

Устройства хранения данных

Луцив Дмитрий Вадимович

Кафедра системного программирования СПбГУ



1 Контроль ошибок (на экзамен не выносится)

2 Энергонезависимая память

- Природа носителя данных
- Диски: жёсткие и не очень
- Оптические носители
- Твердотельные накопители

3 Redundant Array of Inexpensive Disks

Контроль ошибок (на экзамен не выносится)

Код обнаружения (контрольная сумма) — избыточные данные, которые позволяют *не достоверно* контролировать корректность основных

Вычислительные:

- Сумма всех битов по модулю 2
- CRC и другие легко аппаратно реализуемые контрольные суммы

Из «повседневной жизни»:

- Последние символы в ISSN и номерах банковских карт

Код коррекции — избыточные данные, которые позволяют *не достоверно* обнаруживать и исправлять ошибки в основных

- Обнаруживают ошибки в N битах
- Исправляют в $M \leq N$ битах

Пример — **код Хэмминга**  , обнаруживает не более 2 и исправляет не более 1 ошибки.

$$\begin{aligned} D_0 &= \{2^n | n \in [1, 16)\} \\ f_0(x) &= \log_2 x, f_0 : D_0 \leftrightarrow R_0, \\ R_0 &= \{f_0(x) | x \in D_0\} \end{aligned}$$

$$D_0 = \{2^n | n \in [1, 16)\}, f_0(x) = \log_2 x, f_0 : D_0 \leftrightarrow R_0,$$

$$R_0 = \{f_0(x) | x \in D_0\}$$

Вводятся бинарные (групповые) операции \oplus_D, \oplus_R

❶ D – замыкание D_0 относительно \oplus_D

❷ $f : D \rightarrow R (R = R_0), f(x) = f_0(x) \forall x \in D_0$ – расширение f_0

❸ f строится, как гомоморфизм: $f(x_1 \oplus_D x_2) = f(x_1) \oplus_R f(x_2)$

Код Хемминга для исправления 1 ошибки: пример

Отправили: $x_1; f(x_1)$

Пришло: $x'_1 = x_1 \oplus x', x' \in D_0; f_{x_1} = f(x_1)$, причём x' мы не знаем.

Тогда $s = f_{x_1} \oplus f(x'_1) = f(x_1) \oplus f(x_1) \oplus f(x') = f(x')$.

Но $x' = f^{-1}(s)$ – нашли ошибочный бит.

Код Хемминга для обнаружения 2 ошибок

Практика Послали: $x_1; f(x_1)$

Пришло: $x'_1 = x_1 \oplus x'_a \oplus x'_b$, $x'_a, x'_b \in D_0$; $f_{x_1} = f(x_1)$, причём x'_a, x'_b мы не знаем.

Но $\forall x'_a \neq x'_b$ справедливо $f(x'_a \oplus x'_b) \neq 0$,

а значит $f(x_1 \oplus x'_a \oplus x'_b) \neq f(x_1)$,

т.е. ошибка будет обнаружена, хотя сами x'_a, x'_b мы так и не узнаем.

Код Хемминга для обнаружения 2 ошибок

Практика Послали: $x_1; f(x_1)$

Пришло: $x'_1 = x_1 \oplus x'_a \oplus x'_b$, $x'_a, x'_b \in D_0$; $f_{x_1} = f(x_1)$, причём x'_a, x'_b мы не знаем.

Но $\forall x'_a \neq x'_b$ справедливо $f(x'_a \oplus x'_b) \neq 0$,

а значит $f(x_1 \oplus x'_a \oplus x'_b) \neq f(x_1)$,

т.е. ошибка будет обнаружена, хотя сами x'_a, x'_b мы так и не узнаем.

