

Содержание

1 Обзор ОС UNIX: архитектура, вход в систему, файлы и каталоги, ввод и вывод.	5
1.1 Архитектура UNIX	5
1.1.1 Определение ОС	5
1.1.2 Интерфес ядра	5
1.1.3 Коммандная оболочка	5
1.1.4 Обобщение	5
1.1.5 Linux	5
1.1.6 Схема, отражающая структуру UNIX	6
1.2 Вход в систему	6
1.3 Файлы и каталоги	6
1.3.1 Файловая система	6
1.3.2 Имя файла	7
1.3.3 Путь к файлу	7
1.3.4 Рабочий каталог	7
1.3.5 Домашний каталог	7
1.4 Ввод и вывод	8
1.4.1 Дескрипторы файлов	8
1.4.2 Стандартный ввод, стандартный вывод, Стандартный вывод сообщений об ошибках	8
1.4.3 Небуферизованный ввод-вывод	8
1.4.4 Стандартные функции ввода-вывода	8
2 Обзор ОС UNIX: программы и процессы, обработка ошибок, идентификация пользователя.	9
2.1 Программы и процессы	9
2.1.1 Программа	9
2.1.2 Процессы и идентификаторы процессов	9
2.1.3 getpid()	9
2.1.4 Управление процессами	9
2.2 Обработка ошибок	9
2.2.1 errno	9
2.2.2 Вывод сообщения об ошибке	10
2.2.3 Восстановление после ошибок	10
2.3 Идентификация пользователя	10
2.3.1 Идентификатор пользователя	10
2.3.2 Идентификатор группы	11

3	Обзор ОС UNIX: сигналы, представление времени, системные вызовы и библиотечные функции.	11
3.1	Сигналы	11
3.2	Представление времени	12
3.3	Системные вызовы и библиотечные функции	12
4	Стандарты и реализации ОС UNIX: пределы ISO C, пределы POSIX, функции sysconf(), pathconf() и fpathconf(), элементарные системные типы данных.	13
4.1	Стандартизация UNIX	13
4.1.1	ISO C	13
4.1.2	IEEE POSIX	14
4.1.3	Single UNIX Specification	14
4.1.4	FIPS	15
4.2	Реализации UNIX	15
4.2.1	UNIX System V Release 4	15
4.2.2	4.4BSD	15
4.2.3	FreeBSD	15
4.2.4	Linux	16
4.2.5	Mac OS X	16
4.2.6	Solaris	16
4.2.7	Прочие версии UNIX	16
4.3	Пределы ISO C	16
4.4	Пределы POSIX	17
4.5	Функции sysconf, pathconf и fpathconf	17
4.6	Элементарные системные типы данных	18
5	Файловый ввод-вывод: дескрипторы файлов, функция open(), функция creat(), функция close().	19
5.1	Дескрипторы файлов	19
5.2	Функции open и openat	19
5.2.1	Параметр fd в функции openat	19
5.3	Функция creat	20
5.4	Функция close	20
6	Файловый ввод-вывод: Функция lseek(), функция read(), функция write()	20
6.1	Функция lseek	20
6.2	Функция read	21
6.3	Функция write	21
7	Файловый ввод-вывод: эффективность операций ввода-вывода	22

8	Файловый ввод-вывод: совместное использование файлов, атомарные операции, функции dup() и dup2()	23
8.1	Совместное использование файлов	23
8.2	Атомарные операции	24
8.2.1	Добавление данных в конец файла	24
8.2.2	Функции pread и pwrite	24
8.2.3	Создание файла	25
8.3	Функции dup и dup2	25
9	Файловый ввод-вывод: функции sync(), fsync(), fdatasync(), fcntl(), ioctl(), /dev/fd	25
9.1	Функции sync, fsync и fdatasync	25
9.2	Функция fcntl	26
9.3	Функция ioctl	27
9.4	/dev/fd	27
10	Файлы и каталоги: функции stat(), fstat(), lstat(), содержимое struct stat.	27
11	Файлы и каталоги: типы файлов, права доступа к файлу, функция umask().	28
11.1	Типы файлов	28
11.2	Права доступа к файлу	29
11.3	Функция umask	30
12	Файлы и каталоги: функции chmod(), fchmod(), chown(), fchown(), lchown().	30
13	Файлы и каталоги: размер файла, дырки в файлах, усечение файлов, файловые системы, функции link(), unlink(), remove(), rename().	30
14	Файлы и каталоги: символические ссылки, функции symlink() и readlink().	30
15	Файлы и каталоги: временные характеристики файлов, функция utime().	32
16	Файлы и каталоги: функции mkdir() и rmdir(), чтение каталогов, функции chdir(), fchdir(), getcwd().	32
17	Стандартная библиотека ввода-вывода: потоки и объекты FILE, стандартные потоки ввода, вывода и сообщений об ошибках, буферизация.	32

18	Стандартная библиотека ввода-вывода: открытие потока, чтение из потока и запись в поток, функции ввода, функции вывода.	32
19	Стандартная библиотека ввода-вывода: эффективность стандартных операций ввода-вывода, позиционирование в потоке.	32
20	Стандартная библиотека ввода-вывода: форматированный вывод, форматированный ввод, временные файлы.	32
21	Управление процессами: идентификаторы процесса, функция <code>fork()</code> , совместное использование файлов.	32
22	Управление процессами: функция <code>exit()</code> , функции <code>wait()</code> и <code>waitpid()</code> .	32
23	Управление процессами: семейство функций <code>exec()</code> .	32
24	Управление процессами: изменение идентификаторов пользователя и группы, функции <code>setuid()</code> , <code>setgid()</code> , <code>seteuid()</code> , <code>setegid()</code> .	32
25	Управление процессами: интерпретируемые файлы, функция <code>system()</code> .	32
26	Сигналы: концепция сигналов, функция <code>signal()</code> , ненадежные сигналы.	32
27	Сигналы: прерванные системные вызовы, реентерабельные функции.	32
28	Сигналы: функции <code>kill()</code> , <code>raise()</code> , <code>alarm()</code> , <code>pause()</code> .	32
29	Сигналы: надежные сигналы, терминология и семантика, наборы сигналов.	32
30	Сигналы: маска сигналов процесса и функция <code>sigprocmask()</code> , функция <code>sigpending()</code> ,	32
31	Сигналы: функция <code>sigaction()</code> .	32
32	Сигналы: функция <code>sigsuspend()</code> .	32

1 Обзор ОС UNIX: архитектура, вход в систему, файлы и каталоги, ввод и вывод.

