



Операционные системы

Часть 4. Ввод-вывод



Клименков С.В.
Версия 1.0.0
30.08.2020
vk.com/serge_klimenkov



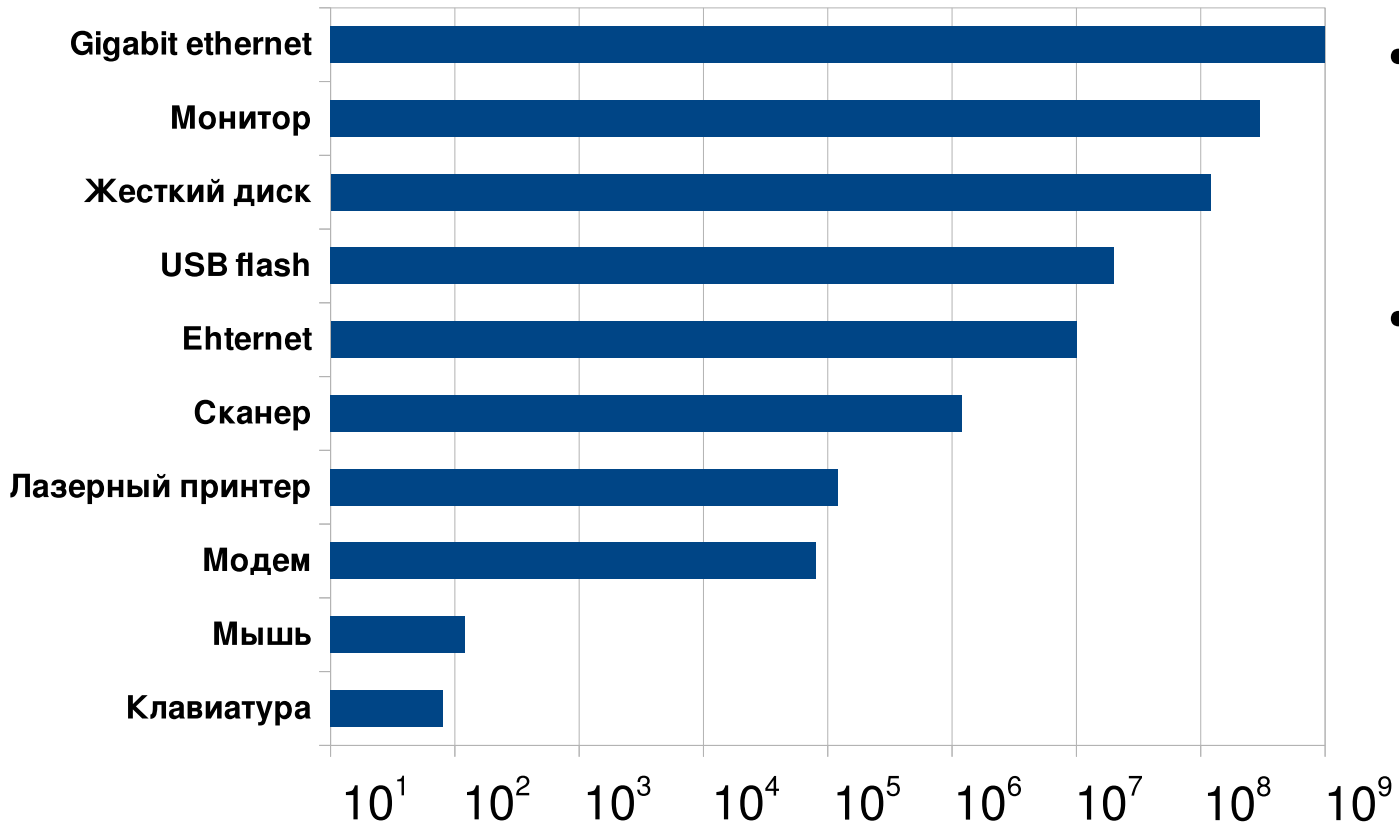
Блочный и поточный ввод-вывод

4.1

- Устройства ввода-вывода
- Режимы работы IO
- Типы буферизации
- Диски и их планирование
- Ввод вывод в SVR4



Устройства ввода-вывода



- **Устройства:**
 - Пользовательские
 - Внутренние
 - Связи
- **Характеристики:**
 - Скорость
 - Применение
 - Сложность управления
 - Единица передачи
 - Представление данных
 - Обстоятельства ошибок