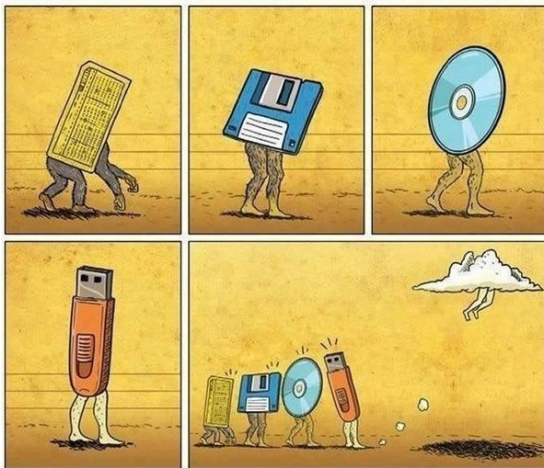
Важно: Мы не сможем узнать, была ли ошибка одна и она была успешно исправлена, или же их было две (или даже больше), и они были исправлены неправильно.

Если ошибок было больше двух, мы вообще можем о них не узнать.

Энергонезависимая память

- Природа носителя данных
- Диски: жёсткие и не очень
- Оптические носители
- Твердотельные накопители

ЭВОЛЮЦИЯ ЗАПОМИНАЮЩИХ УСТРОЙСТВ



- Конспект
- Перфокарта
- Перфолента

- Лента
 - Долго мотать до оглавления
 - Ломается, крошится магнитный слой
 - Высокая скорость
- Барабан
 - Весьма быстрый (на некоторых ЭВМ использовался, как ОЗУ)
 - Громоздкий, сейчас не используется
- Диски

- Сменные быстрые пакеты
 - Пришли на смену барабанам
 - Экранолетные головки ↗
- ГМД
 - Почти 40 лет (1971 – 2010(?))
 - От 80 килобайт (и 8 дюймов) до 2.88 МиБ (3.5 дюйма)
- Сменные магнитно-оптические носители
 - Нагрев до точки Кюри + размазанное магнитное поле. По 100 МиБ Лазер легко сфокусировать

Магнитные устройства: Жёсткие диски (I)

- Компоненты ↗
- Перемещение коромысла «непрерывное»
- Вид изнутри ↗
- Типично использование гидродинамических подшипников ↗

Магнитные устройства: Жёсткие диски (I)

- Компоненты [↗](#)
- Перемещение коромысла «непрерывное»
- Вид изнутри [↗](#)
- Типично использование гидродинамических подшипников [↗](#)

А ещё при помощи жёсткого диска, как и почти любого механического устройства, **можно** воспроизводить звук [↗](#)

- Жесткий — твердый и несменный
- Быстрые, высокая плотность
- Многократная перезапись, обычно экранолетные головки
- Пример сменных — ZIPdrive — не слишком удачный
- История:
 - Первый — **IBM 350 RAMAC** [↗](#), 1956 г., 1 тонна, 5 мегабайт — для мейнфреймов и табуляторов
 - 1980 — 5,25 Winchester, Shugart ST-506, 5 МиБ
 - 2006 — гибридные — диск + несколько ГиБ Flash — во Flash — КЭШ write-back, мотор стоит, ноутбук работает дольше. Промышленный образец — Seagate Momentus XT — 2010/
 - Вектор записи
 - 2006 — перпендикулярная запись
 - 2013 — «**черепичная запись** [↗](#)» — для изменения приходится перезаписать сектора нескольких дорожек

- CD
- DVD
- Blue Ray / HD DVD

- 1979 год
- Дорожка по спирали
- Единица скорости — около 150 КиБ/сек
- Впадины по 0,8 микрона
- Постоянная линейная скорость

RW — пузырьками в пластике

- Единица скорости — около 1350 КиБ/сек
- Впадины по 0,4 микрона (частота луча, больше, теорема Котельникова)
- Многослойные
- Коды регионов

- Blu
 - Четырехслойные до 100 Гиб
- HD
 - Трехслойные до 45 Гиб

Ширпотреб

- Самые долговечные — заводские
- Самые недолговечные — RW
- Долговечность CD/DVD R
 - CD-R и DVD-R — 10–25 лет, органический носитель данных
 - BD-R — 25–50 лет, металлический носитель данных

Ширпотреб

- Самые долговечные — заводские
- Самые недолговечные — RW
- Долговечность CD/DVD R
 - CD-R и DVD-R — 10–25 лет, органический носитель данных
 - BD-R — 25–50 лет, металлический носитель данных

А если хочется подольше?

