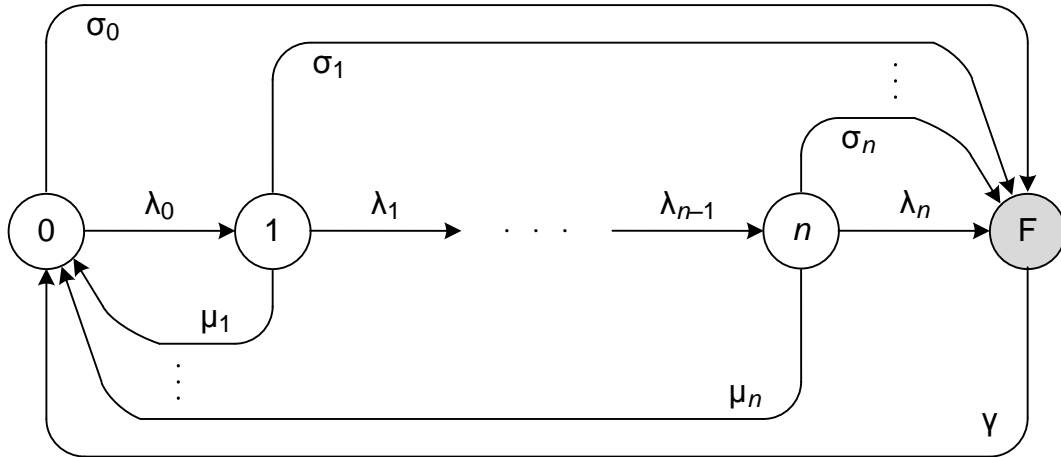


Рис. 2. Граф состояний отказоустойчивой системы (тип 2).

Соответственно, математическая модель (система уравнений Колмогорова-Чепмена) для расчета стационарных вероятностей состояний:

$$\begin{cases} P_0 + P_1 + \dots + P_n + P_F = 1; \\ -(\lambda_0 + \sigma_0)P_0 + \mu_1 P_1 + \dots + \mu_n P_n + \gamma P_F = 0; \\ \lambda_0 P_0 - (\mu_1 + \lambda_1 + \sigma_1)P_1 = 0; \\ \vdots \\ \lambda_{n-1} P_{n-1} - (\mu_n + \lambda_n + \sigma_n)P_n = 0; \\ \sigma_0 P_0 + \dots + \sigma_{n-1} P_{n-1} + (\lambda_n + \sigma_n)P_n - \gamma P_F = 0. \end{cases} \quad (4)$$

Учтем, что стационарный коэффициент готовности системы  $K_\Gamma = \sum_{j=0}^n P_j$ , а также то, что система из аварийного состояния может перейти только в начальное состояние с интенсивностью  $\gamma$ , и, соответственно, среднее время восстановления системы  $T_B = 1/\gamma$ , и, тогда, наконец, среднее время наработки на отказ системы  $T_O = K_\Gamma / (\gamma(1 - K_\Gamma))$ .

В результате исследований автору удалось вывести аналитическое решение системы, и, соответственно, стационарный коэффициент готовности, среднее время наработки на отказ и среднее время восстановления системы определяются следующим образом:

$$\begin{cases} M = \sum_{q=0}^n \left( \frac{1}{\lambda_q} \prod_{j=1}^{n-q} \left( 1 + \frac{\mu_{q+j} + \sigma_{q+j}}{\lambda_{q+j}} \right) \right); \\ D = 1 + \sum_{q=0}^n \left( \frac{\sigma_q}{\lambda_q} \prod_{j=1}^{n-q} \left( 1 + \frac{\mu_{q+j} + \sigma_{q+j}}{\lambda_{q+j}} \right) \right); \\ K_\Gamma = \frac{\gamma M}{\gamma M + D}; \quad T_O = \frac{M}{D}; \quad T_B = \frac{1}{\gamma}. \end{cases} \quad (5)$$

Особо отметим, что в дальнейшем нам понадобится особый случай, когда  $\lambda_n = 0$ , при котором мы сталкиваемся с делением на нуль при вычислении коэффициентов  $M$  и  $D$ , в то же время в дроби  $M/D$  параметр  $\lambda_n = 0$  можно сократить.

Поэтому для случая  $\lambda_n = 0$  автором были выведены отдельные формулы, чтобы избежать деления на нуль при вычислении коэффициентов  $M$  и  $D$ :

$$\begin{cases} M = 1 + (\mu_n + \sigma_n) \sum_{q=0}^{n-1} \left( \frac{1}{\lambda_q} \prod_{j=1}^{n-1-q} \left( 1 + \frac{\mu_{q+j} + \sigma_{q+j}}{\lambda_{q+j}} \right) \right); \\ D = \sigma_n + (\mu_n + \sigma_n) \sum_{q=0}^{n-1} \left( \frac{\sigma_q}{\lambda_q} \prod_{j=1}^{n-1-q} \left( 1 + \frac{\mu_{q+j} + \sigma_{q+j}}{\lambda_{q+j}} \right) \right); \end{cases} \quad (6)$$

$$K_{\Gamma} = \frac{\gamma M}{\gamma M + D}; \quad T_O = \frac{M}{D}; \quad T_B = \frac{1}{\gamma}.$$