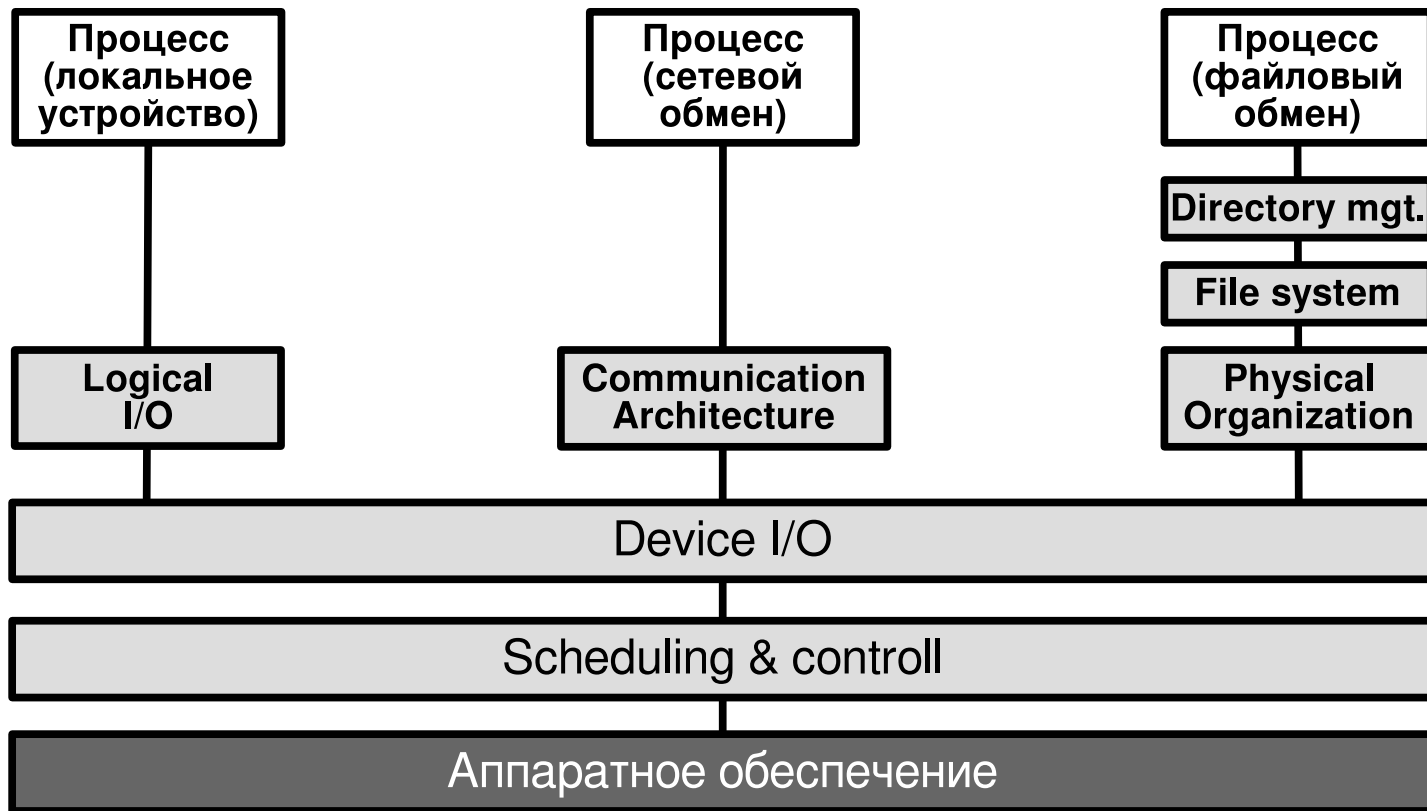


Развитие ввода вывода

- Программируемый ввод-вывод
 - Процессор непосредственно управляет периферийным устройством через его шину и регистры ...
 - ... или контроллер, который подключается к шине и имеет набор управляющих регистров
- Ввод-вывод с использованием прерываний
 - В контроллере добавляются прерывания, исключая ожидания
- Прямой доступ к памяти (Direct Memory Access)
 - В контроллере добавляются регистры и счетчики для обеспечения переноса области буфера в контроллере в область памяти
 - Контроллер превращается в отдельный вычислительный модуль (канал ввода-вывода) с процессором и системой команд
 - В канал ввода-вывода добавляется доп. оборудование и микропрограммы и контроллер становится отдельным вычислительным устройством полностью берущим на себя управление вводом-выводом с группой устройств (процессор ввода-вывода)



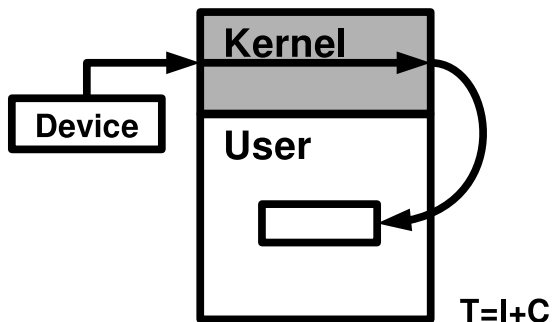
Логическая структура ввода-вывода



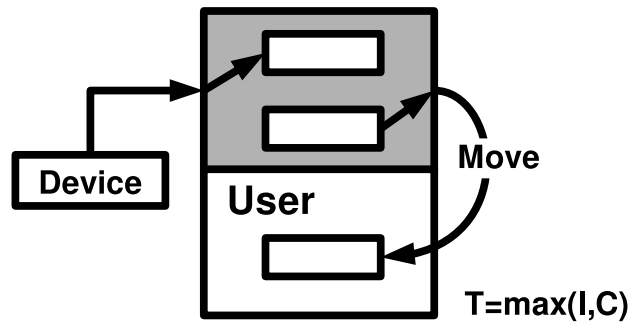


Буферизация ввода-вывода

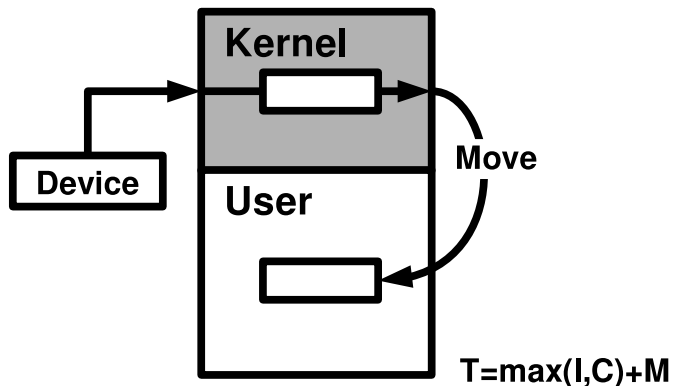
- Устройства:
 - Блочные
 - Символьные (потокковые)
- Легенда:
 - I — время ввода-вывода
 - C — время обработки
 - M — время пересылки



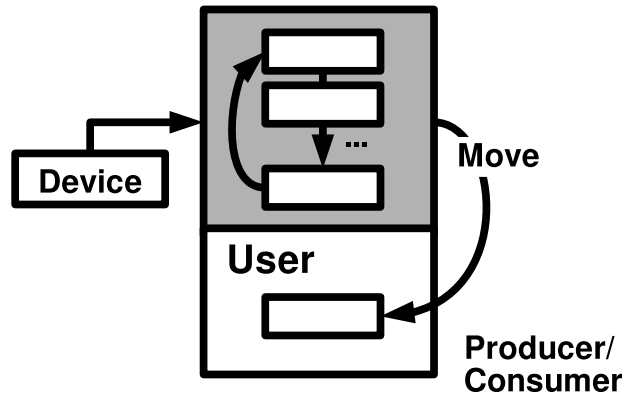
а) без буферизации



в) Двойная буферизация



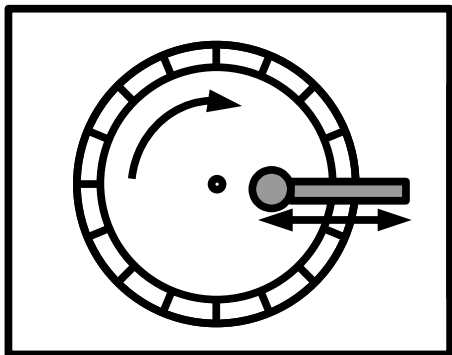
б) Одиночная буферизация



г) Кольцевая буферизация

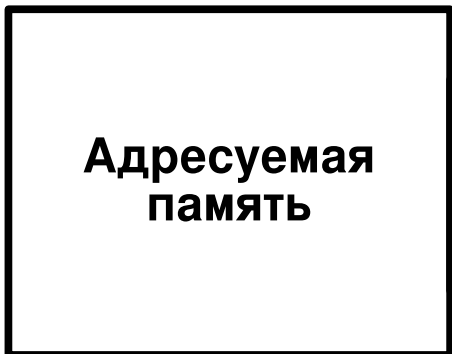


Диски



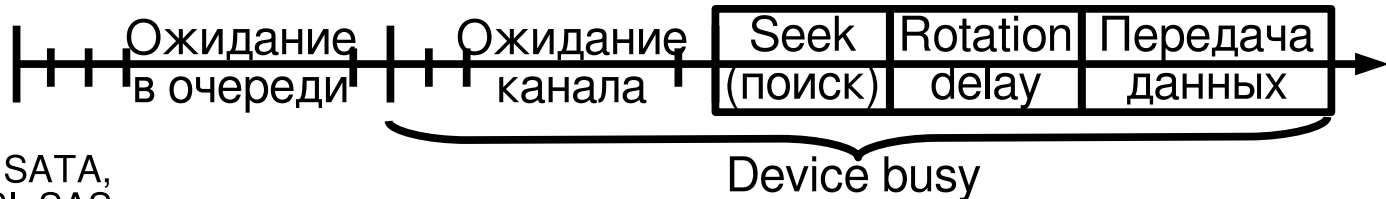
Hard Disk

IDE, SATA,
SCSI, SAS
interfaces



Solid State Disk

SATA, SAS,
NVMe
interfaces



$$T_{\partial} = T_o + T_k + T_n + \frac{1}{2r} + \frac{b}{rN}$$

- T_{∂} — время доступа
- T_o — время ожидания в очереди ОС
- T_k — время освобождения канала
- T_n — время поиска
- r — скорость вращения RPM
- N — количество байт на дорожке
- b — количество байт в запросе