1.1 Архитектура UNIX

1.1.1 Определение ОС

Операционная система (ОС) – это программное обеспечение (ПО), которое управляет аппаратными ресурсами компьютера и предоставляет среду выполнения прикладных программ (*application*). Обычно это ПО называют ядром (*kernel*), т.к. оно имеет относительно небольшой объем и составляет основу системы (см. рисунок ниже).

1.1.2 Интерфес ядра

Интерфес ядра – это слой ПО, называемый системными вызовами (*system calls*). Библиотеки функций общего пользования (*library routines*) строятся на основе интерфеса системных вызовов, но прикладная программа (*application*) может свободно пользоваться как теми, так и другими. Т.е. прикладная программа может использовать как системные вызовы, так и библиотечные функции.

1.1.3 Коммандая оболочка

Коммандая оболочка (*shell*) – это особое приложение, которое предоставляет интерфес для запуска других приложений.

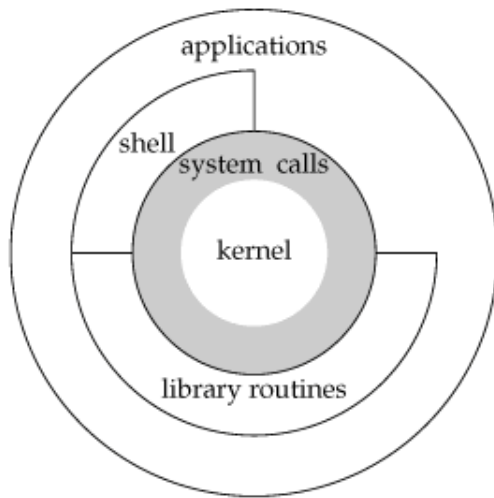
1.1.4 Обобщение

В общем, ОС – это ядро и всё остальное ПО, которое делает компьютера пригодным к использованию. В состав этого ПО входят системные утилиты (например *ls*, *cat*), прикладные программы (например *GIMP*), коммандые оболочки (например *bash*, *fish*), библиотеки функций общего пользования (например *stdio* для *C*) и т.п..

1.1.5 Linux

Linux – ядро ОС GNU (или же GNU/Linux). Отдельно Linux – это ядро, называть его ОС не совсем корректно.

1.1.6 Схема, отражающая структуру UNIX



1.2 Вход в систему

При входе в систему UNIX мы вводим имя пользователя и пароль. После того система отыскивает введенное имя в файле паролей; обычно это файл `/etc/passwd`. Файл паролей содержит записи, каждая из которых состоит из семи полей, разделенных двоеточиями: имя пользователя, зашифрованный пароль, числовой идентификатор пользователя (`205`), числовой идентификатор группы (`105`), поле комментария, домашний каталог (`/home/sar`) и командный интерпретатор (`/bin/ksh`).

Пример записи:

```
sar:x:205:105:Stephen Rago:/home/sar:/bin/ksh
```

Все современные системы хранят пароли в отдельном файле.

После входа в систему пользователь получает доступ к командной оболочке (например к `bash` или `sh`).

1.3 Файлы и каталоги

1.3.1 Файловая система

- Файловая система UNIX представляет собой иерархическую древовидную структуру, состоящую из каталогов и файлов. Начинается она с каталога, который называется корнем (`root`), а имя этого каталога представлено единственным символом - `/`.
- Каталог представляет собой файл, в котором содержатся каталожные записи. Логически каждую такую запись можно представить в виде струк-

туры, состоящей из имени файла и дополнительной информации, описывающей атрибуты файла.

- Атрибуты файла - это такие характеристики, как тип файла (обычный файл или каталог), размер файла, владелец файла, права доступа к файлу (есть ли у других пользователей доступ к файлу), время последней модификации файла.

Для получения атрибутов файла используются функции `stat()` и `fstat()`.

1.3.2 Имя файла

Имена элементов каталога называются именами файлов. Имя файла не может содержать слэш `/` или нулевой символ `0`. Файл «точка» `.` – это текущий каталог, файл «точка-точка» `..` – родительский. При создании нового каталога в нём создаются файлы «точка» и «точка-точка». Для корневого каталога «точка-точка» – это то же самое что и «точка».

Современные ОС позволяют создавать файлы с длиной имени не менее 255 символов.

1.3.3 Путь к файлу

Последовательность одного или более имен файлов (каталог – это тоже файл), разделенных слэшами, образует строку пути к файлу. Эта строка может также начинаться с символа слэша, и тогда она называется строкой **абсолютного пути**, в противном случае – строкой **относительного пути**. В случае относительного пути маршрут начинается от текущего каталога.

Также в пути могут присутствовать «точка» и «точка-точка», например путь `../qwe` – это файл `qwe` в родительском каталоге.

1.3.4 Рабочий каталог

У каждого процесса имеется свой рабочий каталог, который иногда называют текущим рабочим каталогом. Это каталог, от которого отсчитываются все относительные пути, используемые в программе. Процесс может изменить свой рабочий каталог с помощью функции `chdir`.

1.3.5 Домашний каталог

Когда пользователь входит в систему, рабочим каталогом становится его домашний каталог. Домашний каталог пользователя определяется в соответствии с записью в файле паролей.

1.4 Ввод и вывод

1.4.1 Дескрипторы файлов

Дескрипторы файлов - это, как правило, небольшие целые положительные числа, используемые ядром для идентификации файлов, к которым обращается конкретный процесс. Всякий раз, когда процесс открывает существующий или создает новый файл, ядро возвращает его дескриптор, который затем используется для выполнения над файлом операций чтения или записи.

1.4.2 Стандартный ввод, стандартный вывод, Стандартный вывод сообщений об ошибках

По принятым соглашениям все командные оболочки при запуске новой программы открывают для нее три файловых дескриптора: файл стандартного ввода (`stdin`), файл стандартного вывода (`stdout`) и файл стандартного вывода сообщений об ошибках (`stderr`).

`ls > file.out` – перенаправление вывода `stdout` в файл `file.out`

`ls 2> file.err` – перенаправление вывода ошибок `stderr` в файл `file.err`

1.4.3 Небуферизованный ввод-вывод

Небуферизованный ввод-вывод осуществляется функциями `open`, `read`, `write`, `lseek` и `close`. Все эти функции работают с файловыми дескрипторами.