### 3. МОДЕЛИ ОТКАЗОУСТОЙЧИВЫХ СИСТЕМ ХРАНЕНИЯ ДАННЫХ НА БАЗЕ КАСКАДНЫХ ДИСКОВЫХ МАССИВОВ RAID-10 И RAID-01

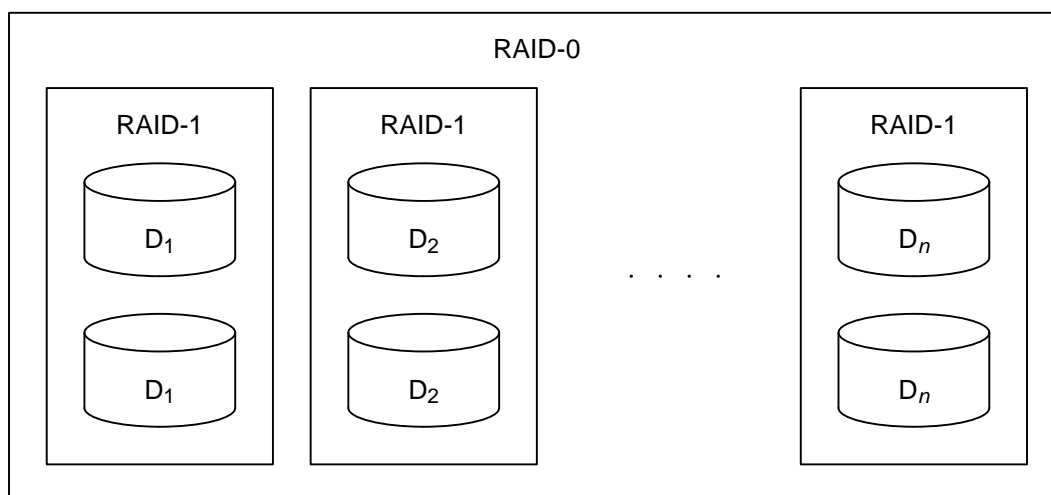
Отказоустойчивые дисковые массивы RAID-10 и RAID-01 являются перспективными системами хранения данных, сочетающими в себе плюсы технологии чередования данных (RAID-0), обеспечивающей высокую скорость чтения данных, и технологии зеркалирования данных (RAID-1), обеспечивающей высокую отказоустойчивость.

Технология RAID-0 позволяет объединять множество из  $n \geq 2$  дисков (массивов), суммируя общее пространство и обеспечивая высокую скорость чтения за счет чередования данных. При операциях записи данные разбиваются на блоки фиксированного размера и записываются в «карусельном» порядке на различные диски (массивы), и при операциях чтения конкретного файла, его блоки часто оказываются на различных дисках (массивах) и считываются гораздо быстрее чем, если бы они были на одном диске. Однако, технология RAID-0 обладает нулевой отказоустойчивостью и повышенным риском потери данных, поскольку при отказе любого из дисков (массивов) все данные теряются, и вероятность отказа в группе дисков (массивов) значительно выше, чем у отдельного диска (массива).

Технология RAID-1 позволяет объединять множество дисков (массивов), обеспечивая работоспособность системы хранения данных до тех пор, пока работоспособен хотя бы один диск (массив) и потеря данных происходит только при отказе всех дисков (массивов) – это обеспечивает максимальную отказоустойчивость. Все диски (массивы) хранят одни и те же данные. При операциях записи данные записываются на все диски одновременно, при чтении – данные считываются с одного из дисков. Соответственно, с ростом числа дисков (массивов) доля полезного дискового пространства обратно пропорционально снижается, и при любом количестве дисков (массивов) полезная емкость всегда равна емкости одного диска (массивов). Поэтому на практике в RAID-1 ограничиваются 2 дисками (массивами).

**RAID-10.** Каскадный массив RAID-10 представляет собою массив RAID-0, объединяющий  $n$  массивов RAID-1 по 2 диска в каждом (рис. 3). Массив RAID-10 строится из четного количества ( $2n$ ) дисков. Доля полезного дискового пространства составляет 50% в силу использования множества двухдисковых массивов RAID-1.

В лучшем случае каскадный массив RAID-10 сохраняет работоспособность при отказах вплоть до  $n$  дисков, если все они оказываются в различных составляющих его массивах RAID-1. В таком случае только отказ  $n+1$ -го диска приведет к неизбежной потере данных, поскольку один из массивов RAID-1 потеряет работоспособность, а RAID-0 сам по себе не обладает никакой отказоустойчивостью. В худшем случае отказ даже двух дисков, оказавшихся в рамках одного массива RAID-1, приводит к потере данных. Соответственно, каскадный массив устойчив к отказу любого одного диска и к некоторым сочетаниям отказов большего числа дисков, но не более  $n$  дисков.



**Рис. 3.** Структурная схема двухуровневого дискового массива RAID-10.

Рассмотрим теперь модель надежности отказоустойчивой системы хранения данных на базе каскадного массива RAID-10. Пусть задана система хранения данных на базе каскадного массива RAID-10, состоящего из  $2n$  идентичных дисков, и резервной копии данных на некотором внешнем хранилище (мы его оставляем за рамками модели).

Интенсивность отказов дисков составляет  $\lambda$ . Диски могут отказывать независимо. Если система находится в состоянии 0 (все диски исправны), то отказ любого из  $2n$  дисков приводит в состояние 1. В состоянии 1 возможен отказ любого из  $2n - 1$  дисков, причем отказ любого из  $2n - 2$  дисков, не затрагивающих RAID-1 массив, в котором содержится ранее отказавший диск, приводит систему в состояние 2, а отказ 1 диска, являющегося единственным исправным в RAID-1 массиве, приводят систему в аварийное состояние F. В состоянии 2 возможен отказ любого из  $2n - 2$  дисков, причем отказ любого из  $2n - 4$  дисков, не затрагивающих RAID-1 массивы, в котором содержатся ранее отказавшие диски, приводит систему в состояние 3, а отказ любого из 2 дисков, являющихся единственными исправными в RAID-1 массивах, приводят систему в аварийное состояние F. И так далее, вплоть до состояния  $n$ . После отказа  $n$  дисков, находящихся в различных массивах RAID-1, система оказывается в состоянии  $n$  и все еще сохраняет работоспособность, но в каждом из RAID-1 массивов содержится один неисправный диск, поэтому отказ любого из оставшихся  $n$  дисков приведет систему в аварийное состояние F.