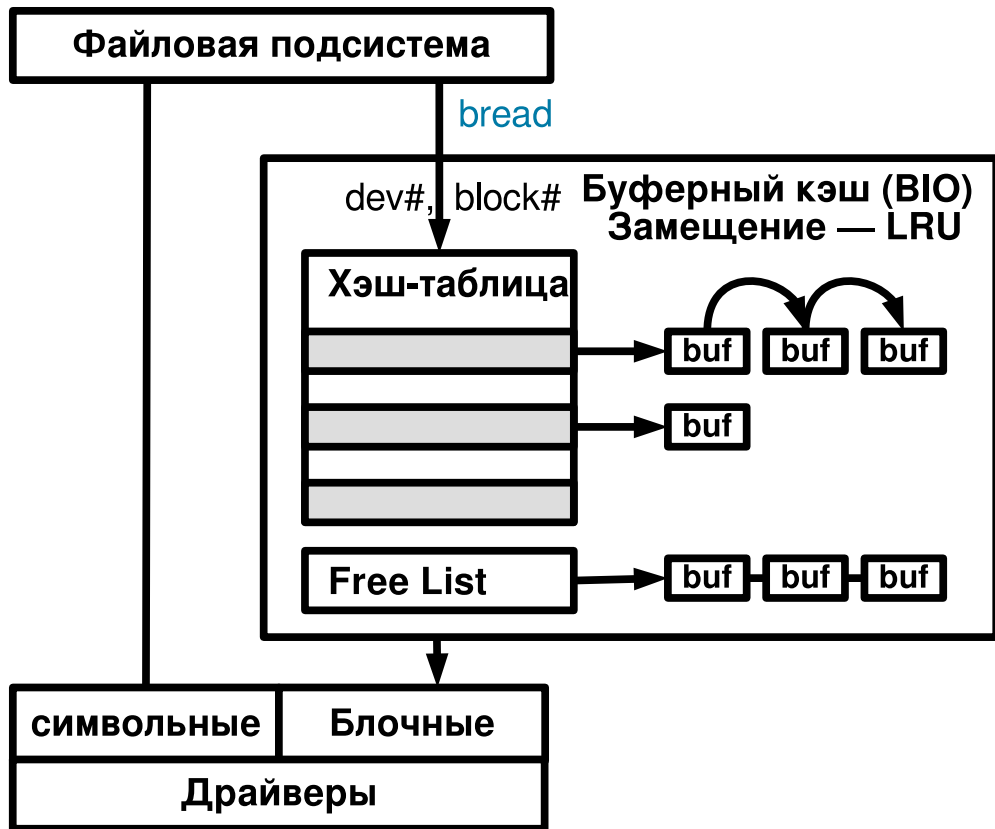


Дисковое планирование

- FIFO — «справедливый метод»
 - Все процессы получают одинаковый доступ к диску
 - Выгоден для небольшого потока запросов, на большом превращается в случайный доступ (издержки большие)
- PRI — на основе от приоритета процесса
 - Выгоден с точки зрения ОС (см. Feedback) для коротких заданий, длинным плохо.
- LIFO — Использует преимущества локальности данных
 - Хороша для транзакционных систем
- А если учитывать при планировании текущее состояние (н.р. - дорожка) диска?
- SSTF — Shortest Service Time First (Минимизация времени поиска)
- SCAN (elevator algorithm) — Ездим туда-сюда по диску и обслуживаем запросы
 - Предпочитает центр диска и плохо «относится» к запросам для только что пройденных дорожек
- C-SCAN — Ездим по диску во время операций в одну сторону, быстрый возврат
- N-step-SCAN — разделяет очередь на подочереды длиной N, 1 подочередь за 1 SCAN, если в подочереды запросов меньше N, то выполнить ее обработку в следующий SCAN.
- FCSCAN — две подочереды, пока одна обрабатывается, вторая заполняется.



Ввод-вывод в SVR4



- Два вида:
 - Буферизированный ввод-вывод (через bio)
 - Небуферизированный ввод-вывод (через DMA)
- Для медленных устройств (терминалы...)
 - Используются отдельные символьные буферы

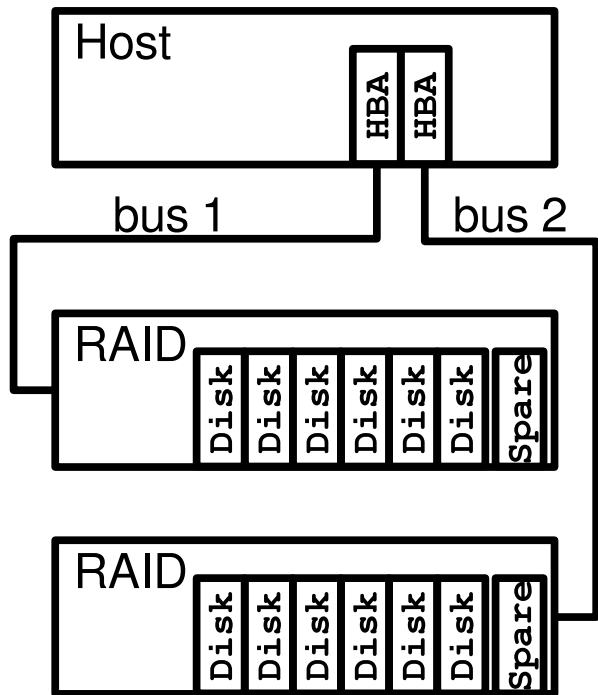


RAID

4.2

- Концепции RAID
- Concatenation & Striping
- Зеркалирование
- Использование ECC в RAID массивах
- Использование контроля четности в RAID массивах

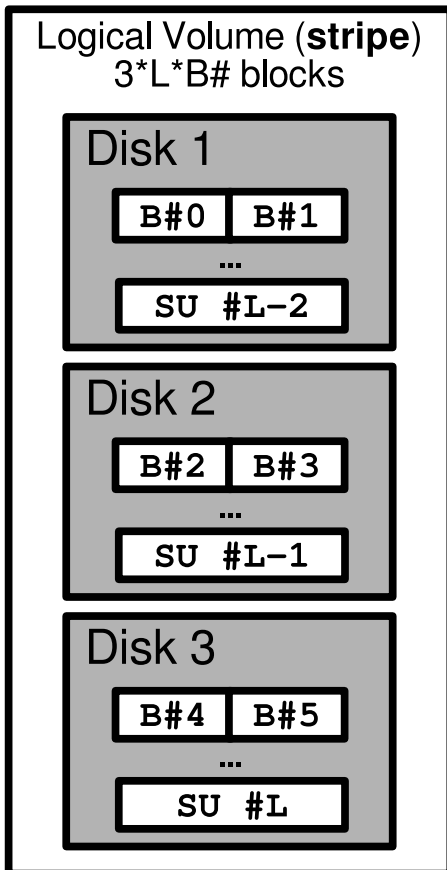
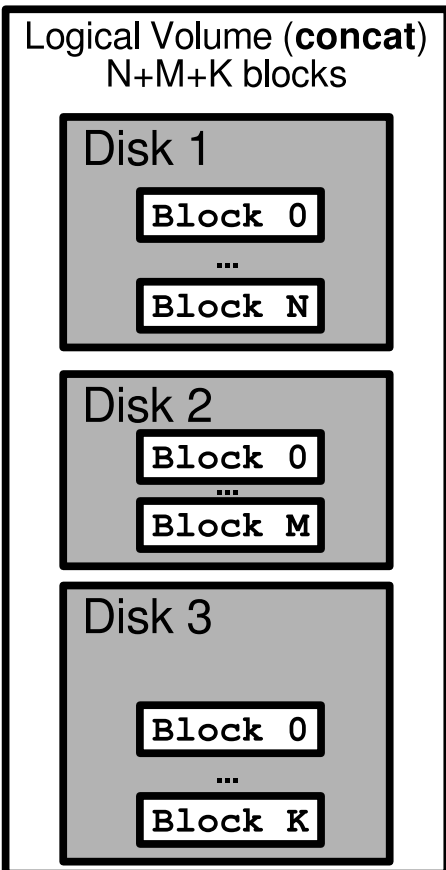
RAID (Redundant Array of Independent Disks)