Заголовочный файл `<unistd.h>`, подключаемый из файла `apue.h`, и константы `STDIN_FILENO` и `STDOUT_FILENO` являются частью стандарта **POSIX**.

Константы `STDIN_FILENO` и `STDOUT_FILENO`, определенные в файле `<unistd.h>`, устанавливают дескрипторы файлов стандартного ввода и стандартного вывода. Обычные значения этих констант - соответственно 0 и 1.

1.4.4 Стандартные функции ввода-вывода

Стандартные функции ввода-вывода предоставляют буферизованный интерфейс к функциям небуферизованного ввода-вывода. Использование стандартных функций ввода-вывода избавляет нас от необходимости задумываться о выборе оптимального размера буфера.

Другое преимущество стандартных функций ввода-вывода в том, что они значительно упрощают обработку пользовательского ввода (например `fgets()` или `printf()`).

`<stdio.h>` содержит прототипы всех стандартных функций ввода-вывода.

2 Обзор ОС UNIX: программы и процессы, обработка ошибок, идентификация пользователя.

2.1 Программы и процессы

2.1.1 Программа

Программа – это исполняемый файл, размещенный на диске. Программа считывается в память и затем выполняется ядром через вызов одной из шести функций семейства `exec`.

2.1.2 Процессы и идентификаторы процессов

Программа, находящаяся в процессе исполнения, называется процессом. В некоторых ОС для обозначения выполняемой в данный момент программы используется термин задача.

UNIX обеспечивает присвоение каждому процессу уникального числового идентификатора, который называется идентификатором процесса. Идентификатор процесса – всегда целое неотрицательное число.

2.1.3 `getpid()`

Получить собственный идентификатор процесса можно при помощи функции `getpid()` (get process id)

2.1.4 Управление процессами

Три основные функции отвечают за управление процессами:

- `fork()` – создает копию вызывающего процесса
- `exec()` – замещает дочерний процесс некоторой программой. Функция `exec` имеет шесть разновидностей
- `waitpid()` – ждёт завершения некоторого процесса с определенным `pid`

2.2 Обработка ошибок

2.2.1 `errno`

Очень часто при возникновении ошибки в любой из функций системы UNIX эта функция возвращает отрицательное число, а в глобальную переменную `errno` записывается некоторое целое, которое несёт дополнительную информацию о возникшей ошибке.

Например, функция `open` возвращает либо файловый дескриптор – неотрицательное число, либо `-1` в случае возникновения ошибки. Некоторые функции следуют иному соглашению. Например, большинство функций, которые должны возвращать указатель на какой-либо объект, в случае ошибки возвращают пустой указатель (`NULL`).

Определения переменной `errno` и констант всех возможных кодов ошибок находятся в заголовочном файле `<errno.h>`. Имена констант начинаются с символа `E` (например `EACCES` – возникли проблемы с правами доступа, например при открытии файла).

Стандарты POSIX и ISO C определяют `errno` как символ, раскрывающийся в изменяемое выражение `lvalue` (то есть выражение, которое может стоять слева от оператора присваивания) целого типа.

2.2.2 Вывод сообщения об ошибке

Для вывода сообщений об ошибках стандарт C предусматривает две функции:

- `char *strerror(int errnum);` – преобразует `errnum` (равный `errno`) в строку сообщения об ошибке и возвращает указатель на нее
- `void perror(const char *msg);` – на основе `errno` выводит сообщение об ошибке с префиксом `msg:`, т.е. `msg: error_text`. Вывод заканчивается символом перевода строки.

2.2.3 Восстановление после ошибок

Ошибки, определенные в `<errno.h>`, могут быть разделены на две категории – фатальные и нефатальные. Восстановление нормальной работы после фатальных ошибок невозможно. Самое лучшее, что мы можем сделать, – это вывести сообщение об ошибке на экран или записать его в файл журнала и завершить работу приложения. Нефатальные ошибки допускают нормальное продолжение работы. Большинство нефатальных ошибок по своей природе носят временный характер (например, нехватка ресурсов), и их можно избежать при меньшей загрузке системы.

2.3 Идентификация пользователя

2.3.1 Идентификатор пользователя

Идентификатор пользователя из записи в файле паролей представляет собой числовое значение, которое однозначно идентифицирует пользователя в системе. Идентификатор пользователя назначается системным администратором при создании учетной записи и не может быть изменен пользователем. Как правило,

каждому пользователю назначается уникальный идентификатор. Ядро использует идентификатор пользователя для проверки прав на выполнение определенных операций.

Пользователь с идентификатором `0` называется суперпользователем, или `root`. В файле паролей этому пользователю обычно присвоено имя `root`.

Если процесс имеет привилегии суперпользователя, большинство проверок прав доступа к файлам просто не выполняется. Некоторые системные операции доступны только суперпользователю. Суперпользователь обладает неограниченной свободой действий в системе.

2.3.2 Идентификатор группы

Кроме всего прочего, запись в файле паролей содержит числовой идентификатор группы. Он также назначается системным администратором при создании учетной записи.

Обычно группы используются для распределения пользователей по проектам или отделам. Это позволяет организовать совместное использование ресурсов, например файлов, членами определенной группы.

В системе существует файл групп, в котором указаны соответствия имен групп числовым идентификаторам. Обычно этот файл называется `/etc/group`. Каждый пользователь может состоять в нескольких группах.

3 Обзор ОС UNIX: сигналы, представление времени, системные вызовы и библиотечные функции.

3.1 Сигналы

Сигналы используются, чтобы известить процесс о наступлении некоторого состояния. Например, если процесс попытается выполнить деление на ноль, он получит уведомление в виде сигнала `SIGFPE` (floating-point exception – ошибка выполнения операции с плавающей точкой). Процесс может реагировать на сигнал тремя способами:

- Игнорировать сигнал
- Разрешить выполнение действия по умолчанию
- Определить функцию, которая будет вызвана для обработки сигнала (такие функции называют перехватчиками (handler) сигналов).

`Ctrl+c` – генерирует сигнал прерывания.

При помощи функции `kill` процесс может послать сигнал другому процессу.

При этом процесс который посылает сигнал должен быть владельцем процесса которому посылается сигнал.