Массивы RAID-1 независимы по восстановлению, и в них может одновременно проходить регенерация (процедура rebuild) данных на замененных дисках путем простого копирования информации, хранящихся на соответствующих вторых дисках. Интенсивность регенерации информации на замененном диске составляет  $\mu$ . Будем считать, что время замены отказавшего диска несущественно по сравнению со временем регенерации информации за счет использования технологии автоматической горячей замены дисков (hot-spare). После завершения процедуры регенерации очередного массива RAID-1 система переходит из состояния  $j = 1 \dots n$  в состояние  $j - 1$ .

Кроме того, пусть при регенерации данных в массиве RAID-1, на диске, с которого идет копирование информации, помимо базовой интенсивности отказов также становится существенной интенсивность ошибок чтения  $\varepsilon$  диска, и эта интенсивность добавляется к базовой интенсивности отказов диска.

Кроме того, пусть в системе возможны критические виды ошибок RAID-контроллера, которые переводят систему из любого работоспособного состояния  $j = 0 \dots n$  напрямую в аварийное состояние. Интенсивность критических ошибок контроллера составляет  $\sigma$ .

Наконец, пусть в случае наступления аварийного состояния требуется замена всех неисправных дисков, повторная инициализация дискового массива (процедура recreate), и восстановление информации из резервной копии внешнего хранилища. Интенсивность восстановления системы из аварийного состояния составляет  $\gamma$ .

Тогда с учетом всего вышесказанного марковская модель отказоустойчивой системы хранения данных на базе массива RAID-10 выглядит следующим образом (рис. 4):

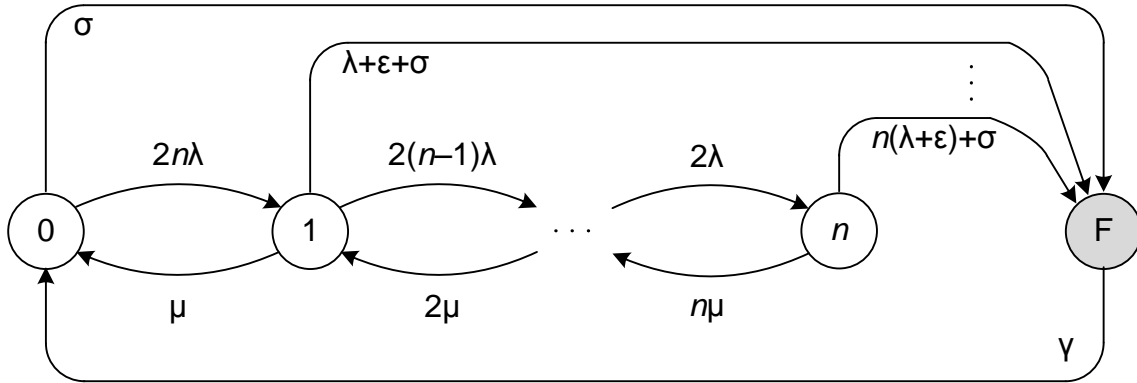


Рис. 4. Граф состояний отказоустойчивой дисковой системы на базе массива RAID-10.

Модель отказоустойчивой дисковой системы получается из рассмотренной выше специальной модели 1-го типа путем следующей подстановки параметров надежности:

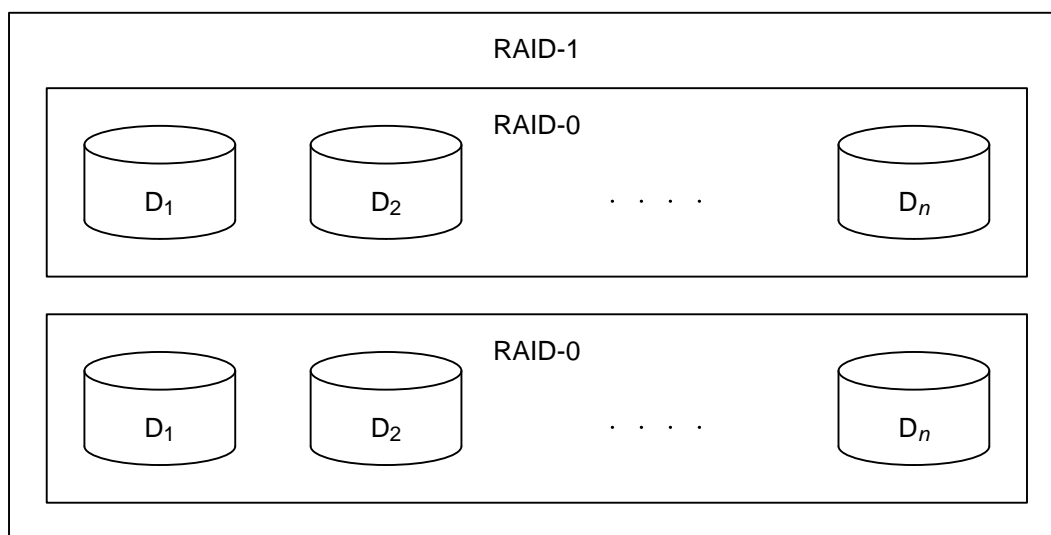
$$\begin{cases} \lambda_0 = 2n\lambda; & \lambda_j = 2(n-j)\lambda; \\ \sigma_0 = \sigma; & \sigma_j = \sigma + j(\lambda + \epsilon); \\ \gamma = \gamma; & \mu_j = j\mu; \\ & j = 1 \dots n. \end{cases} \quad (7)$$