- Ввод вывод к нескольким дискам может происходить параллельно
- С увеличением количества дисков надежность падает
- RAID — несколько дисков, которые ОС логически объединяет в один
 - Данные распределены по всем дискам
 - Может использоваться дополнительная (избыточная) емкость для хранения контрольной информации
 - Один блок данных может находится на нескольких дисках
 - Один запрос может также может быть выполнен параллельно
 - Необходимо учитывать производительность адаптеров и шин подключения к адаптерам, а также шины I/O (PCI, PCIe)



RAID 0 — concatenation/striping

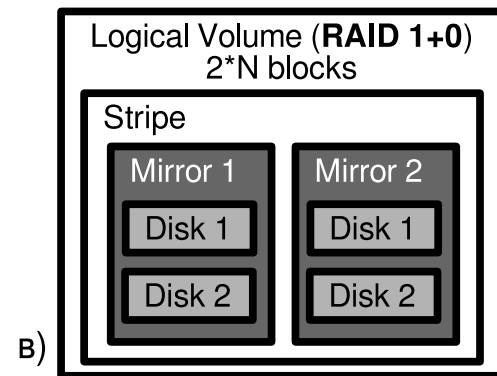
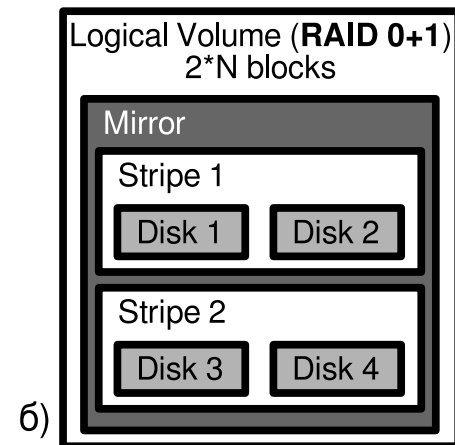
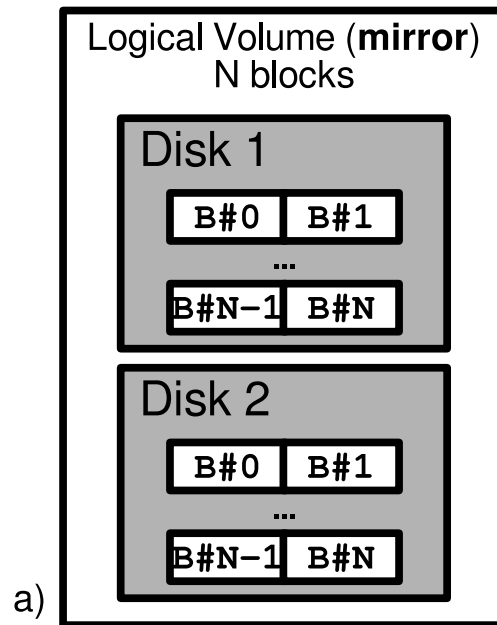


- Объединяет диски в один том большего размера
- Конкатенация — можно склеивать диски разного размера
- Страйп — одинакового размера или размера меньшего диска
 - Stripe Unit — объединение блоков общим размером Stripe Size (16-128Кб)
 - SU зависит от конфигурации при создании
- Характеристики
 - Низкая надежность
 - Высокая пропускная способность
 - Высокая скорость обработки запросов для чтения и записи (stripe)



Raid 1 - Зеркалирование

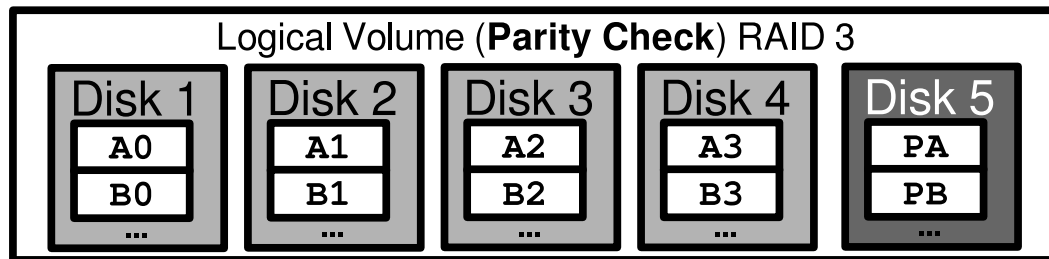
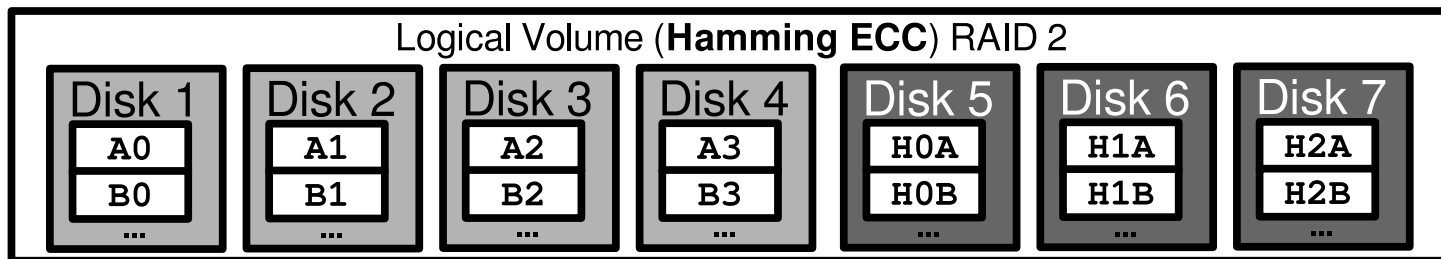
- Объединяет диски, копируя блоки
- Разные конфигурации:
 - RAID 0+1 (б) RAID 10 (в)
- Двойная (или больше) избыточность
- Характеристики (а):
 - Высокая надежность
 - двойная пропускная способность на чтение и одинарная на запись
 - Двойная скорость обработки запросов для чтения и одинарная на запись
- Характеристики у (б) и (в) комбинируются с RAID 0.