3.2 Представление времени

Исторически в системе UNIX поддерживается два различных способа представления временных интервалов:

- Календарное время. Значения в этом представлении хранят число секунд, прошедших с начала Эпохи: 00:00:00 1 января 1970 года по согласованному всемирному времени (Coordinated Universal Time - UTC)
- Время работы процесса. Оно еще называется процессорным временем и измеряет ресурсы центрального процессора, использованные процессом. Значения в этом представлении измеряются в тактах (ticks). Исторически сложилось так, что в различных системах в одной секунде может быть 50, 60 или 100 тактов. Для хранения времени в этом представлении используется тип данных `clock_t`.

При измерении времени выполнения процессасистема UNIX хранит три значения для каждого процесса:

- Общее время (**Clock time**) – это отрезок времени, затраченный процессом от момента запуска до завершения
- Пользовательское время (**User CPU time**) – это время, затраченное на исполнение машинных инструкций самой программы
- Системное время (**System CPU time**) – это время, затраченное на выполнение ядром машинных инструкций от имени процесса

Сумму пользовательского и системного времени часто называют процессорным временем.

3.3 Системные вызовы и библиотечные функции

Любая операционная система обеспечивает прикладным программам возможность обращения к системным службам. Во всех реализациях UNIX имеется строго определенное число точек входа в ядро, которые называются системными вызовами (system calls, см. картинку в начале).

В ОС Linux имеется от 240 до 260 системных вызовов в зависимости от версии. В ОС FreeBSD около 320 системных вызовов.

В системе UNIX для каждого системного вызова предусматривается одноименная функция в стандартной библиотеке языка C. Пользовательский процесс вызывает эту функцию стандартными средствами языка C.

С точки зрения разработчика системы между системным вызовом и библиотечной функцией имеются коренные различия. Но с точки зрения пользователя эти различия носят непринципиальный характер. В контексте нашей книги и системные вызовы, и библиотечные функции можно представлять как обычные функции языка C.

4 Стандарты и реализации ОС UNIX: пределы ISO C, пределы POSIX, функции `sysconf()`, `pathconf()` и `fpathconf()`, элементарные системные типы данных.

4.1 Стандартизация UNIX

4.1.1 ISO C

ISO C, ANSI C – стандарт языка C, опубликованный Американским национальным институтом стандартов (American National Standards Institute – ANSI). ISO C включает в себя следующие заголовочные файлы (библиотеки):

- `<assert.h>` – Проверка программных утверждений
- `<complex.h>` – Поддержка арифметики комплексных чисел
- `<ctype.h>` – Типы символов
- `<errno.h>` – Коды ошибок
- `<fenv.h>` – Окружение операций с плавающей запятой
- `<float.h>` – Арифметика с плавающей запятой
- `<inttypes.h>` – Преобразования целочисленных типов
- `<iso646.h>` – Альтернативные макросы операторов отношений
- `<limits.h>` – Константы реализации (см. ниже, пределы ISO C)
- `<locale.h>` – Классы региональных настроек (локалей)
- `<math.h>` – Математические константы (например π)
- `<setjmp.h>` – Нелокальные переходы

- `<signal.h>` – Сигналы (см. ниже)
- `<stdarg.h>` – Списки аргументов переменной длины
- `<stdbool.h>` – Логический тип и значения
- `<stddef.h>` – Стандартные определения
- `<stdint.h>` – Целочисленные типы
- `<stdio.h>` – Стандартная библиотека ввода/вывода
- `<stdlib.h>` – Функции общего назначения
- `<string.h>` – Операции над строками
- `<tgmath.h>` – Макроопределения математических операций
- `<time.h>` – Время и дата
- `<wchar.h>` – Расширенная поддержка многобайтных символов («широкие» символы)
- `<wctype.h>` – Классификация и функции преобразования многобайтных символов

4.1.2 IEEE POSIX

POSIX – это семейство стандартов, разработанных организацией IEEE (Institute of Electrical and Electronics Engineers). POSIX (Portable Operating System Interface – переносимый интерфейс операционных систем) – набор стандартов, описывающих интерфейсы между операционной системой и прикладной программой (системный API), библиотеку языка C и набор приложений и их интерфейсов. Стандарт POSIX.1 не определяет понятие суперпользователя. Вместо этого он требует, чтобы некоторые операции были доступны только при наличии «соответствующих привилегий», но определение этого термина POSIX.1 оставляет на усмотрение конкретной реализации.

4.1.3 Single UNIX Specification

Single Unix Specification (Единая спецификация UNIX) — это надмножество стандарта POSIX.1 и определяет дополнительные интерфейсы, расширяющие возможности, предоставляемые базовой спецификацией POSIX.1. Стандарт POSIX.1 является эквивалентом раздела Base Specification (базовые спецификации) спецификации Single UNIX Specification.

Расширение X/Open System Interface (XSI) определяет дополнительные интерфейсы POSIX.1, которые должны поддерживаться реализацией, чтобы она получила право именоваться «XSI-совместимой». В их число входят: синхронизация файлов, адрес и размер стека потока, синхронизация потоков между процессами и символьная константа `_XOPEN_UNIX`.

4.1.4 FIPS

Аббревиатура FIPS означает Federal Information Processing Standard (Федеральный стандарт обработки информации). Этот стандарт опубликован правительством США, которое использовало его при покупке компьютерных систем. На данный момент стандарт FIPS отменён.

4.2 Реализации UNIX

ISO C, IEEE POSIX и Single UNIX Specification – это стандарты которые описывают только **спецификации интерфейса**. На основе этих спецификаций реализуются реальные ОС.

4.2.1 UNIX System V Release 4

Версия UNIX System V Release 4 (SVR4) была выпущена подразделением AT&T. Версия SVR4 объединила функциональность AT&T UNIX System Release 3.2 (SVR3.2), SunOS – операционной системы от Sun Microsystems, 4.3BSD, выпущенной Калифорнийским университетом, и Xenix – операционной системы от корпорации Microsoft – в единую операционную систему.

Исходные тексты SVR4 были опубликованы в конце 1989 года, а первые копии стали доступны конечным пользователям в 1990 году.

4.2.2 4.4BSD

Версии Berkeley Software Distribution (BSD) разрабатывались и распространялись Computer Systems Research Group (CSRG) – Группой исследования компьютерных систем Калифорнийского университета в Беркли.

Изначально BSD-системы содержали исходный код, запатентованный AT&T, и подпадали под действие лицензий AT&T. Чтобы получить исходный код BSD-системы, требовалась лицензия AT&T на UNIX.