Тогда, расчет показателей надежности отказоустойчивой дисковой системы на базе массива RAID-10 осуществляется по рассмотренной выше рекуррентным формулам 3 с подстановкой соответствующих исходных параметров надежности системы по формулам 7:

$$\begin{cases} U^{(1)} = 2n\lambda; & V^{(1)} = 1; & M^{(1)} = 1; \\ W^{(1)} = \sigma; & D^{(1)} = 2n\lambda + \sigma; \\ & r = 1 \dots n; \\ & U^{(r+1)} = 2(n-r)\lambda U^{(r)}; \\ V^{(r+1)} = (\sigma + r(\lambda + \epsilon))M^{(r)} + r\mu V^{(r)} + U^{(r)}; \\ M^{(r+1)} = 2(n-r)\lambda M^{(r)} + V^{(r+1)}; \\ W^{(r+1)} = (\sigma + r(\lambda + \epsilon))D^{(r)} + r\mu W^{(r)}; \\ D^{(r+1)} = 2(n-r)\lambda D^{(r)} + W^{(r+1)}; \\ M = M^{(n+1)}; & D = D^{(n+1)}; \\ K_{\Gamma} = \frac{\gamma M}{\gamma M + D}; & T_O = \frac{M}{D}; & T_B = \frac{1}{\gamma}. \end{cases} \quad (8)$$

**RAID-01.** Каскадный массив RAID-01 представляет собою массив RAID-1, объединяющий 2 массива RAID-0 по  $n$  дисков в каждом (рис. 5). Массив RAID-01 строится из четного количества ( $2n$ ) дисков. Доля полезного дискового пространства составляет 50% в силу объединения двух массивов RAID-0 в массив RAID-1.

В лучшем случае каскадный массив RAID-01 сохраняет работоспособность при отказах вплоть до  $n$  дисков в случае, если все они оказываются в рамках одного массива RAID-0. В таком случае только отказ  $n+1$ -го диска приведет к неизбежной потере данных, поскольку тогда уже второй массив RAID-0 также окажется разрушенным, поскольку сам по себе он не обладает никакой отказоустойчивостью. В худшем случае отказ даже двух дисков, оказавшихся в разных массивах RAID-0, приводит к потере данных. Соответственно, каскадный массив устойчив к отказу любого одного диска и к некоторым сочетаниям отказов большего числа дисков, но не более  $n$  дисков.



**Рис. 5.** Структурная схема двухуровневого дискового массива RAID-01.

Рассмотрим теперь модель надежности отказоустойчивой системы хранения данных на базе каскадного массива RAID-01. Пусть задана система хранения данных на базе каскадного массива RAID-01, состоящего из  $2n$  идентичных дисков, и резервной копии данных на некотором внешнем хранилище (мы его оставляем за рамками модели).

Интенсивность отказов дисков составляет  $\lambda$ . Диски могут отказывать независимо. Если система находится в состоянии 0 (все диски исправны), то отказ любого из  $2n$  дисков приводит в состояние 1. В состоянии 1 возможен отказ любого из  $2n - 1$  дисков, причем отказ любого из  $n - 1$  дисков, находящихся RAID-0 массиве, в котором содержится ранее отказавший диск, приводит систему в состояние 2, а отказ любого из  $n$  дисков, находящихся в другом RAID-0 массиве, приводят систему в аварийное состояние F. В состоянии 2 возможен отказ любого из  $2n - 2$  дисков, причем отказ любого из  $n - 2$  дисков, находящихся RAID-0 массиве, в котором содержатся ранее отказавшие диски, приводит систему в состояние 3, а отказ любого из  $n$  дисков, находящихся в другом RAID-0 массиве, приводят систему в аварийное состояние F. И так далее, вплоть до состояния  $n$ . После отказа  $n$  дисков, находящихся в рамках одного массива RAID-0, система оказывается в состоянии  $n$  и все еще сохраняет работоспособность, но отказ любого из оставшихся  $n$  дисков (все они находятся в другом массиве RAID-0) приведет систему в аварийное состояние F.

При отказе дисков, находящихся в рамках одного массива RAID-0, массив RAID-1 после замены неисправных дисков запускает регенерацию информации путем простого копирования данных из второго массива RAID-0, причем считываются данные всех  $n$  дисков. Интенсивность регенерация информации на замененном диске составляет  $\mu$ . Будем считать, что время замены отказавшего диска несущественно по сравнению со временем регенерации информации за счет использования технологии автоматической горячей замены дисков (hot-spare). Информация на дисках в рамках массива RAID-0 регенерируются параллельно в рамках единого процесса, и этот процесс завершается одновременно для всех дисков, поэтому будем упрощенно считать, что интенсивность регенерации данных во всем массиве RAID-0 также равна  $\mu$ . В случае если во время регенерации отказывает очередной диск из массива RAID-0, на котором идет регенерации информации, будем упрощенно считать, что процесс регенерации сбрасывается, и он начинается заново после замены неисправного диска. После успешного завершения процедуры регенерации система переходит в исходное полностью исправное состояние 0.

Кроме того, пусть при регенерации информации в массиве RAID-0, на дисках второго массива RAID-0, с которых идет копирование информации, помимо базовой интенсивности отказов также становится существенной интенсивность ошибок чтения  $\varepsilon$  диска, и эта интенсивность добавляется к базовой интенсивности отказов диска.