RAID 2 и RAID 3

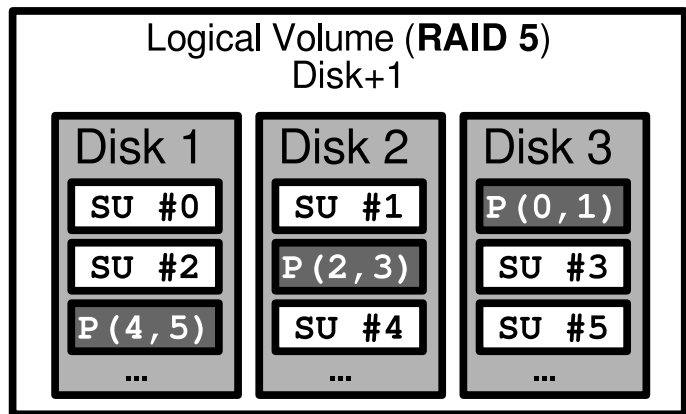
- Диски синхронизированы между собой по позициям головок
- Может использоваться очень маленький stripe unit (байт или слово)
- На практике используются редко
- Запись на все диски одновременно



- RAID 2:
 - Необходимо большое количество дисков
 - Коды Хемминга — коррекция одиночных ошибок, обнаружение двойных.
 - Минимум 7 дисков
- RAID3:
 - Подсчет четности
 - Исправление одиночной ошибки
 - Избыточность Disk+1



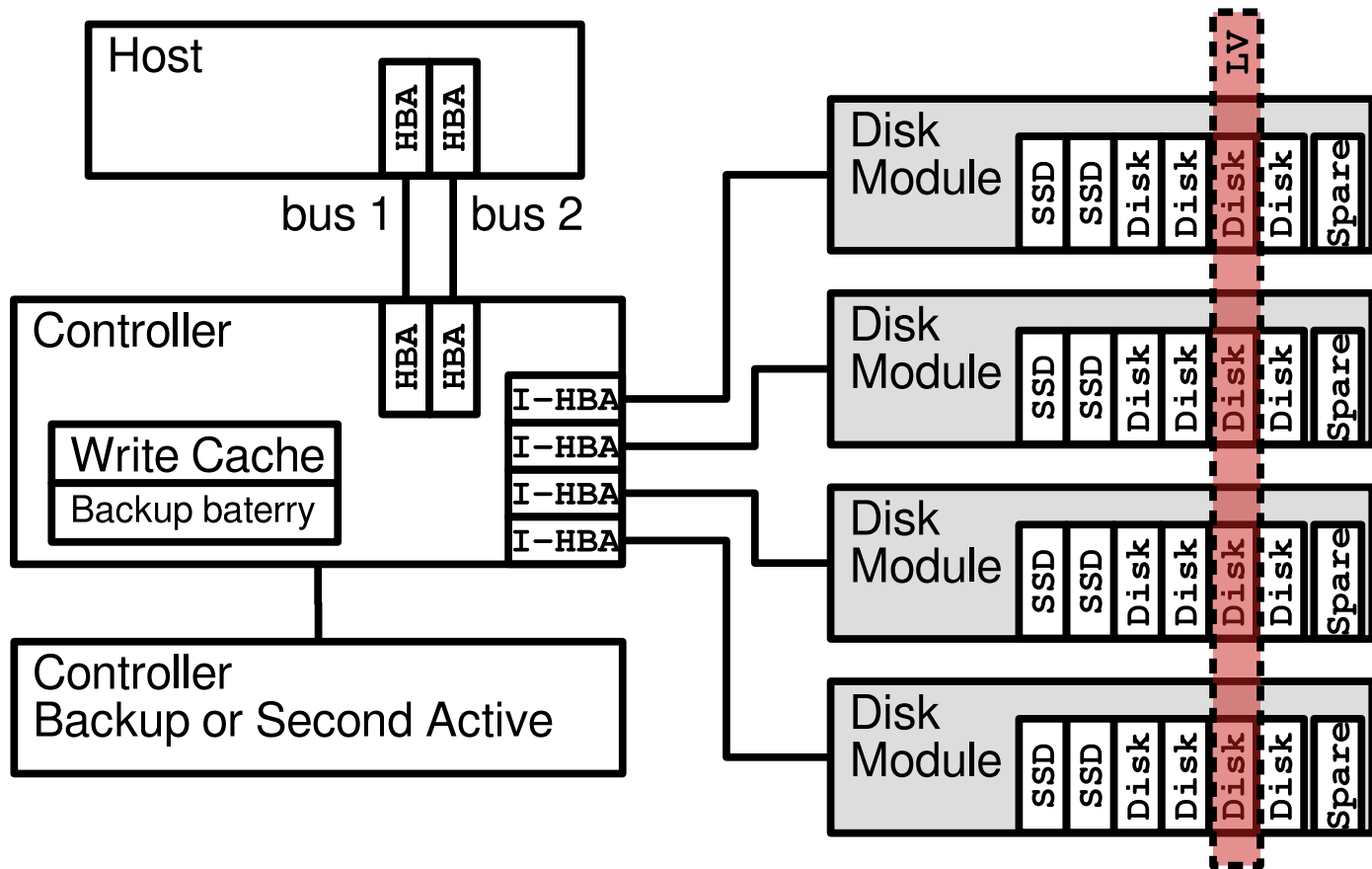
RAID 4, RAID 5, RAID 6



- Подсчитывает четность для SU
 - $P(0,1) = SU(0) \oplus SU(1)$
 - Если сбой диска 2 $SU(1) = P(0,1) \oplus SU(0)$
- Характеристики (а):
 - Высокая надежность, избыточность Disk+1
 - Дополнительные вычисления при записи, и при чтении в случае сбоя диска
 - Пропускная способность на чтение как у RAID 0 и меньше, чем одинарная на запись
 - Скорость обработки запросов для чтения как у RAID 0 и меньше, чем одинарная на запись
- RAID 4 — выделенный диск с четностью
- RAID 6 — двойная размазанная четность, по разным алгоритмам, избыточность Disk+2



Аппаратные дисковые массивы



- Сложные устройства с различной архитектурой
- Очень высокие показатели производительности и отклика
- Отказоустойчиво!
- Дорого!
- Большой объем хранения данных
- Внешнее управление — http, ssh, ...



4.3



Ввод вывод в Windows

4.4



Файлы и файловые системы

4.5

- Поле, запись, файл
- Задачи ОС
- Управление файлами UNIX
- Каталоги файлов
- Размещение файлов
- DOS FAT, UNIX UFS



- Коллекция данных со следующими свойствами:
 - Наличие структуры (может быть сложной, может быть несколько в одном файле)
 - Возможность долгосрочного существования
 - Возможность совместного использования процессами
- Основные файловые операции
 - Создание, удаление, открытие, закрытие, чтение, запись, выбор позиции, контрольные операции, блокирование



Поле, запись, файл, база данных

- Поле — одиночный элемент записи, обычно одно значение.
- Запись — набор полей
- Файл — совокупность записей, относящихся к однородному набору данных
- База данных — файл со сложной взаимозависимой структурой полей и записей
- Операции, проводимые с файлом, могут влиять на его структуру
- Файл может не иметь структуры (например, представлять собой поток символов)