Ширпотреб

- Самые долговечные — заводские
- Самые недолговечные — RW
- Долговечность CD/DVD R
 - CD-R и DVD-R — 10–25 лет, органический носитель данных
 - BD-R — 25–50 лет, металлический носитель данных

А если хочется подольше?

M-DISC

Полностью — [Millennial Disc](#) 

- DVD- и BD-R
- Слой с данными специального состава, обещают сотни лет
- Требуется более мощный лазер для записи

Тоже почти ширпотреб: болванки дороже раз в пять, привод — раза в 2–3

Твердотельные — накопители без движущихся механических деталей

- «Исключительно высокая» скорость позиционирования (позиционирования, как такового, нет)
- Достаточно высокая скорость записи
- Высокая скорость чтения

Твердотельные — накопители без движущихся механических деталей

- «Исключительно высокая» скорость позиционирования (позиционирования, как такового, нет)
- Достаточно высокая скорость записи
- Высокая скорость чтения
- Не боятся перегрузок
- Имеют конечный ресурс из-за свойств элементов памяти
- Удельная цена за бит выше, чем у механических

Flash-память

Самый распространённый и дешёвый вид — NAND. Построена на транзисторах с плавающими затворами, низкий ресурс, десятки тысяч переключений

- Старые флешки можно было за несколько часов испортить, если записывать в одну и ту же область
- Довольно быстро флешки стали умнее и научились переназначать области: если в какую-то область часто записывали, контроллер менял её местами с другой, менее «популярной»

Твердотельный накопитель

- Фактически — параллельно несколько из флешек, что обеспечивает высокие скорость и ёмкость
- Подключается через SATA или PCI Express

Для ускорения переноса «популярных» областей в протоколе SATA есть команда IDLE. Накопитель не знает формата файловой системы и не в курсе, занята область, или свободна, но ОС может подсказать, что область свободна. Тогда накопитель при переносе данных в эту область не будет переносить данные из неё ещё куда-то, что хорошо для скорости и ресурса.

Проф. В.М. Нестеров (тогда директор СПб отделения ЕМС, профессор в СПбГУ),
конференция СПИСОК-2016:

К 2020 г. ожидается стирание грани между твердотельной энергонезависимой памятью и оперативной памятью в плане производительности и цены ↗ , время: 42:50

Проф. В.М. Нестеров (тогда директор СПб отделения ЕМС, профессор в СПбГУ), конференция СПИСОК-2016:

К 2020 г. ожидается стирание грани между твердотельной энергонезависимой памятью и оперативной памятью в плане производительности и цены ↗ , время: 42:50

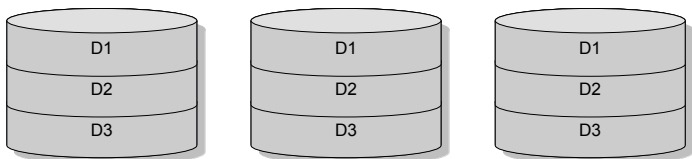
Высказывались и более сильные предположения, например, что скоро вся память будет энергонезависимой

- Успехи твердотельных накопителей на рынке впечатляют, но пока до этого далеко
- Тем не менее в перспективе это реально
 - Серьёзный вызов архитекторам ПО и ЭВМ — это многое может изменить

Redundant Array of Inexpensive Disks



- N устройств
- Сектора адресуются последовательно на разных винчестерах
- Избыточности нет
- Скорость чтения/записи и емкость растут в N раз
- Надежность падает