Кроме того, пусть в системе возможны критические виды ошибок RAID-контроллера, которые переводят систему из любого работоспособного состояния  $j = 0 \dots n$  напрямую в аварийное состояние. Интенсивность критических ошибок контроллера составляет  $\sigma$ .

Наконец, пусть в случае наступления аварийного состояния требуется замена всех неисправных дисков, повторная инициализация дискового массива (процедура `rescreate`), и восстановление информации из резервной копии внешнего хранилища. Интенсивность восстановления системы из аварийного состояния составляет  $\gamma$ .

Тогда с учетом всего вышесказанного марковская модель отказоустойчивой системы хранения данных на базе массива RAID-01 выглядит следующим образом (рис. 6):

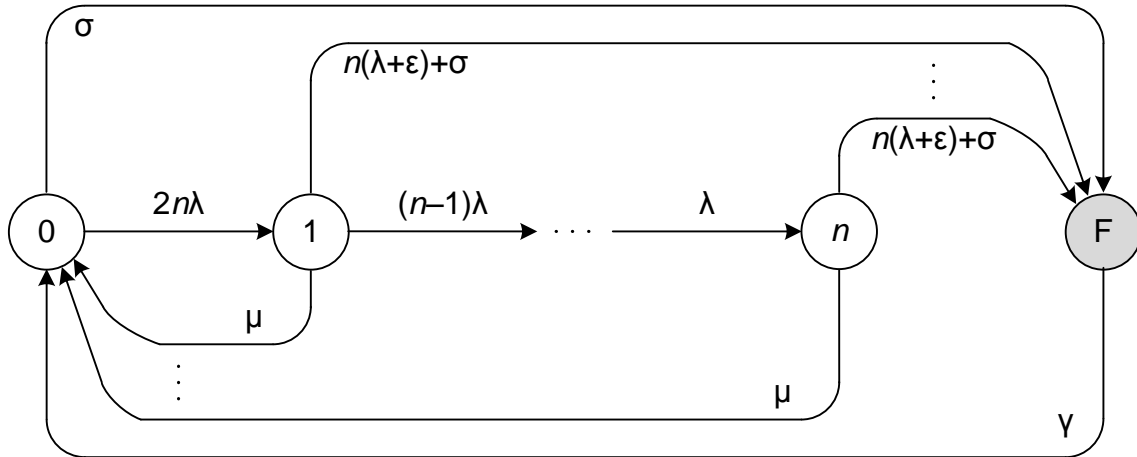


Рис. 6. Граф состояний отказоустойчивой дисковой системы на базе массива RAID-01.

Модель отказоустойчивой дисковой системы получается из рассмотренной выше специальной модели 2-го типа путем следующей подстановки параметров надежности:

$$\begin{cases} \lambda_0 = 2n\lambda; & \lambda_j = (n-j)\lambda; \\ \sigma_0 = \sigma; & \sigma_j = \sigma + n(\lambda + \epsilon); \\ \gamma = \gamma; & \mu_j = \mu; \\ & j = 1 \dots n. \end{cases} \quad (9)$$

Тогда, расчет показателей надежности отказоустойчивой дисковой системы на базе массива RAID-01 осуществляется по рассмотренным выше формулам 6 (учитывая  $\lambda_n = 0$ ), с подстановкой соответствующих исходных параметров надежности системы по формулам 9:

$$\begin{cases} M = 1 + (\mu + \sigma + n(\lambda + \epsilon)) \times \\ \times \sum_{q=0}^{n-1} \left( \frac{1}{((2 - \min(1, q))n - q)\lambda} \prod_{j=1}^{n-1-q} \left( 1 + \frac{\mu + \sigma + n(\lambda + \epsilon)}{(n - q - j)\lambda} \right) \right); \\ D = \sigma + n(\lambda + \epsilon) + (\mu + \sigma + n(\lambda + \epsilon)) \times \\ \times \sum_{q=0}^{n-1} \left( \frac{\sigma + \min(1, q)n(\lambda + \epsilon)}{((2 - \min(1, q))n - q)\lambda} \prod_{j=1}^{n-1-q} \left( 1 + \frac{\mu + \sigma + n(\lambda + \epsilon)}{(n - q - j)\lambda} \right) \right); \\ K_{\Gamma} = \frac{\gamma M}{\gamma M + D}; \quad T_0 = \frac{M}{D}; \quad T_B = \frac{1}{\gamma}. \end{cases} \quad (10)$$

#### 4. ОЦЕНКА ПАРАМЕТРОВ НАДЕЖНОСТИ ДИСКОВ И КОНТРОЛЛЕРА

Для расчета показателей надежности рассмотренной выше отказоустойчивых систем хранения данных на базе каскадных дисковых массивов помимо количества дисков  $2n$  требуются пять параметров надежности  $\lambda, \mu, \varepsilon, \sigma, \gamma$ . Остановимся на них подробнее.

Интенсивность отказов дисков  $\lambda$  нетрудно оценить на основе параметра  $MTBF$  (среднего времени наработки на отказ), предоставленного производителем дисков или полученного из практического опыта эксплуатации. Следует отметить, что производители часто завышают  $MTBF$ , указывая более миллиона часов. Практика же показывает, что  $MTBF$  диска лежит в пределах 50-300 тысяч часов. Соответственно, интенсивность отказов:

$$\lambda = 1 / MTBF_{\text{disk}}. \quad (11)$$

Интенсивность регенерации данных  $\mu$  для массивов RAID-1 зависит от емкости диска  $V$  (в байтах), средней скорости записи  $v_{\text{write}}$  на диск (в байт/с) и средней скорости чтения  $v_{\text{read}}$  данных (в байт/с):

$$\mu = \frac{3600 v_{\text{read}} v_{\text{write}}}{V(v_{\text{read}} + v_{\text{write}})}. \quad (12)$$

Например, для диска емкости  $10^{12}$  байтов, скорости записи  $v_{\text{write}} = 50 \cdot 10^6$  байт/с и скорости чтения  $v_{\text{read}} = 80 \cdot 10^6$  байт/с, интенсивность регенерации составит  $\mu \sim 1/9$  час<sup>-1</sup> (в среднем регенерация длится 9 часов).