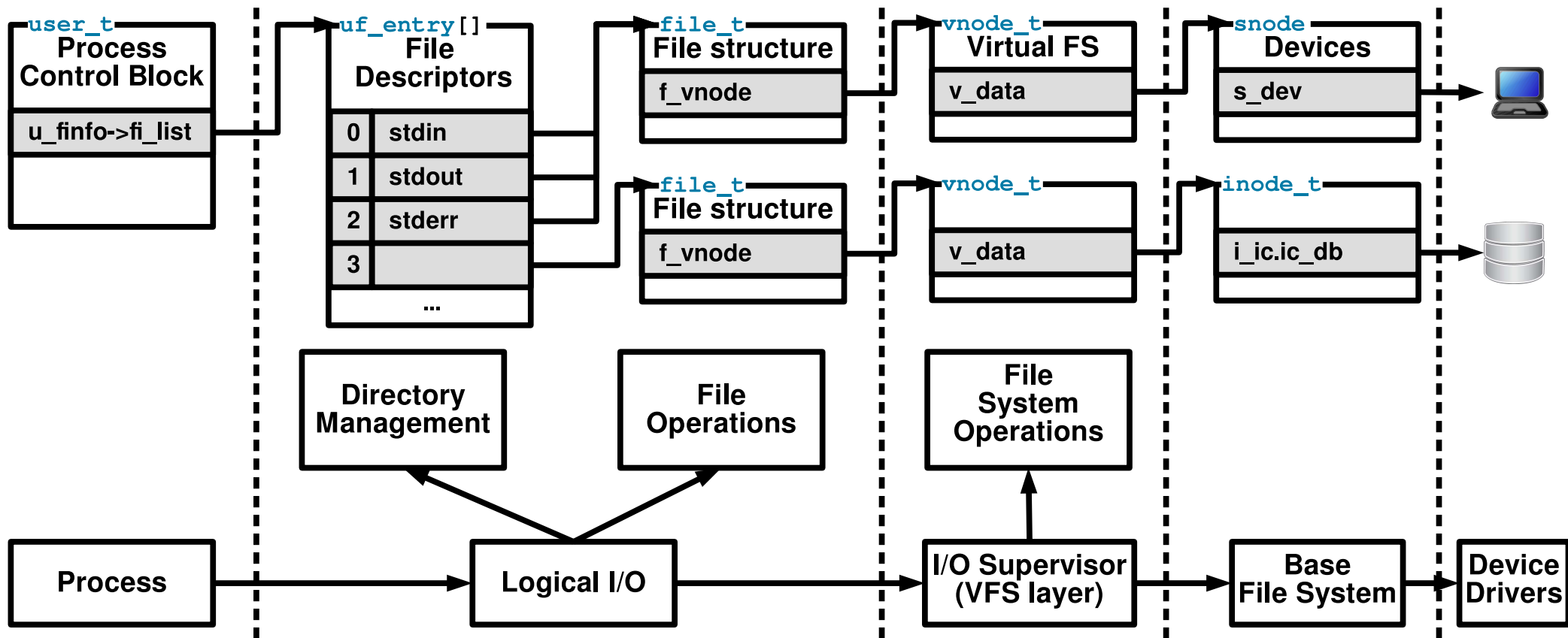


Подсистема управления файлами. Задачи ОС

- Возможность хранить файлы и выполнять пользователю над ними операции:
 - Создание, удаление, чтение и изменения файлов
 - Управление доступом к файлам для себя и других пользователей
 - Перемещение данных между файлами
 - Выполнение резервного копирования и восстановления файлов
 - Обеспечение возможности работы с файлами по именам, удобным для пользователя
- Гарантия корректности данных
- Обеспечение приемлемой производительности
- Поддержка различных типов устройств хранения
- Минимизация или исключения потерь и повреждения данных
- Обеспечение базового набора функций ввода-вывода
- Обеспечение совместного использования файлов



Управление файлами UNIX SVR4





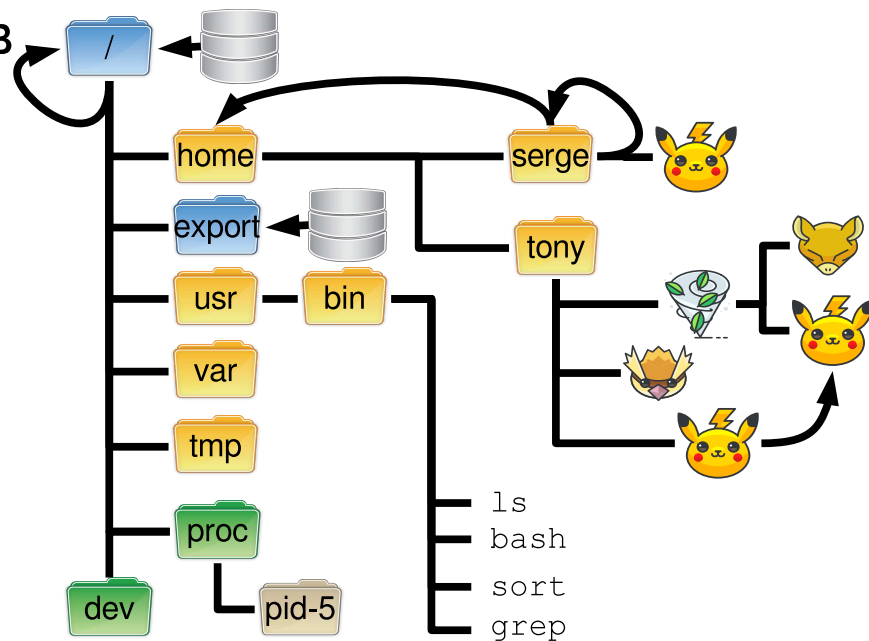
Каталоги файлов. Элементы каталога.

- Основные:
 - Имя файла, Тип файла (бинарный, текстовый, ...)
 - Организация файла — организация внутренней структуры
- Адресная информация
 - Том (носитель) — указатель на физическое устройство
 - Начальный адрес — номер первого блока файла
 - Занимаемый размер и зарезервированный размер— количество байт в файле и максимальное количество байт
- Информация об управлении доступом
 - Владелец — пользователь, который может управлять файлом
 - Разрешенные действия (чтение, запись, исполнение, поиск, и т.п.)
- Информация об использовании
 - Создатель и дата создания — создатель не обязательно может быть владельцем
 - Дата последнего чтения и последний пользователь, прочитавший файл
 - Дата последнего изменения и последний пользователь, изменивший файл
 - Дата последней резервной копии на другом устройстве
 - Текущее использование — информация о текущих действиях с файлом



Каталог, операции ОС

- Часть элементов каталога может находится в записи о файле, уменьшая его размер
- Каталог может быть полностью или частично загружен в основную память
- Простейшая структура — список записей фиксированной длины
- Основные операции:
 - Поиск файловых записей,
 - Создание и удаление записей о файле
 - Получение списка записей
 - Обновление каталога, изменение записи
- Обычно структура каталогов — n-уровневая иерархия, дерево, связанный граф



- Именованние:
 - Имя файла
 - Имя, включающее полный путь от корня
 - Имя, относительно текущего каталога

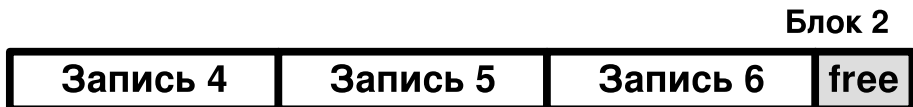


Совместное использование файлов

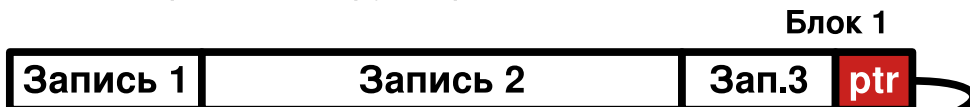
- Владелец — пользователь, имеющий все права на файл, и распределяющий эти права:
 - Отсутствие прав — другие пользователи не знают о существовании файла
 - Знание — пользователи знают о существовании файла, и могут запросить права у владельца
 - Выполнение — пользователь может загрузить и выполнить программу
 - Чтение, добавление в конец файла, модификация, удаление
 - Изменение прав доступа
- Доступ к файлу может быть передан разным наборам пользователей:
 - Конкретному пользователю
 - Списку пользователей или группе пользователей
 - Всем пользователям
- Блокировка одновременного доступа
 - Всего файла
 - Отдельных записей