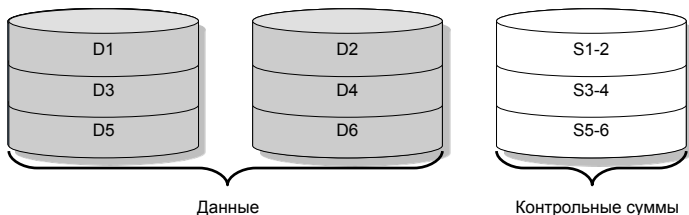


- M устройств
- «Зеркальные» диски
- Есть избыточность
- Надежность растет
- Скорость чтения растет в M раз, скорость записи старая

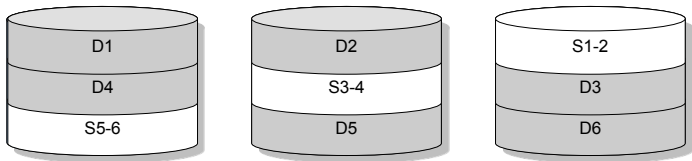
RAID 0+1

- $N \times M$ устройств
- Комбинирует 0 и 1 (0 поверх 1 или 1 поверх 0 — «коммутируют»)
- Надежность в целом растет
- Скорость чтения растет в $N \times M$ раз, записи — в N раз, емкость в N
- Избыточность есть

RAID 2, 3 и 4

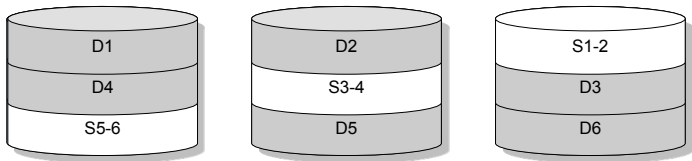


- 2 – Данные разбиваются по битам (8 накопителей) + еще несколько на коды восстановления
- 3, 4 – Примерно как RAID 2, только блоки больше
- Недостаток: высокая нагрузка на диски с кодами восстановления



- Коды восстановления информация поочередно появляется на разных дисках
- При выходе из строя одного диска сильно нагружается — восстанавливаются данные

А что там за коды восстановления?

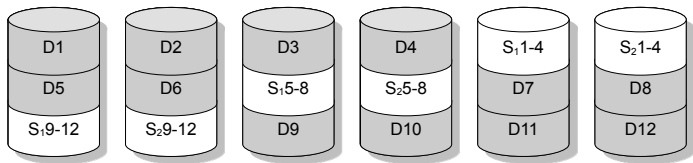


- Коды восстановления информация поочередно появляется на разных дисках
- При выходе из строя одного диска сильно нагружается — восстанавливаются данные

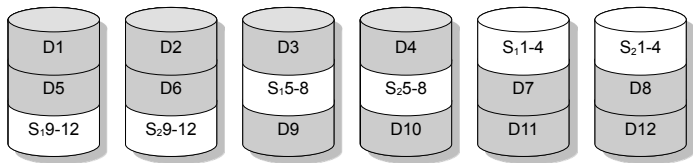
А что там за коды восстановления?

Элементарно!

$$S(i..j) = \bigoplus_{k=i}^j D(k)$$



А что там за коды восстановления?



А что там за коды восстановления?

Не элементарно: коды Рида-Соломона и всякая довольно задумчивая алгебра...

Вообще RAID можно сделать из чего угодно:

- Из настоящих флешек [↗](#)
- Из дисководов с дискетами [↗](#)

Вопросы

- Что такое коды обнаружения и коррекции?
- Назовите исторические и современные виды энергонезависимой памяти
- Назовите основные характеристики жёстких дисков
- Перечислите типы и опишите модели размещения данных RAID разных уровней
- Какими свойствами в сравнении с механическими жёсткими дисками обладают твердотельные накопители?
- Для чего в протоколе SATA есть команда IDLE?

Упражнения

- Попробуйте произвольно «извлечь» звук из Floppy-дисковод (или из чего-нибудь другого =))
 - На «высоком уровне» можно попытаться это сделать, отформатировав дискету, записав на неё файл на весь объём, и обращаясь к разным его участкам
 - На низком уровне это можно сделать, изучив программный интерфейс соответствующего драйвера

Вопросы



[EDU.DLUCIV.NAME](https://edu.dluciv.name) 