Интенсивность ошибок чтения  $\varepsilon$  диска можно определить на основе параметра  $P_{\text{UER}}$  (вероятности невосстанавливаемой ошибки чтения бита), предоставленного производителем дисков или полученного из практического опыта эксплуатации, емкости диска  $V$  (в байтах) и среднего времени регенерации информации, равного  $1/\mu$  (в часах). Для дисков персональных компьютеров параметр  $P_{\text{UER}}$  составляет  $\sim 10^{-14}$ , для дисков серверных систем  $\sim 10^{-15}$ .

Тогда, учитывая, что при регенерации данных требуется считывать весь диск размером  $8V$  битов, то вероятность ошибки чтения  $Q = 1 - (1 - P_{\text{UER}})^{8V}$ . С другой стороны полагая, что время наработки на ошибку – экспоненциально распределенная случайная величина с параметром  $\varepsilon$ , и регенерация длится в течение  $1/\mu$  часов, имеем равенство  $Q = 1 - e^{-\varepsilon/\mu}$ . Тогда, из двух тождеств получаем  $\varepsilon = -8V\mu \ln(1 - P_{\text{UER}})$ . Тогда, учитывая, что  $P_{\text{UER}}$  очень малая величина, и  $\ln(1 - P_{\text{UER}}) \sim -P_{\text{UER}}$ , окончательно получаем:

$$\varepsilon = 8V\mu P_{\text{UER}}. \quad (13)$$

Например, для диска емкости  $V = 10^{12}$  байтов, интенсивности регенерации данных  $\mu = 1/24$  час<sup>-1</sup> и вероятности невосстанавливаемой ошибки чтения бита  $P_{\text{UER}} = 10^{-14}$ , интенсивность ошибок чтения составит  $\varepsilon = 1/300$  час<sup>-1</sup>.

Интенсивность ошибок контроллера  $\sigma$  можно оценить на основе параметра  $MTTE$  (среднее время наработки на ошибку), предоставленного производителем контроллера дисков или полученного из практического опыта эксплуатации. Практика же показывает, что  $MTTE$  контроллера обычно составляют миллионы часов. Тогда, интенсивность ошибок:

$$\sigma = 1 / MTTE_{\text{con}}. \quad (14)$$

Наконец, интенсивность полного восстановления системы  $\gamma$  из аварийного состояния с восстановлением данных из резервной копии зависит от времени, требуемого для создания дискового массива, размера резервной копии, скорости записи данных в дисковый массив, и может быть оценена на основе среднего времени  $MTTR$ , полученного из практики (занимает от нескольких часов до нескольких суток):

$$\gamma = 1 / MTTR_{\text{sys}}. \quad (15)$$

## 5. ПРИМЕР РАСЧЕТА ПОКАЗАТЕЛЕЙ НАДЕЖНОСТИ ДИСКОВЫХ МАССИВОВ

Имеется  $2n$  одинаковых дисков емкости  $V = 10^{12}$  байтов. Среднее время наработки на отказ диска составляет  $MTBF_{\text{disk}} = 120000$  часов. Вероятность невозстанавливаемой ошибки чтения бита  $P_{\text{UER}} = 10^{-14}$ . Средняя скорость чтения данных  $v_{\text{read}} = 80 \cdot 10^6$  байт/с. Средняя скорость записи данных  $v_{\text{write}} = 50 \cdot 10^6$  байт/с.

Имеется дисковый контроллер, поддерживающий каскадные дисковые массивы RAID-10 и RAID-01. Среднее время наработки на критическую ошибку контроллера составляет  $MTTE_{\text{con}} = 1200000$  часов.

При отказе двух или более дисков так, что среди отказавшихся дисков есть диски, оказавшиеся в одном RAID-1 массиве каскадного массива RAID-10 или в разных массивах RAID-0 каскадного массива RAID-01, или критическом отказе RAID-контроллера система переходит в аварийное состояние с потерей всех данных. Для предотвращения безвозвратной потери данных (в аварийном случае) выполняется периодическое резервное копирование данных на внешнее хранилище. Среднее время полного восстановления дискового массива, включая восстановление данных из резервной копии, составляет  $MTTR_{\text{sys}} = 72$  часа.

Вычислить и сравнить коэффициенты готовности и средние времена наработки на отказ систем хранения данных на базе каскадных массивов RAID-10 и RAID-01 с резервной копией данных с количеством дисков  $2n = 4, 6, 8, 10, 12, 14, 16$ .

Решение. Оценим параметры надежности  $\lambda, \mu, \varepsilon, \sigma, \gamma$ , необходимые для расчета показателей надежности отказоустойчивых систем хранения данных.