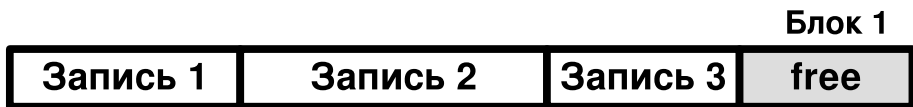
Записи и блоки



а) Фиксированное группирование



б) Группирование переменной длины со сцеплением



в) Группирование переменной длины без сцепления

- Запись — логическая единица доступа к структурированному файлу
 - ОС осуществляет ввод-вывод блоками
- Какой должен быть размер блок относительно размера записи?
 - Блок больше → Больше записей передано за одну операцию ввода-вывода
 - При случайном доступе — передаем записи, которые не используются
 - Большим блокам нужны большие буферы ОС



Размещение файлов

Номер блока на дорожке

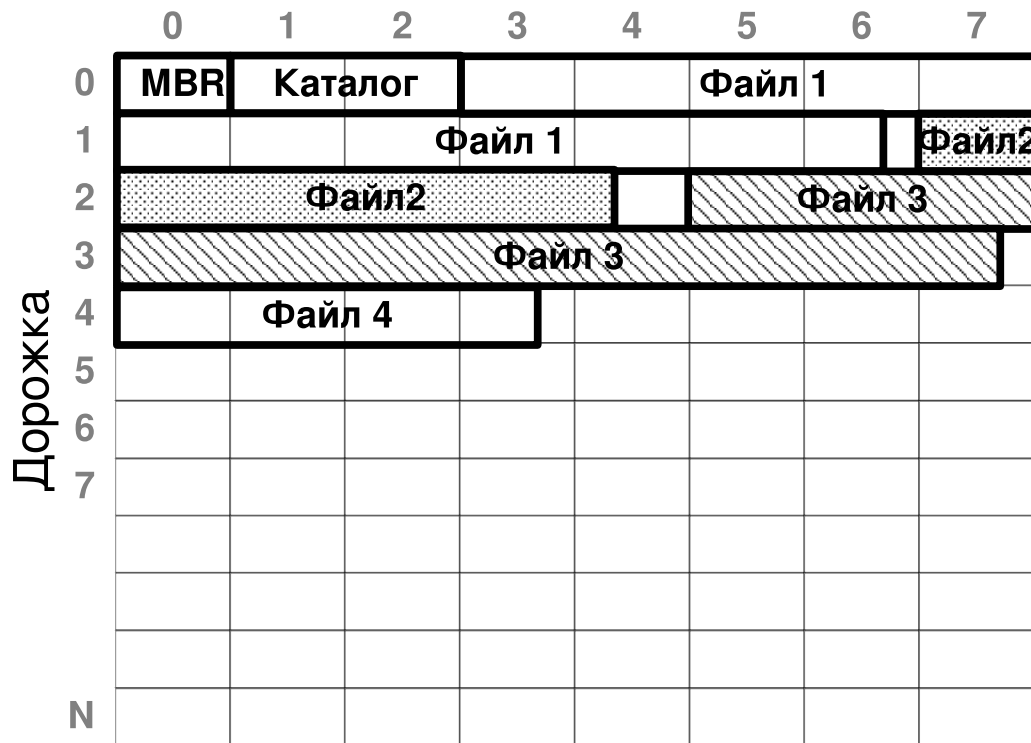
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| 0 | | | | | | | | |
| 1 | | | | | | | | |
| 2 | | | | | | | | |
| 3 | | | | | | | | |
| 4 | | | | | | | | |
| 5 | | | | | | | | |
| 6 | | | | | | | | |
| 7 | | | | | | | | |
| N | | | | | | | | |

- Как на физической структуре диска, состоящего из блоков разместить файлы?
 - Предварительное выделение пространства для всего файла?
 - Размер порции файла?
 - Фиксированный или переменный размер порции?
- Решения влияют на характеристики:
 - Внутреннюю и внешнюю фрагментацию
 - Частота выделения (allocation) блоков для всей файловой системы
 - Время выделения блоков
 - Размер служебной информации о размещении файлов



Непрерывное последовательное размещение файлов (ОС RT-11)

Номер блока на дорожке



- Особенности

- Предварительное выделение пространства для всего файла.
- Один уровень каталога
- Размер блока 512 байт

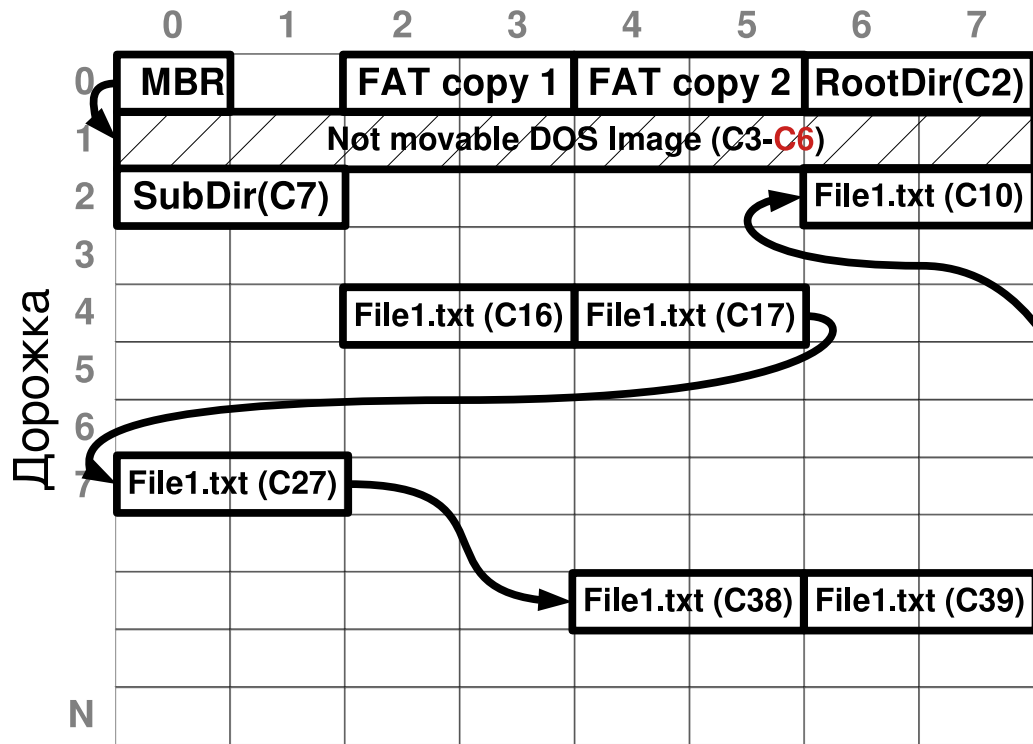
- Характеристики:

- Внутренняя фрагментация мала
- Внешняя может быть большой — необходимо обязательное сжатие ФС
- Выделение блоков для файла производится однократно
- Невысокое время выделения блоков — если ФС пуста, то минимальное, если заполнена, то возможны ошибки размещения
- Минимальный размер служебной информации о размещении файлов



Цепочечное размещение файлов (DOS FAT)

Номер блока на дорожке



• Особенности

- Выделение пространства для файла по мере необходимости
- Размер порции файла — кластер (C) — от 512 (FAT12) до 32Кб (FAT 32)
- Запись каталога содержит имя (8+3), размер файла, ..., номер начального кластера: File1.txt
- FAT (File Allocation Table) — Содержит таблицу аллокации цепочки кластеров