4.2.3 FreeBSD

Операционная система FreeBSD базируется на 4.4BSD-Lite. Проект FreeBSD образован с целью дальнейшего развития линейки BSD-систем после того, как в

Беркли было принято решение о прекращении работ над BSD-версиями операционной системы UNIX и проект 386BSD оказался заброшенным.

Все программное обеспечение, разработанное в рамках проекта FreeBSD, является свободно распространяемым как в исходных текстах, так и в виде бинарных дистрибутивов.

4.2.4 Linux

Linux – это операционная система, которая предоставляет все богатства программного окружения UNIX и свободно распространяется в соответствии с Общественной лицензией GNU (GNU Public License).

ОС Linux была создана Линусом Торвальдсом в 1991 году в качестве замены ОС MINIX.

4.2.5 Mac OS X

Ядро Mac OS X – «Darwin» представляет собой комбинацию ядра Mach, ОС FreeBSD и объектно-ориентированного фреймворка для драйверов и других расширений ядра.

4.2.6 Solaris

Solaris — это разновидность ОС UNIX, разработанная в Sun Microsystems. Основанная на System V Release 4, она совершенствовалась инженерами из Sun Microsystems более 10 лет. Это единственный коммерчески успешный потомок SVR4, формально сертифицированный как UNIX-система.

4.2.7 Прочие версии UNIX

- AIX, версия UNIX от IBM
- HP-UX, версия UNIX от Hewlett-Packard
- IRIX, UNIX-система, распространяемая компанией Silicon Graphics
- UnixWare, версия UNIX, которая происходит от SVR4 и продается корпорацией SCO

4.3 Пределы ISO C

Все пределы, определяемые стандартом ISO C, являются пределами времени компиляции. Пределы ISO C определены в заголовочном файле `<limits.h>`.

4.4 Пределы POSIX

Стандарт POSIX.1 определяет многочисленные константы, связанные с предельными значениями. Пределы POSIX также определены в `<limits.h>`, какие-то пределы определены всегда, какие-то только при определенных условиях.

4.5 Функции sysconf, pathconf и fpathconf

Функции sysconf, pathconf и fpathconf – позволяют получить значение пределов на этапе выполнения программы. Все три функции возвращают значение соответствующего предела в случае успеха или -1 в случае ошибки.

```
1 #include <unistd.h>
2 long sysconf(int name);
3 long pathconf(const char *pathname, int name);
4 long fpathconf(int fd, int name);
```

Различие между двумя последними функциями состоит в том, что первая получает в аргументе строку пути к файлу, а вторая – файловый дескриптор.

Пример использования `sysconf()`:

```
1 void print_sysconf(const char *con_name, int name)
2 {
3     long val;
4
5     errno = 0;
6     if ((val = sysconf(name)) < 0) {
7         if (errno != 0) {
8             if (errno == EINVAL) {
9                 printf("%s is not supported\n", con_name);
10            } else {
11                perror("Sysconf error: ");
12            }
13        } else {
14            printf("%s is unlimited\n", con_name);
15        }
16    } else {
17        printf("%s = %li\n", con_name, val);
18    }
19 }
20
21
22 int main(int argc, char* argv[])
23 {
24     #ifdef _SC_ARG_MAX
25         print_sysconf("ARG_MAX", _SC_ARG_MAX);
26     #else
27         printf("_SC_ARG_MAX is not defined.\n");
28     #endif
29 }
```

4.6 Элементарные системные типы данных

Заголовочный файл `<sys/types.h>` определяет ряд зависящих от реализации типов данных, которые называются элементарными системными типами данных:

- `clock_t` – Счетчик тактов системных часов (время работы процесса, раздел 1.10)
- `comp_t` – Счетчик тактов в упакованном виде
- `dev_t` – Номер устройства
- `fd_set` – Набор файловых дескрипторов
- `fpos_t` – Позиция в файле
- `gid_t` – Числовой идентификатор группы
- `ino_t` – Номер индексного узла (i-node)
- `mode_t` – Тип файла, режим создания файла
- `nlink_t` – Счетчик ссылок для записей в каталоге
- `off_t` – Размер файла и смещение в файле (со знаком)
- `pid_t` – Числовой идентификатор процесса и идентификатор группы процессов
- `pthread_t` – Числовой идентификатор потока выполнения
- `ptrdiff_t` – Разность двух указателей (со знаком)
- `rlim_t` – Предельное значение для ресурса
- `sig_atomic_t` – Тип данных, доступ к которому может выполняться атомарно
- `sigset_t` – Набор сигналов
- `size_t` – Размер объекта (например, строки) (без знака)
- `ssize_t` – Возвращаемый функциями результат, представляющий счетчик байтов (со знаком)
- `time_t` – Счетчик секунд календарного времени
- `uid_t` – Числовой идентификатор пользователя
- `wchar_t` – Может представлять символы в любой кодировке

5 Файловый ввод-вывод: дескрипторы файлов, функция `open()`, функция `creat()`, функция `close()`.

Большинство операций файлового ввода/вывода в UNIX можно выполнить с помощью всего пяти функций: `open`, `read`, `write`, `lseek` и `close`.

5.1 Дескрипторы файлов

Все открытые файлы представлены в ядре файловыми дескрипторами. Файловый дескриптор – это неотрицательное целое число. Когда процесс открывает существующий файл или создает новый, ядро возвращает ему файловый дескриптор.

В соответствии с соглашениями командные оболочки UNIX ассоциируют файловый дескриптор 0 (`STDIN_FILENO`) с устройством стандартного ввода процесса, 1 (`STDOUT_FILENO`) – с устройством стандартного вывода и 2 (`STDERR_FILENO`) – с устройством стандартного вывода сообщений об ошибках.

Под файловые дескрипторы отводится диапазон чисел от 0 до `OPEN_MAX-1`.

5.2 Функции `open` и `openat`

Файл создается (если не существует) или открывается функцией `open` или `openat`.

```
1 #include <fcntl.h>
2 int open(const char *path, int oflag, ... /* mode_t mode */);
3 int openat(int fd, const char *path, int oflag, ... /* mode_t mode */);
```

Последний аргумент обозначен многоточием (...), таким способом стандарт ISO C указывает, что количество остальных аргументов и их типы могут варьироваться.

`path` – имя файла

`oflag` – комбинация флагов (целых чисел) `O_RDONLY`, `O_WRONLY`, `O_RDWR`, `O_EXEC`, `O_APPEND`, `O_CREAT` и др. при помощи объединения ИЛИ (`OR`, `|`)

Функции `open` и `openat` гарантируют, что возвращаемый ими дескриптор файла будет иметь наименьшее неиспользуемое положительное числовое значение.