Интенсивность отказов диска:

$$\lambda = 1/MTBF_{\text{disk}} = 1/120000 \text{ час}^{-1}.$$

Интенсивность регенерации информации в массивах RAID-1:

$$\mu = \frac{3600 v_{\text{read}} v_{\text{write}}}{V(v_{\text{read}} + v_{\text{write}})} \sim 1/9 \text{ час}^{-1}.$$

Интенсивность ошибок чтения данных диска:

$$\varepsilon = 8V\mu P_{\text{UER}} \sim 1/300 \text{ час}^{-1}.$$

Интенсивность критических ошибок контроллера:

$$\sigma = 1/MTTE_{\text{con}} = 1/1200000 \text{ час}^{-1}.$$

Наконец, интенсивность полного восстановления системы из аварийного состояния с восстановлением данных из резервной копии:

$$\gamma = 1/MTTR_{\text{sys}} = 1/72 \text{ час}^{-1}.$$

Теперь, имея все исходные параметры надежности и используя формулы 8 и 10 для отказоустойчивых систем хранения данных на базе каскадных дисковых массивов RAID-10 и RAID-01, вычисляем коэффициент готовности и среднее время наработки на отказ.

Таблица 1

**Коэффициент готовности системы хранения данных**

	$2n = 4$	$2n = 6$	$2n = 8$	$2n = 10$	$2n = 12$	$2n = 14$	$2n = 16$
RAID-10	0,9998699	0,9998349	0,9997999	0,9997649	0,9997299	0,9996949	0,9996599
RAID-01	0,9998039	0,9996423	0,9994252	0,9991569	0,9988416	0,9984827	0,9980837

Таблица 2

**Среднее время наработки на отказ системы хранения данных (в часах)**

	$2n = 4$	$2n = 6$	$2n = 8$	$2n = 10$	$2n = 12$	$2n = 14$	$2n = 16$
RAID-10	553584	436120	359780	306185	266487	235902	211615
RAID-01	367103	201227	125183	85328	62080	47381	37500

Нетрудно заметить, что с ростом количества дисков показатели надежности обоих каскадных массивов снижаются. Кроме того, при равных количествах дисков показатели надежности массива RAID-01 ниже, чем у массива RAID-10, и с ростом количества дисков разрыв в показателях быстро увеличивается. Учитывая, что у обоих каскадных массивов доля полезного дискового пространства составляет 50%, то очевидно, что массив RAID-10 по показателям надежности гораздо более предпочтителен, нежели чем RAID-01.

## ЗАКЛЮЧЕНИЕ

Таким образом, в рамках статьи рассмотрены предложенные автором специальные марковские модели и выведенные формулы для расчета показателей надежности системы, состоящей из множества идентичных восстанавливаемых элементов.

Также рассмотрены модели отказоустойчивых систем хранения данных на базе каскадных избыточных дисковых массивов RAID-10 и RAID-01 с резервной копией данных, а также примеры расчета показателей надежности.

Полученные теоретические результаты использовались в многолетней практике проектирования и эксплуатации систем хранения, обработки и передачи данных НИУ МЭИ (ТУ), Балаковской АЭС, ОАО «Красный Пролетарий» и ряда других предприятий.

## СПИСОК ЛИТЕРАТУРЫ

1. **Черкесов Г. Н.** Надежность аппаратно-программных комплексов. СПб.: Питер, 2005. [ G. N. Cherkesov, *Reliability of Hardware and Software Systems*, (in Russian). Saint-Petersburg: Piter, 2005. ]
2. **Половко А. М., Гуров С. В.** Основы теории надежности. 2-е изд. СПб.: БХВ-Петербург, 2006. [ A. M. Polovko and S. V. Gurov, *Basis of Reliability Theory*, (in Russian). Saint-Petersburg: BHV-Petersburg, 2006. ]
3. **Martin L. Shooman.** Reliability of computer systems and networks. John Wiley & Sons Inc., 2002. [ Martin L. Shooman, *Reliability of computer systems and networks*, John Wiley & Sons Inc., 2002. ]
4. **Israel Koren, C. Mani Krishna.** Fault-Tolerant Systems. Morgan Kaufmann Publishers Inc., 2007. [ Israel Koren, C. Mani Krishna, *Fault-Tolerant Systems*, Morgan Kaufmann Publishers Inc., 2007. ]
5. **Elerath J. G.** 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. [ Elerath J. G., *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. **Каяшев А.И., Рахман П.А., Шарипов М.И.** Анализ показателей надежности избыточных дисковых массивов // Вестник УГАТУ. 2013. Т. 17, № 2 (55). С. 163-170. [ A.I. Kayashev, P.A. Rahman, M.I. Sharipov, "Reliability analysis of redundant disk arrays," (in Russian), *Vestnik UGATU*, vol. 17, no. 2 (55), pp. 163-170, 2013. ]
7. **Каяшев А.И., Рахман П.А., Шарипов М.И.** Анализ показателей надежности локальных компьютерных сетей // Вестник УГАТУ. 2013. Т. 17, № 5 (58). С. 140-149. [ A.I. Kayashev, P.A. Rahman, M.I. Sharipov, "Reliability analysis of local area networks," (in Russian), *Vestnik UGATU*, vol. 17, no. 5 (58), pp. 140-149, 2013. ]
8. **Каяшев А.И., Рахман П.А., Шарипов М.И.** Анализ показателей надежности двухуровневых магистральных сетей // Вестник УГАТУ. 2014. Т. 18, № 2 (63). С. 197-207. [ A.I. Kayashev, P.A. Rahman, M.I. Sharipov, "Reliability analysis of two-level backbone networks," (in Russian), *Vestnik UGATU*, vol. 18, no. 2 (63), pp. 197-207, 2014. ]