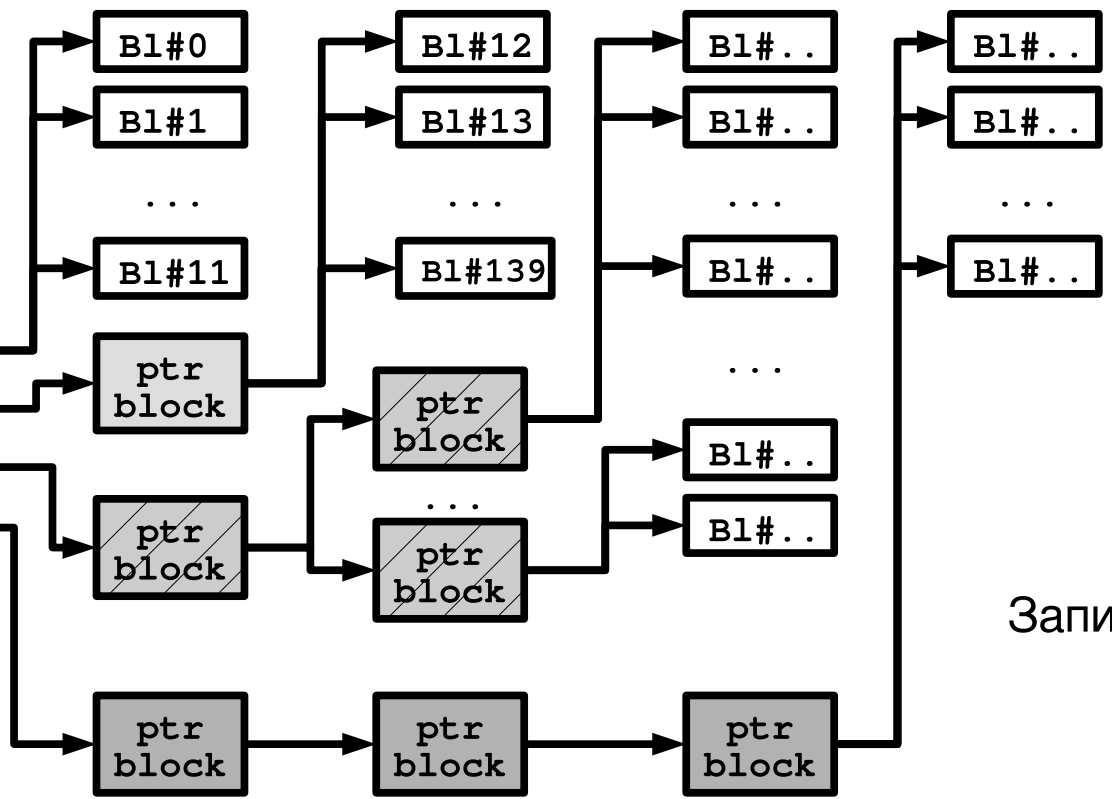
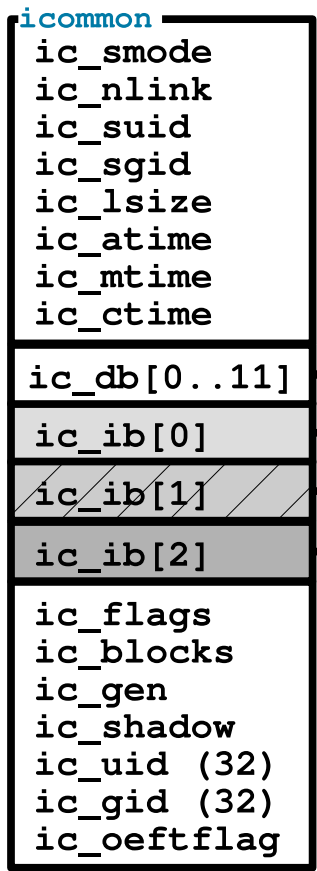
| | | | | | | |
|----------|------|----|----|----|----|----|
| CL# | 10 | 16 | 17 | 27 | 38 | 39 |
| Next Cl# | FFFF | 17 | 27 | 38 | 39 | 10 |

• Характеристики:

- Внутренняя фрагментация мала
- Внешняя может быть большой — файлы размещены непоследовательно, необходимо обязательное дефрагментирование ФС для увеличения скорости
- Высокое время выделения блоков файла
- Размер служебной информации о размещении файлов ограничен FAT

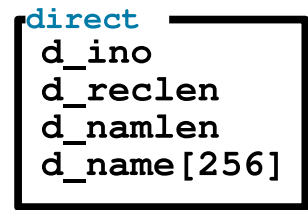


Файловая система UNIX SRV4. Индексированные записи файлов (inode)



- Типы файлов:
 - Обычные файлы
 - Директории
 - Блочные устройства
 - Символьные устройства
 - Именованные каналы (pipes)
 - Жесткие ссылки
 - Символические ссылки

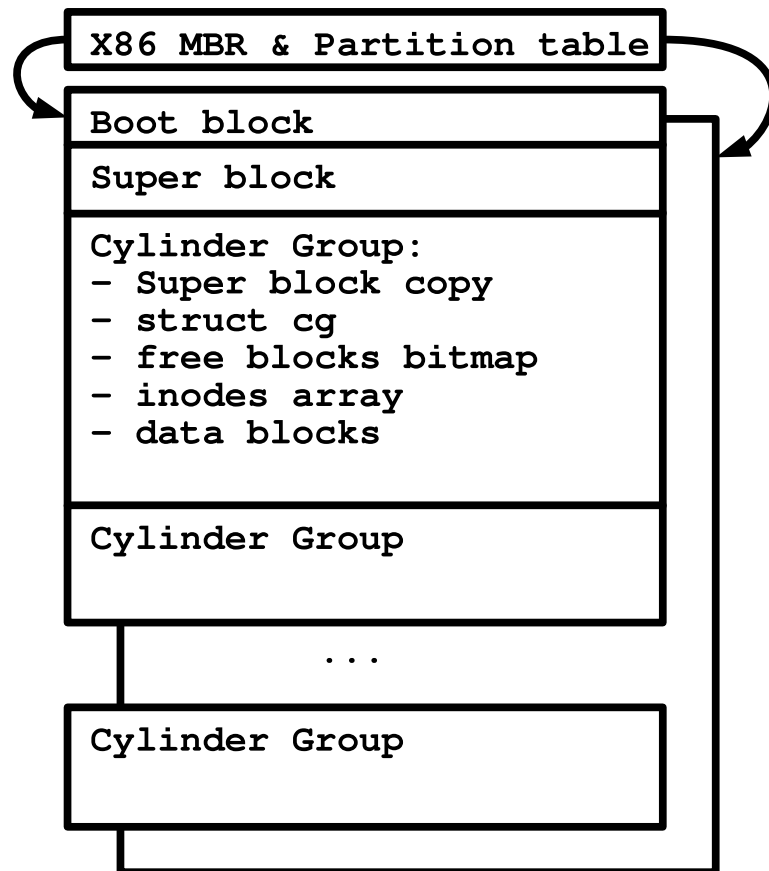
Запись в директории





Структура файловой системы UNIX SVR4 UFS

- Файловая система содержит:
 - Загрузчик ОС
 - Суперблок — геометрия и служебные параметры файловой системы
 - Цилиндровые группы — информация о свободных блоках, массив записей о файлах (inode), блоки данных — равномерно распределены по всему дисковому пространству





Управление файлами LINUX

4.6



СИСТЕМНЫЕ ВЫЗОВЫ СО СТОРОНЫ ПОЛЬЗОВАТЕЛЯ

- Для пользователя можно выполнять операции:
 - `int open(const char *pathname, int flags, mode_t mode);`
 - `int creat(const char *pathname, mode_t mode);`
 - `int close(int fd);`
 - `int unlink(const char *pathname);`
 -



Управление файлами Windows

4.7