5.2.1 Параметр `fd` в функции `openat`

Возможны три разных варианта:

- Параметр `path` содержит строку абсолютного пути. В этом случае параметр `fd` игнорируется и `openat` действует подобно функции `open`.

- Параметр `path` содержит строку относительного пути, а параметр `fd` содержит дескриптор файла, определяющего местоположение в файловой системе, откуда будет откладываться относительный путь.
- Параметр `path` содержит строку относительного пути, а параметр `fd` содержит специальное значение `AT_FDCWD`. В этом случае путь начинается от текущего рабочего каталога и функция `openat` действует подобно функции `open`.

5.3 Функция `creat`

Новый файл можно также создать с помощью функции `creat`.

```
1 #include <fcntl.h>
2 int creat(const char *path, mode_t mode);
```

Эта функция эквивалентна вызову

```
1 open(path, O_WRONLY | O_CREAT | O_TRUNC, mode);
```

5.4 Функция `close`

Закрытие открытого файла производится вызовом функции `close`.

```
1 #include <unistd.h>
2 int close(int fd);
```

Закрытие файла приводит также к снятию любых блокировок, которые могли быть наложены процессом. При завершении процесса все открытые им файлы автоматически закрываются ядром.

6 Файловый ввод-вывод: Функция `lseek()`, функция `read()`, функция `write()`

6.1 Функция `lseek`

С любым открытым файлом связано такое понятие, как текущая позиция в файле. Как правило, это неотрицательное целое число, которым выражается смещение в байтах от начала файла. По умолчанию при открытии файла текущая позиция инициализируется числом 0, если не был установлен флаг `O_APPEND`. Явное изменение текущей позиции в файле выполняется с помощью функции `lseek`.

```
1 #include <unistd.h>
2 off_t lseek(int fd, off_t offset, int whence);
```

Интерпретация аргумента `offset` зависит от значения аргумента `whence`:

- **SEEK_SET** – offset интерпретируется как смещение от начала файла
- **SEEK_CUR** – offset интерпретируется как смещение от текущей позиции в файле. В этом случае offset может принимать и положительные, и отрицательные значения.
- **SEEK_END** – offset интерпретируется как смещение от конца файла. В этом случае offset может принимать и положительные, и отрицательные значения

Поскольку в случае успеха функция `lseek` возвращает новую текущую позицию в файле, в аргументе `offset` можно передать 0, чтобы узнать текущую позицию:

```
1 off_t currpos;
2 currpos = lseek(fd, 0, SEEK_CUR);
```

Таким же способом можно определить, поддерживается ли свободное перемещение текущей позиции файла.

6.2 Функция `read`

Чтение данных из открытого файла производится функцией `read`.

```
1 #include <unistd.h>
2 ssize_t read(int fd, void *buf, size_t nbytes);
```

Возвращает количество прочитанных байтов, 0 – если достигнут конец файла, -1 – в случае ошибки.

6.3 Функция `write`

Запись данных в открытый файл производится функцией `write`.

```
1 #include <unistd.h>
2 #include <unistd.h>
3 ssize_t write(int fd, const void *buf, size_t nbytes);
```

Возвращает количество записанных байтов в случае успеха, -1 – в случае ошибки.

7 Файловый ввод-вывод: эффективность операций ввода-вывода

Пример копирования из стандартного ввода в стандартный вывод

```
1 #include "apue.h"
2 #define BUFSIZE 4096
3
4 int main()
5 {
6     int n;
7     char buf[BUFSIZE];
8
9     while((n = read(STDIN_FILENO, buf, BUFSIZE)) > 0)
10         if(write(STDOUT_FILENO, buf, n) != n)
11             err_sys("write error!");
12
13     if(n < 0)
14         err_sys("read error!");
15
16     exit(0);
17 }
```

Размера буфера `buf` влияет на производительность, т.к. переход ядра в привилегированный режим для чтения и записи занимает определенное время.

Производительность операции чтения с различными размерами буфера в ОС Linux (пример из учебника):

BUFSIZE	Пользовательское время (секунды)	Системное время (секунды)	Общее время (секунды)	Количество циклов
1	20,03	117,50	138,73	516 581 760
2	9,69	58,76	68,60	258 290 880
4	4,60	36,47	41,27	129 145 440
8	2,47	15,44	18,38	64 572 720
16	1,07	7,93	9,38	32 286 360
32	0,56	4,51	8,82	16 143 180
64	0,34	2,72	8,66	8 071 590
128	0,34	1,84	8,69	4 035 795
256	0,15	1,30	8,69	2 017 898
512	0,09	0,95	8,63	1 008 949
1024	0,02	0,78	8,58	504 475
2048	0,04	0,66	8,68	252 238
4096	0,03	0,58	8,62	126 119
8192	0,00	0,54	8,52	63 060
16 384	0,01	0,56	8,69	31 530
32 768	0,00	0,56	8,51	15 765
65 536	0,01	0,56	9,12	7883
131 072	0,00	0,58	9,08	3942
262 144	0,00	0,60	8,70	1971
524 288	0,01	0,58	8,58	986

Наилучшее системное время – при размере буфера равном размеру дискового блока (`st_blksize`).

Для повышения производительности большинство файловых систем поддерживают опережающее чтение. Обнаружив ряд последовательных операций чтения, система пытается прочесть больший объем данных, чем было запрошено приложением, предполагая, что программа вскоре продолжит чтение.

8 Файловый ввод-вывод: совместное использование файлов, атомарные операции, функции `dup()` и `dup2()`

8.1 Совместное использование файлов

ОС UNIX поддерживает совместное использование открытых файлов несколькими процессами.