**Приложение 1. Расчет показателей надежности систем хранения данных на базе каскадного массива RAID-10 по модели 1 в математическом пакете Maple.**

```

> restart; unprotect(gamma);
> with(student):
> # Storage System Markov Model Type 1 (serial repairs)
  # for RAID-10 analytic formulas Pavel A. Rahman, 2014.
> # Analytic formulas via recurrent algorithm by Pavel A. Rahman

# Mean Time To Fatal Failure
MTTFF:= proc(n,lambda,epsilon,mu,sigma) local i,r,res,U,V,M,W,D:
  if (n >= 1) then
    U:= 2 * n * lambda: V:= 1: M:= 1:
    W:= sigma: D:= 2 * n * lambda + sigma:
    for r from 1 to n do
      V:= (sigma + r * (lambda + epsilon)) * M + r * mu * V + U:
      U:= (2 * (n - r) * lambda) * U:
      M:= (2 * (n - r) * lambda) * M + V:
      W:= (sigma + r * (lambda + epsilon)) * D + r * mu * W:
      D:= (2 * (n - r) * lambda) * D + W:
    end do:
    res:= M / D:
    return(res):
  end if:
end proc:

# Mean Time To Full Repair
MTTFR:= (n,gamma) -> 1/gamma:

# Stationary availability factor
KS:= (n,lambda,epsilon,mu,sigma,gamma) ->
  MTTFF(n,lambda,epsilon,mu,sigma)
  / (MTTFF(n,lambda,epsilon,mu,sigma)+MTTFR(n,gamma)) :

> # Multiple calculation examples
> nmax:= 8: # Max number of RAID-1 pairs
DISKMTBF:= 120000: DISKMTTRE:= 300: MIRRORMTTR:= 9:
CTRLMTTE:= 1200000: SYSMTTR:= 72:
> lambda:= 1/DISKMTBF: epsilon:= 1/DISKMTTRE: mu:= 1/MIRRORMTTR:
sigma:= 1/CTRLMTTE: gamma:= 1/SYSMTTR:

> TA:= Matrix(2,nmax,fill=`--`):
TA[1,1] := `2N`: TA[2,1]:= `RAID-10`:
for i from 2 to nmax do TA[1,i]:= 2*i end do:
for i from 2 to nmax do TA[2,i]:=
  trunc(evalf(MTTFF(i,lambda,epsilon,mu,sigma))): end do:
print(TA):

      2N      4      6      8      10      12      14      16
RAID-10 553584 436120 359780 306185 266487 235902 211615

> KA:= Matrix(2,nmax,fill=`--`):
KA[1,1] := `2N`: KA[2,1]:= `RAID-10`:
for i from 2 to nmax do KA[1,i]:= 2*i end do:
for i from 2 to nmax do KA[2,i]:=
  evalf(KS(i,lambda,epsilon,mu,sigma,gamma)): end do:
print(KA):

      2N      4      6      8      10      12      14      16
RAID-10 0.9998699554 0.9998349354 0.9997999181 0.9997649036 0.9997298919 0.9996948829 0.9996598766

```

**Приложение 2. Расчет показателей надежности систем хранения данных на базе каскадного массива RAID-01 по модели 2 в математическом пакете Maple.**

```

> restart; unprotect(gamma);
> with(student):
> # Storage System Markov Model Type 2 (entire repairs)
  # for RAID-01 analytic formulas Pavel A. Rahman, 2014.
> # Analytic formulas via recurrent algorithm by Pavel A. Rahman

# Mean Time To Fatal Failure
MTTFF:= (n,lambda,epsilon,mu,sigma) -> (1+(mu+(sigma+n*(lambda+epsilon)))
  *add((1/((n*(2-min(1,q))-q)*lambda))*mul(1+((sigma+n*(lambda+epsilon))+mu)
  /((n-q-j)*lambda),j=1..n-1-q),q=0..n-1)/((sigma+n*(lambda+epsilon))
  +(mu+(sigma+n*(lambda+epsilon))))*add(((sigma+n*(lambda+epsilon))*min(1,q)
  /((n*(2-min(1,q))-q)*lambda))*mul(1+((sigma+n*(lambda+epsilon))+mu)/((n-q-j)
  *lambda),j=1..n-1-q),q=0..n-1)):

# Mean Time To Full Repair
MTTFR:= (n,gamma) -> 1/gamma:

# Stationary availability factor
KS:= (n,lambda,epsilon,mu,sigma,gamma) ->
  MTTFF(n,lambda,epsilon,mu,sigma)
  / (MTTFF(n,lambda,epsilon,mu,sigma)+MTTFR(n,gamma)):

> # Multiple calculation examples
> nmax:= 8: # Max number of disks in RAID-0
DISKMTBF:= 120000: DISKMTTRE:= 300: MIRRORMTTR:= 9:
CTRLMTTE:= 1200000: SYSMTTR:= 72:
> lambda:= 1/DISKMTBF: epsilon:= 1/DISKMTTRE: mu:= 1/MIRRORMTTR:
sigma:= 1/CTRLMTTE: gamma:= 1/SYSMTTR:

> TA:= Matrix(2,nmax,fill=`--`):
TA[1,1] := `2N`: TA[2,1]:= `RAID-01`:
for i from 2 to nmax do TA[1,i]:= 2*i end do:
for i from 2 to nmax do TA[2,i]:=
  trunc(evalf(MTTFF(i,lambda,epsilon,mu,sigma))): end do:
print(TA):

      2N      4      6      8      10      12      14      16
RAID-01 367103 201227 125183 85328 62080 47381 37500

> KA:= Matrix(2,nmax,fill=`--`):
KA[1,1] := `2N`: KA[2,1]:= `RAID-01`:
for i from 2 to nmax do KA[1,i]:= 2*i end do:
for i from 2 to nmax do KA[2,i]:=
  evalf(KS(i,lambda,epsilon,mu,sigma,gamma)): end do:
print(KA):

      2N      4      6      8      10      12      14      16
RAID-01 0.9998039085 0.9996423236 0.9994251766 0.9991569180 0.9988415579 0.9984827187 0.9980836809

```