Ядро использует три структуры данных для представления открытого файла, а отношения между ними определяют взаимовлияние процессов при совместном использовании файлов:

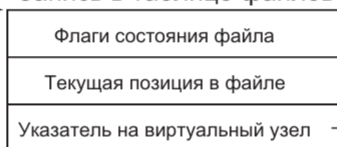
1. Каждому процессу соответствует запись в таблице процессов. С каждой записью в таблице процессов связана таблица открытых файловых дескрипторов. Для каждого дескриптора хранится следующая информация:
 - флаги дескриптора
 - указатель на запись в таблице файлов
2. Все открытые файлы представлены в ядре таблицей файлов. Каждая запись в таблице содержит:
 - флаги состояния файла
 - текущая позиция в файле
 - указатель на запись в таблице виртуальных узлов (`v-node`)
3. Каждому открытому файлу (или устройству) соответствует структура виртуального узла (`v-node`) с информацией о типе файла и указатели для функций, работающих с файлом

В Linux отсутствует понятие виртуальных узлов (`vnode`). Вместо него используются структуры индексных узлов (`inode`). Хотя реализация их различна, концептуально они представляют собой одно и то же. В обоих случаях индексный узел хранит информацию, специфичную для конкретной файловой системы. Ситуация, когда два независимых процесса открывают один и тот же файл:

Таблица дескрипторов процесса



Запись в таблице файлов



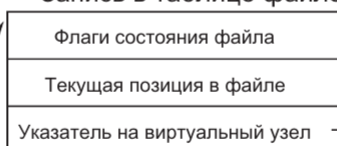
Запись в таблице виртуальных узлов



Таблица дескрипторов процесса



Запись в таблице файлов



Каждый процесс, открывающий файл, создает собственную запись в таблице файлов, но двум этим записям соответствует единственная запись в таблице виртуальных узлов.

8.2 Атомарные операции

8.2.1 Добавление данных в конец файла

ОС UNIX выполняет операцию `write` атомарно (только один процесс может получить доступ), если указан флаг `O_APPEND`.

8.2.2 Функции `pread` и `pwrite`

`pread` и `pwrite` – это расширения XSI для стандарта Single UNIX Specification. Эти расширения позволяют процессам атомарно выполнять ввода/вывода.

```

1 #include <unistd.h>
2 ssize_t pread(int fd, void *buf, size_t nbytes, off_t offset);
3 ssize_t pwrite(int fd, const void *buf, size_t nbytes, off_t offset);

```


Вызов функции `pread` эквивалентен двум последовательным вызовам функций `lseek` и `read` со следующими отличиями:

- При использовании `pread` нет возможности прервать выполнение этих двух операций
- Значение текущей позиции в файле не изменяется

Вызов функции `pwrite` эквивалентен двум последовательным вызовам функций `lseek` и `read` с аналогичными отличиями.

8.2.3 Создание файла

При указании флага `O_CREAT` функция `open()` будет атомарно проверять существование и создавать файл.

8.3 Функции `dup` и `dup2`

Дубликат дескриптора существующего файла можно создать с помощью одной из следующих функций:

```
1 #include <unistd.h>
2 int dup(int fd);
3 int dup2(int fd, int fd2);
```

Возвращают новый дескриптор файла или `-1` — в случае ошибки.

Функция `dup` гарантирует, что возвращаемый ею новый файловый дескриптор будет иметь наименьшее возможное значение.

Вызывая функцию `dup2`, мы указываем значение нового дескриптора в аргументе `fd2`. Если дескриптор `fd2` перед вызовом функции уже был открыт, он предварительно закрывается.

Если значения аргументов `fd` и `fd2` равны, функция `dup2` вернет дескриптор `fd2`, не закрывая его.

9 Файловый ввод-вывод: функции `sync()`, `fsync()`, `fdatasync()`, `fcntl()`, `ioctl()`, `/dev/fd`

9.1 Функции `sync`, `fsync` и `fdatasync`

Традиционные реализации UNIX имеют в своем распоряжении буферный кэш, или кэш страниц, через который выполняется большинство дисковых операций ввода/вывода. При записи данных в файл они, как правило, сначала помещаются ядром в один из буферов, а затем ставятся в очередь для записи на диск в более позднее время. Этот прием называется отложенной записью.

Ядро обычно записывает отложенные данные на диск, когда возникает необходимость в повторном использовании буфера. Для синхронизации файловой системы на диске и содержимого буферного кэша существуют функции `sync`, `fsync` и `fdatasync`:

```
1 #include <unistd.h>
2 #include <unistd.h>
3 int fsync(int fd);
4 int fdatasync(int fd);
5 void sync(void);
```

Функция `sync` просто ставит все измененные блоки буферов в очередь для записи и возвращает управление — она не ждет, пока данные фактически запишутся на диск.

Функция `sync`, как правило, вызывается периодически (обычно каждые 30 секунд) из системного демона, часто называемого `update`.

Функция `fsync` применяется только к одному файлу, который определяется файловым дескриптором `fd`, кроме того, она ожидает завершения физической записи данных на диск, прежде чем вернуть управление.

Функция `fdatasync` похожа на функцию `fsync`, но воздействует только на содержимое файла. (При использовании `fsync` также синхронно обновляются атрибуты файла).

9.2 Функция `fcntl`

С помощью функции `fcntl` можно изменять свойства уже открытого файла.

```
1 #include <fcntl.h>
2 int fcntl(int fd, int cmd, ... /* int arg */);
```

Функция `fcntl` используется в пяти различных случаях:

1. Создание дубликата существующего дескриптора (`cmd = F_DUPFD` или `F_DUPFD_CLOEXEC`)
2. Получение/установка флагов дескриптора (`cmd = F_GETFD` или `F_SETFD`)
3. Получение/установка флагов состояния файла (`cmd = F_GETFL` или `F_SETFL`)
4. Проверка/установка владельца для асинхронных операций ввода/вывода (`cmd = F_GETOWN` или `F_SETOWN`)
5. Получение/установка блокировки на отдельную запись в файле (`cmd = F_GETLK`, `F_SETLK` или `F_SETLKW`)

9.3 Функция `ioctl`

Функция `ioctl` всегда была универсальным инструментом ввода/вывода. Все, что невозможно выразить с помощью функций, описанных в этой главе, как правило, делается с помощью `ioctl`. Возможности этой функции чаще всего использовались в операциях терминального ввода/вывода.

```
1 #include <unistd.h>      /* System V */
2 #include <sys/ioctl.h>    /* BSD и Linux */
3 int ioctl(int fd, int request, ...);
```

9.4 `/dev/fd`

В современных операционных системах имеется каталог `/dev/fd`, в котором находятся файлы с именами 0, 1, 2 и т. д. Открытие файла `/dev/fd/n` эквивалентно созданию дубликата дескриптора с номером n, при условии, что дескриптор n открыт.

В некоторых системах имеются файлы `/dev/stdin`, `/dev/stdout` и `/dev/stderr`, эквивалентные файлам `/dev/fd/0`, `/dev/fd/1` и `/dev/fd/2` соответственно.

10 Файлы и каталоги: функции `stat()`, `fstat()`, `lstat()`, содержимое `struct stat`.

Функции `stat()`, `fstat()`, `lstat()` возвращают информацию об указанном файле в структуру `stat`. В случае успеха возвращается ноль. При ошибке возвращается -1, а переменной `errno` присваивается номер ошибки

```
1 #include <sys/types.h>
2 #include <sys/stat.h>
3 #include <unistd.h>
4
5 int stat(const char *file_name, struct stat *buf);
6 int fstat(int filedes, struct stat *buf);
7 int lstat(const char *file_name, struct stat *buf);
```

`stat()` возвращает информацию о файле `file_name`.

`lstat()` идентична `stat()`, но в случае символьных ссылок она возвращает информацию о самой ссылке, а не о файле.

`fstat()` идентична `stat()`, только возвращается информация об открытом файле, на который указывает `filedes` (дескриптор файла).

Структура `stat` содержит следующие поля:

```

1 struct stat {
2     dev_t      st_dev;      // устройство
3     ino_t      st_ino;      // inode
4     mode_t     st_mode;     // режим доступа
5     nlink_t    st_nlink;    // количество жестких ссылок
6     uid_t      st_uid;      // идентификатор пользователя-владельца
7     gid_t      st_gid;      // идентификатор группы-владельца
8     dev_t      st_rdev;     // тип устройства
9                     // (если это устройство)
10    off_t       st_size;     // общий размер в байтах
11    blksize_t   st_blksize;  // размер блока ввода-вывода
12                     // в файловой системе
13    blkcnt_t    st_blocks;   // количество выделенных блоков
14    time_t      st_atime;    // время последнего доступа
15    time_t      st_mtime;    // время последней модификации
16    time_t      st_ctime;    // время последнего изменения
17 };

```

11 Файлы и каталоги: типы файлов, права доступа к файлу, функция `umask()`.

11.1 Типы файлов

Большинство файлов в UNIX являются либо обычными файлами, либо каталогами, но есть и другие типы файлов:

- Обычный файл – наиболее распространенный тип файлов, хранящих данные в некотором виде. Ядро UNIX не делает различий между текстовыми и двоичными файлами. Интерпретация содержимого файла полностью зависит от прикладной программы, обрабатывающей файл. Одно из наиболее известных исключений из этого правила – выполняемые файлы.
- Файл каталога. Файлы этого типа содержат имена других файлов и ссылки на информацию о них. Любой процесс, обладающий правом на чтение каталога, может прочитать его содержимое, но только ядро обладает правом на запись непосредственно в файл каталога.
- Специальный файл блочного устройства. Этот тип файлов обеспечивает буферизованный ввод/вывод фиксированными блоками для таких устройств, как жесткие диски.
- Специальный файл символьного устройства. Этот тип файлов обеспечивает небуферизованный ввод/вывод для устройств с переменным размером

блока. Все устройства в системе являются либо специальными файлами блочных устройств, либо специальными файлами символьных устройств.

- FIFO, или именованный канал. Этот тип файлов используется для организации обмена информацией между процессами.
- Сокет. Этот тип файлов используется для обмена информацией между процессами через сетевые соединения. Сокеты можно также применять для обмена информацией между процессами на одной и той же машине.
- Символическая ссылка. Файлы этого типа являются ссылками на другие файлы.

Тип файла хранится в поле `st_mode` структуры `stat`. Определить тип файла можно с помощью макроопределений:

- `S_ISREG()` – Обычный файл
- `S_ISDIR()` – Каталог
- `S_ISCHR()` – Специальный файл символьного устройства
- `S_ISBLK()` – Специальный файл блочного устройства
- `S_ISFIFO()` – Канал
- `S_ISLNK()` – Символическая ссылка
- `S_ISSOCK()` – Сокет

11.2 Права доступа к файлу

Поле `st_mode` содержит также биты прав доступа к файлу.

Права доступа к файлу определяются девятью битами. Пример:

`110 100 100 = 6 4 4`

`rw- rw- rw-`

`rw- r-- r--` – `rw`, `read-write` для владельца, `r`, `read` для группы и остальных. `x`, `execute` – не разрешено

Биты прав доступа из файла `<sys/stat.h>`:

Пользователь, владелец:

- `S_IRUSR`
- `S_IWUSR`
- `S_IXUSR`

Группа:

- S_IRGRP
- S_IWGRP
- S_IXGRP

Остальные:

- S_IROTH
- S_IWOTH
- S_IXOTH

11.3 Функция `umask`

Функция `umask()` устанавливает маску режима создания файлов для процесса и возвращает предыдущее значение.

```
1 #include <sys/stat.h>
2 mode_t umask(mode_t cmask);
```

Аргумент `cmask` – это набор констант из списка выше объединённых при помощи ИЛИ (`OR`, `|`)

12 Файлы и каталоги: функции `chmod()`, `fchmod()`, `chown()`, `fchown()`, `lchown()`.

13 Файлы и каталоги: размер файла, дырки в файлах, усечение файлов, файловые системы, функции `link()`, `unlink()`, `remove()`, `rename()`.

14 Файлы и каталоги: символические ссылки, функции `symlink()` и `readlink()`.

- 15 Файлы и каталоги: временные характеристики файлов, функция `utime()`.
- 16 Файлы и каталоги: функции `mkdir()` и `rmdir()`, чтение каталогов, функции `chdir()`, `fchdir()`, `getcwd()`.
- 17 Стандартная библиотека ввода-вывода: потоки и объекты `FILE`, стандартные потоки ввода, вывода и сообщений об ошибках, буферизация.
- 18 Стандартная библиотека ввода-вывода: открытие потока, чтение из потока и запись в поток, функции ввода, функции вывода.
- 19 Стандартная библиотека ввода-вывода: эффективность стандартных операций ввода-вывода, позиционирование в потоке.
- 20 Стандартная библиотека ввода-вывода: форматированный вывод, форматированный ввод, временные файлы.
- 21 Управление процессами: идентификаторы процесса, функция `fork()`, совместное использование файлов.
- 22 Управление процессами: функция `exit()`, функции `wait()` и `waitpid()`.
- 23 Управление процессами: семейство функций `exec()`.
- 24 Управление процессами: изменение идентификаторов пользователя и группы, функции `setuid()`, `setgid()`, `seteuid()`, `setegid()`.